Evaluating culture metrics as predictors of corporate performance

by

Juliana Peres Hernandes Sanches

(**j.phsanches@gmail.com** - AM 2017015017)

Committee:
- Prof. Grigorouris Evangelos (supervisor)
- Assoc. Prof. Tsafarakis Stelios
- Prof. Vassilis Moustakis

April 12th, 2023

Table of Contents

Foreword – Acknowledgments

I would like to express my deepest feelings of gratitude towards my supervisor and professor of the School of Production Engineering and Management of the Technical University of Crete (TUC), Professor Evangelos Grigoroudis for his academic guidance, his encouragement and trust throughout the stages of the preparation of this thesis. Without his mentoring presence and tolerance, completing this thesis would have been impossible.

Chania, April 2023

Abstract

A culture comprises a number of features that include "a pattern of basic assumptions" that the group members adopt in order to cope with the internal and external organizational challenges.

The "Organizational Culture", also noted as "Corporate Culture", has been extensively researched in the modern literature, in order to examine a wide range of aspects, like teaming behaviors, accountability, ownership, leadership, and the emerging habits, beliefs, and norms cultivated by group dynamics. In today's world all of these traits eventually come into shaping a distinct, autonomous, legally recognized entity "born out of statute", the corporation which is manifested as a group of people, processes and tools aligned under a common expression of vision and strategy.

From a utilitarian standpoint of view, such an organized expression needs to be measured and controlled by practices that can increase both the employee's sentiment and the financial performance of the group (i.e., the corporation). The establishment of such processes can be primarily facilitated through the institutionalization of the main areas of contribution.

When considering the combination of financial and non-financial identifiers in order to drive definitive statements about corporate performance, it seems that employees are especially affected by their business leaders, the expectations around career opportunities, the actual compensation and various perks and benefits.

The hypothesis for undertaking the current work lays on the anticipation for the immersion of interesting results when considering metrics like the employee "happiness" and/or "leader appreciation" to satisfactorily predict a company's future performance.

Structure of Work

This study explores the insights regarding the effective utilization of indicative non-financial predictors which can offer good predictors and be used in organizational decision-making. The research can facilitate the identification of the best practices for the application and the interpretation of various culture metrics in the corporate context, leading to ultimately enabling corporations and organizations to generate educated decisions regarding employee satisfaction and engagement and in turn to improve the organizational performance and to promote new emerging and sustainable business practices.

Chapter 1 provides an overview, discussing about the present status of corporate performance evaluation metrics. It proceeds by outlining the problem statement, the hypothesis, the methodology, showcasing the datasets, and mentioning the actual purpose of the study. This chapter lays the foundational base for the entire thesis and sets the stage for the chapters to follow.

Chapter 2 is delving into the discussion subject around common key performance indicators, the potential of financial metrics, and a plethora of factors that can (and do) affect corporate performance. It provides definitions around basic concepts like the corporation archetype, discussing their common properties, and various differentiators among the nature of corporations. An important issue that can be quite challenging in prediction regards the impact of the IPOs, the versatility of scale and the effects in capital vs. labor vs. technology-intensive schemas, and the relevance of the indexes resulting from their evaluation. The chapter is complete by presenting various challenges regarding data interpretation and the nature of the evaluation algorithms.

Chapter 3 is exploring the role of the dimensions that constitute what is defined as company culture, and more specifically its linguistic and semantic orientations in relation to corporate performance. The chapter covers various ethical aspects, political stands and practices, the code of conduct in a workplace and the employee's morale. Then it proceeds by discussing asymmetrical transactions,

explaining which are the related corporate acts, or other misleading indicators, like the ones that are created for greenwashing and survey manipulation. The chapter concludes by the examination of the indicative metrics around culture and morale in businesses.

Chapter 4 introduces the effectuation concept and analyzes the relevance of that notion to firm growth and fitting scores. It continues by covering a plethora of aspects regarding data analysis, by referring to data types, the main stages of data mining, explaining acquisition, cleaning, and exploratory data analysis. If completes by explaining what data modeling is, presenting the regression functions to be used, discussing, challenging, and finally measuring their forecasting strengths.

Chapter 5 begins by providing a data analysis, starting with preprocessing and data reviewing. Then the chapter proceeds by depicting the various findings through feature processing and exploratory analysis. Additionally, it covers various aspects of modeling, issues regarding datasets definition, the importance of scaling data. It concludes with a fruitful comparison of the different datasets used on different model functions and discusses the results and any potential implications on corporate performance prediction.

Finally, Chapter 6 is acting as a summarization of the key findings as they emerge from the results of Chapter 5, providing both direct and indirect useful conclusions. There are various important points to be considered and highlighted per case when evaluating the performance of the predictors over the nature and origin of the specific dataset. The chapter completes the thesis by proposing various future research directions. It also adds by mentioning alternative approaches that can improve prediction, based on the insights that are gained by the study.

1. Introduction

The concept of culture was first introduced as an operational discipline in 1951 by Elliot Jaques, to semantically support the generally shared values of a team as a means of framing the common wishes, desires, and aspirations of workers under an organizational context in his book "The Changing Culture of a Factory" [1]. By his definition, such a culture comprises several features, including "a pattern of basic assumptions" that the group members adopt to cope with internal and external organizational challenges.

Ever since, "Organizational Culture", also noted as "Business Culture" or "Corporate Culture", has been extensively researched in the modern literature to examine a wide range of aspects, like teaming behaviors [2, 3, 4], the ideas of accountability, ownership, leadership [5, 6, 7], and the emerging idiosyncrasies, habits, beliefs, and norms cultivated by group dynamics [8, 9, 10]. In today's world, all of these traits eventually come into shaping a distinct (single), the autonomous, legally recognized entity "born out of statute," the corporation [11]. This morphing is generally expressed under a shared vision and strategy, which is manifested through the aforementioned term of Corporate Culture.

However, this term is mostly used to denote the internal way in which the employees and the management interact on a daily basis [12]. When it comes to the expression that formalizes an outward 'identity figure' in order to denote the umbrella that covers the transactions norm of the corporation with the public, the Company Culture is substituted by the Corporate Identity term [13]. As expected, this outward expression is not necessarily valid since a corporation benefit when it illustrates the positive traits it wants to be associated with its brand. In contrast to a natural person, the Corporate Identity mainly reflects a tool of communicating to the public a value of differentiation, its intentions, and beliefs, but not its actual social behavior.

Although a solid corporate identity turns out to be quite beneficial for any business, in practice, the liability of a corporation may, per case, be "limited," which,

in a tort, remains controversial. This situation occurs since the third parties do not agree to waive the right to pursue shareholders, i.e., any natural persons who may hold ethically responsible for the actions of a vague schema of corporate transactions [14]. This seemingly individualistic dimension of a corporation, which can be loosely defined as a Corporate Personality, raises strong ethical issues for both the members of the group (internally regulated by the compliance under a namely "Code of Conduct" for the employees) and the public (which is usually externally appreciated as the requirement for Customer Satisfaction).

From a utilitarian standpoint, both the internal and external expressions of the ethical behavior of an organization need to be measured and controlled with practices that can – in general - increase both the employee's sentiment and the financial performance of the group (i.e., the corporation). Establishing such processes can be primarily facilitated through the institutionalization of Social Responsibility Reporting (SRR) [15]. In this sense, SRR aims at estimating in either monetary on non-monetary units the effect of the activities of an organization, both on the firm and/or on those served/affected by the firm, including social costs and benefits. The scope of the SRR spans across five main areas of contribution (net income, human resources, public, environment, and product/service).

1.1. Brief review of current status

Since the Human Resource Contribution is usually the one that is harder to appreciate, the methods for measuring and affecting employee morale are constantly evolving, but still, this is not an easy target to achieve. The promotion of the ethical corporate culture depends on timely commitment, passionate engagement, and sufficient budget. As mentioned by the authors of [16] a report, that was published by the Chartered Institute of Internal Auditors (IIA), states that although 94 of the FTSE 100 companies claimed to support ethics in their 2014 annual reports, it was found that only 23 of them had devised a way of measuring the extent of their ethical behavior.

On 22nd of March 2021, after conducting a thorough internal investigation for its 160.000 employees around the world, Microsoft published a report [17] with 7 warning signs, mentioning in 3 of them that the "leaders are out of touch with employees and need a wake-up call", that "high productivity is masking an exhausted workforce" and that "Gen Z is at risk and will need to be re-energized".

On the 13th of April same year, Fortune states that "Happy employees = happy shareholders", in an attempt to quickly establish the notion that there is a direct connection between the harboring of happy employees and the profitability of the companies that "went above and beyond for employees" [18]. These results suggest that, in regards to employee satisfaction and commitment, the 100 top companies in Fortune's respective list were able to meet the needs of their employees during the Covid-19 pandemic and their actions seem to be paying off already.

This seems to be backed up by the work of [19] which studies the correlation of employee satisfaction and corporate performance based on Glassdoor.com review data. The work concludes that under 9 aspects in total, the Retailing industry is heavily concentrated in Quality (since such companies deal with customers) when in contrast the Technology industry is favored by a balanced mix of Teamwork, Innovation, Respect, and Quality, with Innovation being the most essential aspect. Finally, the financial industry is also found to be quite concentrated on Quality but also includes the aspect of Innovation and Respect.

Regarding the job satisfaction data reviews placed on Glassdoor.com, the survey of [20] examines the construct validity of the ratings and reaches the conclusion that these ratings are very informative and highly recommended.

1.2.   Problem Statement

However, this is far from running to the generalized conclusion that such correlations are governed by causality, mostly due to the complexity that is induced by the stochastic nature of the data, or other inherited biasing properties, which to

some – unknown extent – may relate to the human nature or to hidden third-party externalities.

Regarding the employee satisfaction (and other non-financial indicators), the complexity is enhanced by the multi-disciplinary activities that any worker exercises and is involved with, like recruiting, training, gaining experience, job rotation, wages, promotion levels and special bonuses, benefit plans, union relations, layoff and recalling practices, imposed stress, compliance policies, work conditions and physical environment. Additionally, the complexity is further stretched by the peculiarities of an employee's personality and job description deviations that may be due to the differentiation of the individual's: skills, acquired knowledge, life goals, self-actualization, mutual trust and confidence, perceived idea of security, work-life balance, personal value system, commonly accepted merits, cognitive biases etc.

Regarding the financial indicators that describe the performance of the organization, the annual effects need to be – to some estimated extent – normalized against the effects of globally induced fluctuations, like government legislation, GDP fluctuations, the market capitalization of a specific industry at times of crisis, company merges, etc.

Another important systemic issue relates to the action-and-effect hysteresis, where a number of good organization practices may take time to be acknowledged and on the contrary, some bad organization practices may stand out once a critical mass of misbehaving has been accumulated, and in the worst-case scenario, at times that the current management is truthfully loyal and caring towards the employees' needs. The latter would not only produce an unrightful reaction of the employees' unions, but it would also hurt the performance of the corporation and it would signal the wrong idea to its management, thus cancelling or distorting the direction of any future decision.

The action-and-effect hysteresis is evident to popular Customer Satisfaction Metrics that are shown to correlate with revenue growth like the Net Promoter Score (NPS) and Customer Satisfaction Score (CSAT) which are notorious for their ability

to predict future business performance [21]. Similarly, the same can happen with other metrics like the Customer Effort Score (CES), the Customer Trust Index (CTI) and the customer Churn which need to be applied with caution, especially during times of crisis or volatility. The same may be the case of employee engagement in companies, but it is yet to be shown, as this is not necessarily a conclusive result. For example, the survey of the authors of [22] indicates that recent prescriptions which dictate that "*managers should abandon customer satisfaction monitoring and to focus solely on customer recommendation metrics*" are misguided and potentially harmful. This shows that inferring any statement about the distant case of employee engagement, when considering similar to customer satisfaction metrics, is far-fetched and possibly erroneous.

1.3.   Hypothesis and Proposition

When considering the combination of financial and non-financial identifiers in order to drive definitive statements about corporate performance, it seems that employees are significantly affected by their business leaders, attributing a personal character to the company they work for, especially to the ones that seem to be embedding purpose and ethics (like for example by addressing injustice) [23]. The accepted leaders are chosen based on their responsible decision making and the ones who act by promoting a consistent, clear, and certain strategy [23]. Similar are the findings of the study done by Glassdoor.com in [24, 25], which suggest that the CEOs of more profitable companies generally are rewarded with higher Approval Ratings. Other significant predictors seem to be career opportunities, compensation and benefits.

The hypothesis for undertaking the current work lays on the anticipation for the immersion of interesting results when considering metrics connected to the employee "happiness" and/or "leader appreciation" to satisfactorily predict a company's future performance. The proposition is to examine the data that has been

mined from Indeed.com and to investigate the usability of similar non-financial predictors towards that goal.

1.4.   Methodology

The approach will be based on methodologies that adopt ideas of papers like [26-31] and more, which hold very useful insights on how to properly measure interlinkages among various dimensions, the innate characteristics of corporate culture, and how to focus on metrics that may stand as valid performance predictors, including the necessary compensation on contingencies and macroeconomic fluctuations.

1.5.   Datasets

The favored datasets for this work can be found in [32] and [33].

1.6.   Purpose of the study

The work in this thesis primarily aims **to provide a comprehensive understanding** of the, usually complex in nature, relationships among **corporate performance** and **various culture metrics**, while considering a number of factors that can influence financial indicators and employee satisfaction, and vice versa. By examining any intricacies presented by employee morale and its relation to organizational performance, the study aspires to ascertain the origin of the observed correlations among the selected variables and whether they are causal indeed or a mere result of other underlying factors. The same aspiration lays to the hope that this study can contribute to the ongoing debate regarding the importance of the non-financial indicators and their ability to measure the otherwise abstract entities that shape the working environment of a company, and to predict the future company's performance. The primary focus is given in examining the role that **employee happiness** has over **leader appreciation** and vice versa, as there are

indications that it can predict a company's future performance. As stated before, the first part is dedicated into investigating the complex relationship between various indicators, aiming to determine the extent to which employees' perception of their business leaders and their personal sense of fulfillment (within the corporation/organization) can impact the performance of both (employees and company). The investigation of that matter will be based upon research of the existing literature, highlighting the importance of responsible guidance and decision-making by leaders, as well as applying clear strategies in general, can have when aiming to promote a positive work experience and to drive the teams to business success.

Secondly, this thesis is standing as an explorational guide that **searches for the potential** of non-financial predictors which are mined from online databases into assessing various types of performance. The research to follow on Chapters 4 and 5 will receive dataset insights and evaluate the predictive power of such metrics, respectively, in order to determine the prediction ability of specific models, and their suitability per case, by data analysis on various dimensions related to employee satisfaction, to leader approval ratings, and to other (seemingly) relevant factors. Towards that goal there is a need to consider the phenomenon of **action-and-effect hysteresis,** especially in the context of various corporate culture, let alone financial, metrics. This investigation will keep an eye on any time lag that may exist between the implementation of organizational policies/practices and their observable effect or impact on measurable variables (which may be related to performance) as well as the potential for concluding to erroneous comments, because of this delay. Examining the impact of hysteresis helps avoiding pitfalls and drawing a more comprehensive/accurate picture of organizational success. Thus, the study can provide valuable **validity** and **reliability** insights regarding such datasets and predictors which are important in times, especially during company crisis or times of high volatility.

Thirdly, this thesis can **contribute as a guidance** into the practices of effective utilization of non-financial predictors and to the identification of best practices for

the applying and interpreting culture metrics in the corporate analysis context. The identification of the most relevant and thus reliable culture metrics that are related to employee happiness and, as already discussed, most probably leader appreciation and all other factors to be unveiled, the research will provide credible recommendations for corporations/organizations that seek to improve their performance and one justifiable way will be through employee engagement that leads to employee satisfaction (and vice versa). By fortifying the establishment of the existing literature and studied cases, this work will unveil the potential risks as well as the benefits that are associated with the use of various culture metrics and it will proceed by offering recommendations for their effective utilization. The resulting development of sustainable strategies can help businesses balance outstanding considerations based on such predictors, which can help them to reach long-term success and resilience, regardless of market fluctuations and other limiting or threatening external factors. The goal is encouraging corporations ανδ organizations to the making of more informed decisions taking into consideration the monitoring and thus the management of employee engagement and satisfaction, ultimately leading to an improved organizational performance and to more sustainable business policies / practices.

2. Conventional Predictors of corporate performance

As mentioned in the first section, this study examines the relationship that exists between corporate culture metrics, and more particularly the complex links between employee happiness, leader appreciation, and corporate performance. The study seeks to understand if there are useful correlations to be observed between these variables and whether these are causal or appear due to other factors. The research aspires to contribute to the ongoing debate regarding the importance of various (indicative) non-financial indicators and their ability to predict corporate performance and, in turn, to provide insights about the effective utilization of constructed non-financial predictors to be applied in organizational/corporate decision-making processes.

In order to pay tribute to the several studies surrounding this ongoing debate, and in order to demonstrate the role of non-financial indicators in this endeavor, facilitating towards a comprehensive understanding of company behavior, a few indicative cases are presented in the following paragraph.

The authors of [34] introduce the balanced scorecard which is an approach that emphasizes the importance of non-financial indicators alongside financial ones. This creates a far more comprehensive understanding of what consists of organizational performance. On a similar note, the article of [35] discusses the potential of non-financial indicators, and more particularly the ones related to sustainability and environmental, social, and governance (ESG) factors, in order to predict corporate performance and drive strategic innovation. The work of [36] is investigating the relationship between corporate social responsibility (CSR) activities and financial performance, while [37] is researching the CSR in relation to employee satisfaction, and [38] is examining the relationship employee vs. satisfaction and stock market performance. Similarly, the authors of [39] explore what links customer satisfaction and financial performance. The references span across a timeline which is wider than 30 years, during which timeline any type of researchers have been contributing to the ongoing debate about the relevance of the

non-financial indicators and their success (ability) in predicting corporate performance.

2.1. Common key performance indicators

Conventional Key Performance Indicators (KPIs) are able to form valuable predictors of corporate performance. They often involve various financial metrics and ratios.

2.1.1. Most Popular Purely Financial Performance Indexes

The economic performance indexes that have a high popularity can be divided into three primary subsets: **financial performance**, **efficiency**, and **growth**. The lines to follow include a brief discussion of indexes belonging in that type, including literature reference(s) that support and justify their selection:

The most popular financial performance indexes are defined as profitability ratios and these are the Return on Assets (**RoA**), the Return on Equity (**RoE**), and the **Net Profit Margin**. These ratios are used because they provide insights into a corporation's ability to generate income in relation to its **assets**, **equity**, and **revenue**. The efficiency indexes are defined to measure the ability of a company to utilize its resources. Indicative examples are the asset turnover and inventory turnover ratios. Finally, growth indexes are defined to evaluate a company's expansion and they consider factors like the revenue growth rate and the Earnings per Share (**EpS**) growth. These three performance indexes will be more extensively discussed in paragraph 2.1.2 that follows.

The authors of the book found in [40] have studied a plethora of economic indexes and created an extensive overview discussing the importance of financial ratios and (mainly economic) performance indicators, stating their contribution when the aim is to evaluate a company's overall performance. They have followed a classification scheme which led to three main categories: financial performance, efficiency, and growth. Their intention has been to provide a comprehensive guide

in understanding the reasons that lead to using a variety of economic performance indexes, the common rationale that follows the definition of each, and the financial management and analysis capabilities that emerge from the usage of each one.

2.1.2. Other Popular Purely Financial Performance Indexes

What follows is a list of common financial performance indexes (mostly ratios):

a. **Return on Assets (RoA):** Net Income / Total Assets:

It compares a corporation's net income in respect to its total assets and stands as a measure of profitability. Companies with higher ROA may present better efficiency when using their assets to generate profit [41].

b. **Return on Equity (RoE):** Net Income / Shareholder's Equity:

It compares a corporation's net income in respect to its shareholders' equity and stands as a measure of profitability. Companies with higher ROE may present better efficiency in shareholder profit generation [42].

c. **Earnings per Share (EpS):** Net Income / Number of Outstanding Shares:

It is the portion of a company's profit that is allocated to each one of the outstanding shares of common stock and stands as a measure of profitability [43].

d. **Price-to-Earnings Ratio (P/E):** This index results by dividing the market price of a share by its earnings and stands as a valuation ratio. The calculated high values of P/E ratio are indicators of stock overpricing. Conversely, a calculated low P/E ratio may be an indication of undervaluation [44].

e. **Dividend Yield:** This is a financial ratio which calculates the annual dividend income per share in relation to the market price of that share. Corporations with higher dividend yield are usually more attractive investments [45].

f. **Current Ratio:** This index measures a company's ability to pay short-term and long-term obligations, so it stands as a liquidity ratio. When a current ratio is greater than 1 this suggests that the respective corporation has sufficient assets that can cover its liabilities [46].

g. **Debt-to-Equity Ratio:** This ratio compares a corporation's total debt in relation to its total shareholders' equity. In theory a lower debt-to-equity ratio also indicates a company that presents lower financial risk [47].

h. **Gross Margin:** This measures the profitability of a corporation. It is being calculated as the difference between sales revenue and the cost of goods sold, and it is expressed as the percentage of the sales revenue [48].

i. **Operating Margin:** This index compares a corporation's operating income in relation to its net sales and stands as a measure of profitability. Companies with higher operating margin present better managing and higher operating expense efficiency [49].

j. **Net Profit Margin:** This index compares the net income against the net sales of a company and stands as a measure of profitability. A corporation with a higher net profit margin most likely suggests faster and better generating profits and higher overall efficiency [50].

k. **Asset Turnover:** This index compares the net sales of a corporation in relation to its total assets and stands as a measure of efficiency. The corporations with higher asset turnovers indicate cases that exercise better utilization of assets and thriving when generating revenue [51].

l. **Inventory Turnover:** This index compares a corporation's cost of goods been sold to the corporation's average inventory and stands as a measure of efficiency. As expected, a higher inventory turnover value suggests a more appropriate inventory management [52].

m. **Quick Ratio (Acid-Test Ratio):** This index measures a corporation's ability to pay its current liabilities by exploiting its assets with higher liquidity (excluding its inventory) and stands as a liquidity performance ratio. A ratio that is greater than 1 usually suggests a healthy (at least for a short-term) financial position [53].

n. **Times Interest Earned (Interest Coverage Ratio):** This index measures a corporation's ability to cover the outstanding expenses of its interest using its

operating income and stands as a solvency performance ratio. As expected, a high interest coverage ratio indicates a healthy financial entity that services debt adequately [54].

o. **Total Shareholder Return (TSR):** This index is a measure of the total return on the shareholders' investment (including any capital gains and all dividend payments). The corporations with a high TSR indicate a better overall return for their investors [55].

The financial metrics can be quite indicative, especially when used in combination, of an entity's financial performance, however, the relevance of the selected metrics and their contribution weights to a predictor, can produce results that vary according to the applied context (like the nature of the industry, the corporation's cyclic changes), their business model, strategic objectives and more.

2.2.   Financial metrics and their potential

Although the financial metrics (indexes) are constructed in order to measure the performance of a corporation, there are a few of them which may, if the datasets are well refined, present correlations with non-financial metrics such as culture and employee happiness. It goes without saying that the relationships among such metrics may essentially be strong but not necessarily casual, and the strength of the correlation is dependent among context and other specificities. A complementary work to this thesis would be credible and quite suggestive if it could discover credible correlations between one of the following three financial metrics:

- Return on Assets (ROA): The corporations that present higher ROA are more efficient in asset investment and profit generation. A company performing well can invest more time and effort in their employees' development and subsequently to pragmatic workplace improvements, in order to foster a positive work environment and culture that leads to higher employee satisfaction [56].

- Return on Equity (ROE): The corporations that have a higher ROE may present higher performance and efficiency in financial terms, which translates to profits for shareholders. As expected, the entities with a strong financial position are able to prioritize their employee's well-being and to offer a supportive work environment, thus contributing to their employees' happiness [57].

- Net Profit Margin: The corporations have present higher net profit margins usually exercise their business model with an overall higher profit generation efficiency. Similarly, to the justifications of the other two metrics, such corporations can withhold the necessary resources to invest in welfare programs for their employees, thus creating a positive work environment, quality management, better processes and higher employee satisfaction.

This is, however, a very difficult task. If any potential correlations are identified in such a case, they need to be interpreted with extreme caution, there is always a chance that the relationship between any of such financial metrics and metrics that reflect employee happiness may not exist, or otherwise, if no correlations are observed, it is important to realize that they may not be straightforward. There are many factors, like the size of a company, the nature of the industry, even the various internal organizational policies that can influence (positively or negatively) the correlation degree.

2.3.   The basic corporation archetype

In order to fully grasp the organizational principles that govern a corporation, regardless of its size, industry, revenue levels or business model, it is useful to define its basic characteristics which are usually disguised as a set of commonalities amongst all corporations. This idea can be regarded as the introduction of a notion related to the corporation's archetypical behaviors. Given this conceptual setup, a corporation archetype can generally refer to what usually comes to mind as the traditional model of a for-profit company. Such a company primarily focuses on

shareholder value maximization and financial prosperity (returns). This archetype is setting the path that most companies are expected to operate surrounded by a capitalistic system, as they give emphasis on financial performance rendering it the primary measure of success. Fortunately, an increasing awareness is setting limitations to this model as there is a growing interest in extra human-centric approaches which tend to prioritize the stakeholders over the (financial) shareholders of a company.

The corporations are mostly centered around the sole pursuit of raising financial performance, with their primary goal being the maximization of a what their considered as a virtue, which is mostly measured in profits and, as expected, controlled according to the shareholder value. This focus on financial performance has led to a prioritization of short-term gains, potentially at the expense of long-term sustainability, ethical considerations, and broader stakeholder interests.

It is almost widely accepted that corporations are a direct product of the capitalist economic system, which is a consideration on a belief driven by the idea that the profit pursuit challenges competition and innovation. It is also accepted that capitalism has, so far, been quite successful in creating more wealth and has raised our living standards; however, capitalism has also been criticized, fairly or not per case but most probably correctly in the general case, for its contributions towards income inequality, global environmental degradation, and for nurturing a disregard around social and ethical concerns. In this context, the traditional idea around the corporation model is faced as a major contributor to the aforementioned issues.

Creating a response to the issues, limitations and constructive criticisms regarding the basic corporation archetype, the community has presented a growing interest in the human-centric approaches towards what is simply referred as "business". These approaches emphasize the significance of trying to consider a wide range of stakeholders (like employees, clients, suppliers) and most importantly, the environment. The Triple Bottom Line (TBL) mentioned in [58] can

be regarded as a good example approach (which focuses on social, environmental, and financial performance), the construct of B Corporations (which means businesses that create a balance between purpose and profit) [59], and additionally, the Stakeholder Theory as shown in [60] related to corporate governance, arguing that all corporations should, by definition, serve the interests of all stakeholders (and not just the shareholders, as already mentioned).

As far as in concerns the integration of human-centric approaches into the fundamental corporation archetype, it can as a proactive act, increase awareness and lead to a more appropriate alignment between a corporation's goals and our societal values. Subsequently this integration creates a stronger, more sustainable, and more equitable form of capitalism. This approach can also work as a deliberate shift in focus from the short-term strictly financial gains to the long-term creation of value (and virtue), by the inclusion of non-financial performance indicators that, when adopted, can reflect a broader range of interests for all people involved. Besides, it is quite straightforward to anticipate companies fostering growth and enhancing their long-term economic performance when they address the most prominent social and/or environmental challenges and when they increase their employee inclusion, by considering the well-being of their employees, clients, and affiliate communities.

So, in conclusion, although the basic corporation archetype has had a major effect in shaping what is today understandably featured as a corporate behavior nurtured by a capitalist system, by paying strong focus on terms of financial performance the growing interest of communities in the more human-centric approaches, can generate a shift that will potentially re-balance the way that most corporations operate. This can be achieved by placing greater emphasis on the broader social and environmental effects of such activities. When these approaches are incorporated the companies can align their strategic goals with commonly accepted societal values, to create a long-term value, and to contribute onto a more sustainable/equitable capitalistic market structure.

2.4.   The common properties of corporations

All corporations share many common properties which define their **structure**, **function**, and **role** as viewed inside the economic landscape. Some of these common properties include the following: Limited Liability [61], Legal Personhood [62], Ownership by Shareholders [63], Governance by a Board of Directors [64], and Profit Orientation [65]. The references that accompany them discuss a series of important aspects regarding corporate properties and also their implications within the broader frame that encompasses the economic and legal context. The definition and understanding of these properties can provide further insights into the nature of most corporations and, as already discussed, to explain their role in modern capitalism.

Limited Liability: it is defined as a fundamental principle in corporate law, which is intended to protect the personal assets of the shareholders of a corporation from being exploited in order to satisfy the debts and the obligations of the company. Limited liability, in essence, can ensure that the shareholders' financial responsibility will be restricted up to the amount that they have invested in the corporation. Following this principle, we can encourage any potential investments by reducing the risks that the shareholders may face. As a shareholder, confidently knowing that any personal assets (like savings, homes or even cars) will never be at the scope of financial risk and that there will be no legal liabilities, enables investing at a high level with minimal exposure. The constitution of limited liability is additionally fostering the global economic growth by facilitating capital flow into numerous businesses thus promoting any risky entrepreneurial activities because it allows all individuals to engage into educated risks without jeopardizing the entirety of their net worth.

Legal Personhood, that in literature is also widely known as corporate personhood, has to do with the concept where a corporation is regarded as a separate legal entity, i.e. apart from its owners, its shareholders, and its managers. As a legal entity, a corporation adopts many of the same rights and the same

obligations as an actual person, an individual. One example is related to the ability to enter as an entity into contracts, to act as a property owner, to be eligible to sue (and to be sued), and, never the less, to make transactions and pay taxes. Legal personhood is a very essential property that can facilitate the functions of corporations, as it enables them to operate independently and in an efficient way in the marketplace. This way ensures the separation of individual/personal affairs of people involved. Additionally, the status of legal personhood is enabling the corporations to last more than their founders and to continue existing even if or when the ownership or the top management changes, thus ensuring that the continuity and the stability of this business entity through time.

Ownership by Shareholders is a characteristic that defines the ownership status of corporations, when this ownership is being divided into numerous shares or units, namely stocks. The shareholder entities may be individuals, or institutions, or any other entities who can claim to own such stocks, thus they are granted a right over the corporation's assets and earnings. The shareholders of a company often have voting rights, which enables them to have a say and thus influence the corporation's future decisions, and they may also receive dividends for the case that the corporation is distributing its profits to its owners. As to the corporations' ownership structure, it actually allows for an efficient capital allocation, for risk diversification, and for liquidity rights given to shareholders, who can, if need be, buy and sell their shares in what is called "secondary markets".

Governance by a Board of Directors is another important aspect applied in the operation of corporations, since it provides a basic structure that can act as a decision-making and oversight platform. The election body ot the board of directors is the shareholders who are responsible for the protection of their interests, the setting of the strategic direction of their corporation, and the monitoring of the management team performance. The board consists, in the typical case, of both executive directors (people who are also part of the management team) and of independent or non-executive directors (who provide their external expertise and thus foster objectivity). This results to a structure that can help the members of the

corporation to maintain a balance of accountability vs. power, thus ensuring that the interests of all related members are considered.

Profit Orientation is a key feature of corporations as defined in the capitalistic economies. The primary objective of this is the generation of profits primarily for their shareholders. This a) pushes corporations to be efficient as a scheme which focuses on profitability and b) drives their competitive and innovative nature in a pursuit of numerous market opportunities. By maximizing their profits, the corporations can actually a) increase their share price and b) manage to provide instant returns in the form of capital revenue and dividends to their shareholders. Since the profit orientation has been almost the sole driving force behind the economic growth and the creation of wealth, it is being criticized for leading societies to find virtue in short-termism, and for neglecting all broader social and environmental concerns, maximizing income inequality.

2.4.1. The entrepreneurial nature of corporations

The title of this paragraph refers to the ability of corporations to innovate, to take risks, and to seize the emerging market opportunities, which are regarded as critical factors for business growth and financial success in competitive market environments. One way to analyze and to predict the entrepreneurial performance of a corporation is by the facilitation of culture metrics, which refer to measurements of the internal cultural environment of a corporation. These metrics include aspects like the leadership style, the values, the norms, and the behaviors that usually influence how an organization is functioning.

The recent literature research has shown that a number of specific cultural traits can inculcate the entrepreneurial spirit within the organizations, such as the openness to change, the adaptability, and the empowerment of employees. These important traits can create a human-friendly environment where employees feel encouraged to take educated risks, to generate new ideas, and to pursue innovation. The entrepreneurial nature of corporations will ultimately drive them towards growth and will foster their financial performance.

A renowned expert in literature on the scientific field of organizational culture, Edgar Schein, explores how the culture is shaping the behaviors and the attitudes of various people (either as individuals and/or in groups) within organizations. In his work in [66] he argues that a culture which is strong and supportive can serve as an important catalyst for enhancing the effects of proper administration in entrepreneurship, in innovation, and in its overall performance. He also supports the argument that the entrepreneurial spirit of corporations/organizations can flourish through the strategic management of the corporations' cultural environment.

2.4.2. Most Common Company Characteristics

Regarding company characteristics, the most common of them, like the **organizational structure**, the **leadership style**, and the **communication practices**, may product a significant effect on work ethics and on employee engagement. Work ethics is a term that refers to the set of moral principles and values that are guiding and encouraging employees to raise their creativity and performance. The work ethics are enabling and increasing the employee engagement (which is the emotional commitment/involvement of their labor/service to the object of their work) for their team/organization.

In general, a positive organizational culture can promote trust, can open communication, while immersing into collaboration principles that may foster strong work ethics combined with high levels of employee engagement. Conversely, a negative (toxic or unsupportive) organizational structure directly affects the work environment, leading to poor work ethics and employee disengagement, thus ultimately affecting overall company performance.

In the work of [67], Daniel Denison studies the raw relationship expression among the culture of a corporation, its organizational characteristics, and its performance outcomes. He contributes by highlighting the importance of assisting the alignment of specific corporate characteristics that enhance the employee work ethics and cherish their engagement in order to expect a positive business outcome.

Denison's research is emphasizing how significant it is to create a supportive and inclusive environment at work, which encourages ethical behavior and personal commitment, thus ultimately leading to more appropriate policies and organizational performance.

In a more recent literature study, found in the work of [68], Masa'deh et al. have explored the impact that various leadership styles and organizational cultures may have on widely accepted as important human-centered virtues like the sharing of knowledge, the performance in a job, and the respective firm's performance. The research results to findings that support the notion of cultivating positive company characteristics, like transformational leadership and enhancing any act originating from supporting culture, to positively influence work ethics and employee engagement. This attitude will in turn lead to improvements in the actual jobs and subsequently the performance of the respective firms. This work emphasizes the significance of creating and then supporting a positive work environment and explains how effective leadership can guarantee high levels of ethical behavior and thus employee engagement.

2.4.3. Idiosyncratic Differentiators among corporations

Idiosyncratic refers to unique **characteristics**, to **strategies**, or to **practices** that differentiate a company from its competitors. There are various aspects of such differentiators that encompass a plethora of aspects, like the style of management, the innovational capabilities, the organization structure, and other initiatives like Corporate Social Responsibility (CSR).

Organizational culture, as already mentioned, may have a key role in a corporation's competitiveness and success. Numerous distinctive cultures may endorse employee creativity, loyalty, and motivation, eventually leading to higher performance [69]. As regards **management styles**, like the servant style of leadership or its transformational version, may positively influence the employees' engagement and satisfaction, turning it into a company's competitive advantage [70].

Additionally, firms with **innovation capabilities** often display traits of great adaptability and resilience, which allows them to navigate through market uncertainties and to stay on top of most competitors [71].

Lastly, **CSR initiatives** can help corporations to walk on the paths of constructive differentiation from competitors, by stepping to address any (global or local) environmental, social, or governance (ESG) concerns, rendering the company as an entity that exercises responsible business practices [72], thus building an appealing brand that supports its stakeholders' growing demands.

2.5. Initial Public Offerings and their impact on financial index evaluations

Initial public offerings (IPOs) are created by a process where private companies are transitioning to publicly trading on some stock exchange. This is the event which allows corporations to accumulate capital by selling their shares (i.e. making their evaluation available) to the general public. These entities (IPOs) are useful to the extent that they provide various growth opportunities and increase liquidity for both the corporations and its shareholders. On the other hand, IPOs expose the firms to extra scrutiny and frequent regulatory audits. The decision to "go public" involves several factors, among which a vital one relates to the firm's financial health. The rest reflect the ever-changing market conditions, and the drive for additional capital to fund a corporation's growth, (or as the authors of [73] state, "the primary motivation for going public is to facilitate acquisitions").

Some quite recent studies on the IPO format have explored numerous aspects, like the role of venture capital (VC) backing [74], and the underpricing effects (i.e., the difference between the offering price and the first-day closing price) on the IPO performance [75]. Both of these studies contribute to a wider understanding regarding the factors that influence the IPO outcomes and that support the corporations into making more educated decisions about going public.

2.5.1. The impact of IPOs on various evaluations indexes

As it is quite anticipated that IPOs can, to some extent, impact the evaluation of financial performance indexes, the same is true for their correlation with work ethics and employee engagement, which can be both affected in several ways. For example, during the IPO process, corporations may focus on increasing their financial performance in order to attract numerous investors, which might to some extent, even only temporarily, overshadow the significance of the human-centered work ethics approach. Complementary to this, as the authors of [76] claim "*Despite the best efforts of policy makers and platform designers, there will surely be spectacular failures*" suggesting that the employee morale may get a hit. Additionally, transitioning to a public ownership can easily generate extra pressure to begin prioritizing short-term financial benefits over the long-term value creation (which may also lead to an increment of performance), subsequently potentially leading to a misalignment between the organization's goals and employees' motivation [77].

A quite recent study by C. Stone [78] is focused on the analysis of the relationship between employee engagement and firm performance, specifically in the context of IPOs. The two authors concluded that employee engagement is closely and positively associated with the post-IPO performance, which means that the success in maintaining a well-defended and unshakeable focus on work ethics and employee engagement it can crucially determine the long-term success of a newly going public corporation. Another, also quite recent study by G. Serdar in [79] researched the IPOs impact on CSR which is an idea that includes related aspects like the work environment and employee satisfaction. The authors of this work found that, although the respective IPO event may initially lead to a decrease of metrics regarding corporate social performance, this decrement is most often a temporary one, and companies in due course proceed and succeed in improving their focus, enhancing their attention around employee-related matters.

2.6.   Capital vs. Labor vs. Technology Intensive Schemas

The term **capital-intensive** refers to corporations that require great investments in physical (fixed) assets, like expensive machinery, heavy equipment, and wide infrastructure, in order to produce goods or services [80]. Such corporations typically present high fixed costs and low labor costs, as they mostly rely on the capital rather than the human labor.

In contrast, the term **labor-intensive** characterizes the dependence of corporations on human labor, where a higher proportion of the operating costs are being allocated directly to employee wages, benefits (and sometimes the overhead). For such cases, the human resources (HR) management can have a huge effect on the corporations' performance and, vice versa, as works like [81] suggest as they have proved that some macro-factors and to some extent have impact on the HR management.

Lastly, the term **technology-intensive** describes the corporations that primarily focus on creating (or developing, producing, marketing) technologically advanced products or services, and subsequently, are being characterized by high levels of research and development (R&D) spending to gain and maintain specialized knowledge [82]. Usually such knowledge, or know-how is distributed in structures and processes controlled by software, exotic hardware and high-tech tools.

2.6.1. The impact of each schema on various evaluations indexes

Each one of the aforementioned schemas can impact the evaluation of the corporation's performance indexes (financial or non-financial) and also affect their correlation with work ethics and employee engagement.

The capital-intensive firms usually can experience a lower labor productivity because of the nature of automation and mechanization, which leads to a weak correlation between employee engagement and the respective corporation's financial performance [80].

On the other hand, as already mentioned, the labor-intensive companies may present a stronger correlation between employee engagement and the respective corporation's financial performance because of its dependence on human labor [81].

Finally, the technology-intensive firms usually experience a more nuanced relationship between the employees' engagement and the respective corporation's financial performance. This is because the highly skilled employees are regarded essential for boosting innovation and fostering value creation, on the one hand, while on the other, the landscape where these organizations lay upon can easily be subjected to rapid changes and progress uncertainty [82].

As regards the main differences between the three corporate schemas (capital-intensive, labor-intensive, and technology-intensive) have to do with their primary reliance on different sources/inputs, which is capital, labor, or technology.

The capital-intensive corporations often present high fixed costs and low variable costs, while the labor-intensive companies present lower fixed costs and higher variable costs. The technology-intensive companies, lay in between as they are prioritizing innovation and thus usually experience high R&D costs but at the same time they benefit from the economies of scale (given that their products can rapidly and successfully reach the commercialization level) [82].

As regards the financial performance indexes, as expected, the capital-intensive corporations mostly focus on metrics like RoE and asset turnover ratios [80]. The labor-intensive firms are prioritizing labor (employee) productivity combined with cost efficiency (lean management) [81]. Finally, the technology-intensive corporations mostly emphasize the R&D spending level, the innovation rates, and the market share [82]. This means that differing priorities for each schema impact the evaluation of the metrics (financial and non-financial) and the role of the employee engagement under different dynamics.

In other words, the varying emphasis that is put upon capital, labor, and technology in every one of these schemas will most probably disturb the proper evaluation of indexes/metrics and their correlation with work ethics and employee

engagement. For example, and complementary to what has been discussed up to this point, in [83], Zhao et al. found that firms characterized as capital-intensive that follow policies which contribute to high employee engagement, show better financial performance, while the labor-intensive firms had produced mixed results. Similarly, in another study, conducted by Friesenbichler, K, et al. in [84], it was found that the impact of employee engagement on firm performance is more significant in firms that belong in labor-intensive industries. Finally, as far as it concerns the technology-intensive firms, the work of [85] confirms that the relationship employee engagement vs. financial performance can be mostly influenced by dominating factors like innovation capabilities and rapid pace of change.

2.7. Challenges in interpretation and inference

The interpretation of the correlation among the financial performance indexes and the work ethics and employee engagement may be challenging due to a number of factors:

Multidimensional nature of work ethics and engagement: The work ethics and the employee engagement are quite complex multidimensional constructs that make it difficult to establish concise, direct, and linear relationships with the financial performance indexes ([86]; [87]). There are many underlying factors, like the organizational culture, the leadership styles, and the job design, that can greatly influence the extent to which these employee morale factors translate into economic/financial performance.

Diverse financial performance indexes: Most corporations are quite often using a variety of financial performance indexes, like as RoA, RoE, and net profit margin, in order to assess their performance [88]. Each one of these three indexes can capture several different aspects of economic/financial performance, thus leading to the realization of potential inconsistencies that lay in the correlation with the work ethics and engagement of employees.

Confounding variables: There are several other variables, such as the industry, the firm size, and the market conditions that can influence both the economical/financial performance indexes and the employee morale factors (work ethics and employee engagement). In alignment to what the authors of [89] state, "*high levels of employee engagement in extra-role behaviors, regular employee attendance, and low voluntary turnover intentions based on affective commitment to an organization are likely to benefit performance outcomes across industries, business strategies, and cultures*". These are all factors that can potentially confound the existing correlation between the two groups, making it difficult for any decision supporting tool to accurately assess their relationship.

Temporal dynamics: Additionally, any relationships between the work ethics, the employee engagement, and the economic/financial performance indexes may change over time since companies evolve and thus adapt to the changing market conditions and societal expectations [90]. This is why temporal complexity can challenge the establishment of a stable, long-term correlation between otherwise strongly related variables.

Measuring work ethics and engagement: The accurate measurements of the work ethics and employee engagement can be quite challenging due to the subjective nature of such constructs. There are several methods, like the self-report surveys, the supervisor evaluations, and the behavioral observations, that are used to assess such factors that potentially lead to discrepancies in results [91], [92]. These measurements are challenging and can affect the validity of the existing correlations between the economical/financial performance indexes and the employee morale factors.

Directionality of the relationship: In some cases, it may actually be difficult to determine if a strong correlation between economic/financial performance indexes and employee morale factors is due to the employees' work ethics and engagement that drive the economic/financial performance or vice versa [93]. This is why disentangling any causal relationship that exists between variables can be quite a

challenging task that requires caution and to some extent experience, since they may be variables that are mutually reinforcing, with the financial success motivating the employees and the strong work ethics, in turn, contributing to a better economic/financial performance.

Cultural and contextual factors: Any cultural differences and other contextual factors may easily affect the existent correlation between economic/financial performance indexes and the employee morale factors. As an example, the employee morale factors may in some cases play a more significant role in economic/financial performance in specific collectivist cultures that actually value teamwork and the cohesion of groups [94]. Additionally, it is important to note that the industry's nature, like the degree of competition and related regulation, may to some extent, influence the strength and the direction of this relationship, too [95].

Non-linear relationships: The relationships that exist between economic/financial performance indexes and the employee morale factors may not be linear, with possible diminishing returns or unexpected threshold effects [96]. So, it is plausible that such a non-linearity will complicate the interpretations emerging from such correlations. Such cases require the use of advanced tools and special statistical techniques in order to model and understand the challenges of so peculiar underlying relationships.

Impact of external factors: The existing correlations between the economic/financial performance indexes and employee morale factors may be influenced by several external factors, like economic conditions, global or local political (in)stability, and various market trends, which can either independently or collaboratively affect a company's economic/financial performance [97]. These external influencing factors may incommode the isolation of the direct impact of employee morale factors on economic/financial performance indexes.

Heterogeneity among companies: Corporations vary in size, in industry, in organizational structure, and in management practices, which can lead to various differences in the way employee morale factors actually affect their

economic/financial performance [98]. Thus, just a single correlation cannot accurately capture the strength of any existing relationships between such variables across so diverse contexts and it may require a tailored analysis (pre-selection of subcategories) for specific company types.

Time lags and dynamic relationships: The relationship among economic/financial performance indexes and employee morale factors may not be instantaneous and may include time lags that affect the observed correlation [99]. Additionally, any relationship between such variables may be to some extent dynamic and actually change over time, since companies adapt their practices, and the context of ethics, morale and engagement evolve.

Multiple dimensions of work ethics and engagement: The employee morale factors encompass various dimensions, such as the employees' job satisfaction, their organizational commitment, and many other work-related values [100]. These multiple features (dimensions) may have different effects on the economical/financial performance indexes and subsequently, not accounting for their existing distinct impacts per scenario (depending on what needs to be asked, predicted and/or interpreted) may easily lead to some oversimplified interpretations about the correlation between employee morale factors and economic/financial performance.

Competing theories: The are different theoretical perspectives that may offer various alternative explanations for the existing relationship between economic/financial performance indexes and employee morale factors [101]. For example, the resource-based or human capital theories are emphasizing on the role of the employees' skills and technical knowledge, while the social exchange or psychological contract theories mostly focus on the quality (level of balance) expressed by the employee-employer relationships. These are competing theories which can lead to differentiated predictions and to interpretations of the same observed correlation, which requires careful consideration, especially when trying to properly and justifiably evaluate the relationships between such variables.

2.8.   The respective algorithms

In order to explore the potential correlation between a corporation's economical/financial performance and employee morale factors, there are various statistical and machine learning (ML) algorithms that can be utilized in order to identify the various relationships and underlaying patterns within the given data. In the lines to follow we include the most common algorithms for this purpose:

Multiple Linear Regression: This is a widely used statistical method that is able to establish the potential relationships that exist **between multiple (seemingly) independent variables** and a single (carefully chosen) dependent variable. This is a method that can be applied to examine the impact of employee happiness/morale/engagement and other factors may have on the respective company's economical/financial performance [99].

Partial Least Squares (PLS) Regression: This is a multivariate statistical technique which **is able to handle multiple predictor and response variables**. This technique is particularly useful when the researchers are dealing with multicollinearity issues in their dataset [102]. The PLS is a method that can be employed, per case, to gradually analyze (when exploring various parts of the dataset) the potential relationships that may exist between economic/financial performance and the employee morale factors while accounting for other factors, like the organizational culture or the leadership styles.

Support Vector Machines (SVM): This is a machine learning algorithm which is capable of performing **the functions of classification, regression, and outlier detection**. SVM can be utilized in order to model any nonlinear relationships that may exist between the economic/financial performance and the employee sub-dimensions in their reviews about happiness, as well as other factors [103].

Random Forest: This is **an ensemble learning method**, i.e., a suite of distinct tree models, which combines them in multiple ways to improve their predictive performance. Random Forest can be used in order to assess the importance of

several factors, including employee morale ones, when predicting the possibility that this can affect economic/financial performance [104].

Hierarchical Linear Modeling (HLM): This is a statistical technique that is able to account for **nested data structures**, such as employees within a company. The HLM technique is useful in understanding the impact of the individual-level employee morale factors on a company-level economic/financial performance [105].

Artificial Neural Networks (ANN): The ANN is an ML technique which is inspired by the structure and **the functioning of biological neural networks**. This method can be used to model complex, non-linear relationships like the ones that can potential exist between economical/financial performance of a corporation and its humanitarian values (work ethics) or, as mentioned, company culture, expressed, as always, as quantifiable employee morale factors [106].

Structural Equation Modeling (SEM): This is a **multivariate statistical method** which can analyze the potential causal relationships between latent constructs (like employee morale factors) and their impact on a company's interesting dimension (like economical/financial performance). As the authors of [107] state "*with SEM, we can examine relationships between latent constructs much as we examined the relationships between dependent and independent variables in multiple regression analysis*").

Bayesian Network: The Bayesian Networks are **probabilistic graphical models** which represent the existing dependencies among multiple variables. These networks can be used in our dataset in order to understand any causal relationships that may exist between employee morale factors, and economical/financial performance, by taking into account the potential uncertainty in the data [108].

3. Company Culture Metrics

3.1.   Definitions of ethics per discipline

There are various disciplines across which we can define ethics in the corporate landscape. A few of such definitions are given in the lines to follow:

Business Ethics: This term regards principles and standards which guide the behavior of entities like individuals and/or organizations when conducting business matters. It covers issues mostly related to corporate governance, to insider trading, to bribery, to discrimination, to corporate social responsibility, and to legal responsibilities [109].

Organizational Ethics: This term regards the study of moral/ethical values, on principles, and on standards of an organization thus guiding decision-making and expediting proper conduct among its members. Organization ethics is an area mostly concerned with the way that organizations are developing and maintaining the ethical culture which encourages personal ethical behavior [110].

Human Resource (HR) Ethics: This term regards the ethical aspects related to the management of human capital within corporations/organizations, including functions like the recruitment, the selection, the compensation, the performance management, the diversity and inclusion, and the employee well-being. HR ethics takes special care in ensuring the fair and equitable treatment of all employees, and in promoting the ethical behavior among all of them [111].

Environmental Ethics: In the corporate context this term is concerned with the ethical obligations that businesses have towards the environment. This actually includes practices like resource conservation, pollution reduction, waste management, and any other related sustainable practices. As a conduct of corporate acting for the environment, it addresses the moral relationship between humans and the environment (i.e., and other humans), emphasizing the very important responsibilities of organizations in order to protect and preserve an adequate level

of provision, when designing/implementing company related matters in any area of their business (marketing, procurement, development or even human resources), to supports its relations, as a collective entity, to its immediate environment specifically and to future generation in general [112].

Marketing Ethics: This term concerns the moral principles that need to guide the marketing activities of corporations/organizations, including the advertising, the promotions, the pricing, the product development, and the distribution. It also examines issues like deceptive advertising or manipulation of consumer behavior, including the use of customer data in a responsible and ethical manner [113].

Financial Ethics: This term addresses the moral principles and the values that underpin the economical/financial decision-making processes in corporations / organizations. Financial ethics covers a wide range of important topics, related to corporate governance, the financial reporting, the risk management, and the investment practices. It is also considering what constitutes the typical ethical behavior of various financial professionals, like the accountants, the auditors, and the financial analysts [114].

Supply Chain Ethics: This term focuses on the ethical aspects of the management of the entire process of sourcing, of producing, and of distributing goods and services. It encompasses various issues such as the labor rights, the fair trade, the environmental impacts, and the responsible sourcing. This discipline encourages corporations/organizations to consider the potential ethical implications that may exist in their supply chain decisions and encourages them to strive for a responsible stand that favors a sustainable global economy [115].

Information Technology (IT) Ethics: This term refers to the ethical aspects of using (exploiting) and managing the information systems, the data, and the technology within organizations. This discipline is exploring the moral capping stone protecting any principles and guidelines that are related to privacy, to security, to intellectual property, and to the digital divide. It also considers the

ethical implications from any emerging technologies, like automation and artificial intelligence [116].

All the aforementioned disciplines are highlighting the wide range of ethical considerations that the corporations/organizations should address in a variety of aspects related to their operations. These disciplines offer many different perspectives on what is called "ethics" in the corporate environment. This way they are demonstrating the complexity and the variety of the ethical considerations that organizations must navigate through. Each of these disciplines depicts a unique perspective on the ethical challenges and the responsibilities that the companies are facing in today's complex and vastly interconnected business landscape.

The work of this thesis directly and indirectly addresses almost all the disciplines found in this section. However, it is mostly focusing on work ethics as viewed from the scope of 'Business', 'Organizational' and 'Human Resources' and 'Financial' ethics.

3.2. Impact of work ethics and social movement

The impact of work ethics and social movements in the corporate environment is quite significant, since these factors can influence the way that organizations are conducting business and are interacting with various stakeholders. The principles of work ethics are shaping the behavior of the employees and the overall organizational culture, while social movements can drive the change of business practices and can promote much greater corporate social responsibility.

The existing relationship between work ethics, the social movements, and the corporate environment is considered complex, and its approaches are quite multifaceted. When we delve deeper into the subject, we can productively understand the strength of the dynamics at play and the nature of the implications that such relationships pose on businesses and the respective stakeholders.

The term of work ethics is commonly accepted to include the moral principles and the values that guide (the majority of) the employees in their professional

conduct at their lives. So, the notion of well-established and strong work ethics can, itself, positively impact the environment of corporations by enabling and fostering the trust, the cooperation, and the commitment that is so much required to expedite excellence for all employees. The corporations that are prioritizing the ethical behavior and are encouraging the employees to adhere to various ethical standards are the ones that are also more likely to experience the highest increment of the anticipated employee satisfaction, the productivity, and the organizational / personal success. This is also projected to go beyond the usual employment practices to the case of Moral Development of Society. As the authors of [117] state "*A new norm is created because, as discussed above, once a better ethical norm is followed, it contributes to a better society*".

The strong work ethics can be manifested in several ways. These include both personal and professional conduct, as they complement each other and are mainly attributed to traits like honesty, responsibility, integrity, and diligence. A company that is prioritizing ethical behavior is actually preparing an environment where the leading employees will feel more empowered to express themselves, to speak about issues and concerns around ethical matters, will decide upon what aligns their behavior and duties to the organization's values. This way of thinking and acting creates a trusting culture, enhances transparency, and turns into higher levels of employee engagement, employee retention, and overall productivity performance [118].

The uprising social movements begin to have a profound and impactful effect on the corporate environment as they raise the awareness around modern social, economic, and environmental issues which, in turn, affect business matters and subsequently their stake- (and share-) holders. The social movements usually call for enhancements in corporate accountability, in transparency and in sustainability which pushes the corporations/organizations to proceed in reevaluation of their business policies and practices and to address their social (branding) and environmental impact. Most companies respond in a proactive way which subsequently strengthens their reputation, either through marketing, or employee

online reviews or even by word-of-mouth. This enhances the trust with the shareholders and in turn enhances the long-term value creations of the whole organizational entity [119].

Most of the social movements are emerging in immediate response to important societal issues or injustices, trending in a fashion that can mobilize public opinion and proceed the advocating for change. In most of the corporate environments, the social movements have the strength to influence the respective businesses by spotlighting all outstanding pressing issues, like climate change, income inequality, and racial injustice. The employees of such corporations are engaging with these movements in order to push for change, to make these entities socially responsible by incorporating their demands, aligning them with their strategies and operations. The corporations that are highly responsive, manage to mitigate potential risks, enable the capitalization of new opportunities, discover new sustainable ways of practicing HR management, quality and process management which subsequently fosters an environment which encourages the development of sustainable products or the entering new markets [120].

So, summarizing, work ethics and the social movements do play a critical role in shaping both specific and generic dimensions in a corporate environment. The bunch of the companies that are prioritizing ethical behavior and that respond promptly and positively to raising social movements can benefit by experiencing numerous changes, including an improved view on their stakeholder relationships, an increased and more authentic employee engagement, and subsequently an enhancement towards the practices that foster long-term value creation.

As to the interplay between the virtues of work ethics, the awareness of social movements, and business making in the corporate environments, there is a debate which is highly relevant in today's markets landscape. The corporations / organizations that are prioritizing the ethical conduct and that are engaging with the social movements seem to succeed in achieving a competitive advantage while strengthening their online (but also offline) reputation, building real trust with a

plethora of stakeholder types, and fostering an encouraging culture that expedites acts of innovation, engagement, and adaptability. Such companies contribute to a sustainable and equitable society by actively addressing outstanding ethical concerns and by embracing several demands project by social movements.

3.3.   Responsible practices and political stands

The role of responsible practices and political stands within the corporate environment is constantly gaining attention in recent years. As all businesses begin to recognize the importance of the social and the environmental responsibility, they are also beginning to take more active stances across various political issues and to incorporate several responsible policies and practices into their operations.

Such responsible practices are encompassing a very wide range of various activities which demonstrate a corporation's commitment to the social and the environmental sustainability which include a) ethical sourcing, b) fair labor standards, and c) reducing of the operations with great environmental impact. The implementation of responsible practices is able to enhance a corporation's reputation, to foster positive relationships with all stakeholders, and to contribute to an aspiring, long-term business success. The research work found in [121] can offer a comprehensive (and also an up-to-date) analysis of the corporate social responsibility (CSR) idea and to further explain this with several related concepts which include various responsible practices. In this work its authors mention and also take special care to explain various aspects of what is called "responsible practices", like the aforementioned ones (i.e., the ethical sourcing, the fair labor standards, and the environmental sustainability) while stressing the importance of such terms (notions) in a global business context. By the adaptation of such responsible practices, the corporations are able to enhance their reputation, to build positive relationships with the stakeholders, and to contribute to the long-term success of businesses.

As the time passes the corporations are gradually taking political stands on various environmental and social matters, like the climate change, the LGBTQ+ rights (identity politics) and the issues emerging from racial inequality. The businesses can, through their engagement with various political issues, shape the public discourse and influence policies and/or demonstrate their (in some cases full) commitment to CSR. The authors of [122] suggest that this attitude can a) strengthen the brand of the respective corporations, b) can foster the loyalty of the customers and c) encourage the existing or push new ones to prioritize ethical consumption.

There is a clear intersection between responsible practices and various political stands (also expressed in the corporate environment) which is specifically existent because it is reflecting a growing emphasis on the actual role of the newly-created businesses in addressing the challenges related to the environmental and social matters. The companies that are actually creating some valuable service or product for their stakeholders are the ones that integrate responsible policies and practices into their operations – mostly, as already discussed – by officially taking meaning political stances in order to contribute to a just and sustainable society. This is also a beneficial stance from a practical and economical point of view as it enhances competitiveness and resilience.

3.4.   Code of Conduct in the workplace

The so called "Code of Conduct" in the workplace is a very well structured set of guidelines and principles (usually set by the Quality Monitor and Control department of a company) that outlines the behavior to be expected by the employees and the management for all cases and events taking place in the professional environment. It is usually comprised in a series of rules and/or examples that serve as a reference point for the employees to understand their duties, their responsibilities, their rights, and the common standards that they all need to maintain (and protect) at work. The "Code of Conduct" usually covers topics

like the definitions of ethical behavior, the workplace diversity and inclusion, the cases of harassment and discrimination, the need for confidentiality, and the various conflicts of interest.

The authors of the [123] article stress the importance and the implications that the "Code of Conduct" brings in the workplace. Their work highlights the high significance, for a company, to create and maintain a strong stance on ethical corporate culture, as this includes the development and the implementation of a comprehensive "Code of Conduct" for all employees, originating from a single source of origin and a well-respected figure of authority that allows for the alignment of all stakeholders. The authors identify the core elements of what constitutes an ethical corporate culture and provide useful insights into the paths that lead towards the effective communication and the reinforcement of the "Code of Conduct" within a corporation/organization.

The authors of the study of [124] explain that the "Code of Conduct" role in fostering psychological safety, learning behavior, and overall performance in teams. Their findings are complementing the wok of [123] as they suggest that a well-prepared, well-communicated and properly enforced "Code of Conduct" can create a workplace environment that will be enjoyable and rightful for all stakeholders which, in turn, encourages the openness in communication, the drives for collaboration, and the continuous employee joy in curiosity and learning, that is ultimately contributing to an enhanced team performance.

Although the direct correlation between the economical/financial performance metrics and a well-prepared, well-designed "Code of Conduct" may be difficult to communicate and establish, there is sufficient evidence that suggesting this kind of mindset to corporations/organizations, enriched with strong positions regarding the ethical cultures and the adherence to a "Code of Conduct" tend to produce better results and an increased overall performance, including economical/financial performance.

In the work that the authors of [125] examine, there lays an investigation about the potential relationship between the organizational ethical culture (which includes a well-prepared, well-designed "Code of Conduct"), CSR, and economical/financial performance. The findings of the authors indicate that the corporations/organizations that pose a strong ethical culture stance, actually tend to present a better economical/financial performance, which is partially due to the company's commitment to CSR.

The research of the article [126] is examining the potential relationship among the corporate ethical values (incl. the role of a "Code of Conduct"), CSR, the state control, and the economical/financial performance, specifically for Chinese companies. The research reaches to the conclusion that the corporate ethical values are usually positively associated with economical/financial performance, which suggests that a well-prepared, well-designed "Code of Conduct", can actually produce the anticipated results by contributing to better economical/financial outcome, when regarded as a part of a broader ethical culture.

3.5. Employee morale in workplace

One quite important aspect of the corporate landscape is what is widely defined as "employee morale", which is a common term that has been extensively used in the literature and mentioned for a plethora of time in the previous paragraphs. The action following the employee morale (and its factors) can significantly impact several aspects of corporate/organizational performance, like productivity and/or employee retention, and subsequently the overall success of a company. It has also been documented that the high levels of employee morale often lead to increased satisfaction, to improved work quality, and to better teamwork. Conversely, a low employee morale level can result in an increased turnover, reduced productivity, and negative work environment.

In the study of [127] the authors investigate the potential relationships among person vs. organizational fit vs. job satisfaction vs. work engagement, and employee

morale (independent variable). The authors of the study find that when the employees perceive a high ranking (degree) of fitness with their organization, then the employees are experiencing a higher level of job satisfaction and of work engagement, which subsequently, leads to the improved morale of the employee. The authors also discuss about the terms called "transformational leadership", about "Work engagement", about "Personal-job fitness", etc., and they conclude (in a discussion focused on practical implications), that, especially for the non-western contexts: "*It would benefit organizations if they introduce development programmes to build transformational leadership capability amongst their leadership population and embed them in practice through related HR agendas.*"

Finally, in the article of [128] the authors explore the way that corporations / organizations build any resilience and maintain their employees' morale and engagement, especially during times of challenges, where exists a volatile, uncertain, quite complex, and ambiguous (VUCA) world. The authors of this work are proposing that the corporations/organizations can, to some extent, enhance the morale and the employee engagement by creating and nurturing a brand-new culture which will be able to promote the psychological safety, encourage continuous learning, and the adaptation to change.

3.6.   Multi-sided Policies

The term of "multi-sided policies" in the context of the corporate environment refers to the development and the implementation of a set of strategic plans and various initiatives that are designed to meet the needs and the concerns of multiple stakeholders. Such policies' goal is to balance the needs and the interests of a variety of parties, like employees, like customers, like investors, like suppliers, and like the broader community, ultimately cultivating a sense of mutual collaboration, of trust, and of shared value, thus contributing to a sustainable and a responsible business form. There are various practices and policies which can be observed in a

variety of aspects in corporate operations, like the decision-making, the resource allocation, and the strategic planning.

In the context of the decision-making, the various multi-sided policies encourage the organizations to carefully consider the implications of the choices that are applied as relevant to all stakeholders. It is important to note that this has to take into account the economic, social, and environmental impacts of all corporate decisions, as well as weighing any potential benefits (and also risks) for each one of the groups involved.

The proper resource allocation is an extra area where any multi-sided policies can easily play a crucial role. The corporations can guarantee that their business operations will create great value not only for their shareholders but also for the stakeholders (employees, customers, suppliers) and the respective community of the industry and beyond, as long as they distribute their resources equitably and among various groups. The approach is essential as it can increase the company's brand and reputation, it can strengthen the relationships with an extensive set of related people and professions and can promote actions of long-term sustainability.

The strategic planning that follows up the aforementioned work, incorporates several multi-sided policies that require a comprehensive understanding (sometimes passed by the experienced board members to the rest of the company's governance) of the position that the corporation/organization holds within its broader ecosystem. The companies need to understand how to identify their valuable stakeholders, to properly evaluate their needs, to project and to meet their future expectations, and to develop strategies that will facilitate towards the creation of mutually beneficial outcomes. This may actually involve forming new partnerships and/or investing in social and environmental initiatives and/or adopting various innovative business models.

In recent years, the raise of advanced **multi-sided platforms** (MSPs) has emphasized the vital importance of the multi-sided policies. The MSPs can facilitate the various transactions and the interactions between different kinds of user

groups, often through innovative digital technology. Some examples include the ride-sharing services, the e-commerce marketplaces, and the social media platforms. In order to succeed, the MSPs must try to strike a balance among the various interests of differentiated parties, thus ensuring that the given platform will remain attractive to both the providers and the consumers.

In overall, the multi-sided policies in the corporate environment are quite essential for creating extra shared value, for fostering trust, and for promoting any sustainable business practices. Corporations are able to achieve a long-term success and to provide an equitable and responsible capitalism form when they push policies (or even regulations) that will be considering the interests of all stakeholders, at all actions and at all times.

The study of the work in [129] is presenting a comprehensive review based on the business model of the multi-sided platforms. It explores how corporations / organizations are able to successfully balance the interests of a variety of stakeholders while also maintaining a sustainable and competitive advantage.

The research work in [130] has examined the strategic options that are available to the multi-sided platform firms, highlighting how important it is to nurture strong relationships with a set of diverse stakeholders and to create value for all involved parties.

Another article, as presented in [131], has focused on the investigation of the strategies that are characterized by competition and corporation (as adopted by various multi-sided platform firms). The article emphasized the need to balance both approaches in order to achieve its goal and to be regarded as a successful corporation in a (quite rapidly) changing business environment.

The research of the work found in [132] has explored the specificities (to some extent) role of the multi-sided policies, as expressed, for the dematerialization platforms. In more detail it focused on the effect that various platform firms had in (re)shaping the existence of tight-interdependencies between users' activities at project level. The authors of this work argued that "*an organization may have to deal*

with interdependent affiliation costs that directly depend on other users' adoption decisions.".

3.6.1. Fairness

The term fairness as a part of the "multi-sided policies" refers to the need for equitable treatment among every one of the several stakeholders comprising and framing the corporate environment. The term refers to the balanced attribution of related but 'antagonistic' properties like the benefits, the risks, and the responsibilities among the members of different groups, like the employees, the customers, the suppliers, the investors, and the broader community. Given what has already been discussed in the previous paragraph, it is understandable that it is of vital importance to guarantee that fairness in "multi-sided policies" can actually lead to more sustainable policies and practices in business, to enhanced stakeholder relationships, and to improved overall corporate reputation.

Transparency is one of the aspects (sub-dimensions) of the "multi-sided policies". Corporations need to discuss their practices and policies, their objectives, and their decision-making processes in an open discussion, that allows all stakeholders to comprehend the logic (the rationale) behind their reasons and drives for acting in a specific manner and to hold them (ethically and preferable legally, too) **accountable** for their choices (acts). Such a transparency will be able to enhance trust, to promote in-team collaboration, and to mitigate any potential conflicts among the various stakeholders [133].

Another component of the "multi-sided policies" which seems to be a quite critical sub-dimension of fairness is **inclusiveness.** This term refers to the stakeholders' participation in the decision-making processes of the workplace and any related environment. This is also vital as the corporations need to actively engage with their stakeholders, for consolidating and soliciting helpful feedback, and for incorporating a series of diverse perspectives in order to co-formulate their strategic planning. It is important to understand that inclusiveness is an important sub-dimension and an important part of our human nature in general. The

facilitation of its proper application can (and must) result in an enjoyable and educated, thus also effective, decision-making, and foster a sense of ownership and personal commitment among the various stakeholders [134].

Additionally, the notion of fairness in multi-sided policies also involves the idea of **balanced distribution of resources and benefits**. This means that it can entail the guarantee of fairness in wage attribution and of the proper working conditions for the employees, thus offering (the anticipated by all) equitable pricing and (the anticipated by all) quality products for customers. Also, it guarantees the maintenance of responsible and ethical relationships with all suppliers. The corporations are encouraged to demonstrate their commitment to all responsible business practices and to long-term sustainability by giving priority to a fair treatment to all related stakeholders.

3.6.2. Flexibility

The evolution of the various needs (requirements) and expectations of the various stakeholders of a corporation asks for the utmost responsiveness from the latter, which translates to the notion of **flexibility** (another sub-dimension of the "multi-sided policies") which relates to a company's adaptability to create (innovate) or adapt new policies and processes that cover and/or favor the former. The embracement of flexibility can help corporations/organizations to apply (or navigate through) changes in its business adventure (environment), while successfully addressing the concerns of the various stakeholders in an effective way, thus fostering a culture that favors improvement and innovation.

An important key aspect that allows for all of this to be actionable and effective is the embracement of **organization learning**. The corporations that are gathering useful feedback from its employees (or stakeholders in general) and are monitoring the outcomes of their (quality) processes' application are using this feedback as information that can refine (and even improve) their decision-making. The orientation (focus) of companies towards adaptation (learning) in a procedural manner (organizational) helps them identify any useful potential challenges right

from the get-go, and thus to properly adjust their policy making accordingly. In turn, this drives various ongoing improvements enhancing their performance metrics [135].

In addition to the aforementioned aspects, there is one more important sub-dimension of flexibility which usually involves the idea of nurturing a new culture of **adaptability** within the corporation/organization. Paraphrasing what has already been discussed in the previous paragraph, this time focusing on the complementary idea of flexibility, the idea of adaptability entails the directed encouragement towards stakeholders to embrace (and not be afraid of something new, something unknown or even risky), and in other words, to embrace "change", as long as they are provided with the necessary knowledge, know-how and resources (i.e., training) to adapt (seamlessly and effortlessly) to new processes / policies. Subsequently, this cultivation leads to the creation of a supportive environment that values the diversity, the new perspectives and, in general, values innovation. By cultivating the culture of adaptability, corporations can both manage their stakeholder needs in an effective way and also enhance their resilience (and competitiveness) in overall, regardless of the ever-changing and challenging market conditions [136].

To conclude, the flexibility in "multi-sided policies" plays is an extremely crucial tool towards the empowerment of the corporations to adjust their business model to the evolving requirements of their employees, in order to tackle emerging matters, and in order to promote ongoing improvement in their processes, always aspiring for high performance. The embracement of "organizational learning" and by cultivating a culture of "adaptability", almost all companies will be able to improve their "multi-sided policies" remaining responsive and effective in the ever-changing and challenging markets.

3.6.3. Meritocracy

The virtue of **meritocracy** as a sub-dimension of "multi-sided policies" refers to the common practice of promoting individuals by rewarding them within a corporation/organization according to their skills, their abilities, and their performance, rather than other (irrelevant to their competency) factors like personal connections, networking, or seniority. The corporations that embrace and follow meritocratic principles by properly designing their processes/policies, can generate enjoyable working conditions (i.e., a fair and inclusive workplace). Subsequently, they can improve the employee motivation and the labor engagement, leading to the increment of the company's overall performance and competitiveness.

There is an aspect of meritocracy in the "multi-sided policies" which involves establishing **transparent performance evaluation** systems which allow employees to understand the various criteria by which they are assessed and the related opportunities that their actions can unveil regarding their professional growth within the corporation/organization. The enforced **transparency** in that front (and any contractual agreements that include such an approach) can benefit both parties since it contributes to a sense of fairness. This transparency reduces the false perceptions of bias or favoritism, and thereby it is promoting a more inclusive and vastly more collaborative workplace culture [137].

One last sub-dimension of meritocracy involves the **provision for equal opportunities (for all employees)** in order to take advantage of their existent skills and to develop extra ones, to have access resources, and to advance in their careers preferably according to their wishes. This provisional way of approaching equal employee opportunities may imply the inclusion of diversity implementations and similar inclusion initiatives, by the offerings of specialized mentoring and/or training programs, and by ensuring several promotion and reward systems in a fair and unbiased (digital) tool/platform. By nurturing an environment where all of the employees have the chance to succeed, under even rules, and solely based on their merits, the interested corporations can attract and also retain top talents and

eventually even redeem their investment through the increments of innovational drive and subsequently of economical/financial performance [138].

As a result, the incorporation of meritocratic values in the "multi-sided policies" can be regarded as a platform that enables companies (through the creation of fair and inclusive work environments), to promote the engagement and the motivation of their employees, and to eventually reach a level where the enhancement of their overall performance is organically re-created (expanded). The organizations can guarantee that their processes/practices/policies are not only fair but also rich and effective by definition (by creating and applying proper and transparent evaluation tools).

3.7.   Asymmetrical Transactions

The asymmetrical transactions (as given in the context of the corporate environment) are referring to naturally emerging situations where one of the parties that are being involved in the transaction, has more power and/or information and/or resources than the other party(ies), thus leading to a systemic imbalance in this newly formed relationship. This kind of asymmetries can quite easily occur in a variety of contexts, like the interactions between the employers and the employees, the businesses, and their customers, or even between businesses found in differentiated positions (and interests) within business setup (e.g., a supply chain). Asymmetrical transactions may pose great implications on the fairness, the efficiency, and the sustainability of business operations, and subsequently, to the overall market dynamics and, in turn, the consumer welfare.

3.7.1. Informational Asymmetry

There is a notable manifestation regarding asymmetrical transactions and this is the vital concept of **information asymmetry**, where one of the parties has special access to a broader, more recent or of higher quality information that the other(s). This asymmetry can be easily formed and can also easily drive things to adverse

selection or to serious moral hazard problems. Such a development may affect the quality (or even the direction of the accountability weight) of decision-making and can potentially generate vast market failures [139]. The corporations/organizations need to address the occurrence (or risk of occurrence) of potential information asymmetry by nurturing transparency, e.g., by applying legal actions, by managing the creation and circulation of disclosure agreements, by adopting standardized reporting practices, or by utilizing digital technology in order to encrypt/decrypt and monitor the edits of open and shareable information more efficiently [140].

There is another important aspect regarding asymmetrical transactions which directly relates to the **existing imbalance of power (or resources) between parties**. The emergence of such imbalances can, as already mentioned, quite easily result, if tolerated, in unfair results due to a) the unequal bargaining positions and b) due to the unfair act of exercising intended exploitation. For example, some large corporations may exert their power in the market fields in order to negotiate for favorable deals with suppliers that are of common reach to their competitors or can easily dictate special (unfavorable and unfair) terms towards the smaller businesses, thus

    a) leading to polarizing effects and

    b) gradually reducing the welfare of the market, in total.

Most of the policymakers and regulators are well aware of such issues and try to mitigate these risks through the implementation of tough "antitrust and unfair competition" laws, in order to encourage and promote fair trading practices, and to support the development of ethical and honest cooperative structures which will allow to the smaller players to strengthen their position and their bargaining power [141].

In more detail, the existence of asymmetry in the information access and/or ownership among parties can significantly affect the interpretation and the analysis of several performance indices in the corporate landscape either between employees or businesses. As we have already mentioned, the information

asymmetry is higher when a single party possesses more, faster and/or better-quality information than the other(s), which will most certainly lead to skewed results, and thus false decision-making, and in turn inefficient resource allocation, and/or even frequent market failures. When viewed from the context of performance indices, the information asymmetry, as we have already seen, can undermine the validity of the comparisons run (between companies, strategies, employees of the same of or across various industry sectors) thus creating and nurturing the most challenging conditions for the interpretation of the metrics and the generation of any meaningful conclusions.

One of the main and more vibrant effects of the existence of information asymmetry is related to the performance indices that can quite easily lead **to biased or incomplete assessments of the corporate performance**. The corporations that attempt to disclose information in a selective, cryptic, or non-standardized way, are actually using fake reporting practices, making it difficult (if not impossible) for any external stakeholder to accurately evaluate the performance indexes related to that entity and to compare them vs. other industry peers [142]. In order to address this issue, various researchers and regulators have been advocating for the adoption of specialized **(standardized) reporting frameworks**, like the Global Reporting Initiative (GRI) or the Integrated Reporting Framework (IRF), which are aiming towards the application and improvement of comparability and transparency of the shareable corporate performance data [143].

A very important effect related to **information asymmetry** is the mere potential for **adverse selection**, according to which the investors struggle in an attempt to differentiate their investments between the high-performing and the low-performing corporations due to the lack of transparency in the reporting practices of the latter study [144]. This practical problem leads to various inefficiencies in proper capital allocation and can potentially hinder any expected growth of the companies that in any other case would be deserving an unbiased, rich and quick investment.

Another effect that emerged through information asymmetry is related to the potential for **manipulation or misrepresentation of performance indices**. The study by Cohen, Dey, and Lys [145] examines the effects of the corporate governance reforming on factors like the executive incentives, the investments, and the exercise of risk-taking. The authors are highlighting the potential of the information asymmetry emergence leading to various manipulative practices, like earnings management, that vastly distorts the pragmatic financial health of a corporation. The authors proceed with the analysis suggesting that the toughest corporate governance mechanisms are required in order to mitigate any appearing risks through the alignment of all executive incentives with any long-term shareholder value(s) and through the reduction of any possibility of manipulation emergence or any case for the misrepresentation of the performance indices.

A recent article authored by DeFond and Zhang [146] is discussing the role of the audit committees and their independence by providing an extensive overview regarding (robust) internal controls, and other practices falling under the category of "corporate governance mechanisms" in order to mitigate the risks that are associated with the existence of information asymmetry. The authors emphasize the fact that these mechanisms are contributing to the **integrity of performance indices through the promotion of transparency and through the reduction of any manipulation or misrepresentation likelihood**.

In summary, the asymmetrical transactions in the context of the corporate environment can quite easily produce great impact on fairness, on efficiency, and on sustainability, especially when examining a variety of performance indices in the corporate environment. This is to state that addressing similar matters through a standardized reporting process, supported by strong corporate governance, guarantees that the various stakeholders can raise awareness, enhance transparency, challenge comparability and introduce reliability to the performance assessments. This is also why addressing the information asymmetry and any power imbalances protects the smaller corporations/organizations and helps the

bigger ones to contribute to the maintenance of a fair, equitable and competitive market landscape.

3.8. Corporate acts and misleading indicators

3.8.1. Greenwashing

Many false or misleading indications of positive corporate social responsibility (CSR) occur by corporations that engage in practices of "greenwashing" or other deceptive actions in order to **present an image of social and/or environmental responsibility that may not (accurately) reflect their actual practices, beliefs or impact**. Such acts are undermining any credibility of the authentic CSR ambitions hindering the assessment progress of stakeholders that want to support companies that run under genuine aspirations regarding their social and environmental impact.

A primary effect regarding the false (fake or misleading) CSR indication is related to the immersive erosion of what we call "trust" between various corporations and their stakeholders (incl. their investors, their consumers, and their regulators [147]). These stakeholders become skeptical and much less supportive, and they are less likely to engage, let alone support any similar future initiatives, as they realize the true intentions of companies that misrepresent their CSR initiatives.

A secondary effect that relates to the deceptive CSR is the absolute distortion of various market signals, which lead to the shaping of an uneven playing field for all corporations that aspire to be genuinely committed to social and/or environmental responsibility activities [148]. The corporations that are engaging in deceiving CSR practices are enjoying only short-term benefits (and favoritism) in the expense of eventually undermining any long-term values, minimizing the effectiveness of the CSR initiatives across all adjacent industries.

3.8.2. Misinterpretations of Engagement

Any false, fake, or misleading indications of any positive employee engagement occurs when corporations **exaggerate or misrepresent their efforts to foster a supportive and inclusive work environment**. This directly leads to a distorted (malformed) perception regarding the origin, the events, and the levels of employee engagement and/or satisfaction or the respective corporations. Subsequently, this comes with very negative consequences impacting, as we have already discussed, employee retention, the productivity, and the overall corporate performance.

Another strong effect is the gap (requiring the design of a good contingency plan) that is being created between an employee's expectations and his/her actual experiences around the given workplace [149]. This produces a disconnection which can be quite impactful and can lead to a lower job satisfaction and to an increased turnover, since employees face their disillusionment about the faking around the company's unfulfilled promises.

One last (but not least) effect related to the misleading of the employees' engagement is undermining the credibility of any genuine efforts that a) are aiming towards the creation of a positive workplace environment and b) are encouraging practices that enhance employee engagement [150]. It is important to note that when corporations are misrepresenting their engagement initiatives, their employees will eventually (and sometimes quickly) turn into skeptical and less engaged participants that will not fall for supporting any similar future actions, further lowering the chances of the respective company to attract and retain top talent.

3.8.3. Survey Manipulation, Tokenism and Facades

As we have already mentioned, historically various corporations have actually employed several methods in order to create a false, fake and misleading series of acts in order to attract employees and create a brand their workplace enjoying an environment of positive work ethics and high employee morale. As expected, some

of the tactics that have been used, included manipulation of (seemingly realistic but actually fake) surveys and subsequently results and/or metrics, of tokenism, and of propagating a facade of an ethical culture without pragmatic substance.

Manipulation of surveys or metrics: Some corporations have manipulated past employee engagement and/or satisfaction surveys in order to create an artificially positive image of their corporate identity. This was achieved by the selective reporting of only the favorable results, by cherry-picking specific data and/or by pressuring their employees in order to provide positive feedback [151].

Tokenism: Some corporations have engaged in tokenism, which is a practice that involves the showcasing of a few well-selected initiatives or of specific personas (employees) to be used as "evidence" of a strong commitment to their work ethics, team/leadership values and morale, without really making any major improvements to the broader organizational culture or policies/practices of theirs [152].

Facade of ethical culture: Some corporations have created a facade of an ethical culture proceeding to the establishment of faked codes of conduct, non-existent ethics training programs, and other "symbolic" initiatives which were lacking any true commitment, professionalism or even enforcement. These were superficial efforts that were aiming at the creation of a fake positive image rather than a genuine promotion of the so-needed ethical behavior and high morale movements among the company's employees [153].

Such deceptive tactics only had negative consequences for the respective corporations in the long run, like eroding trust, like fostering cynicism, and like perpetuating all kinds of unethical behavior within the corporation/organization.

3.9.   Indicative Culture & Morale Metrics in Businesses

The culture and morale metrics in the context of doing business play a very important role in the assessment and the clear monitoring of the overall health and success of any corporation/organization. There are multiple key indicators that can

help corporations/organizations to understand the current (and according to this thesis potential even the future) state of their culture and employee morale. It is quite straightforward to indicate that sub-dimensions of these metrics include the employee engagement, the metrics related to job satisfaction, the retention rates, and anything that relates in that sense to psychological safety.

Employee engagement: The measurement of this dimension can result to fruitful insights regarding the level of employee commitment, the level of enthusiasm, and the quantification of aspects like the contribution that the employees are bringing to their work. The high engagement of employees is most likely (and stands to be proven) correlated with dimensions related to the respective company's productivity and its organizational success [154].

Job satisfaction: As regards the sub-dimension of job satisfaction, this involves the evaluation of employee satisfaction in relation to the quality of their working environment, their duties (job responsibilities), and the overall company's organization. High levels of job satisfaction can, according to [155] result in a minimization of the company's turnover and to improved employee performance.

Retention rates: As to retention, these rates reveal corporation's/organization's ability to retain the working (and performing) talent over time. Usually, any high values in retention rates are enough to infer that the respective company's employees are thoroughly (or more than adequately) satisfied and thus committed, while any low-valued rates actually signal that there are underlying issues (still) existent in the workplace environment and/or corporate culture [156].

Psychological safety: This is a term that relates to the extent on which most employees feel comfortable enough to express their inner thoughts and ideas without fearing for any retribution or ridicule. This is very important as the encouragement of psychological safety nurtures open communication, easy of brainstorming and innovation, and learning within the corporation / organization [157].

The frequent examination of these metrics over a sufficient amount of time (period) can result in valuable information regarding the quality and peculiarities that may govern the overall corporate culture and the employee morale of a company, proving leaders with useful insights in order for them to identify any risky areas that may require closer monitoring and improvement, and to foster a much more positive, more inclusive, and more productive work environment.

4. Data Mining and Processing

In section we will delve into the most critical role of the data mining concepts and the processing that we will follow in order to uncover all the valuable insights and patterns possible in order to help us (in the same manner that this is done for corporations/organizations) to understand the inner existing relationships between the cultural metrics and the overall performance. The section is formed in a way that will enhance the exploration capabilities on a plethora of aspects regarding the data analysis and the meta-processing.

We begin with the idea of effectuation and its (casual) relevance, especially for the context of evaluating company culture metrics. We then move on discussing how will the fitting scores be exploited in order to properly assess the impact of various culture metrics on the firm growth [158]. In section 4.3 we will dive further into the peculiarities of data analysis (as a knowledge area that is worth exploring before actually applying its tools) in order to understand the reasons leading to the specific data type definitions, the stages of data mining (like data acquisition and/or cleaning, data preparation, and data transformation) [159]. Furthermore, we define and present the data modeling process, the original ideas and the mechanisms that support classification, and the importance of inferring the right conclusions when dealing with the forecasting strengths and the challenges of data handling (its exploration, its analysis and/or its required "manipulation").

This last step helps us understand and then directs us towards the paths that corporations/organizations also need to follow, including the obstacles that we all need to overcome when evaluating the abstract notions (the qualitative nature) of employee morale factors (later on discussed as metrics), which is quite important when companies we aspire to face the latter as (credible) predictors of corporate performance [160].

The section is completed by covering the importance of method formalization noting, inter alia, the proper manner of measuring the applicability of such models (by organizing, documenting, and presenting the respective findings), taking into

consideration the idiosyncratic characteristics of each company case in the corporate environment [161].

4.1. The concept of Effectuation

The concept of the effectuation, which was initially introduced by Sarasvathy in 2001 [162], put in simple terms, refers to a specific grouping (a set) of decision-making principles that the entrepreneurs are using in order to navigate through operational uncertainty and through complex (market) situations. The idea of effectuation is mostly focusing on the leverage required to be exercised between the exploitation of the existing resources, the need to embrace contingency, the requirement of forming strong and trustful partnerships, and the limitation of the downside risk rather than to rely on any processes based on prediction and/or optimization. As the authors of [162] poetically put it: "*Effectuation provides useful design principles for transforming extant environments into new futures in the face of ambiguous goals.*" As regards this thesis, and more specifically in the context of evaluating employee morale factors (culture metrics), the aspect of effectuation can be quite insightful on the way that corporations/organizations are developing and adapting their culture (or in general their quality control in order to measure non-financial metrics) in response to quantifying the uncertainty and the change [163].

The effectuation principles can (and in our opinion, should) be applied to the assessment of the various non-financial metrics (here more specifically, but briefly, referred as "culture metrics" or "employee morale factors") by emphasizing the role of existing resources, such as the employees' skills, the knowledge, and the values, in shaping a corporation's culture [164]. The corporations/organizations can better grasp the (group) dynamics that drive the performance and the innovation when studying their available resources. In addition, the effectuation is highlighting the significance of building up strategic partnerships and trustful collaborations to shape and sustain the company's culture, thus acknowledging that such

relationships (close and effective with the various stakeholders) have the power to influence (affect) the cultural outcomes [165].

The incorporation of the effectuation principles when evaluating cultural metrics can also encourage the corporations/organizations in embracing the necessary flexibility and/or adaptability processes that enhance the acknowledgement that unexpected events play a key-role in shaping employee culture [166]. Consequently, the establishment of such a perspective can facilitate the agility and resilience of corporations, especially at times of unprecedented or just unexpected changes, thus nurturing a company culture that will eventually thrive in dynamic and challenging environments. In overall, the effectuation concept builds up a strong framework for comprehending the virtue of properly estimating the role of culture in organizational performance.

4.2.   Fitting scores and firm growth

The fitting scores, which are also known as "fit indices" or "goodness-of-fit measures", are actually statistical tools which are used in order to estimate the existing compatibility among the observed data and a hypothesized (predictive in our case) model or (in general) a theory. These scores can help the researchers to properly assess how well do the available data (or created datasets) support a specific (given) model, and in this case, the potential relationship between culture metrics and firm performance (for example in terms of growth) [167]. The notable concept of the "fitting scores" has (fairly) gained its prominence in the field of organizational research, since it enables the quantification of the effect that cultural factors have on several aspects of a corporation's performance (including growth) [168].

More specifically, in the context of the culture metrics evaluation the "fitting scores" may be employed in order to estimate (measure) the degree that specific cultural variables (factors) are aligning with, and in turn are predicting, a corporation's trajectory of growth. For example, the researchers can use the "fitting

scores" in order to examine how (different in their nature and/or origin) cultural attributes, like collaboration, like innovation, or like leadership styles, can influence the firm growth by examining (analyzing) the potential relationships among such factors and the various key performance indicators (KPIs), i.e., the conventional quantifiers being used in most cases [169].

The tools of "fitting scores" can be calculated (or more typically to be computed) by using several statistical methods and techniques, like structural equation modeling (SEM), like regression analysis, or like other data-driven approaches. The researchers are employing such techniques in order to better understand, as it has already been mentioned, the degree of interplay among cultural factors and company growth, and subsequently to identify which of the various aspects (sub-dimensions) that a company's culture include, has (or have) the strongest effect on its performance [170].

Summarizing, the "fitting scores" is a valuable method for the assessment of the impact of culture metrics on the growth of a firm, always in respect to what is the quantifying approach being used, per case. This enables the comprehension of the innate relationships that exist among the various cultural attributes and the performance outcomes and, in turn, helps corporations/organizations to identify and to prioritize the factors that dominate in their success.

4.3.   Data Analysis

The data analysis (in simple terms) is the process of applying various statistical and logical methods, in a systematic way, in order to extract comprehensible information like patterns and various insights (which can be expressed in mathematic terms) from raw data essentially [171]. Data analysis is a crucial step in the decision-making process and in research in general since it undertakes the task of converting complex datasets into exploitable information and actionable knowledge [172].

There are numerous methods of data analysis. In brief, these are usually individual theories and techniques falling under big generalized sets like the descriptive, the exploratory, the inferential, the predictive, and the causal analyses. Such methods are typically applied across several fields, including the social sciences, several business knowledge areas, the financial research, and the healthcare, in order to help applied scientists, researchers and practitioners to properly infer conclusions and be led to well-informed decisions [173].

Data analysis allows for the identification of the aforementioned key-relationships among the cultural factors and the organizational performance of the respective company. By the employment of such advanced (mostly statistical) techniques, like the regression analysis, the structural equation modeling, or the machine learning algorithms, the experts can isolate the impact of specific sub-dimensions (i.e., cultural features or attributes) on the KPIs and thus, to better understand the various drivers of the companies' success [167].

In addition, the science behind the data analysis can help researchers to identify various trends, to uncover hidden patterns, and to predict future outcomes, which are of high value for corporations that are trying to improve their overall performance [174].

Concluding, the methods of the data analysis stand as a vital toolset for understanding and extracting meaningful information from complex (and big) data. When properly used it has the potential to reveal the hidden relationships among abstract metrics and to quantify results that allow experts to infer useful conclusions and make educated decisions that can optimize their strategies.

4.3.1. Data types in data science

Moving on, in more detail, there are in total four main types (or classes) of data in the data science scientific discipline. In the lines to follow we will check the definitions of each one of these four classes:

Nominal data: This term refers to **categorical data** that do not own any inherent order and/or ranking. The term is used to label or categorize any kind of **items that can be set into distinct groups** or categories. Simple examples of the nominal data type include information about the gender, the nationality, or the hair color.

Ordinal data: This term refers to **categorical data** that has a clear order and/or ranking. The term is used for data that can be sorted or **ranked along a continuum** and the distances between the ranks are not uniform or measurable. Simple examples of ordinal data include the Likert scale responses (e.g., strongly disagree, disagree, neutral, agree, strongly agree) or educational levels (e.g., elementary school, high school, college, postgraduate).

Interval data: The term refers to **numerical data** that has a consistent scale and/or equal intervals between values, but it **lacks a true zero point**. This definition suggests that a comparison between values (or differences) is valid and understandable, but no meaningful statements can be inferred about any ratios. Simple examples of interval data include the temperature measured in Celsius or in Fahrenheit, and the calendar dates.

Ratio data: This term refers to **numerical data** that has a consistent scale, with equal intervals between values, and **a true zero point**. This definition suggests that we can infer meaningful statements about differences, and the same is true for ratios and proportions as well. Simple examples of ratio data include height, weight, age, or income.

4.3.2. Types and Stages of Data Mining

The data mining is defined as a process which is related to discovering various patterns, potential relationships, and valuable insights within large data sets (sometimes from different sources and of various types, like big data) by applying several mathematical, statistical, machine learning, and lately artificial intelligence (and deep artificial intelligence) techniques [175]. The respective knowledge areas have been evolved quite significantly over the past few years, with a plethora of

methods and algorithms being designed and developed in order to sufficiently address the versatility of all types of data and, in respect, their analytical requirements [176].

As already mentioned, there are numerous relevant types of data mining tasks, like classification, like regression, like clustering, like association rule mining, like anomaly detection, and like sequential pattern mining, among many others [177]. The tasks that were mentioned are designed to cover different purposes and to serve different cases. They are usually applied according to the scenario (i.e., the nature of the available data and the research objective).

The science of data mining usually follows a chain of stages which, in brief, can be summarized in the bullets to follow [178]:

- **Business understanding**: The first stage is dedicated to the problem definition and the objective comprehension from the perspective of the given business.
- **Data understanding**: During this second stage the experts need to assess the available data sources, to explore the properties of the various data sets that can be formed and to evaluate the quality and the suitability of the raw data for the purpose of the analysis (fit-for-purpose examination).
- **Data preparation**: During the third stage, researchers need to clean, to transform, and to integrate the raw data in order to create a unified data set which will then be the suitable source for any further (algebraic) analysis.
- **Modeling**: During the fourth stage numerous data mining algorithms (and or special mathematical techniques) are applied on the previous (prepared) data set in order to formulate (to create) the models that will eventually uncover any hidden patterns and valuable insights.
- **Evaluation**: During the fifth stage the researchers assess the quality, the reliability, and the relevance of the created models and then examine (meta-process) the resulting outcomes to ensure that they can be aligned with the objectives of the actual research.

- **Deployment**: During the sixth and final stage, the various insights (that may have been gained by the application of the dataset on the constructed models) are process to be presentable in a meaningful format and then to be easily integrated into the rest of the decision-making process.

By following this process, the experts ensure a very methodological, systematic and effective approach, namely "data mining", that allows for the extraction of valuable information that when properly processed can result in a complete decision-making suggestion and, in turn, improve the corporations' organizational performance [175].

4.3.3. Data Acquisition

Data acquisition is simply regarded as the process of collecting, of gathering, and of measuring any raw data from several sources to be used for data analysis, for decision-making, and for reporting purposes [179]. Data acquisition is a very important step in the process of data analysis, as the quality, the accuracy, and the relevance of the data that are collected. Data acquisition can significantly impact the quality (and quantity) of the potential insights that can be derived from any subsequent analysis [180].

The data acquisition process may typically involve several methods and numerous techniques, depending on the type of data that is needed and the resources that are available. Some very basic and common data acquisition techniques include the ones that follow in the bullets below:

- **Surveys and questionnaires**: These are tools that are extensively used in order to collect data from various individuals or from groups, on a variety of topics, like customer feedback, like employee engagement, or like various market trends [181].
- **Web scraping**: This tools is about data extraction from online (internet) websites and other online sources, which is typically conducted by programming (scripts) that form automated bots (or "spiders"), that wonder

around the inter-"net" and gather what is regarded as "useful" information filtered by a specific topic or an industry [182].

- **APIs (Application Programming Interfaces)**: The APIs are acting as servants that bring useful information to a user upon specific requests. Similar to scraping, they allow for automated data extraction from online platforms, but contrary to scraping this is done by following specific pre-set methods that return properly formalized information. The APIs are usually applied in websites like online databases, like social media, or other web services, enabling for the efficient data collection and integration [183].

- **Direct observation and measurement**: There are cases where the data have to be collected through primary data collection (primary sources). This is usually done through direct observation, through questionnaires, through experiments, or through other forms of empirical research [184].

- **Secondary data sources**: Also, there is the case that researchers will need to gain access and acquire information from various existing (primarily collected) data sources, like databases, like government reports, or like academic publications, which can, to some extent, provide extra useful insights (and support) for the primary data collection efforts, too [180].

The final choice regarding the best data acquisition method usually depends on several factors, like the research objectives, the type of the available data (and the data that are actually required, since these two sets may not be necessarily the same), and also the resources that are available for data collection. Irrespective to the method to be selected, it is important to guarantee the quality, the reliability and the representativeness of the collected data – which is not only an ethical requirements but also a beneficial principle to the success of the data analysis process [179].

4.3.4. Data Cleaning

The process of data cleaning, which is also known as process of "data cleansing" or of "data scrubbing", focuses on the identification and the correction of errors, of

inconsistencies, and of inaccuracies found in the available datasets. It is applied in order to guarantee that the data is of high quality, that it is very accurate, and that it is also very reliable for the next steps to follow in the analysis [179]. This pre-processing is important since the presence of any dirty data will most likely lead to erroneous parsing and eventually to biased or to incorrect results, which will affect the reliability and validity of the final inferred conclusions [185].

The process of data cleaning can involve a number of tasks, such as:

- **Removing duplicate records**: When the available dataset includes duplicate entries, these can skew the anticipated results of the data analysis and can lead to incomplete or incorrect conclusions. So, it is important to identify and eliminate any duplicate records during the data cleaning process [186].

- **Handling missing values**: It is quite common that the "raw" datasets will include missing data which can quite easily result from a variety of issues, like data entry errors or like non-response in the source-files (e.g. surveys). In order to properly handle the missing values, it is important to design actions that involve imputation (which means to replace any missing values with well-estimated values) or in some cases deletion, depending on the extent and the type (nature) of the missing data [187].

- **Correcting inconsistencies and errors**: There can exist various data inconsistencies which can occur when the "raw" data is collected from different origins or include various types of sources or when there are data entry errors accumulated and passed from edit to edit. The identification and the correction of such issues is imperative in order to guarantee the data consistency and information accuracy [179].

- **Outlier detection and treatment**: The outliers are specific data points that can significantly deviate from the norm or the general pattern of the rest of the data. The identification and the addressing of outliers can help experts to prevent their undue (negative) effects on the outcome of the data analysis. Knowing which and why outliers appear can also be quite useful information when properly treated. As the author of [188] states "*an outlier often contains*

useful information about abnormal characteristics of the systems and entities that impact the data-generation process. The recognition of such unusual characteristics provides useful application-specific insights.".

- **Data transformation and normalization**: There can exist cases where the data may require transformation and/or normalization in order to ensure consistency and comparability across several variables, across units, or across scales [185].

The process of data cleaning is an iterative and quite often a very time-consuming process. This has been the case of the cleaning that was required for the available datasets that were tried and used in this thesis. However, it is very crucial to invest the required time and resources (expertise gained through analysis) in this preliminary point in time, as such an investment can ensure that the data is fit for the next steps of the analysis, gradually leading to better processed, more reliable, and more accurate insights [179].

4.3.5. Multicollinearity

Multicollinearity is a statistical phenomenon which emerges when there are two or more predictor variables (usually in a multiple regression model) which are highly correlated. So, in other words, multicollinearity happens when one of the predictor variables seems to be linearly (and almost directly) predicted from the other ones, with a high (definite) degree of accuracy.

The existence of multicollinearity can cause a series of serious problems in the regression analysis:

- For starters, it can drive towards unstable (incorrect) estimates about the regression coefficients. This is because small changes in the data can easily produce significant changes to the (estimated) coefficients, thus creating scenarios where it will be difficult if not impossible to properly interpret the results.

- Also, the standard errors of the model's coefficients will tend to be inflated and in turn this will distort the determination of the statistical importance of any individual predictor variables.
- Finally, the overall **goodness-of-fit** (namely "**R-squared**") of the created model may be unreasonably high, while the real contribution of each one of the individual predictor variables may also be unreasonably difficult to determine.

In order to detect any existing multicollinearity, the experts often have to look at **the correlation matrix** that depicts the various predictor variables or to calculate the **variance inflation factor** (VIF) for each one of the predictor variables. In order to address multicollinearity, the experts may consider removing one (or more) of the highly correlated predictor variables (dimensions), or to combine them into a single variable, or to use techniques like the **principal component analysis** (PCA) or the **ridge regression**.

4.3.6. Exploratory Data Analysis

The exploratory Data Analysis (EDA) is a very important step in the way that the data analysis process is being involved in the use of several methods in order to gain an primary understanding of the nature of the data, its inner structure, and its dominating characteristics before proceeding to the application of any advanced statistical methods or any machine learning algorithms [189]. The key objective of EDA is to manage the pattern identification, the trends, the outliers, and the potential relationships that may exist among variables, as well as to detect various issues and/or anomalies of the data that could affect the results of the subsequent analysis [190].

The EDA can be typically conducted by the facilitation of several approaches, like the ones found in the bullets to follow:

- **Univariate analysis**: This is an important analysis that involves the investigation (examination) of various individual variables, like calculating

descriptive statistics (the mean, the median, or the standard deviation) or like visualizing data distributions using **histograms** or **box plots** [191].

- **Bivariate analysis**: The bivariate analysis is an approach that involves the exploration of the relationships that may exist between two specific variables, like using the **scatter plots** in order to visualize the correlations or the **cross-tabulations** to distinguish (and then examine) any associations found between categorical variables [189].

- **Multivariate analysis**: The multivariate analysis is a method that focuses on the examination of relationships between multiple variables at the same time. For instance, the use of techniques like the principal component analysis (PCA) or **cluster analysis** in order to identify any underlying patterns or any groupings in the given data [190].

The EDA is typically an iterative and an interactive procedure, where data analysts use various visualization tools and several statistical techniques in order to generate impactful insights and thus to enhance and refine their initial understanding of the collected data [191]. The EDA is helping experts to reach accurate, reliable and, most importantly, meaningful conclusions by creating a robust (solid) foundation for the subsequent modeling and the hypothesis testing [189].

4.3.7. Data Modeling, Definitions and Classification

Data modeling is related to the process of creation and more precisely, the creation of proper representation of the various data structures and their innate relationships as these exist in a dataset. This process helps the experts to guide their development and the evaluation of the ML algorithms and/or statistical models towards the right direction [192]. Data modeling typically includes the exercising of various techniques, like regression, like classification, like clustering, or like dimensionality reduction. The selection of the technique depends to a great extent to the nature of the data and to the specific objectives of the data analysis, per case [193].

In more detail, when viewing data modeling from the context of the culture metrics evaluation, the data **classification** is a particularly relevant (and highly important) data modeling technique. The classification is in simple terms a **supervised learning** method which aims to the prediction of the existing discrete class label (or category) of the available (given) data instance, according to its expressed features or attributes (or dimensions) [177]. Some examples of classification algorithms may include the **logistic regression**, the **support vector machines (SVMs)**, the **decision trees**, and the **neural networks**, among many others [193].

In order to perform classification, a part of the original dataset, called **training dataset,** which is associated with known class labels, is used in order to properly "train" the model, which, in turn, learns the existing relationship between the input features (dimensions) and the **output class labels** (the output layer of classifiers) [192]. When the training is complete, the predictor can be applied to new, unseen data (usually the rest of the original dataset) in order to predict its class labels according to the relationships it learned [177]. As to the performance of the classification models, this is usually estimated (assessed) by various metrics of accuracy, of precision, of recall, or of F1-score, which can provide a very good indication of how well the model generalizes to new (unseen) data and how well does it predicts class labels [192].

In conclusion, data modeling, and more specifically the data classification, is playing a key role in the data analysis of culture metrics (or employee morale factors) and to their potential effect on the corporate performance. By using the classification methods described in this section, the experts are able to develop predictive models that can help them to identify and to better understand the potential relationships that may exist among various non-financial and financial metrics. For example, by primarily looking at the innate features that culture metrics hold and by answering which are the most dominant ones and to which extent one or more of their features (dimensions) can be used and by which

performance predictor, the experts can be led to educated decisions and to draw useful conclusions for a company's performance outcomes.

4.3.8. Regression Functions

The regression functions are typically used in order to model the existing relationship(s) between a **dependent variable** (target) and one or more **independent variables** (features or dimensions). There are different regression algorithms that have been historically developed in order to address several types of data and relationships. There is a brief overview of a few basic regression techniques given in the bullets of the list to follow:

- **Linear Regression**: This is a quite simple and widely used technique for modeling the existing relationship between a target variable (as already mentioned) and one or more input features (or dimensions) using a linear function [177]. The structure of the algorithm is designed to seek a solution that will minimize the sum of squared errors expressed between the predicted and the actual target values.

- **KNN Regressor**: This is namely the "k-Nearest Neighbors (KNN) regressor" which is a non-parametric, and "lazy learning algorithm" which can predict the target value according to the average of the **k-nearest training instances** [192]. The optimal k value is typically determined through a cross-validation procedure in order to minimize any prediction errors.

- **SVM Regressor**: The Support Vector Machines (SVM) is an ML technique which is quite versatile. The SVM can be used for regression tasks, known as Support Vector Regression (SVR) [194]. It can use either a **Linear** or **Gaussian** (Radial Basis Function) **kernels**. The SVR is looking for a function that can fit the data within a pre-specified **margin of tolerance** while it minimizes the model complexity.

- **Decision Trees**: The decision tree method of regression includes the **recursive partitioning** of the input space and the fitting of a simple model (e.g., a constant) within each one of the partitions [177]. The parameter called

"**max_depth**" controls the tree depth. The deeper trees present, as expected, more complex models but also have a higher risk of overfitting.

- **Random Forest Regressor**: The random forests are an ensemble (a mix) of various learning functions, like **a combination of multiple decision trees** that aims to improve the prediction performance and to reduce the risk of overfitting [195]. Each one of the trees is being trained on **a bootstrap sample of the data** and then a random subset of the features (dimensions), leads to a diverse set of models which are combined to collectively produce more robust predictions.

- **Gradient Boosting**: This is one more ensemble method which is combining **weak learners** (typically this means shallow decision trees) in **a stage-wise manner** in order to minimize a given loss (optimization) function [196]. This method is able to capture the complex relationships in the data and to achieve a quite high predictive performance by iteratively fitting its models to the residuals of the previously created models.

4.3.9. Other Predictive Models

As it has already been well-established, the various predictive models are using input features in order to create mathematical relationships that map the inputs to the target variable. There exist several types of **predictive models** in literature, including **regression**, **classification**, and **time series models**. In the science of predictive modeling, the final quality of the designed model is being assessed by measuring, as already mentioned, its **accuracy**, its **precision**, its **recall**, its **F1-score**, or other evaluation metrics, based on the nature of the problem type [177]. In order to avoid issues related to overfitting and to improve the generalization, the techniques of **cross-validation**, of **regularization**, and of **early stopping** can be employed [197].

In the bullets to follow there are two extra types of algorithms mentioned, which may be used for building models based on the respective prediction:

Neural networks: These are highly flexible and quite powerful models that consist of many interconnected layers of artificial neurons which create networks and are capable of approximating extremely complex relationships between their inputs and their outputs [197].

Ensemble methods: These models, similar to random forests and to gradient boosting, are designed to combine a variety of weak learners in order to create a much more robust and more accurate predictive model [195], [196].

4.3.10. Forecasting strength and challenges

The forecasting capability is an essential component of the decision-making processes and can be applied on several disciplines and scientific (or commercial) fields like finance, like economics, like supply chain management, and like marketing [198]. The various forecasting models can help the corporations / organizations to reach informed decisions (conclusions), to effectively allocate resources, and to proactively prepare for any potential risks. Nevertheless, there is a variety of factors that can influence the robustness, the strength, and the accuracy of the forecasts, thus leading to various challenges in the process. A brief overview of important forecasting characteristic is given in the bullets to follow:

- **Strength**: Any well-designed forecasting models can supply the experts with very valuable insights into the future trends, thus enabling the businesses to plan and adapt their strategies, accordingly. The recent advances in ML and AI have been leading to the development of very sophisticated forecasting models. These models are able to handle very large amounts of data and to capture the complex relationships between variables [199].
- **Data quality**: The high-quality datasets are an essential part of forecasting accuracy. Conversely, any inaccurate, incomplete, or outdated data can easily lead to false or erroneous predictions and to undermine the effectiveness of the forecasting models [200]. So, ensuring the data quality for the forecasting process is an essential aspect to overcome this challenge.

- **Model selection**: Selecting an appropriate forecasting model is based on numerous factors, including a) the type of data, b) the relationships among the variables, and c) the forecasting horizon. The selection of a wrong model can gradually lead to unseeingly suboptimal forecasts [199]. This is why the cross-validation and the model comparison techniques can be so helpful in the identification of the most suitable model per case (for a given problem).

- **Uncertainty**: The is an **inherent uncertainty to forecasting**, as the future events are being influenced by several factors which are difficult to predict quantified [198]. The quantification of the uncertainty that is associated with forecasts is very important for the comprehension of the limitations of the predictions (per case) and for making educated decisions.

- **Dynamic environments**: Any rapid changes in the business landscape (like market conditions, like technology, and like external factors) make the forecasting a challenging procedure. Nevertheless, the adaptation of special forecasting models that can account for these typical changes is essential in order to maintain the predictions' accuracy [200].

4.4. Formalization

The formalization, in the context of the data mining and the data processing for evaluating culture metrics as predictors of corporate performance, is referring to the procedure that is dedicated to the systematic organization, documentation, and presentation of the several findings that have been obtained by the data analysis [201]. This is an important step which guarantees that the results will be simple, clear and transparent, easily replicable, and that they can be effectively circulated among stakeholders. The formalization, when done correctly, provides a concrete foundation to the experts for decision-making and for the enhancement of strategic development.

The process of (results) formalization involves the following steps in bullets:

- **Meta-processing - Documentation**: The clear documentation of the data mining and the processing methods, which includes the steps of data acquisition, the data cleaning, the data analysis, and the data modeling, is crucial in guaranteeing the **transparency** and the **reproducibility** of any research or study [202]. Such a documentation needs to sufficiently describe the various methodologies that were employed, the respective assumptions that were made, and the limitations that have been encountered (and how they were faced), and finally, any regulatory or ethical considerations involved.

- **Validation**: This is a crucial aspect of formalization as it involves the assessment of the robustness and the accuracy of the findings, typically using cross-validation, or by sensitivity analysis, or by comparisons with similar studies [203]. The step of validation helps experts to confirm the outcome credibility. It also allows them to be confident in the assumptions made and, in the conclusions drawn from the data analysis.

- **Inference - Interpretation**: The results that were collected from the data mining and the data processing need to be interpreted in the context of the research question, which in this thesis regards *the evaluation of culture metrics as predictors of corporate performance*. The interpretation of the results always involves the identification of the key insights and the various implications that the context may pose on the findings (here for the nature of businesses), as well as understanding and expecting any potential limitations or areas that may require further research [201].

- **Presentation - Communication**: The presentation of the results in a clear, concise, and accessible manner is very important as this guarantees that all stakeholders will be able to understand and to act upon the findings of the research. The presentation may include visualizations and/or tables, and/or concise summaries in order to fully convey the main conclusions and to stress their implications for the corporate performance vs. culture metrics (employee morale factors) [202].

Concluding, the formalization is a last but vital step in the data mining procedure and in the overall processing workflow, as it enables and enhances the effective communication and the pragmatic application of the research findings in the context, here more specifically, of the culture metrics evaluation as potential predictors of corporate performance.

5. Correlation Analysis

In the context of evaluating a company's performance, one very important step is to explicitly define the factors that are contributing to a high-rating (of that company) when it is being reviewed by its employees. To determine whether this hypothesis is true, i.e., to guarantee that such ratings can stand as indicative metrics for the purpose of rating a company's performing potential (at least by the employees' point of view), we primarily need to study various datasets that may hold the information of such associations. Since glassdoor.com is one of the biggest websites that focus on conducting this type of company ratings, the focus is paid onto retrieving such a dataset and conducting a correlation analysis on its available dimensions, since such an analysis can play a pivotal role in determining whether there is a linear association between specific cultural factors and the company's overall review performance.

The dominant idea is to succeed in finding a dominant factor (or the smallest set of the most dominant factors and their participation percentages) that can predict whether a company is highly rated in the perception of its most related stakeholders. Once the documented predictor has been consolidated as an emerging representative of such ratings, it will be eligible for further use in more complex analyses, including ones that debate a company's financial performance.

5.1. Preprocessing – Data Loading

The paragraphs to follow present the main stages of the data-mining and analysis conducted in order to reach to a few safe conclusions that can complement the thesis hypothesis, at least for the case of United States (since all data refer to companies based in US). The processing was done using Python 3.7.2, loaded with commonly used libraries for data analysis and visualization like Numpy, Plotly, Pandas, Matplotlib, Seaborn, and Scikit-learn.

- **NumPy** is a library used for numerical computing in Python. It supports working with algebra arrays, matrices, and with various mathematical functions. It is quite useful for computations/operations on large data sets.

- **Plotly** is a famous library that creates interactive, web-based visualizations. It can support a wide range of several chart types, including customizable representations of bar charts, line charts, scatter plots, and more.

- **Pandas** is a well-known and powerful data manipulation and analysis library that is quite popular in Python. It can provide data structures like Series and DataFrame in order to handle and process information in a flexible and efficient fashion. It is widely used to clean, to transform, and to analyze data.

- **Matplotlib** is a well-known plotting library for Python. It provides a wide range of functionalities for static, interactive, and animated visualizations.

- **Seaborn** is a notable statistical data visualization library which is built and running on top of Matplotlib. It can provide a high-level user interface (UI) that facilitates the creation of informative and attractive statistical representations using computer graphics. It includes several built-in themes and color palettes to create aesthetically pleasing visualizations.

- Scikit-learn library is a popular Python library for ML and for data analysis.

The original .csv data is presented on the screenshot of Figure 5.1 below:

Figure 5.1: The original dataset of the Glassdoor.com review ratings (20 columns, 17050 rows).

5.2.   1ˢᵗ pass – Data Reviewing

A brief data research reveals that there exist only a few columns, like the third one (highlighted on Figure 5.1) which is called "description" that are completely populated by unique information (17050 rows). There are many other columns that with missing values. A graphical representation of the issue is shown in Figure 5.2.
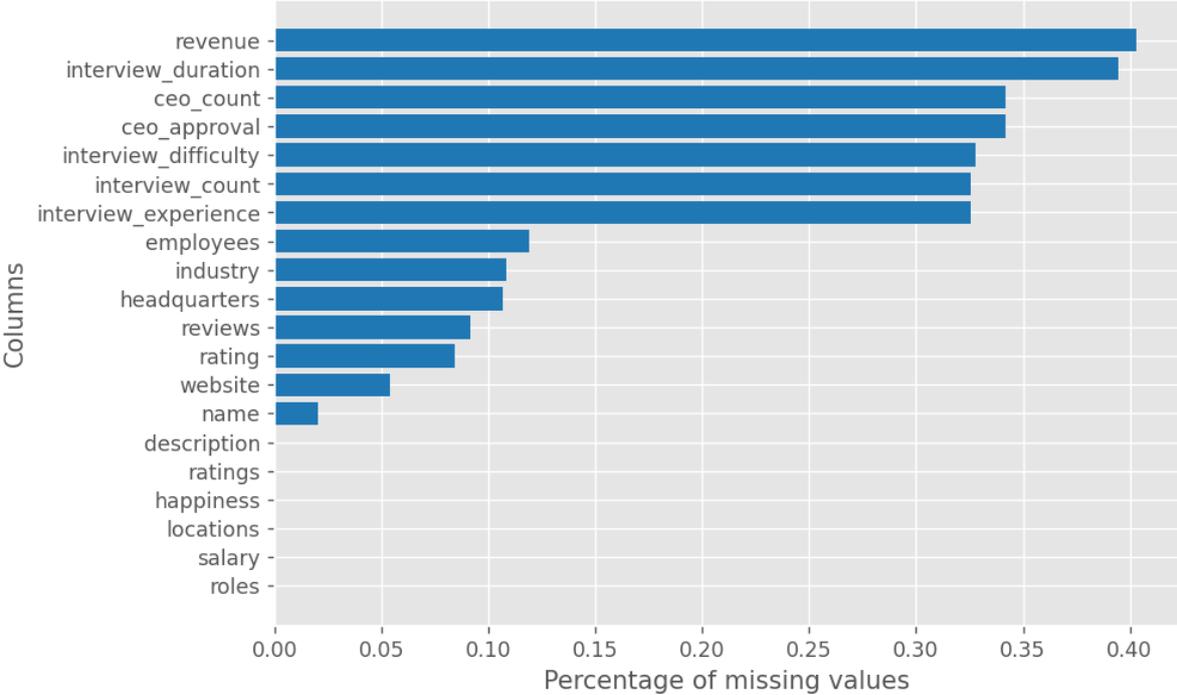
Figure 5.2: The percentage of missing information per column.

It is apparent that "revenue" is the dimension that has more than 40% of its columns unpopulated. In general, missing data can significantly impact the data analysis, leading to either biased or inaccurate results. When data is missing, it can introduce uncertainty into the analysis, reducing the power of the statistical tests and potentially causing to draw incorrect conclusions and to receive mislead decisions. Missing data can also affect the representativeness of the sample, which leads to issues of generalizing the results to a broader population. Moreover, missing data can also challenge the analysis of the potential relationships among the available variables, as it can mask the true (existing) associations, or it can create spurious correlations. Depending on the nature of the missing data (missing completely at random, missing at random, or not missing at random) and the

proportion of 'missingness', there are different techniques like imputation and/or deletion and/or weighted analysis that when employed they can mitigate the impact of missing data on the analysis. Here, we consider that information is missing **at random**. An interesting insight is to try to reveal the allotment of the recorded reviews, by sorting the available companies according to their revenue range. As Figure 5.3. depicts, the higher the revenue of the reviewed company, the higher the reviews volume.
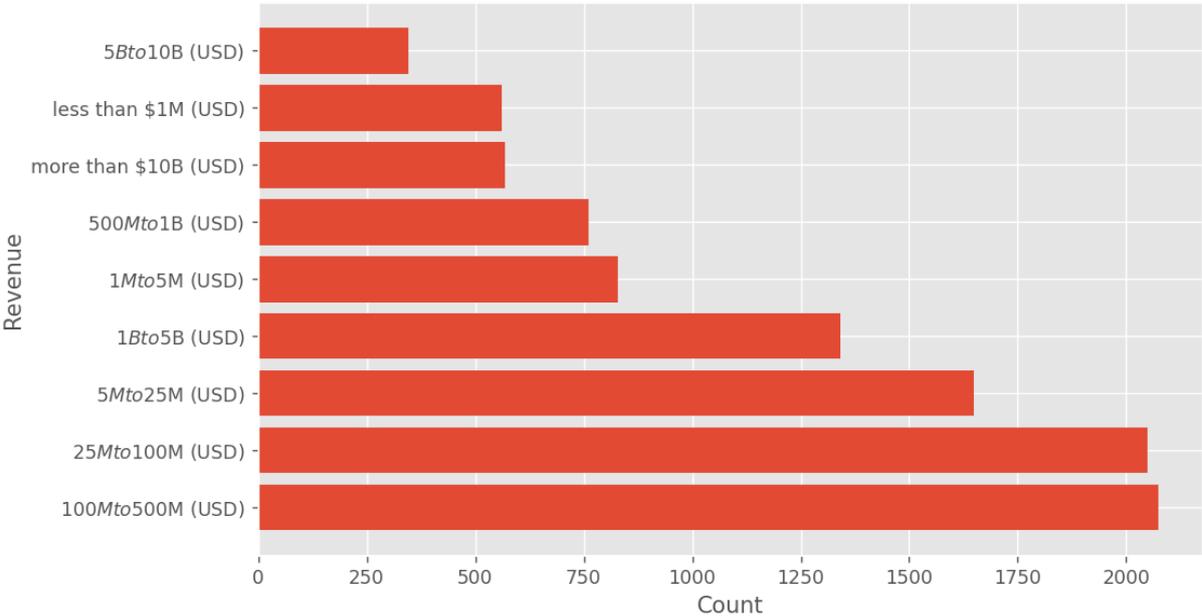
Figure 5.3: The distribution of the reviewed companies based on revenue range.

This result is to be expected to some extent and this is for several reasons; Firstly, most of the **larger and higher revenue-generating companies generally employ more people,** thus increasing the likelihood of having a larger number of reviews. So, more employees will lead to more experiences, and thus opinions and perspectives which can contribute to the overall number of the reviews, per case.

Secondly, the **successful and notable companies often are the ones that receive more public attention**, which leads employees to be further inclined to share their unique experiences when working for such corporations/organizations. Most of these corporations (companies) usually project a stronger and more significant

online presence in dedicated platforms which makes employees willing to leave their reviews.

Thirdly, **higher revenue companies usually have more structured HR departments** and feedback mechanisms encouraging employees for feedback, (either internally or externally). This can result in a higher employee's review volume, as they feel their voices will be taken into account by their corporation.

Lastly, employees at companies with higher revenues can have **higher expectations (regarding their work environment), their benefits, and their career opportunities**. If these expectations are met (or exceeded), the employees are likely to leave a positive review. If not, then employees may leave negative reviews to express dissatisfaction. In both cases, the higher review volume is attributed to higher revenue of the company and the associated employee expectations.

Proceeding with the insights that emerge from a brief data analysis focusing into distributions of various dimensions of the dataset, Figure 5.4 depicts the calculations' result of the distribution of the top-30 companies based on "Industry".
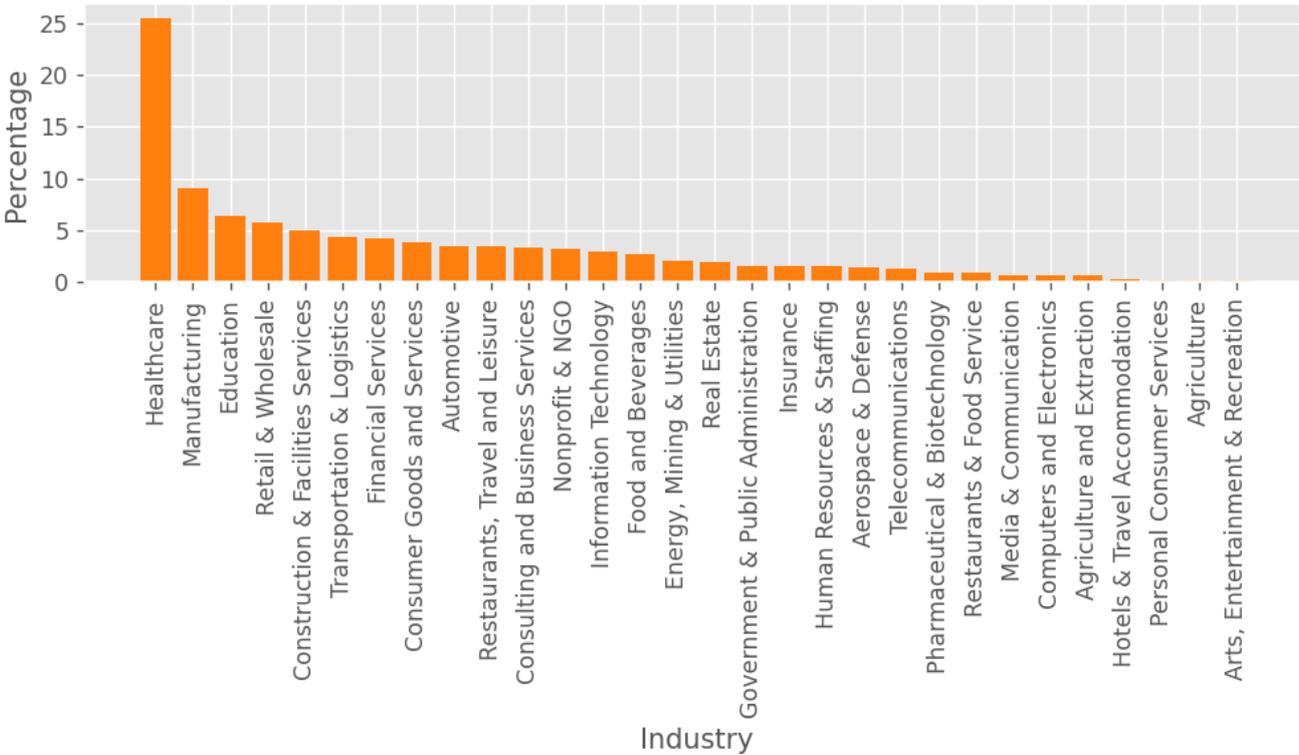
Figure 5.4: The distribution of the reviewed companies based on revenue range.

The identifiable presence of a large number of "healthcare" companies in the company review dataset (Glassdoor.com) may be due to a number of reasons. One possibility is that the respective industry is one of the largest (and fastest-growing) sectors in the United States (US) with millions of stakeholders. As a result, there exist many employees who raise the percentage of those that are interested in sharing their experiences and to seek information for potential employers. In addition, the specific group may be more inclined to use Glassdoor due to the job's nature, as the work environment, the management, etc., favor responsibility and accountability which can significantly impact their job satisfaction and overall well-being.

According to the US Bureau of Labor Statistics, the industry of healthcare is one of the largest employing industries in U. Other major industries include the **retail** trade, the professional and business services, the **manufacturing**, the **education**, the **healthcare services**, the leisure and the hospitality. The distribution of companies in the US varies per industry. Some sectors like retail and professional services have a larger share, while some others such as mining and utilities have a limited share. It is important though to note that the companies' distribution in this dataset may not perfectly represent the one in the US based on industry. The dataset can be influenced by several factors like the user demographics, the company size, and the company practices/policies regarding employee feedback.

Mentioning bias and demographics, the existence of the "locations" column can further facilitate the data analysis research. It can reveal useful information about the spread of the reviews on the US map, either in terms of "industry" load or "revenue" load. After data processing on the "location" column to transform the records from simple strings to enumerated data types (in Python programming) of city and state combinations we generate three different maps, depicted in Figure 5.5, that depict, in color-mapped values:

   a. the most popular industry per state
   b. the most popular revenue range per state

c. average company rating per state
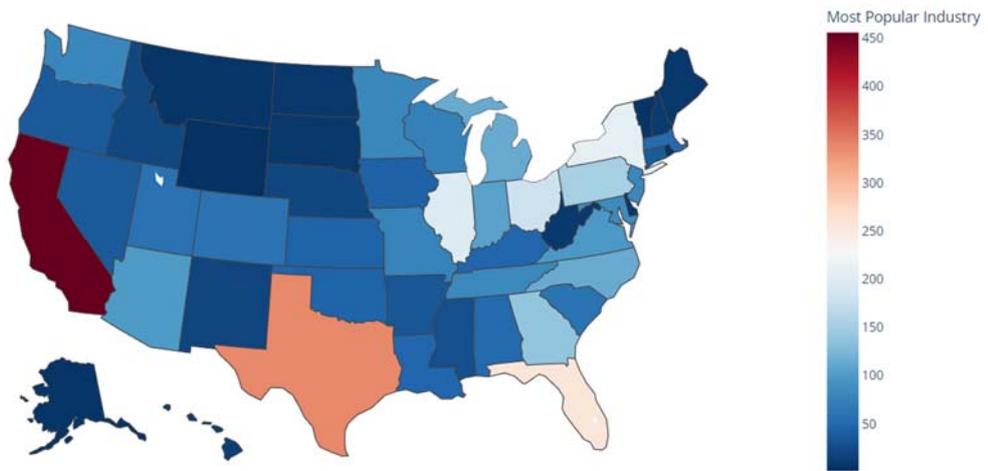
**Figure 5.5a**: Most popular industry per state.
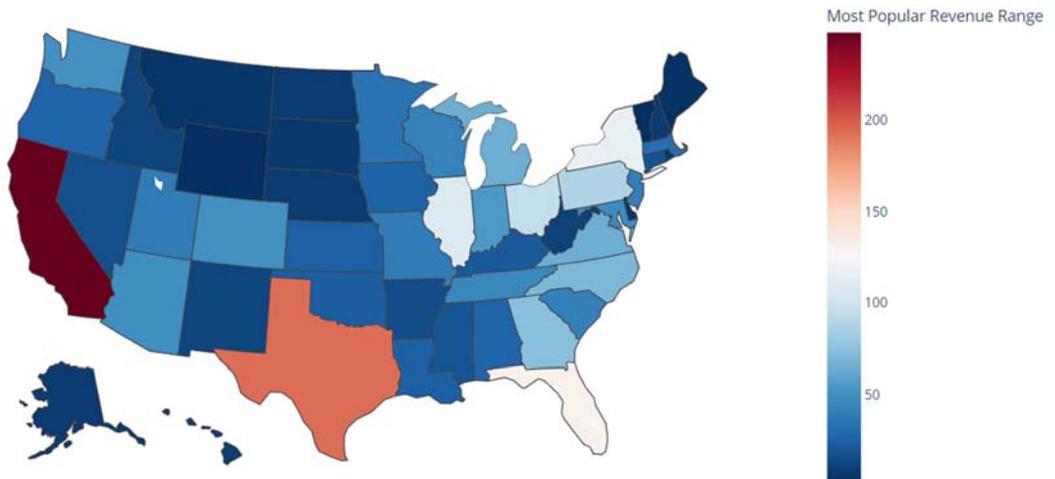
Figure 5.5b: Most popular revenue range per state.
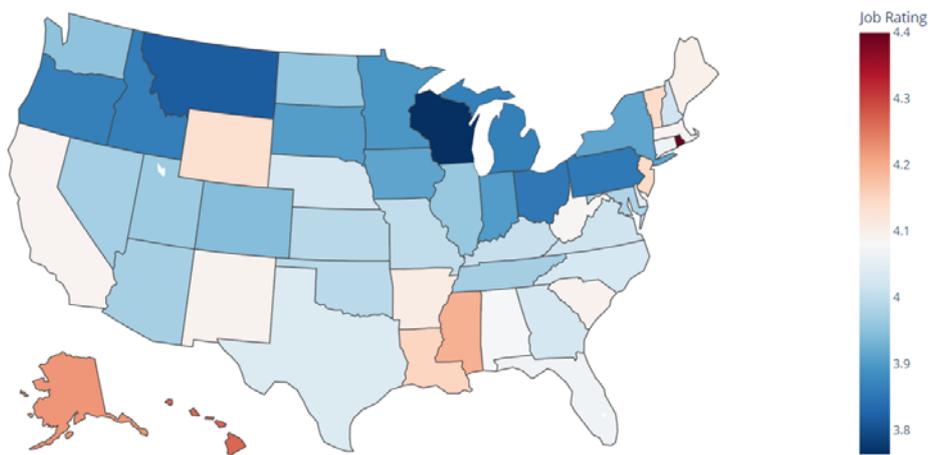
Figure 5.5c: Company review (overall) rating per state.

A look at the color distribution of Figures 5.5a and 5.5b indicate that the differences are very slight and the reason behind that is that the count of reviews of healthcare companies dominates, with the only exception being Alaska where the highest count is assigned to education. The color-matching of Figures 5.5a and 5.5b is also justified to some extent by the information inferred when studying Figures 5.3 and 5.4. Figure 5.5.c dictates that the highest job ratings can be found in Hawaii, Alaska, and Mississippi.

Next, we investigate two other columns of the dataset, called "ceo_approval" and "ceo_count" since the literate review dictates that some useful indicators may exist to be found related to the CEO approval and the employee rating of a company. Such correlations will be examined further in the paragraphs to follow. The ceo_approval in this dataset is expressed as a percentage of approval by each company reviews record and the "ceo_count" is a factor that expressed the number of reviews, or otherwise the weight of that percentage. The plotting of the ceo_count has been applied in a log10 scale since it has large variation (median 41, average 105).
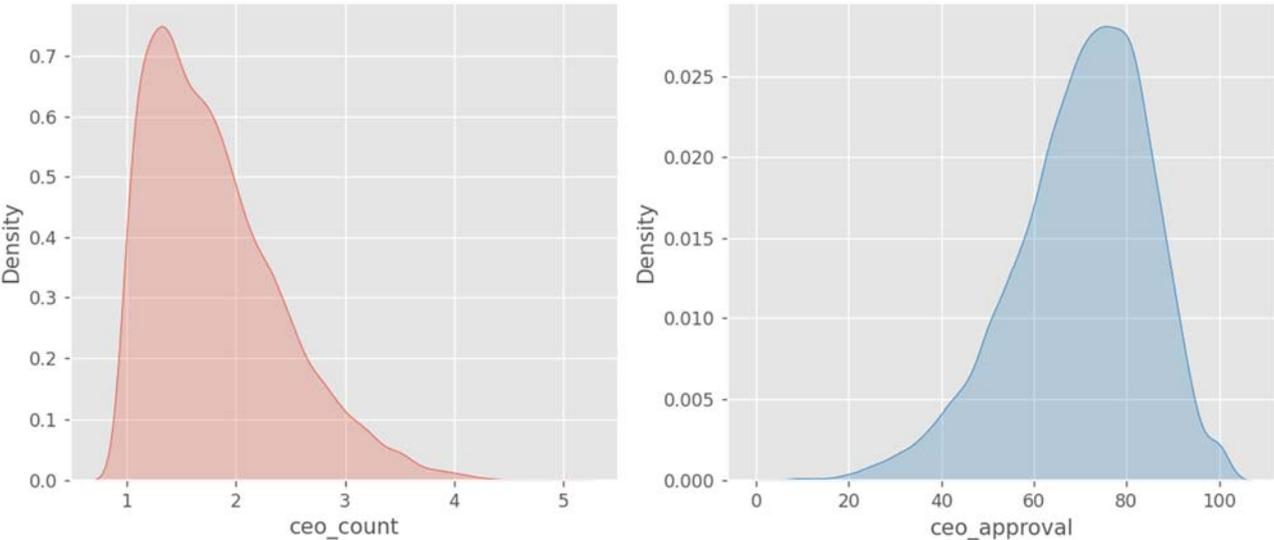
Figure 5.6: Distribution (density) of ceo_count (log10) and ceo_approval.

Another useful insight is to examine the ceo_approval percentage distribution against each industry, which can be done by creating the boxplot of Figure 5.7. The industries shown in the boxplot are the top 30 (most popular of all in terms of number of occurrences).
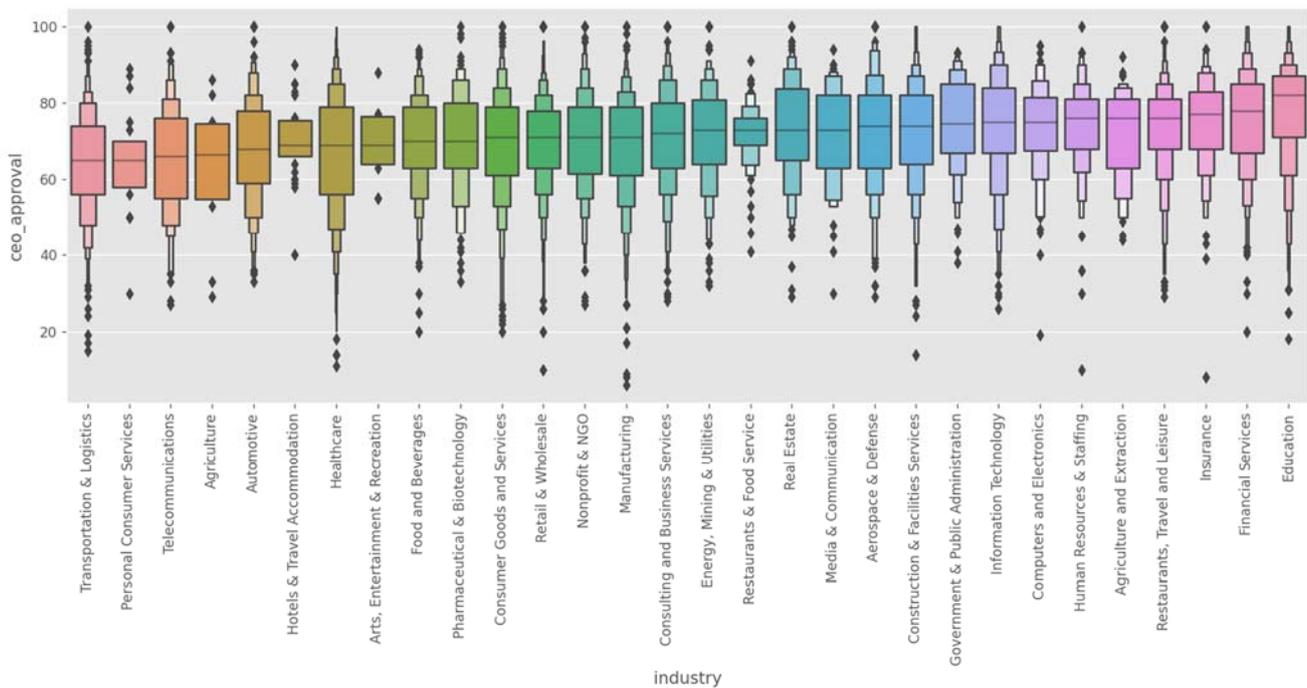
Figure 5.7: Boxplot of CEO approval vs. Industry for the top 30 most popular industries.

By looking at the boxplot, it is obvious that the median of the CEO approval is quite stable across all industries. This could be due to several reasons:

- **General satisfaction**: In most of the industries, the employees may have a general satisfaction level (baseline) with their respective CEOs, which leads to a similar median of approval rates.

- **Central tendency**: The median is regarded as a measure of central tendency, and it is less sensitive to any extreme values (outliers) than the mean. This is a property that can easily lead to a more stable median across the variety of industries even with the existing differences in the spread (spectrum) or the extreme values of the CEO approval ratings.

- **Industry-independent factors**: Due to the vast differentiating characteristics of companies, the factors affecting the CEO approval ratings may be closely related to the company size, the culture, the management style, or the leadership qualities rather than the specific industry itself. Mergers are actually making this an even harder challenge to confine the semantics (the anticipated characteristics) under a sole title. As a result, the median of the CEO approval rate may not vary significantly across various industries.

5.3. 1st pass – Exploratory Analysis

In a similar fashion to what was done for the ceo_approval and ceo_count, we do for the columns of "reviews" and "rating" in order to examine the count of reviews in relation to the distribution of the (overall) rating occurrence. The results of the calculations are depicted in Figure 5.8. The plotting of the reviews has been applied in a log10 scale since it has large variation (median 44, average 120).
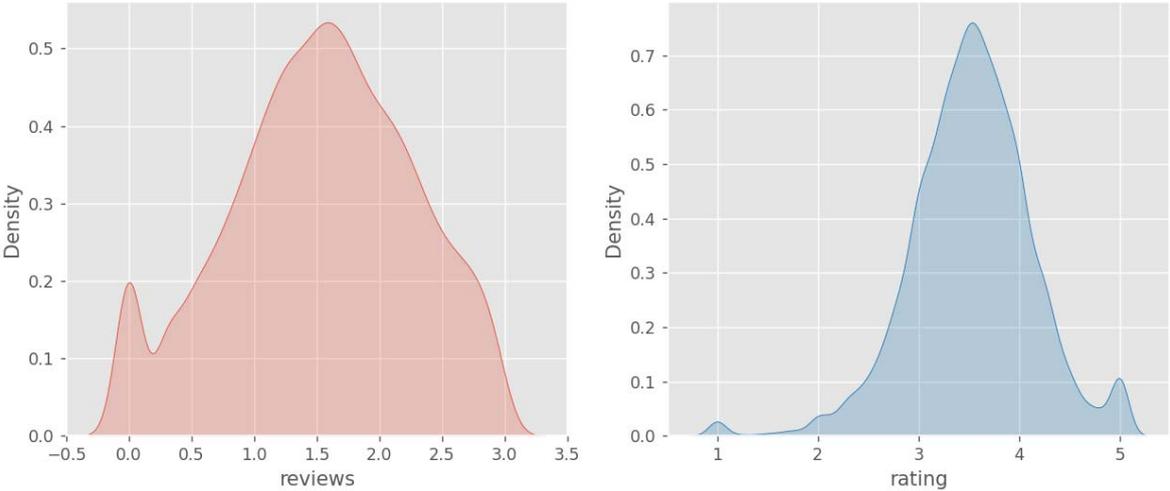
Figure 5.8: Distribution (density) of reviews (log10) and rating.

An important observation is depicted in Figure 5.9, which depicts a scatterplot between the records of the columns ceo_approval and (overall) rating.
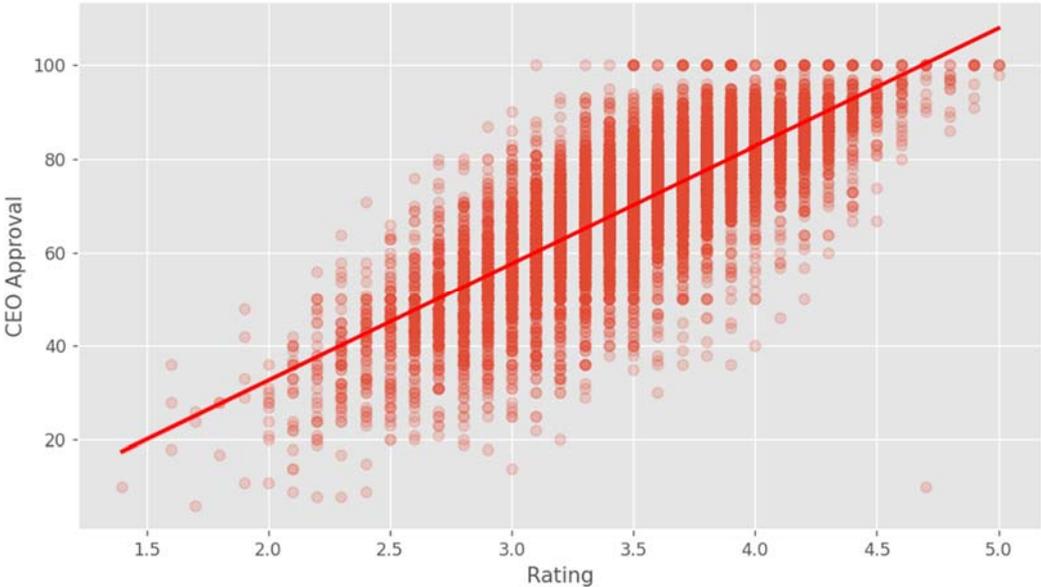
Figure 5.9: Scatterplot depicting the correlation between CEO Approval and (overall) Rating.

The regression line was added to stress out the quite strong positive correlation that seems to exist between the two dataset dimensions. The appearing positive correlation between these two dimensions exists because, as it has already been mentioned in the preceding sections of this work, a CEO's leadership style, vision, and decision-making can have a significant impact on employees' overall experience and satisfaction at work. When a company's employees feel that their CEO is competent, supportive, and transparent, they often feel doing business in a deeply positive work environment, sharing higher employee morale, and presenting better company performance. In turn, employees are more likely to give their companies higher ratings. On the other hand, if a CEO is perceived as quite ineffective or that lacks in leadership qualities, his/her employees may feel less satisfied with their duties in the work place and rate their companies lower. This relationship between the CEO approval and the employee rating is not surprising as the CEO plays a critical role in shaping a company's culture and work experience.

To further examine if the salary plays any significant role to employee satisfaction, and thus review the "rating", we process the unit "salary" data, converting its records to yearly salary and then comparing each row against the "rating", using the scatterplot of Figure 5.10. It is apparent (and justified by the literature review) that there is no strong correlation between these two dimensions.
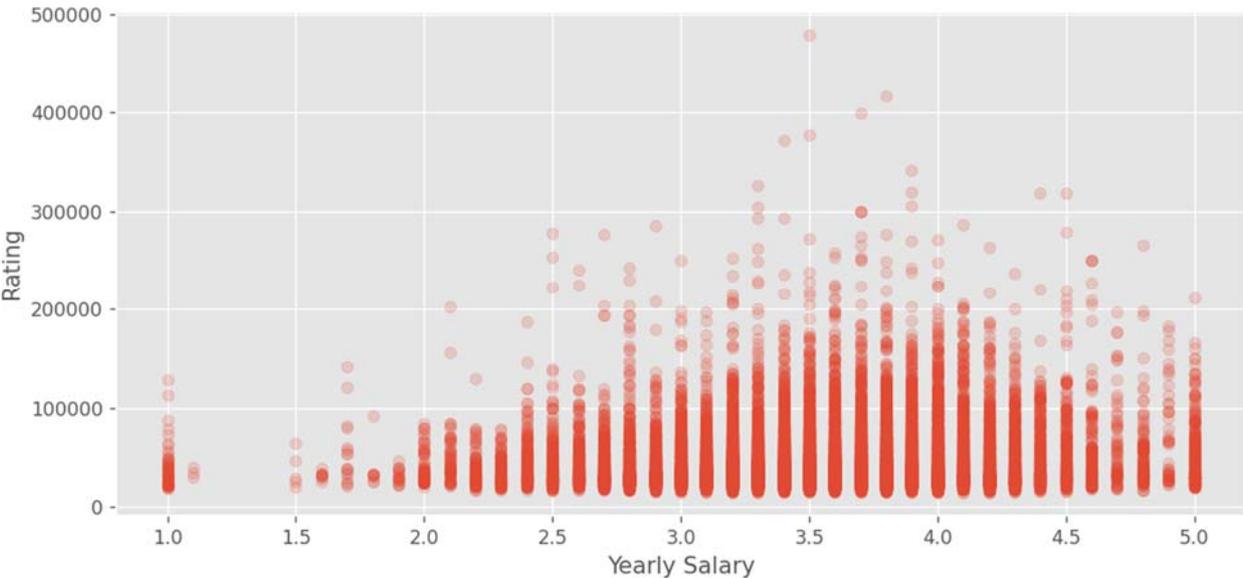
Figure 5.10: Scatterplot depicting the correlation between Yearly Salary and (overall) Rating.

A very important column of the given dataset, that is very much worth diving into, is "Happiness", which is a column that includes tabulated data in the form of free text (string). This string can be separated into 13 distinctive (nominal) sub-categories that, according to the Glassdoor.com schema, intend to record employee happiness.

These 13 sub-dimensions are:

- Work Happiness Score,
- Achievement,
- Learning,
- Flexibility,
- Support,
- Compensation,
- Purpose,
- Appreciation,
- Management,
- Inclusion,
- Energy,
- Trust, and
- Belonging.
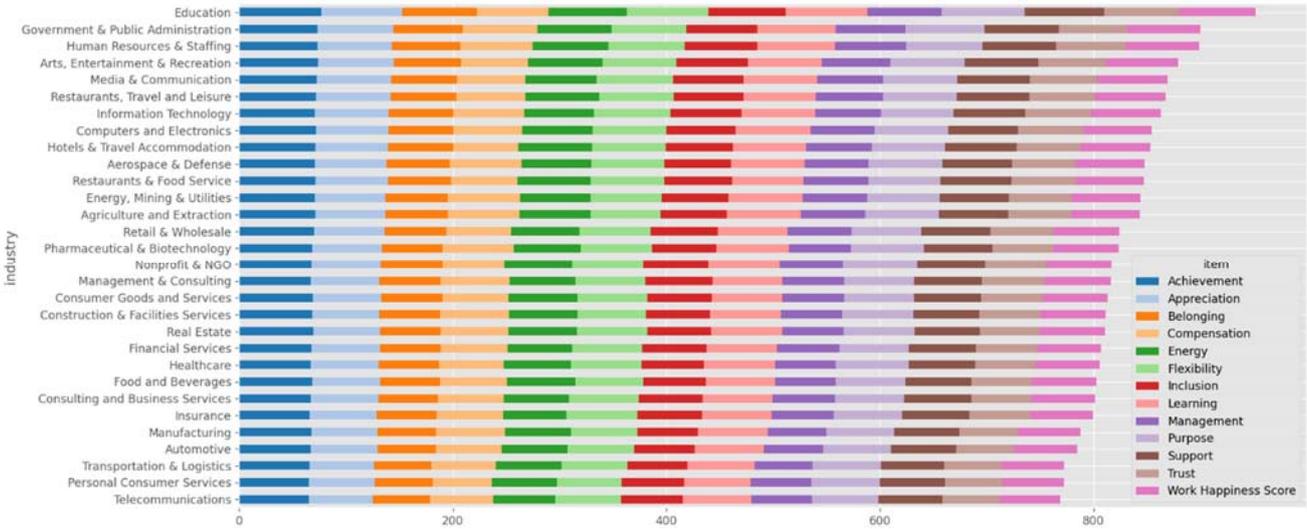
Figure 5.11: Contribution of all Happiness sub-dimensions across the top-30 most popular industries.

The idea is to examine whether these sub-dimensions vary across various fields of industry. In order to present the contribution of each sub-dimension on each

industry and in a fashion that would enable the comparability of all industries, the bar-graphs of Figure 5.11 were created for the 30 most popular industries.

Please note that each sub-dimension is expressed in a percentage, i.e., an integer number value ranging from 0 to 100. Also, please note that the visual result of Figure 5.11 which depicts the contribution (score) of all Happiness sub-dimensions across the top-30 most popular industries, was quite similar for all industries, not just the top-30 in popularity. However, the rest of the industries were not represented in the graph in order to keep an aesthetically pleasing and visually interpretable result.

As regards the particularities of the "happiness" column, it is quite anticipated to observe a positive correlation between its sub-dimensions. We also expect to observe a correlation between each one of "happiness" sub-dimensions and "rating". Figure 5.12 justifies this expectation by the strong positive correlation that is depicted in its scatterplot between the sub-dimension of "happiness" called "Work Happiness Score" against the dimension of "rating".
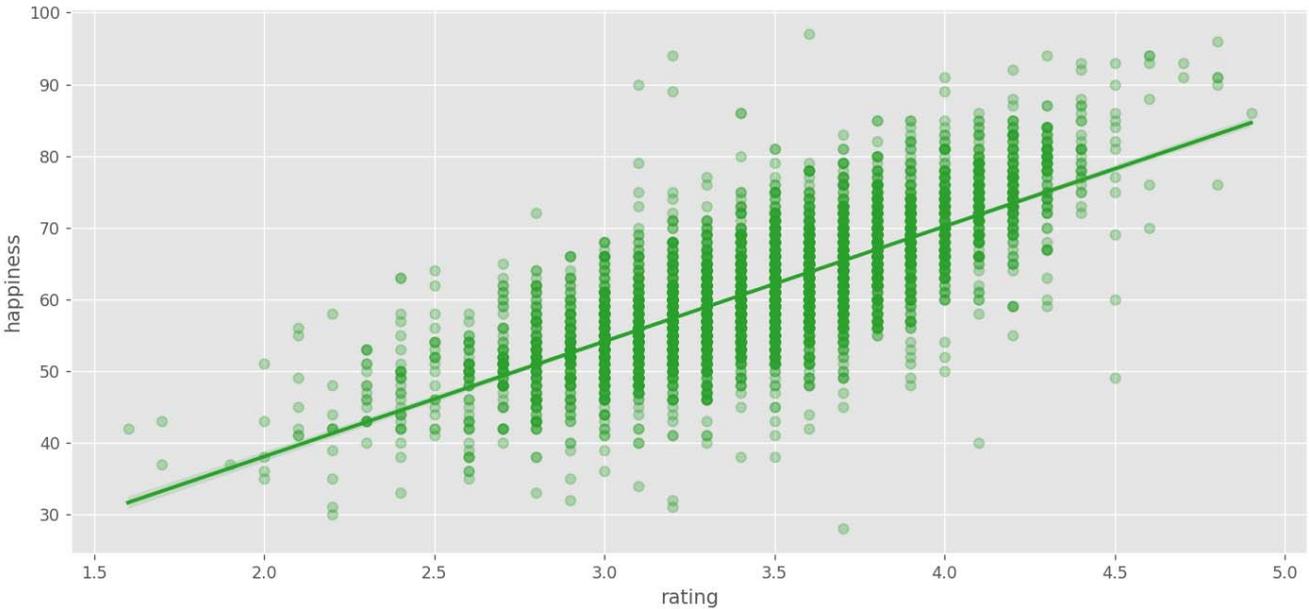
Figure 5.12: Scatterplot depicting the correlation between Work Happiness Score and (overall) Rating.

The two aforementioned dimensions present a strong positive correlation because both of these metrics (in the general case) reflect employee satisfaction and their overall experience at a company. A higher work happiness score will typically

mean that the employees are content with various aspects of their job, like work-life balance, their labor compensation, the work environment, and the career growth opportunities. Similarly, a higher rating given to a company is an indicator of a positive employee perception for the respective company, its management, and its practices/policies. When the employees are happy with the work they produce, they are more likely to reciprocate by rating their company higher. In contrast, if the same employees are unhappy or quite dissatisfied with their job, then they tend to provide lower ratings to their boss or company. This directly leads to the observation of a positive correlation between "work happiness score" and "rating".

Another aspect worth examining is the distribution of happiness ("work happiness score") against various "roles". The "roles" (csv column) is a dimension of the given dataset which includes information about the ratings that various roles (professions) have attributed to their company being reviewed (csv row). The data in this field are given as 'free text', so some pre-processing is important for separating the available roles and for assigning each one to a rating (ratio data type).

After dropping the records (csv cells) that miss information in the respective rows of "happiness" we proceed combining roles (called "titles" here) that are similar across all companies by grouping them together. Then we calculate an average (a mean value) for each job title. The calculations generate a bar plot of the mean "work happiness score" for job titles that have greater than 20 records in the dataset.
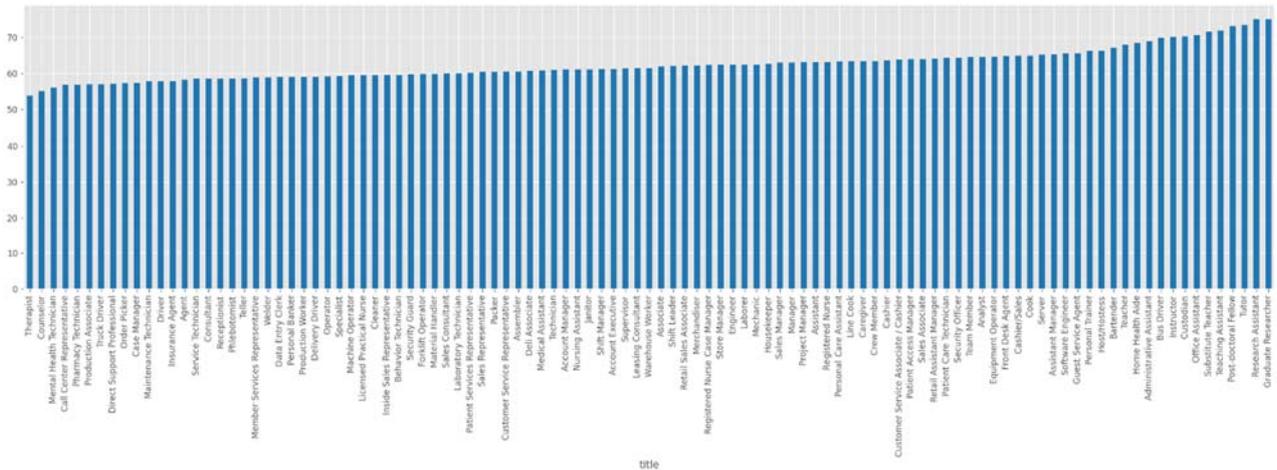
Figure 5.13: Bar graph of the distribution of Work Happiness Score against various Titles (Professions).

The results reveal that employee happiness is higher for the following titles: Graduate Researcher, Research Assistant, Tutor, Post-doctoral Fellow, Teaching-Assistant, Substitute Teacher, Office Assistant, Custodian, Instructor, Bus Driver, Administrative Assistant, Home Health Aide, Teacher, Bartender etc.

The results also reveal that employee happiness is lower for the following titles: Therapist, Counselor, Mental Health Technician, Call Center Representative, Pharmacy Technician, Production Associate, Truck Driver, Direct Support Professional, Order Picker, Case Manager, Maintenance Technician, Driver, Insurance Agent, Service Technician, Consultant, Receptionist, Phlebotomist etc.

5.4.   2nd pass – Feature Processing

After the initial data review and pre-processing pass, two interesting questions are arising, complementary to the main goal that is predicting the company rating by the most indicative dimensions; a) is there any chance that the compensation is an important factor taken into consideration when reviewing a company? b) are there any other factors (like what constitutes a company culture) which play a more significant role to more enthusiastic company reviews? The following paragraphs aspire to shed more light into these questions and that kind of data relations.

Removing Irrelevant Dimensions (columns): Towards that new quest, which eventually aspires to lead to liable company rating predictors, there are a few irrelevant data (columns) blocking the way and which seem best to remove. These are the columns "roles", "salary", "locations", "description", "name", "website", "headquarters", "interview_experience", "interview_difficulty", "interview_duration", and "interview_count". This is because these columns contain data that are not related to rating (which is what contributes to the predictor). After their removal, the original full dataset decreases from 17050 records down to 15494 records.

Another essential extra step that can be taken in order to increase the results' validity is to remove any companies that have a minimum number of reviews. This is attempted so that their weighting contribution, which may be much smaller than

the rest of the group, will not interfere with the most credible data. By removing all of the companies that present less than 20 reviews in the "reviews" column the resulting new dataset records are minimized down to 10916. This dataset still leaves plenty of room for credible correlation calculations and for fruitful post-processing.

Attributing weighted credibility: The next step is to create a weighted measure of the "ceo_approval" field based on the "ceo_count". For this, firstly we replace the non-populated fields of "ceo_count" with a numeric value (0). Then the "ceo_count" percentage is converted into an integer and finally the initial "ceo_approval" value is replaced with its calculated weighted mean (based on "ceo_count"). This data manipulation will rather enhance (instead of skew or corrupt) a) the information of this dimension and b) its inherent data relations with the rest of the dataset.

Extracting extra sub-dimensions: The column "ratings" (plural) includes five (5) more sub-dimensions in the form of free-text, which is formatted as five pairs of nominal data types:

- Work/Life Balance,
- Compensation/Benefits,
- Job Security/Advancement,
- Management, and
- Culture,

Each one of them is delimited with commas and followed by ratio data types (numeric values of one decimal point resolution ranging from 0.0 to 5.0).

Replacing nominal data: The column "industry" although highly populated, includes some fields that have no data (null or NaN). Such fields are replaced with the nominal value (string) 'unknown'.

Removing records (rows) with missing data: Two extra columns that can be used for weighting the results on a later stage are the "employees" dimension (which includes nine (9) nominal values about the company size, given as strings unveiling the range of employee count), and "revenue" dimension (which include nine (9) nominal values about the company's revenue).

The nine nominal values of the "employees" dimension are:

1. 1 employee,
2. 2 to 10 employees,
3. 11 to 50 employees,
4. 50 to 200 employees,
5. 201 to 500 employees,
6. 501 to 1000 employees,
7. 1001 to 5000 employees,
8. 5001 to 10000 employees, and
9. 10000+ employees.

The nine nominal values of the "revenue" dimension are:

1. less than $1M (USD),
2. 1Mto15M (USD),
3. 5Mto525M (USD),
4. 25Mto25100M (USD),
5. 100Mto100500M (USD),
6. 500Mto5001B (USD),
7. 1Bto15B (USD),
8. 5Bto510B (USD), and
9. more than $10B (USD).

Both dimensions are highly populated in the given dataset so the impact for removing any rows (companies) that have NaN in any of these two columns, will be minimal. After the deletion the dataset records decrease to 7823.

5.5.   2nd pass - Exploratory Analysis

The paragraphs to follow aim at summarizing, visualizing, and promoting a better understanding of the main characteristics that dominate in the given dataset. One of the first tools to use in this section is the correlation matrix, which provides a comprehensive view of the linear relationships between pairs of dimensions ("rating", "ceo_approval", "employees", "revenue" and the five sub-dimensions of "ratings"). The correlation matrix of Figure 5.14 contains the correlation coefficients

in graded coloring to visually 'quantify' the strength and direction of these relationships.
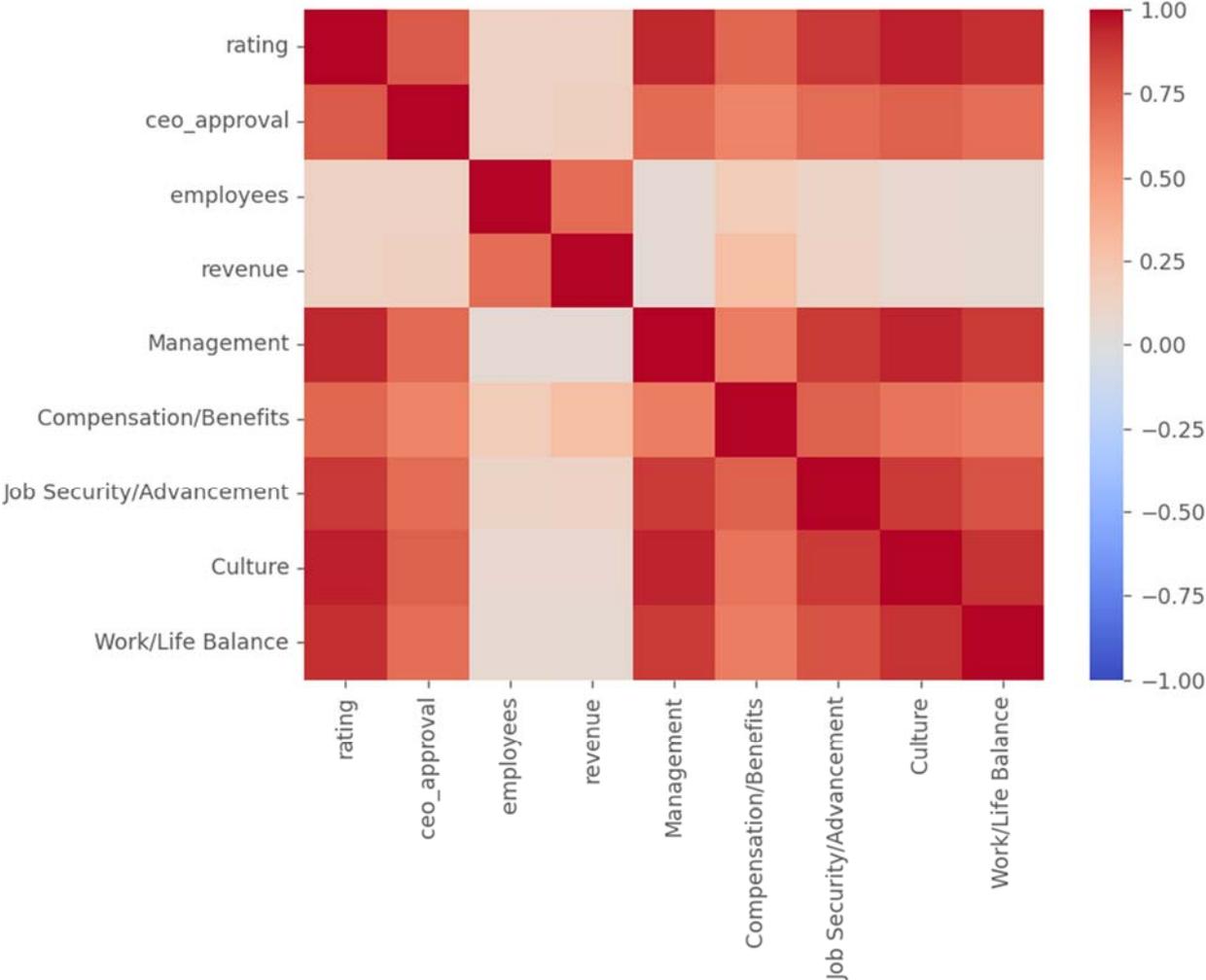
Figure 5.14: Correlation matrix depicting strong and weak relations using a heat-map visualization.

By examining the correlation matrix of Figure 5.14 we gain valuable insights into which variables are positively or negatively correlated. This allows for an easier identification of potentially useful dimensions that can lead to successful predictors, or that are characterized by multicollinearity or/and areas for further investigation.

As we can observe, the five sub-dimensions of "ratings" present a very small correlation with company size ("employees") and the company's revenue ("revenue"). A positive correlation between the other features is also apparent (and stronger) and it will be analyzed in detail in the paragraphs to follow. In overall, such

an exploratory analysis using correlation matrices plays a crucial role in laying the foundation for more advanced data analysis techniques in a pursuit towards well-informed decision-making. Any further processing will now focus in identifying trends, patterns, anomalies. We will proceed to unveil the potential relationships that may exist between the variables we just laid our interest upon, in order to develop a credible model (predictor) and to further generate data-driven questions.

1. 1 employee,
2. 2 to 10 employees,
3. 11 to 50 employees,
4. 50 to 200 employees,
5. 201 to 500 employees,
6. 501 to 1000 employees,
7. 1001 to 5000 employees,
8. 5001 to 10000 employees, and
9. 10000+ employees.
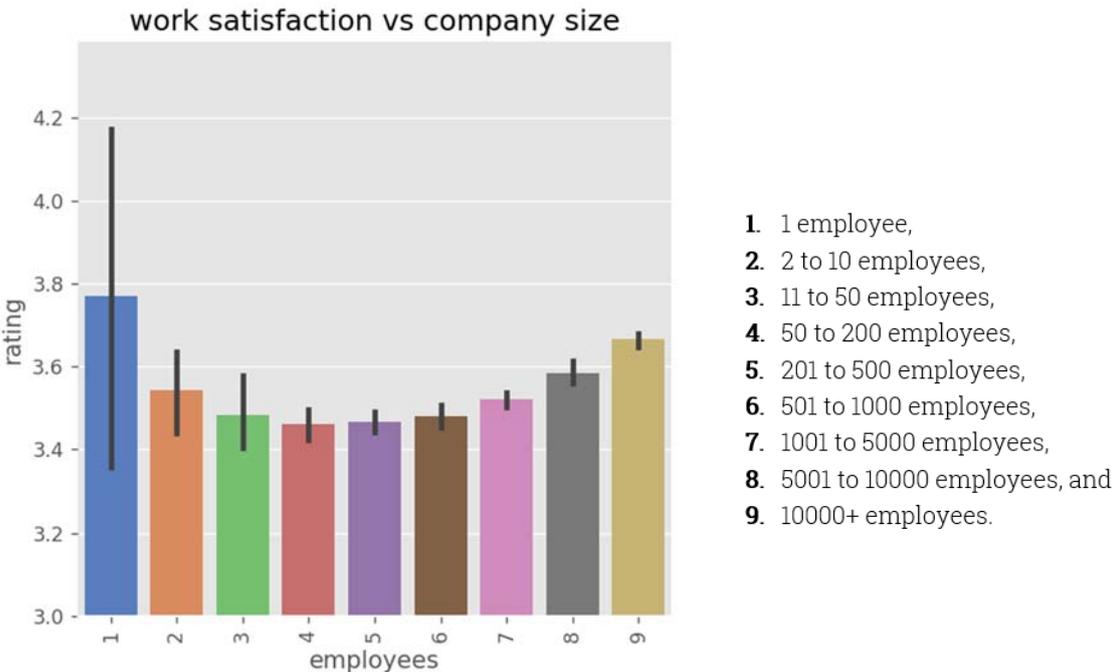
Figure 5.15: Relation of company rating to company size (employees).

work satisfaction vs company revenue

1. less than $1M (USD),
2. 1Mto15M (USD),
3. 5Mto525M (USD),
4. 25Mto25100M (USD),
5. 100Mto100500M (USD),
6. 500Mto5001B (USD),
7. 1Bto15B (USD),
8. 5Bto510B (USD), and
9. more than $10B (USD).
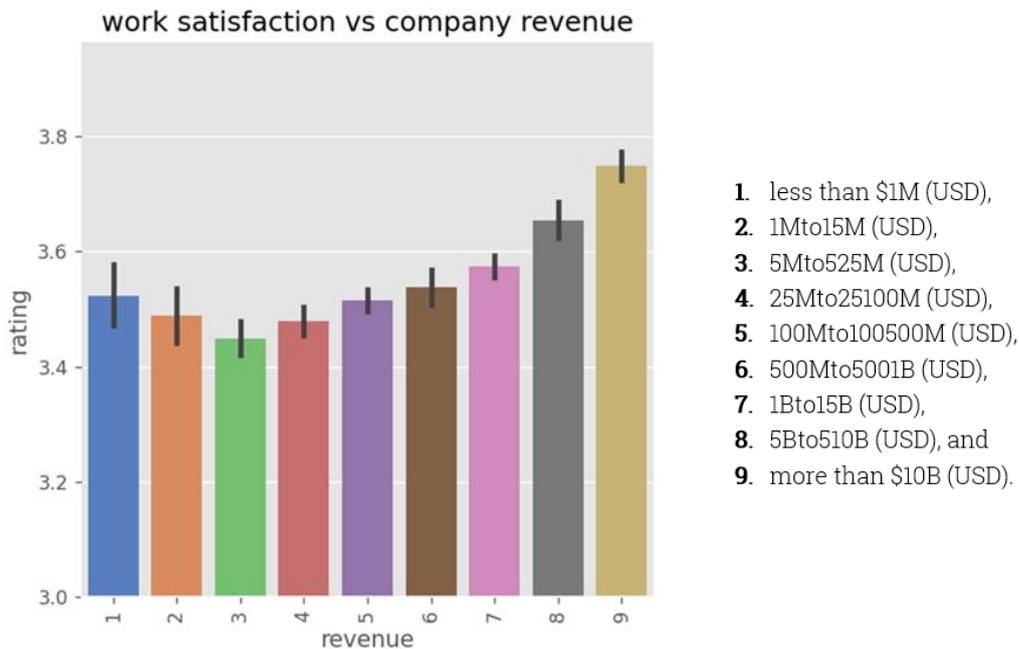
Figure 5.16: Relation of company rating to company revenue.

Since the two dimensions of "employees" and "revenue" present a very small correlation with all other dimensions of "ratings" (plural), it is advisable to further examine the column "rating" in relation to each one of them.

Figures 5.15 and 5.16 present the relation of rating increase in respect to company size and company revenue, respectively. It is important to note that in both figures the y axis begins from value 3, not zero, in order to visually enhance the differences of the bar heights and facilitate the comparisons. In both figures, the bar-graphs begin from the smallest values (ranges) on the left and end to the highest values (ranges) on the right. The texts were replaced with numerical values for Python to be able to properly sort the ranges. The assigned scale mapping from text to numbers can be found on the right of each bar-graph and also lays in the numbered lists (with 9 items each) at the end of section 5.4.

As we can observe, the company rating is higher for small and larger companies and the same goes for the company rating being higher for companies with either too small or too high revenue. The variance of rating is larger in both cases for the small companies and the ones who have lower revenues. In this sense it is quite safe to assume that there is a slight increase in the company rating for companies bigger than 50 employees that present revenues of more than 5M.

The justification for these observations can be attributed to several factors, as explained in the bullets below:

- **Higher ratings for small and large companies**: Most of the smaller companies often make use of close-knit teams, which allows employees to form some strong bonds and to enjoy a more personalized work environment. This, in turn, contributes to higher job satisfaction and thus ratings. On the other hand, the larger of the companies quite often offer more resources, better benefits, and faster growth opportunities, which may also lead to higher ratings.

- **Higher ratings for companies with either low or high revenue**: The companies with the lower revenues, often the smaller ones, tend to provide a more flexible and a more relaxed work environment, which can lead to higher ratings. In contrast, there are companies with quite high revenues which usually have the financial stability to provide better benefits and/or resources and/or opportunities for development and growth, which can also potentially lead to higher employee satisfaction and ratings.

- **The ratings variance is larger for small companies and those with lower revenues**: The smaller companies (also those with lower revenues) may experience fluctuations in performance and/or work environment and/or management practices, which leads to a wider range of employee experiences and satisfaction levels. This can appear as a larger variance in ratings.

- **Slight increase in company rating for companies with more than 50 employees and revenues over 5M**: This is an observation that suggests that as companies grow beyond a certain size and a revenue threshold, they may be able to provide much better benefits and/or resources and/or opportunities for their employees, which seems to lead to a slight increase in company ratings.

In conclusion, the relationship between **company size**, **revenue**, and **company rating** can be attributed to factors like the work environment, the in-work benefits, the extra provided resources, and the opportunities for growth and development. Corporations/organizations can improve their overall employee satisfaction and ratings by understanding the importance of such factors.

Before proceeding to further analyze the other dimensions of the problem, it is recommended to produce some useful insights about the **multicollinearity** of the "ratings" sub-dimensions and their correlation with the "rating" dimension. Figure 5.17 presents the correlation between all these dimensions. Please note that the correlation coefficient color range is selected to be from -0.5 to 1.0 As expected, the dimension Compensation/Benefits is the least correlated with the company rating. On the other hand, the dimensions of Culture and Management seem to be correlated with the company rating quite tightly. These insights are very useful for the model design of the predictors that will be developed in the following section. This finding is important because in linear regression modeling, predictors are the independent variables that are used to estimate the dependent variable or target. So, it is crucial to select the appropriate dimensions, as they significantly influence the model's accuracy, its generalizability, and its interpretability, allowing for a better understanding between the independent and dependent variables' relationships.
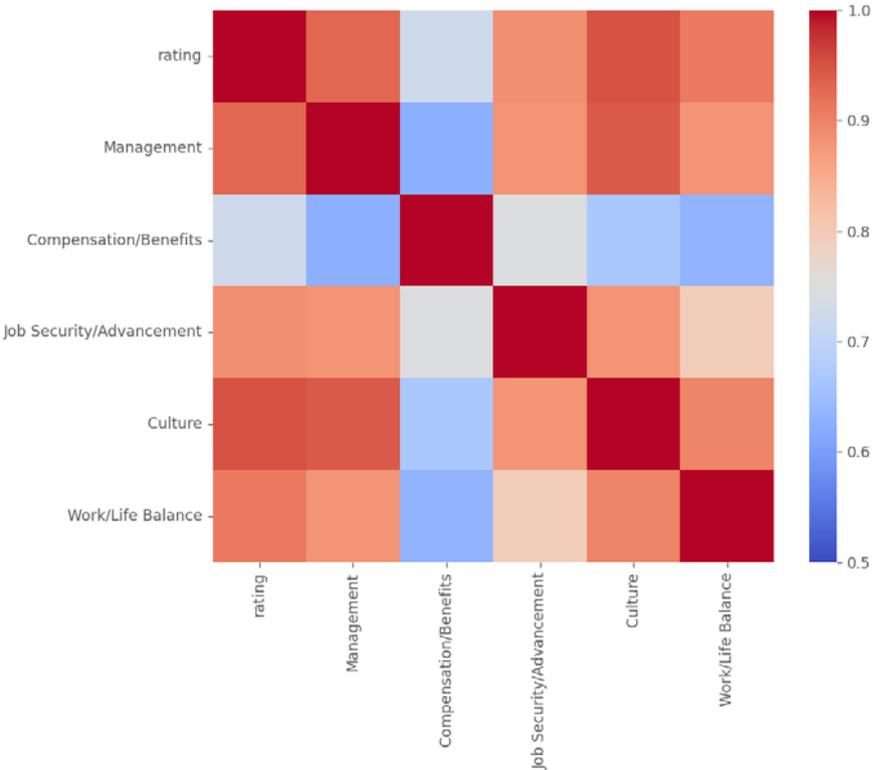
Figure 5.17: Correlation matrix depicting strong and weak relations using a heat-map visualization.

Moving on to other dataset dimensions, it is regarded wise to proceed in examining the relative effect of the "industry" sector on the "rating". In order to

receive some credible results in this front, we will need to ensure that the industries that will be taking part in this representation will be the ones that include 20 or more companies in their (respective) groups. It is more preferable to present data with a sufficient number of items in their sets for both visualization purposes and for drawing safer conclusions, as such a provision can enhance the reliability and the validity of the findings. It is important to mention though that larger data sets can better represent the underlying population and can reduce the influence of outliers, thus ensuring that the conclusions drawn will be robust and statistically significant.

Figure 5.18 and Figure 5.19 include the graphs that show the 15 industry sectors with the highest and the lowest rating, respectively.
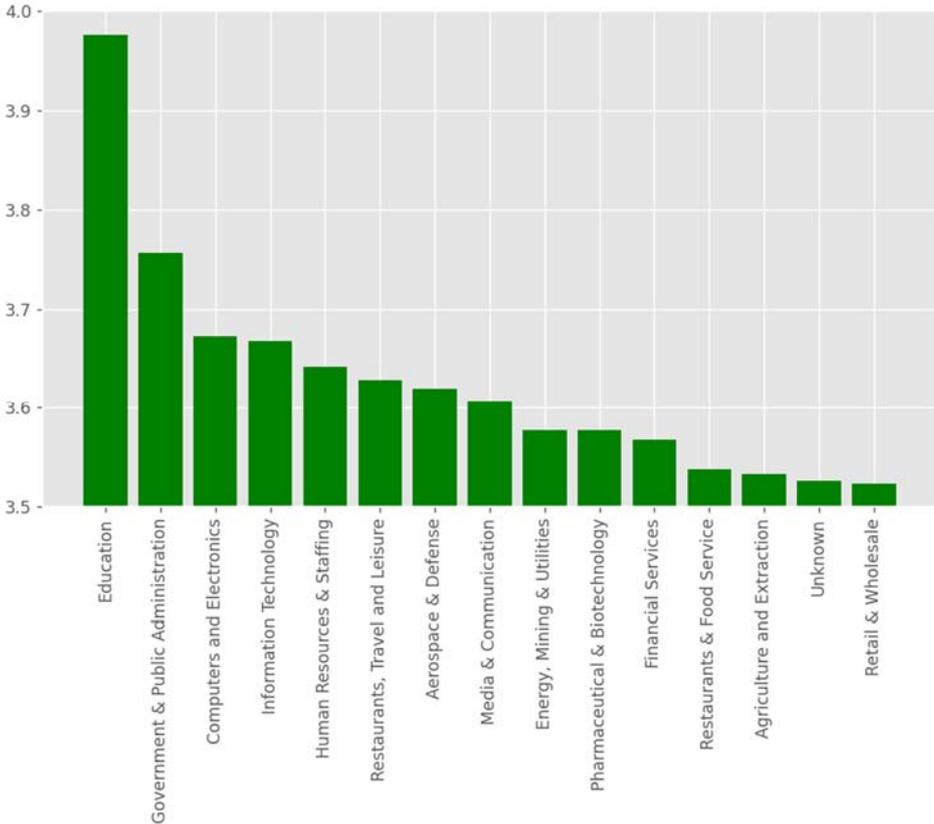
Figure 5.18: The 15 industry sectors with the highest rating (for companies with more than 20 records).
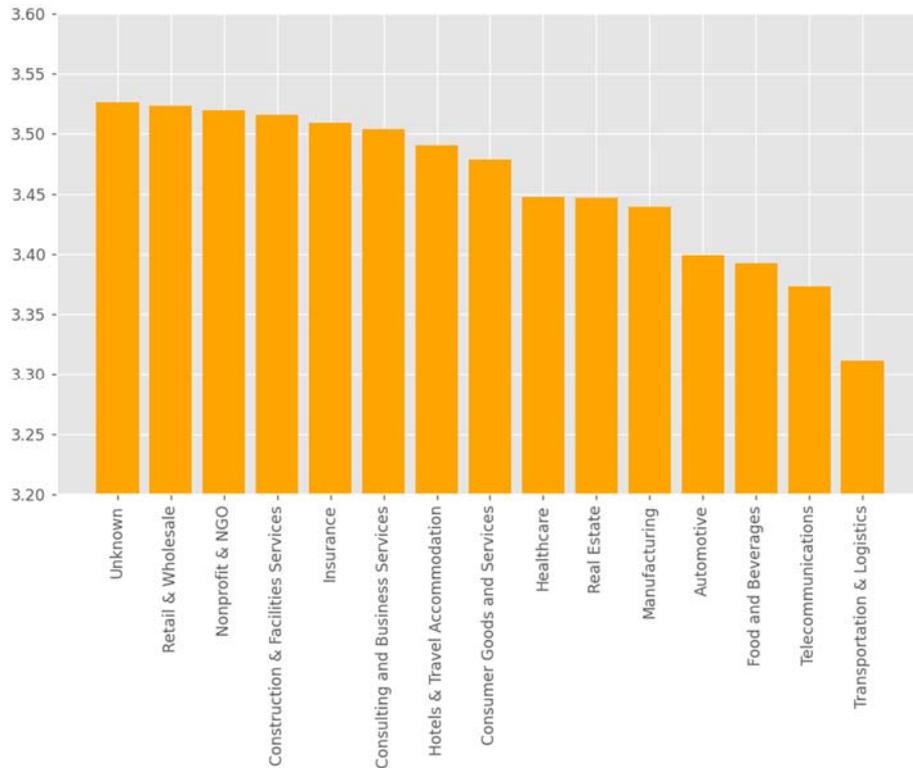
Figure 5.19: The 15 industry sectors with the lowest rating (for companies with more than 20 records).

Next, it is advisable to check the correlation of "ceo_approval" and "rating" in a scatterplot. Similarly, we examine the correlation of the sub-dimension "Compensation/Benefits" and "rating".
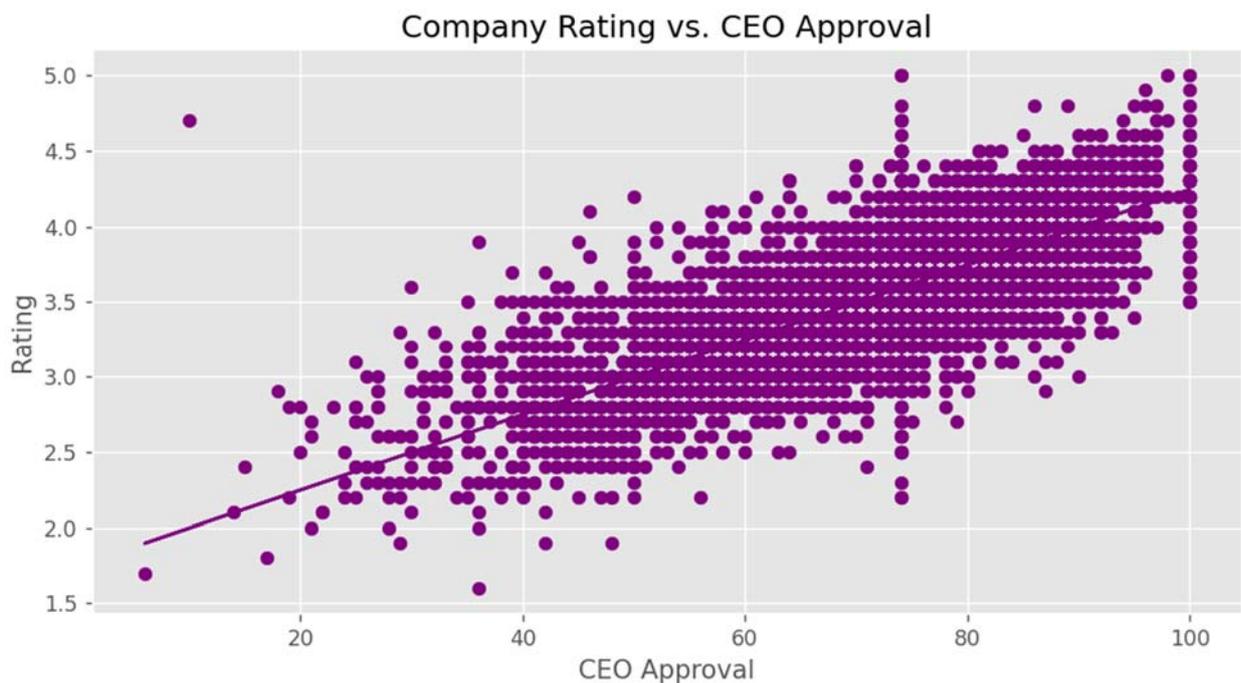
Figure 5.20: Scatterplot depicting the correlation between CEO Approval and (overall) Rating.
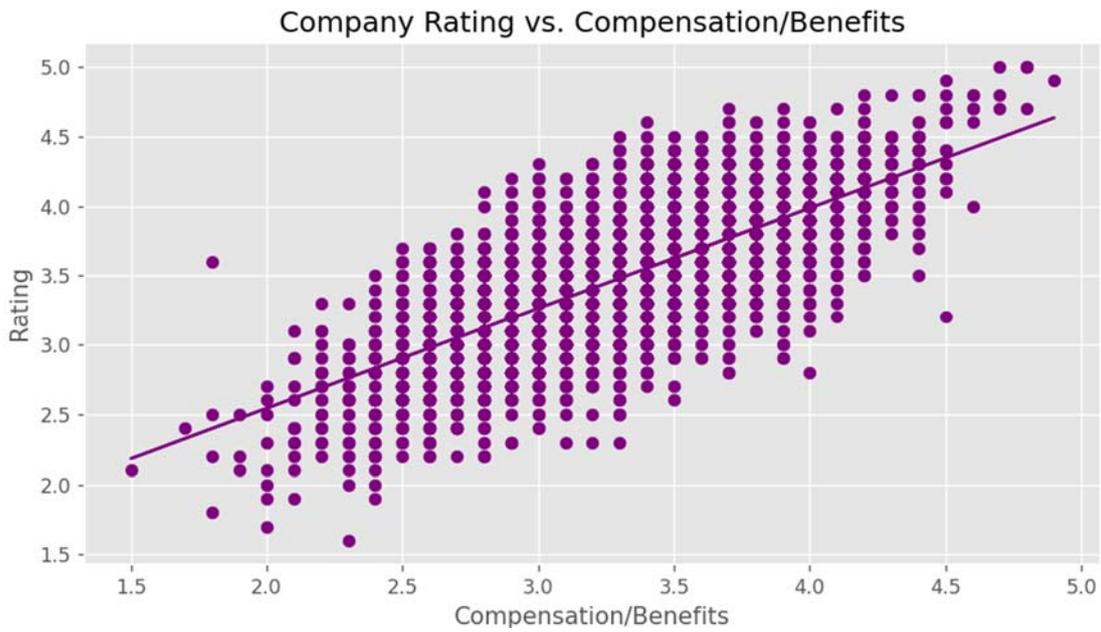
Figure 5.21: Scatterplot depicting the correlation between Compensation/Benefits and (overall) Rating.

Both figures have scatterplots which include an overlaid line (linear polyfit, made by a 1st degree polynomial) in order to visually express the positive correlation between both pairs, although the variance is quite large on both illustrations.

At this point it is advisable to examine the distribution density of "rating" versus the rest of the sub-dimensions of "ratings" (plural), which seem to be correlated to a satisfactory degree.
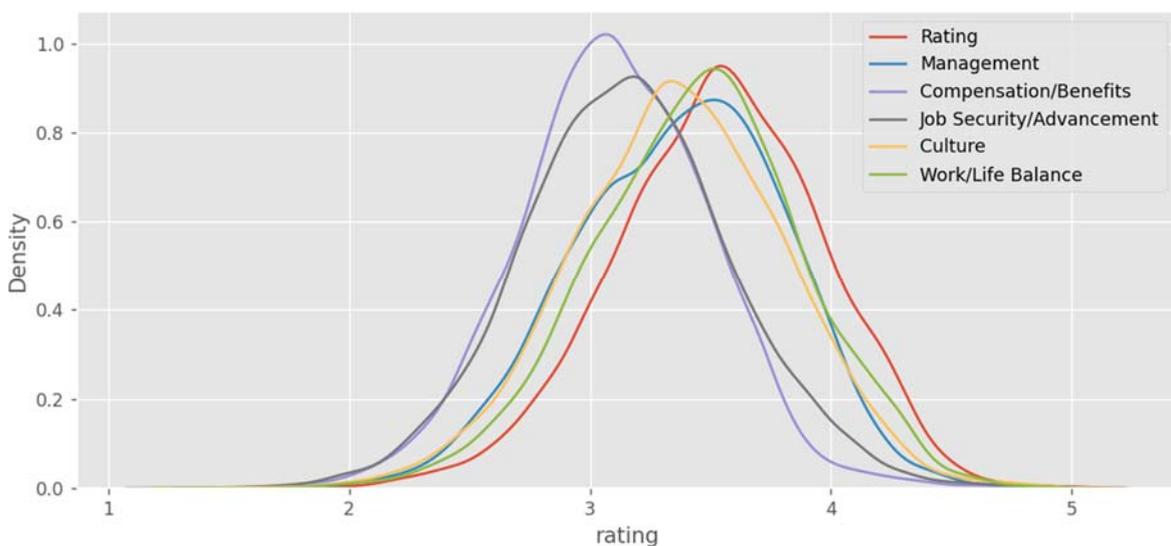
Figure 5.22: Density Distribution of "Compensation/Benefits", "Job Security/Advancement", Culture", "Management", "Work/Life Balance" and overall "Rating".

Based on the curves depicted on graph of Figure 5.22, we can infer the following conclusions:

- The ratings for "Compensation/Benefits" and "Job Security/Advancement" tend to be lower compared to the other factors. This suggests that, on average, employees may be less satisfied with their compensation and job security compared to other aspects of their jobs.

- The ratings for "Culture", "Management", "Work/Life Balance", and overall "Rating" are closely clustered together. This indicates that employees generally have similar levels of satisfaction with these aspects of their jobs, and these factors may be more important drivers of overall employee satisfaction.

- The distribution of ratings for all factors seems to follow a bell curve, suggesting that there is a wide range of satisfaction levels among employees. However, the majority of ratings are concentrated around the mean, with fewer employees at the extreme ends of satisfaction or dissatisfaction.

It is always essential to consider the context and the specific dataset when drawing such conclusions, as these may vary, as we have already witnessed regarding the variance of the extreme cases, across industries (education), companies, or job roles (Researchers, Tutors etc.).

Finally, as shown in section 5.3, "happiness" has 13 distinctive (nominal) sub-categories which would be nice to explore further in order to produce useful insights regarding

    a. the multicollinearity of the "happiness" sub-dimensions and
    b. their correlation with the "rating" dimension.

The "happiness" dimension is populated to about 30% but it is still considered to be an important dataset part that is loaded with great significance in this research. Figure 5.23 presents the correlation of these sub-dimensions and "rating".

The results show that there is a high correlation almost across all the "happiness" sub-dimensions expect the "Compensation". The highest correlation is presented by the pairs:

- "Appreciation" – "Support",
- "Trust" – "Belonging", and
- Work Happiness Score" – "Management".

It is also important to note that the sub-dimension of "happiness" called "Work Happiness Score" and "Appreciation" present quite strong connections with all the others and the same happens with the "Inclusion", "Belonging" and "Management". The interconnections of these notions most probably emerge from the language semantics themselves, i.e., they ontologically inherit a pre-programmed rhetoric representation, the one that is hardwired with our common emotional expressions.
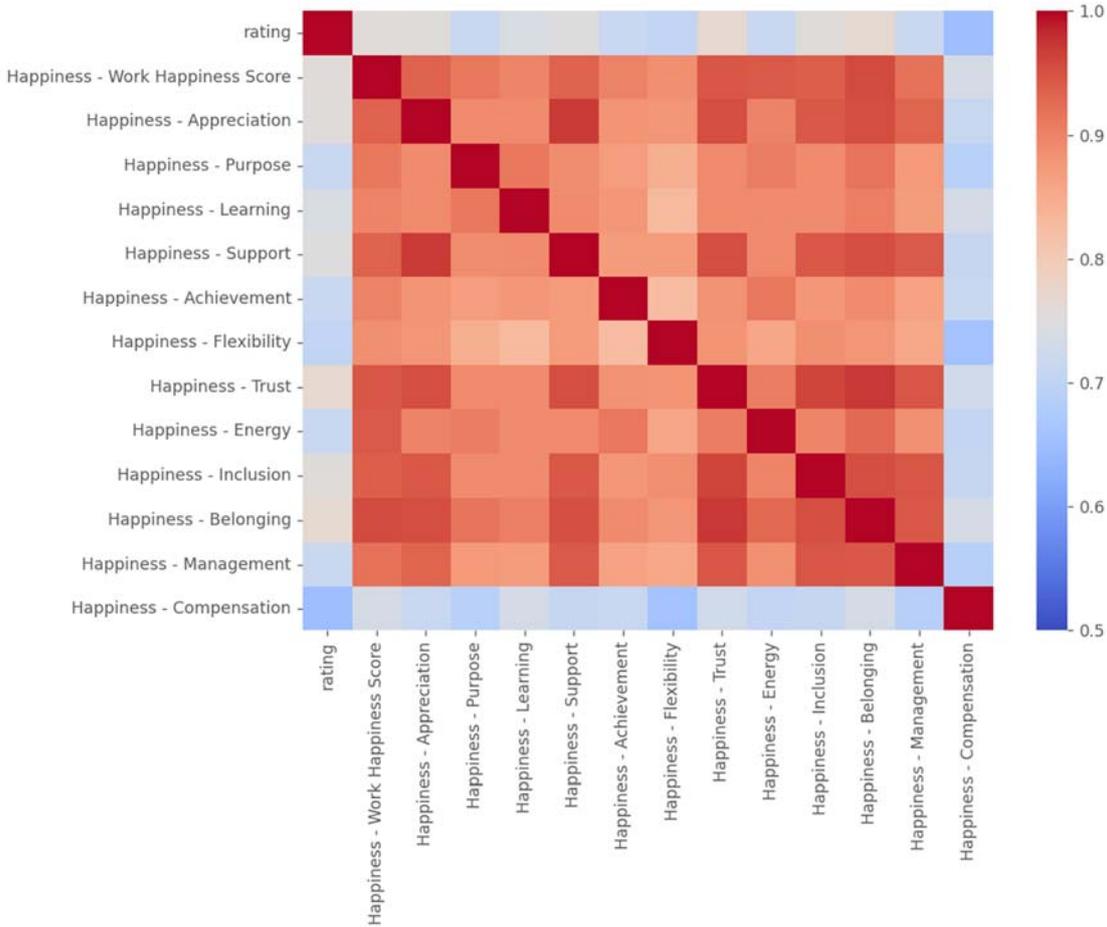
Figure 5.23: Correlation matrix depicting strong and weak relations using a heat-map visualization on all "happiness" sub-dimensions and "rating".

So, the strong correlations that are observed among "Work Happiness Score", vs. "Appreciation", vs. "Inclusion", vs. "Belonging", and vs. "Management" can be attributed to the deeply interrelated nature of these dimensions (aspects, or features) of an employee's work experience. Below we explore the semantic reasons behind these correlations:

- **Work Happiness Score**: This is a composite measure of employee happiness at work which can be influenced by several factors that include appreciation, inclusion, management and belonging. A recorded positive experience in these sub- dimensions is expected to create higher work happiness scores.

- **Appreciation**: When an employee feels directly appreciated and recognized for his/her efforts, this contributes to their overall happiness. This sense of appreciation can foster their sense of belonging and inclusion (team), which in turn, impacts the quality of management and teamwork.

- **Inclusion and Belonging**: An inclusive workplace can nurture a great sense of belonging among employees, as they may feel valued as an integral part of the corporation. These feelings can then directly contribute to happiness and can positively influence their perception regarding the HR management.

- **Management**: Effective management is crucial when building up a positive workplace environment. Good managers will be able to recognize employees' efforts, they will encourage inclusivity, and they will create a sense of belonging, which in turn, significantly impacts employees' happiness and satisfaction.

It is also quite important to note that while these dimensions are quite strongly correlated, **they are not necessarily presenting causal relationships**. However, such relationships do indicate that focusing on these dimensions (features, aspects) can lead to a positive work experience and to higher employee satisfaction in overall.

5.6. Modeling (on the pursuit for a credible predictor)

The goal of this section is to forecast company review ratings using various attributes available in the given dataset, making regression models an appropriate choice. For this reason, the performance of the following models needs to be assessed:

- Linear Regression
- K-Nearest Neighbors Regressor
  (optimizing for the best K value)
- Support Vector Machines Regressor
  (employing both linear and Gaussian kernels)
- Decision Trees (with a maximum depth of 2 and without depth limitation)
- Random Forest Regressor
- Gradient Boosting

In the sub-section to follow we generate five dataset variations which will participate in the exploration for the most appropriate one in modeling.

5.6.1. Datasets Definition

The datasets to be created range from large to small, depending on the count of dimensions (columns) and records (rows) that they hold.

- The 1st one is the full dataset which keeps all its dimensions and sub-dimension intact.
- The 2nd one is a reduced dataset that is comprised of the dimensions that have been analyzed to the aforementioned sections; "ceo_approval", "revenue", "employees", all five (5) sub-dimensions of "ratings", and "industry", but manipulated with one-hot encoding, meaning splitting its categorical data into binary columns where only one element is true (1), per record.
- The 3rd one is a small dataset which holds two dimensions that seem quite related; "Culture" and "Work/Life Balance".
- The 4th one is another small dataset which includes what is conventionally regarded as important; "Compensation/Benefits" and "Job Security/Advancement".

- The 5th one is a medium size dataset that includes "ceo_approval", "revenue", "employees", all five (5) sub-dimensions of "ratings", and all thirteen (13) sub-dimensions of "happiness".

5.6.2. Scaling Data

The next important step is to appropriately scale the dimensions (also called features or aspects) using the MinMaxScaler from scikit-learn in Python, which scales the data so that all the features are ranging from 0 to 1. This is an important action since it guarantees that all features are on the same scale, thus **preventing any single feature dominating the model** due to its larger range of values.

This step is useful because it ensures that the data will be **properly prepared for further modeling**, subsequently allowing for more accurate and reliable model predictions. Properly scaled data can help improve convergence of certain models, especially those that are most sensitive to the scale of the input dimensions, like **Support Vector Machines** (SVM) and **K-Nearest Neighbors** (KNN).

5.6.3. Model Functions, Datasets and Best Predictors per case

The eight model functions to be examined are shown in the bullets below:

- Linear Regression,
- KNN,
- SVM with gaussian kernel,
- SVM with linear kernel,
- Decision Tree regressor with max_depth = 2,
- Decision Tree regressor with unlimited depth,
- Random Forest with depth = 2, and
- Gradient Boosting.

1st Dataset: Using Linear Regression, the effect of the different dimensions (57 in total) is depicted in Figure 5.23. It is easy distinguishing the 9 dimensions that present the highest weights. These 9 dimensions are adopted for the 2nd dataset.

Figure 5.23: 1ˢᵗ dataset: effect of the 57 different dimensions (parameters/features).

In Figure 5.24, the KNN results (Mean Square Error) are being displayed covering a range of K values (from 2 to 12). The optimal result (lowest value) is observed at K=5. This will be used for subsequent future comparisons.

Figure 5.25: 1ˢᵗ dataset: MSE for KNN with various K.

Figure 5.26 compiles and compares all the results that are related to the regressors' performance on the 1st dataset ("full"). For this specific dataset (1st) the best predictors are **Linear Regression**, **both SVM-Rs**, and **Gradient Boosting**.
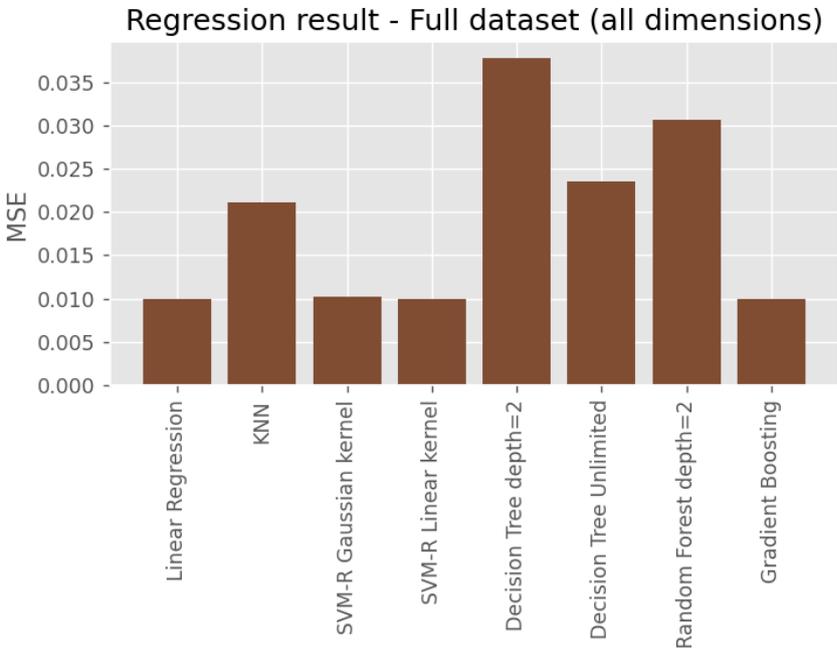
Figure 5.26: Comparison of all regressors for the **1st dataset**.

2nd Dataset: Using Linear Regression, the effect of the 9 dimensions that present the highest weights of Figure 5.23 is shown in the bar-graph of Figure 5.27.
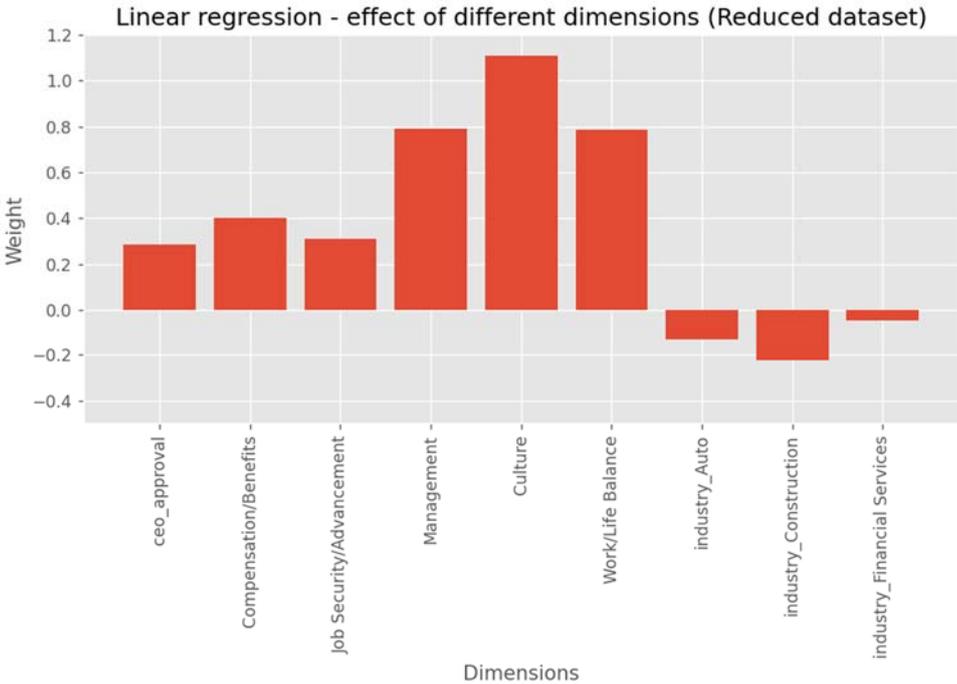
Figure 5.27: 2nd dataset: effect of the 9 different dimensions (parameters/features).

In Figure 5.28, the optimal KNN result (Mean Square Error) is observed at K=11.
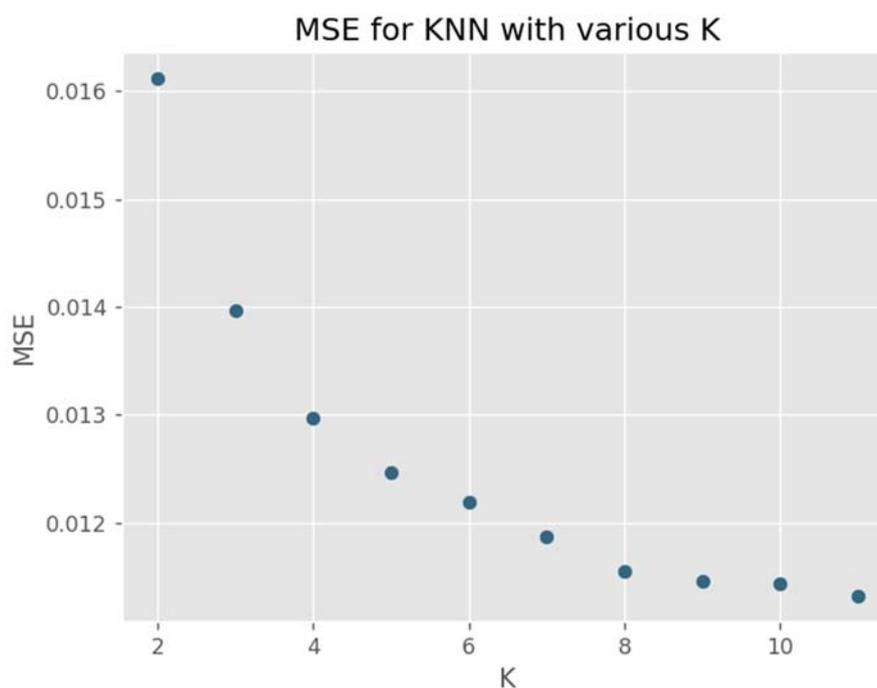
Figure 5.28: 2ⁿᵈ dataset: MSE for KNN with various K.

In Figure 5.29, the 2ⁿᵈ dataset (reduced) best regressors' performance is very similar to the 1ˢᵗ dataset performance. This suggests that the feature reduction had limited effect on the best predictors' performance (**Linear Regression**, **both SVM-Rs**, and **Gradient Boosting**. The only predictor that was negatively affected was KNN.
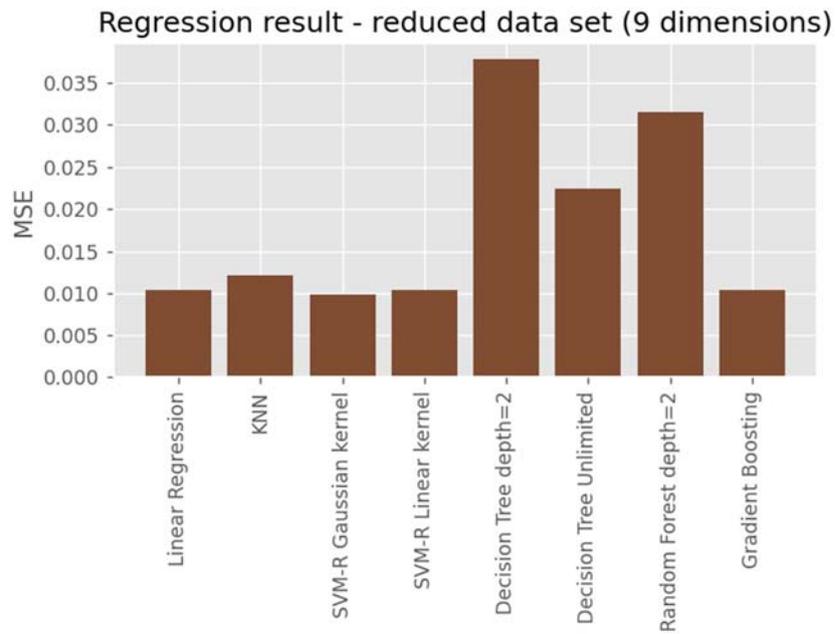
Figure 5.29: Comparison of all regressors for the **2nd dataset**.

3rd Dataset: In Figure 5.30, the optimal KNN result is observed at K=11.
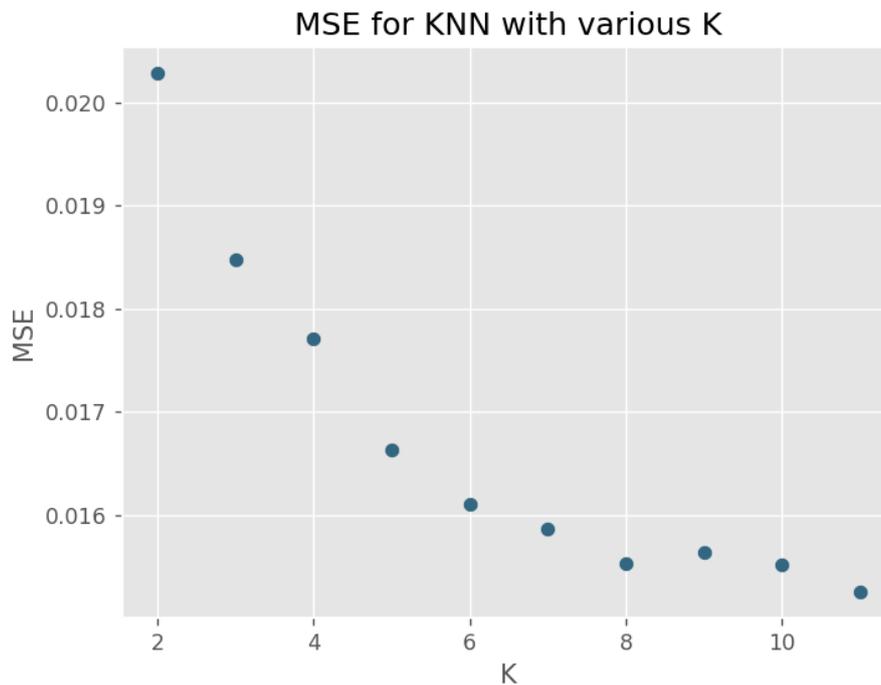
Figure 5.30: 3rd dataset: MSE for KNN with various K.

In Figure 5.31, the 3rd dataset (the first small one), which is based on only two specific dimensions, ("Culture" and "Work/Life balance") presents a very good regression performance (0.015 MSE). The value may not be as good as it is on the bigger data sets, but the prediction is quite acceptable for most of the predictors.
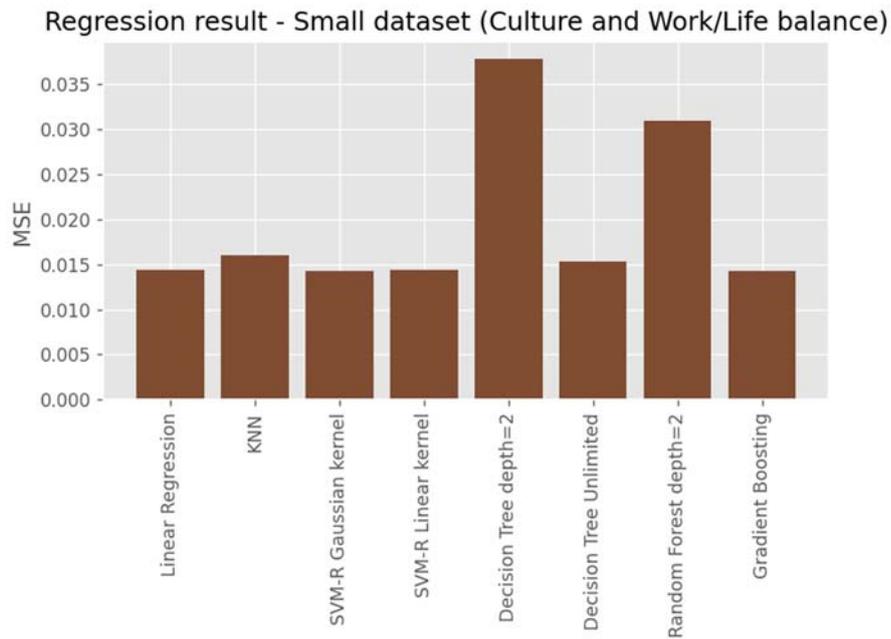
Figure 5.31: Comparison of all regressors for the **2nd dataset**.

4th Dataset: In Figure 5.32, the optimal KNN result is observed at K=11.
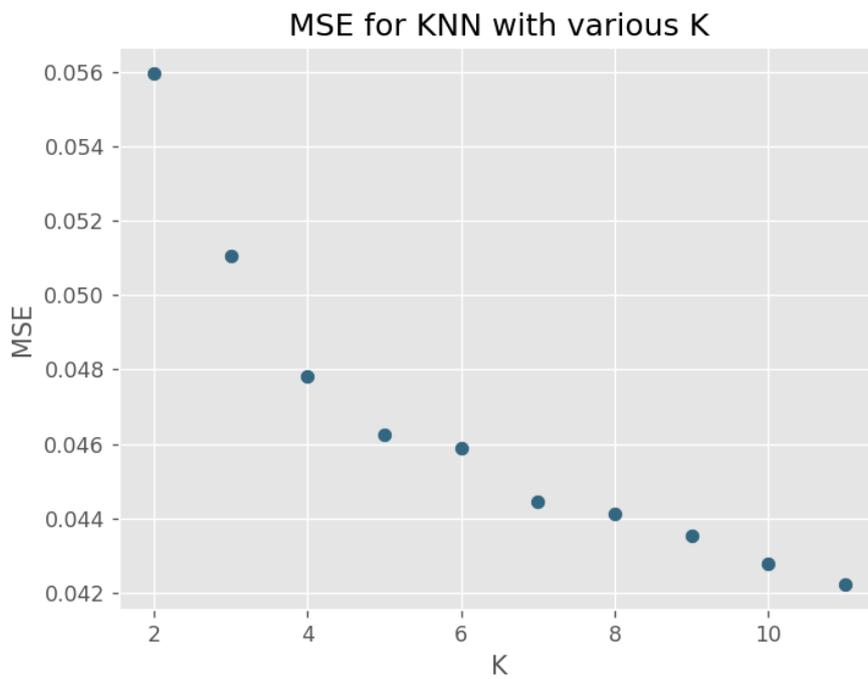
Figure 5.32: 4th dataset: MSE for KNN with various K.

In Figure 5.33, the 4th dataset (small), which is based on only two dimensions, the ones who are commonly regarded as important ("Compensation/Benefits" and "Advancement/Job Security") the best regressors' performance is not that good

(approx. 0.04 MSE). The best predictor in this case is the SVM Regressor with Gaussian kernel.
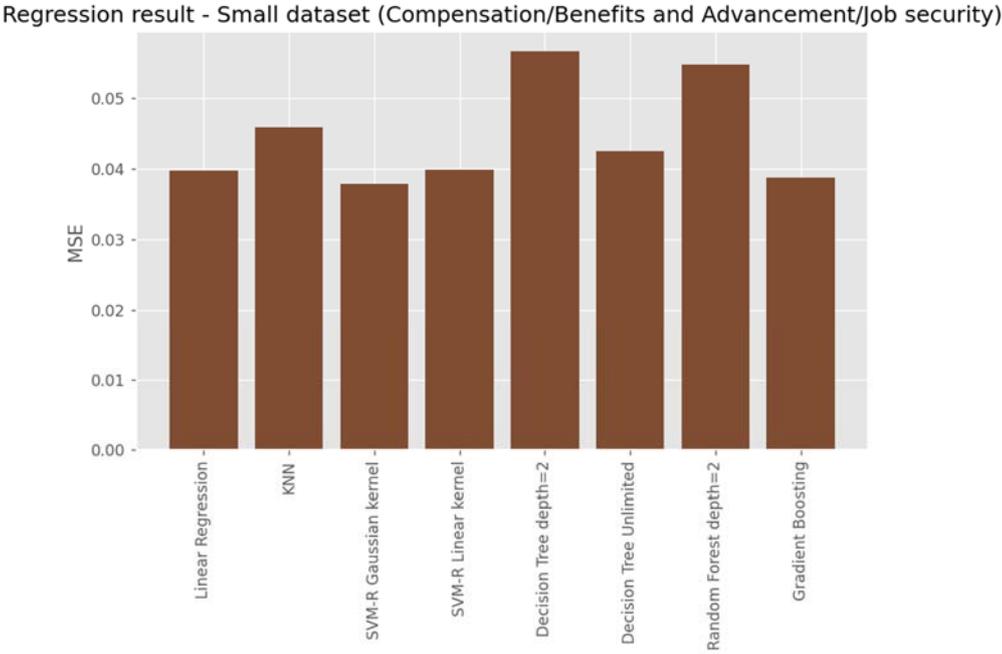
Figure 5.33: Comparison of all regressors for the **4th dataset**.

5th Dataset: Using Linear Regression, the effect of the different dimensions (13+5+3=21 in total) is depicted in Figure 5.34. All 21 dimensions (which are the ones representing the 'warmest' expressions) are adopted for the 5th dataset.

As already mentioned, the dimensions that represent the 'warmest' feelings are reflecting aspects (sub-dimensions) of the work environment that have the most impact on employee satisfaction (happiness). When these dimensions show strong relationships with the overall company review rating, this is a strong indication that the employees are highly sensitive to such workplace aspects. Consequently, incorporating these significant dimensions in the regression models can help us to accurately predict the employee ratings, thus leading to better understanding, decision making and actionable insights that can improve the overall work experience and employee satisfaction.
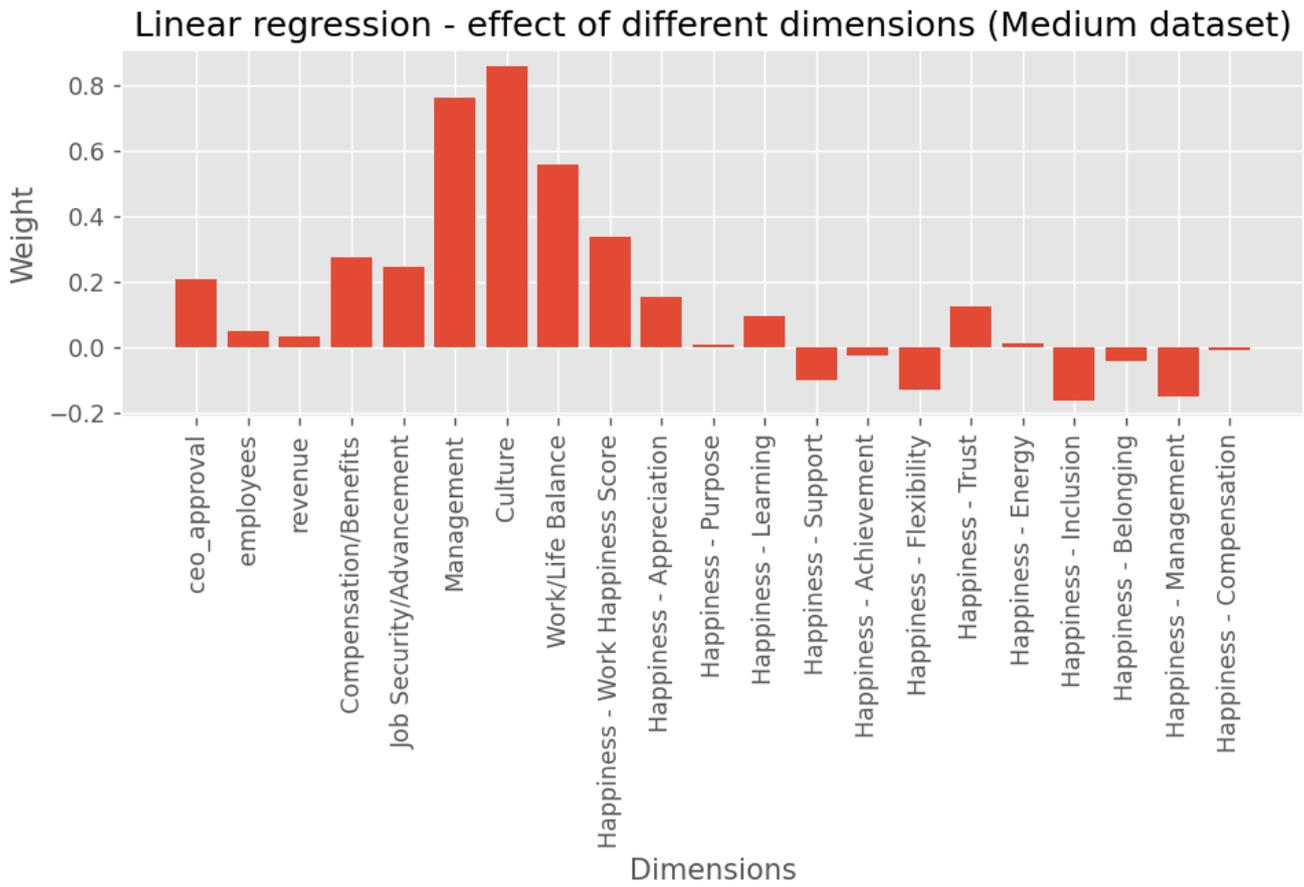
Figure 5.34: 5th dataset: effect of the 21 different dimensions (parameters/features).

5th Dataset: In Figure 5.35, the optimal KNN result is, once again, observed at K=11.
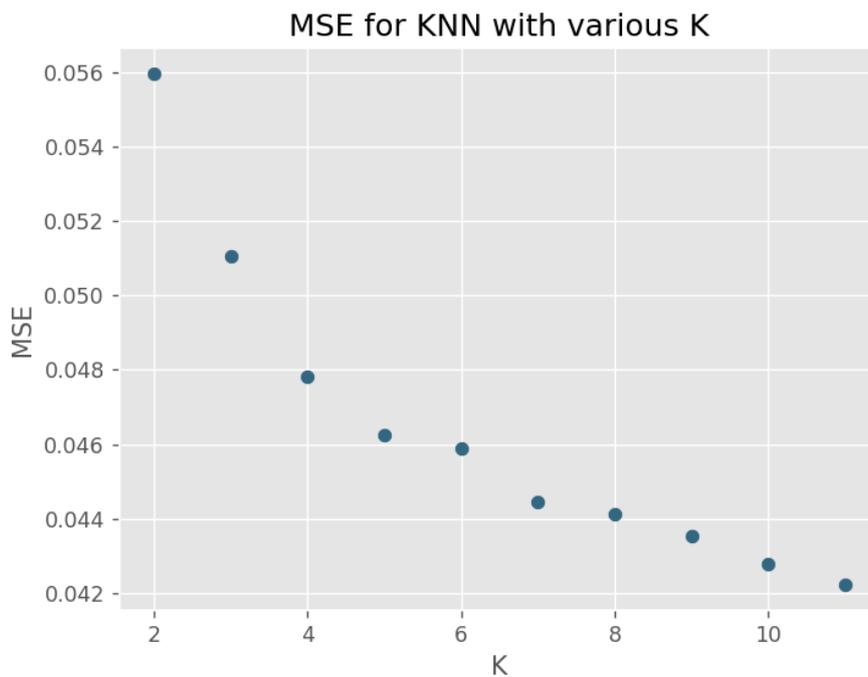
Figure 5.35: 5th dataset: MSE for KNN with various K.

In Figure 5.36, for the 5th dataset (medium), which is based on the 21 dimensions (i.e., the ones that are related to the 'warmest' expressions), the best regressors are the same as of the cases of the big datasets (**Linear Regression**, **both SVM-Rs**, and **Gradient Boosting).** However, the predicted results in this setup produce the smallest MSEs that we have seen so far (with the best being approximately 0.007).
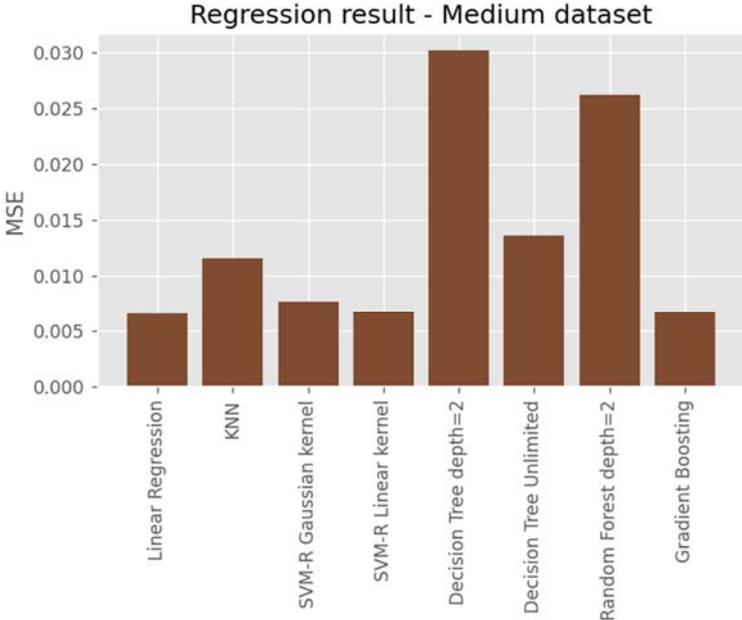
Figure 5.36: Comparison of all regressors for the **5th dataset**.

5.6.4. Datasets comparison

This section is ranking the predictors per dataset. The best predictors are presented by their dataset scores in the results that follow on the table of Figure 5.37.

	1st Predictor	2nd Predictor	3rd Predictor	4th Predictor	5th Predictor	6th Predictor	7th Predictor	8th Predictor
	Linear Regression	KNN	SVM-R Gaussian kernel	SVM-R Linear kernel	Decision Tree depth=2	Decision Tree Unlimited	Random Forest depth=2	Gradient Boosting
1st Dataset	0,0099132	0,0210780	0,0102557	0,0099131	0,0377842	0,0225352	0,0302802	0,0099322
2nd Dataset	0,0104831	0,0121901	0,0099308	0,0104733	0,0377842	0,0222816	0,0310364	0,0103938
3rd Dataset	0,0144356	0,0161118	0,0142934	0,0144436	0,0377842	0,0153479	0,0314517	0,0143218
4th Dataset	0,0398276	0,0458959	0,0379428	0,0399250	0,0566675	0,0425689	0,0550846	0,0387515
5th Dataset	0,0065405	0,0115599	0,0076582	0,0067343	0,0301350	0,0129271	0,0259751	0,0066568

Figure 5.37: Datasets comparison (Best predictor results per dataset case).

6. Conclusions and Future Work Propositions

6.1.   Direct Conclusions

After a very detailed curation on the results that are compiled in Figure 5.37, there are five (plus one) sentences which hold true for each one of the dataset cases. The sentences are highlighted in color in order to stress out in **bold black** the best predictor per case. The second best is highlighted **in blue**, the third best **in green** and the 4th best **in yellow**.

1. For the **1st Dataset** it is best to use **SVM Regressor w. Linear kernel**, or **Linear Regression** (as they differ on the 7th decimal digit!!). **Gradient Boosting** is also performing very well and the same goes for **SVM Regressor w. Gaussian kernel**.

2. For the **2nd Dataset** it is best to use **SVM Regressor w. Gaussian kernel**, or **Gradient Boosting** (as they differ on the 6th decimal digit!). **SVM Regressor w. Linear kernel** is also performing very well and the same goes for **Linear Regression**.

3. For the **3rd Dataset** it is best to use **SVM Regressor w. Gaussian kernel**, or **Gradient Boosting** (as they differ on the 4th decimal digit). **Linear Regression** is also performing very well and the same goes for **SVM Regressor w. Linear kernel**.

4. For the **4rd Dataset** it is best to use **SVM Regressor w. Gaussian kernel**, or **Gradient Boosting** (as they differ on the 4th decimal digit). **Linear Regression** is also performing very well and the same goes for **SVM Regressor w. Linear kernel**.

5. For the **5th Dataset** it is best to use **Linear Regression**, or **Gradient Boosting** (as they differ on the 4th decimal digit). **SVM Regressor w. Linear kernel** is also performing very well and the same goes for **SVM Regressor w. Gaussian kernel**.

6. The coloring on these sentences allows to quickly draw one more conclusion which relates to the fact that the **SVM Regressor w. Gaussian kernel** is in

general the best choice when the dataset is unknown. The second best in general is **Gradient Boosting** and the third best is **Linear Regression**.

6.1.1. Additional points to consider

In the following lines we lay some extra points to consider:

7. We can predict the company ratings that are based on various factors in an effective way by trying to employ diverse linear regression methods.

8. We can reduce the number of features from 57 to 9 and maintain a high prediction performance while ensuring no compromising of model accuracy.

9. The findings of this work highlight that the "Compensation/Benefits" and the "Job Security/Advancement" pose a smaller impact on company ratings than the company Culture, which is a result that aligns with numerous research studies and HR (human resources) articles.

10. The data on "happiness" guarantees further examination and breaking-down into distinct features to enhance our understanding on its influence on company performance ratings.

Additionally, according to the computational complexity of these eight (8) methods, in the lines to follow we list a ranking of them, starting from the lightest to the heaviest:

- **Linear Regression**: This is **the lightest method** when it comes to **computational load**. It is based on matrix operations only in order to find the coefficients.

- **Decision Tree (depth=2)**: A decision tree of depth=2 is computationally light and quite simple.

- **SVM Regressor with Linear Kernel**: The complexity is dependent upon the size of the dataset. However, in general, it is usually **relatively light** because it aims to solve a convex problem on optimization.

- **KNN**: Similar to the rest, KNN's computational complexity depends on the dataset size and dimensionality. During the training phase, it can be quite fast since it just stores data. However, while it runs the prediction phase, it can be computationally heavy because it needs to calculate distances between the test point and all training points.

- **Decision Tree (unlimited depth)**: As the tree grows, its structural and thus computational complexity increases and this is why it can easily be rendered computationally challenging, especially for large datasets that can require deep trees.

- **Random Forest (depth=2)**: The method is facilitated by the inclusion of building multiple trees, thus the overall complexity increases.

- **SVM Regressor with Gaussian Kernel**: The SVM with Gaussian kernel complexity can be quite high, especially when using large datasets, because it involves calculating the matrix of the kernel and for requiring rendering the solution of a convex problem of optimization.

- **Gradient Boosting**: Gradient boosting is an iterative method which is building trees in a sequential manner. This can turn out to be computationally heavy, especially when applied for large datasets and running for many iterations.

The actual computational cost of the aforementioned methods usually varies based on the dataset size, the number of the data features (dimensions), the way of its implementation, and various other factors. However, in the general case, and when given the information regarding the computational loading above, an extra point is added on the conclusions which is expressed in the following paragraph.

6.1.2. Generalized conclusion

11. Considering both the performance AND the computational complexity of the eight aforementioned methods the best choice to use primarily is **Linear Regression**. The second best is to use the **SVM-Regressors** (preferable the

linear for big datasets and the Gaussian for small datasets). The third best is
Gradient Boosting.

6.2. Indirect Conclusions

6.2.1. Based on the dataset nature

1. The reviews volume is (generally) higher for the companies with higher revenues. This is for various reasons, including the size of the company, public attention, structured HR departments, and employee expectations.

2. Healthcare companies dominate the (Glassdoor.com) dataset, possibly due to the industry's large size, and the "high-stakes" nature of the healthcare work.

3. The distribution of companies in the (Glassdoor.com) dataset may not perfectly represent the distribution of companies in the US based on industry.

4. The most popular industry and revenue range per state are quite similar across the US. The dominant industry is healthcare, and the only exception is Alaska that has education as its most popular industry.

5. The highest job ratings can be found in Hawaii, Alaska, and Mississippi.

6. The median of the CEO approval rate is quite stable across all industries. This could be due to the general satisfaction, the central tendency, or other industry-independent factors.

6.2.2. Based on the exploratory analysis

7. There is a (strong) positive correlation between the CEO approval and the overall rating of the company since the CEO seems to play a critical role in the shaping of the company's culture and the employees' overall work experience.

8. No obvious correlation exists between the yearly salary and the overall company rating (employee morale factors).

9. The sub-dimensions of "happiness" are positively correlated (multi-correlated) and also tightly correlated with the overall company ratings. As expected, the "Work Happiness Score" has a strong positive correlation with the overall company rating.

10. The distribution of "happiness" across various job titles reveals that the employee "happiness" is higher for certain roles (like Graduate Researcher, Research Assistant, and Tutor), which seems all related to R&D and Academia. Conversely, the employee "happiness" is lower for some other specific roles, as expected (like Therapist, Counselor, and Mental Health Technician) which seem to include professions with services related with treatment.

11. The contribution of the sub-dimensions of "happiness" across the top-30 most popular industries seems to be quite even. This indicates that these aspects of "happiness" may not vary significantly across different industries.

12. All five sub-dimensions of "ratings" have a very small (positive) correlation with the company size which is measured in employees count ("employees") and the company revenue ("revenue").

13. The company overall rating is higher for the smaller and larger companies, and for companies with either very low or very high revenue. In other words, the company overall rating is higher at the extremes of "size" and "revenue".

14. The variance of the overall rating is larger for the smaller companies and for the companies with the lower revenues.

15. There is a slight increment in the company overall rating for the companies with more than 50 employees and for the ones with revenues of more than 5M.

16. The Compensation/Benefits is the least correlated sub-dimension of "happiness" with the company overall rating. "Culture" and "Management" seem to be more tightly correlated with the company overall rating.

17. The ratings for "Compensation/Benefits" and for "Job Security/Advancement" tend to be lower compared to other factors. The ratings for the "Culture", the

"Work/Life Balance", the "Management", and the overall "Rating" are closely 'clustered' together.

18. There is a high correlation among the sub-dimensions of "happiness" (multi-correlation), except for the "Compensation".

19. There are strong correlations observed between the "Work Happiness Score", the "Appreciation", the "Inclusion", the "Belonging", and the "Management" which could be attributed to the interrelated nature of these aspects (multi-correlation) of the work experience.

6.3.   Future Research Propositions

Based on the literature review content and the findings of this thesis, the following research propositions are suggested for future studies:

- **Investigating the potential relationship between purely financial performance indicators** (e.g., profit margin, return on investment, and earnings per share) **and the employee happiness predictors** (employee morale factors) in order to determine if the economical/financial success of a company has a direct impact on the employee satisfaction/engagement levels, and vice versa.
- Further examining the **impact of non-financial company performance indicators or what in this thesis is being referenced as actions that increase employee morale factors** (e.g., the corporate social responsibility initiatives, the employee training, the personal development programs, and the workplace safety) over the employee happiness predictors (the employee morale factors) in order to **understand the influence of organizational culture on employee satisfaction**.
- Assessing the **role of the various management styles and the numerous leadership approaches** in the shaping of the employee happiness predictors, and then exploring whether certain management practices (or company

policies, quality control measures etc.) are more effective in fostering a positive work environment and/or employee satisfaction and to what extent.

- Analyzing the pragmatic **impact of the remote work arrangements and the flexible work contracts** on the employee happiness predictors, in order to determine whether these arrangements do actually contribute to the improved satisfaction levels and the overall company ratings.

- Investigating the **influence of the industry-specific factors** (either in general or on specifically pre-selected industries (i.e., healthcare) on the employee happiness predictors, in order to understand if these certain industries are more conducive towards employee satisfaction, and if so, why is this the case.

- Exploring the role of the **employee demographics** (e.g., age, gender, sex, and educational background) on the happiness predictors, in order to identify if some certain groups have a different satisfaction level from others and any needs/requirements that should be addressed by the corporations / organizations.

- Examining the potential moderating effects of the **macroeconomic factors** (e.g., economic growth, unemployment rates, and inflation) on the relationship between the financial performance indicators and the employee happiness predictors.

- Studying the **long-term effects of** the **employee happiness** predictors on the company performance, including how do changes in the employee satisfaction levels impact financial results and the competitiveness over time.

All of these research propositions are suggestion that can, to some extent, provide valuable insights into the appreciation and understanding of the factors that actually contribute to the employee happiness, engagement and satisfaction, and how do these factors may, in turn, influence the overall performance and the market/business success of corporations/organizations.

6.3.1. Other plausible solutions / approaches to complement this work

In the lines to follow we have included some additional approaches and solutions that are useful as considerations while researching the most indicative of the predictors for the companies achieving good employee review ratings and for safely predicting the financial success of a company:

- **Machine Learning Models** [193]: These models can leverage machine learning techniques, like decision trees, and/or random forests, and/or deep neural networks, and to identify the most significant features that may contribute to employee review ratings and economical/financial success. Also, they can help uncover complex potential relationships and any existing interactions among the variables that may not be easily detected by traditional tools like linear regression analysis.

- **Feature Engineering** [204]: In order to experiment with the creation of new features that may capture the underlying relationships between the variables in a better and more holistic way. For example, feature engineering can help to consider creating several interaction terms between different predictors or to consider aggregating the existing variables into some higher-level categories that, in some cases, can reveal more meaningful patterns.

- **Feature Selection Techniques**: The idea to employ feature selection techniques like Recursive Feature Elimination (RFE) [205], or Lasso regression [206], or the Feature Importance Ranking from Tree-based models [207], can help identify the most significant of the predictors in order to select the best ones per case, that can lead to good employee review ratings and economical/financial success for the respective company(ies).

- **Longitudinal Analysis**: Longitudinal Analysis [208] when properly conducted, can examine the changes in the employee review ratings and the economical/financial performance over time (if such data dimension exists). This approach will help to uncover the temporal dynamics of these relationships (between dependent and independent) variables and to provide useful insights for the potential causes and the consequences of these

changes in the correlation between the employee morale factors and economical/financial performance of the respective company.

- **Qualitative Research**: In order to supplement our quantitative analysis with some Qualitative Research methods [209], we can select the inclusion of tools like interviews or focus groups, in order to gain a deeper understanding of the factors that contribute to employee satisfaction, happiness and engagement and thus to the economical/financial success of the company they work for. Such an analysis can provide valuable context and facilitate the identification of potential areas for intervention and improvement.

- **External Data Sources**: The idea to consider the incorporation of external data sources, such as various **macroeconomic indicators**, and/or **industry trends**, or markers that measure market trends or the **competitive landscape**, in order to provide additional context for all the stages of the analysis and to potentially uncover more predictors.

- **Cross-Validation**: Another idea is to use cross-validation techniques, like K-fold Cross-Validation [210] or like Holdout Validation [211], in order to ensure that the robustness and the generalizability of our findings is valid. This can also facilitate the mitigation of the overfitting risk and to ensure that the designed models (and subsequent results) can be regarded as reliable and applicable in different kinds of contexts.

- **Sensitivity Analysis**: Finally, another idea is to conduct Sensitivity Analyses [212] in order to assess the robustness of the thesis findings and to explore any potential sources of uncertainty and/or bias in the given data (and metadata). This approach can facilitate the identification of areas where the additional data collection or the methodological improvements are required in order to strengthen the aforementioned results and conclusions.

By incorporating the aforementioned additional approaches, tools and methods as potential solutions into this thesis, we can gain more robust models of predictors and a more comprehensive understanding of the peculiarities of the given data, the nature of the problem itself and the specificities expressed as factors that contribute

to a good employee morale and to the economical/financial success of a company, and thus we may aspire to ultimately develop much more effective strategies tailored to improve employee satisfaction, happiness and engagement and thus fostering and the economical/financial performance of the respective corporations.

References

1. *Jaques, Elliott (1951). The changing culture of a factory. Tavistock Institute of Human Relations. [London]: Tavistock Publications. p. 251.*
2. *Hatch, Mary Jo; Cunliffe, Ann L. (2013) [1997]. "A history of organizational culture in organization theory". Organization Theory: Modern, Symbolic and Postmodern Perspectives (2 ed.). Oxford: Oxford University Press. p. 161.*
3. *Kummerow, Elizabeth (12 September 2013). Organisational culture: concept, context, and measurement. Kirby, Neil.; Ying, Lee Xin. New Jersey. p. 13.*
4. *Ravasi, D.; Schultz, M. (2006). "Responding to organizational identity threats: Exploring the role of organizational culture". Academy of Management Journal. 49 (3): 433–458.*
5. *Schein, Edgar H. (2004). Organizational culture and leadership (3rd ed.). San Francisco: Jossey-Bass. pp. 26–33.*
6. *Schein, Edgar (1992). Organizational Culture and Leadership: A Dynamic View. San Francisco, CA: Jossey-Bass. pp. 9.*
7. *Modaff, D.P., DeWine, S., & Butler, J. (2011). Organizational communication: Foundations, challenges, and misunderstandings (2nd Ed.). Boston: Pearson Education. (Chapters 1–6)*
8. *Deal T. E. and Kennedy, A. A. (1982, 2000) Corporate Cultures: The Rites and Rituals of Corporate Life, Harmondsworth, Penguin Books, 1982; reissue Perseus Books, 2000*
9. *Kotter, J. P.; Heskett, James L. (1992). Corporate Culture and Performance. New York: The Free Press.*
10. *Selart, Marcus; Schei, Vidar (2011): "Organizational Culture". In: Mark A. Runco and Steven R. Pritzker (eds.): Encyclopedia of Creativity, 2nd edition, vol. 2. San Diego: Academic Press, pp. 193–196.*
11. *Hirst, Scott (2018-07-01). "The Case for Investor Ordering". The Harvard Law School Program on Corporate Governance Discussion Paper. No. 2017-13.*
12. *Corporate Culture - https://www.investopedia.com/terms/c/corporate-culture.asp*
13. *Corporate Identity - https://en.wikipedia.org/wiki/Corporate_identity*
14. *Corporation - https://en.wikipedia.org/wiki/Corporation*
15. *Social Reporting: Meaning, Uses and Scope - https://www.yourarticlelibrary.com/accounting/financial-reporting/social-reporting-meaning-uses-and-scope/57360*
16. *How Do You Measure Ethics in Business? Jack Torrance - https://www.managementtoday.co.uk/measure-ethics-business/article/1387257*
17. *The Next Great Disruption Is Hybrid Work-are We Ready? - https://www.microsoft.com/en-us/worklab/work-trend-index/hybrid-work*
18. *Happy Employees = Happy Shareholders - Sheryl Estrada - https://fortune.com/2021/04/13/happy-employees-happy-shareholders/*
19. *Luo, N., Zhou, Y., & Shon, J. (2016). Employee satisfaction and corporate performance: Mining employee reviews on glassdoor.com.*
20. *Landers, R. N., Brusso, R. C., & Auer, E. M. (2019). Crowdsourcing job satisfaction data: Examining the construct validity of glassdoor.com ratings. Personnel Assessment and Decisions, 5 (3), 6.*
21. *5 Customer Satisfaction Metrics To Predict Revenue Growth - https://boast.io/5-customer-satisfaction-metrics-to-predict-revenue-growth/*
22. *Morgan, N. A., & Rego, L. L. (2006). The value of different customer satisfaction and loyalty metrics in predicting business performance. Marketing science, 25(5), 426-439.*

23. *The Ethics Study 2021 − Principia* - https://www.ibe.org.uk/resource/the-ethics-survey-2021-principia.html

24. *Glassdoor research reveals what makes a great CEO, according to employees - MILL VALLEY, CALIF. (Aug 24, 2016) –* https://about-content.glassdoor.com/en-us/glassdoor-research-reveals-great-ceo/

25. *What Makes Great CEO – Glassdoor.com* - https://www.glassdoor.com/research/app/uploads/sites/2/2016/08/FULL-STUDY_WhatMakesGreatCEO_Glassdoor-2.pdf

26. *Garralda, J.M. (2017). Measuring interlinkages between non-financial firms, banks and institutional investors: How securities common identifiers can help? IFC Bulletins chapters, 46.*

27. *Richtermeyer, S. B., Greller, M. M., & Valentine, S. R. (2006). Organizational ethics: Measuring performance on this critical dimension. Management Accounting Quarterly, 7(3), 23.*

28. *Kim, D. S., & Yang, J. (2012, December). Behind the scenes: Performance target setting of annual incentive plans. In AFA 2010 Atlanta Meetings Paper.*

29. *Flamholtz, Eric G.; Randle, Yvonne (2011). Corporate Culture: The Ultimate Strategic Asset. Stanford Business Books. Stanford, California: Stanford University Press. p. 6.*

30. *Case study: Ecobank Group − Integrating metrics to track organizational ethics* - https://www.brvm.org/sites/default/files/rapport_annuel_2018_ecobank_group.pdf

31. *Benjamin Schneider and Karen M. Barbera, The Oxford Handbook of Organizational Climate and Culture. Oxford Library of psychology. Oxford: Oxford University Press. p. 247.*

32. *Company Reviews - Reza*
   - https://www.kaggle.com/search?q=glassdoor+employee+reviews

33. *Financial Data Of 4400+ Public Companies - Jiun Yen*
   - https://www.kaggle.com/qks1lver/financial-data-of-4400-public-companies.

34. *[34] Kaplan, R. S., & Norton, D. P. (2005). The balanced scorecard: measures that drive performance (Vol. 70, pp. 71-79). US: Harvard business review.*

35. *[35] Eccles, R. G., Serafeim, G., Seth, D., & Ming, C. C. Y. (2013). The Performance Frontier: Innovating for a Sustainable Strategy: Interaction. Harvard business review, 91(7), 17-18.*

36. *[36] Singh, A., & Chakraborty, M. (2021). Corporate Social Responsibility and Financial Performance: The'Virtuous Circle'Revisited in India. Available at SSRN 4112346.*

37. *[38] Edmans, A. (2011). Does the stock market fully value intangibles? Employee satisfaction and equity prices. Journal of Financial economics, 101(3), 621-640.*

38. *[39] Ittner, C. D., & Larcker, D. F. (1998). Are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction. Journal of accounting research, 36, 1-35.*

39. *[41] Brigham, E. F., & Ehrhardt, M. C. (2002). Financial management: theory and practice. 10 th. Melbourne: Thomson Learning.*

40. *[42] Palepu, K. G., Healy, P. M., Wright, S., Bradbury, M., & Coulton, J. (2020). Business analysis and valuation: Using financial statements. Cengage AU.*

41. *[43] Penman, S. H., & Penman, S. H. (2007). Financial statement analysis and security valuation (Vol. 3). New York: McGraw-hill.*

42. *[44] Damodaran, A. (2012). Investment valuation: Tools and techniques for determining the value of any asset. John Wiley & Sons.*

43. *[45] Fabozzi, F. J., & Markowitz, H. M. (Eds.). (2011). The theory and practice of investment management: Asset allocation, valuation, portfolio construction, and strategies (Vol. 198). John Wiley & Sons.*

44. [46] Weil, R. L., Schipper, K., & Francis, J. (2013). Financial accounting: an introduction to concepts, methods and uses. Cengage Learning.

45. [47] Brigham, E. F., & Houston, J. F. (2021). Fundamentals of financial management: Concise. Cengage Learning.

46. [48] Jha, L. D. N., Ubale, A., & Jani, N. (1990). Introduction to management accounting. Mumbai, India: Himalaya.

47. [49] Drury, C. M. (2013). Management and cost accounting. Springer.

48. [50] Wahlen, J. M., Baginski, S. P., & Bradshaw, M. (2022). Financial reporting, financial statement analysis and valuation. Cengage learning.

49. [51] McMahon, R., Holmes, S., Hutchinson, P., & Forsaith, D. (1993). Small enterprise financial management: Theory and practice.

50. [52] Horngren, C. T., Datar, S. M., Foster, G., Rajan, M. V., & Ittner, C. (2009). Cost accounting: a managerial emphasis. Pearson Education India.

51. [53] Weil, R. L., Schipper, K., & Francis, J. (2013). Financial accounting: an introduction to concepts, methods and uses. Cengage Learning.

52. [54] Hillier, D., Clacher, I., Ross, S., Westerfield, R., & Jordan, B. (2014). Fundamentals of corporate finance.

53. [55] Rappaport, A. (1999). Creating shareholder value: a guide for managers and investors. Simon and Schuster.

54. [56] Pfeffer, J., & Veiga, J. F. (1999). Putting people first for organizational success. Academy of management perspectives, 13(2), 37-48.

55. [57] Bassi, L., & McMurrer, D. (2007). Maximizing your return on people. Harvard business review, 85(3), 115.

56. [2R23] Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. Academy of management journal, 38(3), 635-672.

57. [58] Arowoshegbe, A. O., Emmanuel, U., & Gina, A. (2016). Sustainability and triple bottom line: An overview of two interrelated concepts. Igbinedion University Journal of Accounting, 2(16), 88-126.

58. [59] Lee, C. K., Dmytriyev, S. D., Rutherford, M. A., & Lee, J. Y. (2023). The impact of B Corporations' certification timing on a firm's economic and social value creation. Social Responsibility Journal.

59. [60] Freeman, R. E., Harrison, J. S., Wicks, A. C., Parmar, B. L., & De Colle, S. (2010). Stakeholder theory: The state of the art.

60. [61] Hansmann, H., Kraakman, R., & Squire, R. (2005). Law and the Rise of the Firm. Harv. L. Rev., 119, 1335.

61. [62] Ripken, S. K. (2009). Corporations are people too: a multi-dimensional approach to the corporate personhood puzzle. Fordham J. Corp. & Fin. L., 15, 97.

62. [63] Blair, M. M. (2012). Corporate law and the team production problem. In Research Handbook on the Economics of Corporate Law. Edward Elgar Publishing.

63. [64] Johnston, A., Veldman, J., Eccles, R. G., Deakin, S., Davis, J., Djelic, M. L., ... & Chabrak, N. (2019). Corporate governance for sustainability: Statement. SSRN.

64. [65] Stout, L. (2012). The shareholder value myth: How putting shareholders first harms investors, corporations, and the public. Berrett-Koehler Publishers.

65. [66] Schein, E. H. (2010). Organizational culture and leadership (Vol. 2). John Wiley & Sons.

66. [67] Denison, D. R., Haaland, S., & Goelzer, P. (2003). Corporate culture and organizational effectiveness: is there a similar pattern around the world?. In Advances in global leadership (Vol. 3, pp. 205-227). Emerald Group Publishing Limited.

67. [68] Masa'deh, R. E., Obeidat, B. Y., & Tarhini, A. (2016). A Jordanian empirical study of the associations among transformational leadership, transactional leadership, knowledge sharing, job performance, and firm performance: A structural equation modelling approach. Journal of management development, 35(5), 681-705.

68. [40] Brigham, E. F., & Ehrhardt, M. C. (2016). Financial management: Theory & practice (15th ed.). Boston, MA: Cengage Learning.

69. [69] Schein, E. H. (2010). Organizational culture and leadership (Vol. 2). John Wiley & Sons.

70. [70] Wang, J., van Woerkom, M., Breevaart, K., Bakker, A. B., & Xu, S. (2023). Strengths-based leadership and employee work engagement: A multi-source study. Journal of Vocational Behavior, 142, 103859.

71. [72] Bhattacharya, C. B., Sen, S., & Korschun, D. (2011). Leveraging corporate responsibility: The stakeholder route to maximizing business and social value. Cambridge University Press.

72. [73] Brau, J. C., & Fawcett, S. E. (2006). Initial public offerings: An analysis of theory and practice. The journal of Finance, 61(1), 399-436.

73. [74] Hochberg, Y. V. (2012). Venture capital and corporate governance in the newly public firm. Review of Finance, 16(2), 429-480.

74. [75] Yalcin, N., & Ünlü, U. (2018). A multi-criteria performance analysis of Initial Public Offering (IPO) firms using CRITIC and VIKOR methods. Technological and Economic development of Economy, 24(2), 534-560.

75. [76] Agrawal, A., Catalini, C., & Goldfarb, A. (2013). Some simple economics of crowdfunding. Innovation Policy and the Economy, 14(1), 63-97.

76. [77] Certo, S. T., Covin, J. G., Daily, C. M., & Dalton, D. R. (2001). Wealth and the effects of founder management among IPO-stage new ventures. Strategic management journal, 22(6-7), 641-658.

77. [78] Stone, C. (2019). Examining the Input, Process, Output Model of Team Effectiveness (IPOMTE), Leadership Styles, and Relational Coordination as Contributors to a Profile of Team Effectiveness.

78. [79] Serdar, G. (2017). Heterogeneity in Institutional Context and Its Impact on Initial Public Offerings and Corporate Social Responsibility (Doctoral dissertation, Syracuse University).

79. [80] Rajnoha, R., & Lorincová, S. (2015). Strategic management of business performance based on innovations and information support in specific conditions of Slovakia. Journal of Competitiveness.

80. [81] Mura, L., Ključnikov, A., Tvaronavičienė, M., & Androniceanu, A. (2017). Development trends in human resource management in small and medium enterprises in the Visegrad Group. Acta Polytechnica Hungarica, 14(7), 69-88.

81. [82] Lee, N., Sameen, H., & Cowling, M. (2015). Access to finance for innovative SMEs since the financial crisis. Research policy, 44(2), 370-380.

82. [83] Zhao, X., Lynch Jr, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. Journal of consumer research, 37(2), 197-206.

83. [84] Friesenbichler, K., & Selenko, E. (2017). Firm performance in challenging business climates: does managerial work engagement make a difference?. Asian Business & Management, 16(1-2), 25-49.

84. [85] Shi, X., Zheng, Z., Zhang, Q., & Liang, H. (2020). External knowledge search and firms' incremental innovation capability: The joint moderating effect of technological proximity and network embeddedness. Management Decision, 58(9), 2049-2072.

85. [86] Macey, W. H., & Schneider, B. (2008). The meaning of employee engagement. Industrial and organizational Psychology, 1(1), 3-30.

86. [87] Shuck, B., & Reio Jr, T. G. (2014). Employee engagement and well-being: A moderation model and implications for practice. Journal of Leadership & Organizational Studies, 21(1), 43-58.

87. [88] Amran, A., Lee, S. P., & Devi, S. S. (2014). The influence of governance structure and strategic corporate social responsibility toward sustainability reporting quality. Business Strategy and the environment, 23(4), 217-235.

88. [89] Kehoe, R. R., & Wright, P. M. (2013). The impact of high-performance human resource practices on employees' attitudes and behaviors. Journal of Management, 39(2), 366-391.

89. [90] Galpin, T., Whitttington, J. L., & Bell, G. (2015). Is your sustainability strategy sustainable? Creating a culture of sustainability. Corporate Governance.

90. [91] Macey, W. H., & Schneider, B. (2008). The meaning of employee engagement. Industrial and organizational Psychology, 1(1), 3-30.

91. [92] Saks, A. M. (2006). Antecedents and consequences of employee engagement. Journal of managerial psychology, 21(7), 600-619.

92. [93] Gruman, J. A., & Saks, A. M. (2011). Performance management and employee engagement. Human resource management review, 21(2), 123-136.

93. [94] Taras, V., Kirkman, B. L., & Steel, P. (2010). Examining the Impact of Culture's Consequences: A Three-Decade, Multilevel, Meta-Analytic Review of Hofstede's Cultural Value Dimensions (vol 95, pg 405, 2010). J. Appl. Psychol, 95(5), 888-888.

94. [95] Cooke, F. L., Saini, D. S., & Wang, J. (2014). Talent management in China and India: A comparison of management perceptions and human resource practices. Journal of World Business, 49(2), 225-235.

95. [96] Kehoe, R. R., & Wright, P. M. (2013). The impact of high-performance human resource practices on employees' attitudes and behaviors. Journal of management, 39(2), 366-391.

96. [97] Kraus, S., Richter, C., Brem, A., Cheng, C. F., & Chang, M. L. (2016). Strategies for reward-based crowdfunding campaigns. Journal of Innovation & Knowledge, 1(1), 13-23.

97. [98] Cascio, W. F. (2014). Leveraging employer branding, performance management and human resource development to enhance employee retention. Human Resource Development International, 17(2), 121-128.

98. [99] Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: a meta-analysis. Journal of applied psychology, 87(2), 268.

99. [100] Burke, R. J. (2006). Supporting women's career advancement: Challenges and opportunities. Human Resource Management International Digest, 14(3).

100. [101] Aguinis, H., & Glavas, A. (2019). On corporate social responsibility, sensemaking, and the search for meaningfulness through work. Journal of management, 45(3), 1057-1086.

101. [102] Monecke, A., & Leisch, F. (2012). semPLS: structural equation modeling using partial least squares.

102. [103] Wang, B., Huang, H., & Wang, X. (2013). A support vector machine based MSM model for financial short-term volatility forecasting. Neural Computing and Applications, 22, 21-28.

103. [104] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R news, 2(3), 18-22.

104. [105] Gooty, J., Gavin, M., & Ashkanasy, N. M. (2009). Emotions research in OB: The challenges that lie ahead. Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior, 30(6), 833-838.

105. [106] Aydin, A. D., & Cavdar, S. C. (2015). Prediction of financial crisis with artificial neural network: an empirical analysis on Turkey. International journal of financial research, 6(4), 36.

106. [37] Chang, C., Lin, H., Tsai, W., Wang, W., & Huang, C. (2021). *Employee Satisfaction, Corporate Social Responsibility and Financial Performance. Sustainability, 13(18), 9996.*

107. [107] Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2021). *Multivariate Data Analysis (9th ed.). Cengage Learning.*

108. [108] Fenton, N., & Neil, M. (2014). *Decision support software for probabilistic risk assessment using Bayesian networks. IEEE software.*

109. [109] Crane, A., Matten, D., Glozer, S., & Spence, L. J. (2019). *Business ethics: Managing corporate citizenship and sustainability in the age of globalization. Oxford University Press, USA.*

110. [110] Trevino, L. K., & Nelson, K. A. (2021). *Managing business ethics: Straight talk about how to do it right. John Wiley & Sons.*

111. [111] Fisher, C. D., & Southey, G. (2005). *International human resource management in the introductory HRM course. The International Journal of Human Resource Management, 16(4), 599-614.*

112. [112] Des Jardins, J. R. (2012). *Environmental ethics. Cengage Learning.*

113. [113] Murphy, P. E., Laczniak, G. R., & Harris, F. (2016). *Ethics in marketing: International cases and perspectives. Taylor & Francis.*

114. [114] Kothari, J., & Barone, E. (2006). *Financial accounting: An international approach. Pearson Education.*

115. [115] Seuring, S., & Müller, M. (2008). *From a literature review to a conceptual framework for sustainable supply chain management. Journal of cleaner production, 16(15), 1699-1710.*

116. [116] Tavani, H. T. (2012). *Ethics and technology: Controversies, questions, and strategies for ethical computing. Wiley Publishing.*

117. [117] Kaptein, M. (2019). *The moral entrepreneur: a new component of ethical leadership. Journal of Business Ethics, 156(4), 1135-1150.*

118. [118] Aguinis, H., & Glavas, A. (2019). *On corporate social responsibility, sensemaking, and the search for meaningfulness through work. Journal of management, 45(3), 1057-1086.*

119. [119] Tian, L., Tse, C. H., Xiang, X., Li, Y., & Pan, Y. (2021). *Social movements and international business activities of firms. Journal of International Business Studies, 52(6), 1200-1214*

120. [120] Hadani, M., Doh, J. P., & Schneider, M. (2019). *Social movements and corporate political activity: Managerial responses to socially oriented shareholder activism. Journal of Business Research, 95, 156-170.*

121. [121] Crane, A., Matten, D., & Spence, L. (Eds.). (2014). *Corporate social responsibility: Readings and cases in a global context. Routledge.*

122. [122] Korschun, D. (2015). *Boundary-spanning employees and relationships with external stakeholders: A social identity approach. Academy of Management Review, 40(4), 611-629.*

123. [123] Schwartz, M. S. (2013). *Developing and sustaining an ethical corporate culture: The core elements. Business Horizons, 56(1), 39-50.*

124. [124] Edmondson, A. (1999). *Psychological safety and learning behavior in work teams. Administrative science quarterly, 44(2), 350-383.*

125. [125] Schneider, B., Ehrhart, M. G., & Macey, W. H. (2016). *Organizational climate and culture. Sage Publications, Inc.*

126. [126] Bai, K., Ullah, F., Arif, M., Erfanian, S., & Urooge, S. (2023). *Stakeholder-Centered Corporate Governance and Corporate Sustainable Development: Evidence from CSR Practices in the Top Companies by Market Capitalization at Shanghai Stock Exchange of China. Sustainability, 15(4), 2990.*

127. [127] Bhatti, A., Batool, I., & Akram, U. (2021). The role of person-organization fit, job satisfaction, and work engagement in promoting employee morale: A serial mediation model. International Journal of Organizational Analysis, 29(1), 157-173.

128. [128] Ramakrishnan, D. (2021). Leading in a VUCA world. Ushus Journal of Business Management, 20(1).

129. [129] Mancha, R., & Gordon, S. (2021). Multi-sided platform strategies for organizations: transforming the business model. Journal of Business Strategy, 43(3), 175-183.

130. [130] Abdelkafi, N., Raasch, C., Roth, A., & Srinivasan, R. (2019). Multi-sided platforms. Electronic Markets, 29, 553-559.

131. [131] Cennamo, C., & Santalo, J. (2013). Platform competition: Strategic trade-offs in platform markets. Strategic management journal, 34(11), 1331-1350.

132. [132] Loux, P., Aubry, M., Tran, S., & Baudoin, E. (2020). Multi-sided platforms in B2B contexts: The role of affiliation costs and interdependencies in adoption decisions. Industrial Marketing Management, 84, 212-223.

133. [133] Rawhouser, H., Cummings, M., & Crane, A. (2015). Benefit corporation legislation and the emergence of a social hybrid category. California Management Review, 57(3), 13-35.

134. [134] Othman, S., Darus, F., & Arshad, R. (2011). The influence of coercive isomorphism on corporate social responsibility reporting and reputation. Social Responsibility Journal, 7(1), 119-135.

135. [135] Sarta, A., Durand, R., & Vergne, J. P. (2021). Organizational adaptation. Journal of management, 47(1), 43-75.

136. [136] Hmieleski, K. M., Cole, M. S., & Baron, R. A. (2012). Shared authentic leadership and new venture performance. Journal of Management, 38(5), 1476-1499.

137. [137] Mulang, H. (2022). Analysis of The Effect of Organizational Justice, Worklife Balance on Employee Engagement and Turnover Intention. Golden Ratio of Human Resource Management, 2(2), 86-97.

138. [138] Carter, D. A., Simkins, B. J., & Simpson, W. G. (2003). Corporate governance, board diversity, and firm value. Financial review, 38(1), 33-53.

139. [139] Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. The quarterly journal of economics, 84(3), 488-500.

140. [140] Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. Journal of political economy, 100(6), 1126-1152.

141. [141] Stucke, M. E., & Grunes, A. P. (2016). Introduction: big data and competition policy. Big Data and Competition Policy, Oxford University Press (2016).

142. [142] Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. Journal of accounting and economics, 31(1-3), 405-440.

143. [143] Eccles, R. G., & Krzus, M. P. (2014). The integrated reporting movement: Meaning, momentum, motives, and materiality. John Wiley & Sons.

144. [144] Cohen, Daniel A. and Lys, Thomas Z., Substitution between Accrual- Based Earnings Management and Real Activities Manipulation – A Commentary and Guidance for Future Research (February 23, 2022).

145. [145] Cohen, D. A., Dey, A., & Lys, T. Z. (2013). Corporate governance reform and executive incentives: Implications for investments and risk taking. Contemporary Accounting Research, 30(4), 1296-1332.

146. [146] DeFond, M., & Zhang, J. (2014). A review of archival auditing research. Journal of accounting and economics, 58(2-3), 275-326.

147. [147] Santos, C., Coelho, A., & Marques, A. (2023). A systematic literature review on greenwashing and its relationship to stakeholders: state of art and future research agenda. Management Review Quarterly, 1-25.

148. [148] Wu, Y., Zhang, K., & Xie, J. (2020). Bad greenwashing, good greenwashing: Corporate social responsibility and information transparency. Management Science, 66(7), 3095-3112.

149. [149] Adil, A., & Kamal, A. (2020). Authentic leadership and psychological capital in job demands-resources model among Pakistani university teachers. International Journal of Leadership in Education, 23(6), 734-754.

150. [150] Pandita, D., & Ray, S. (2018). Talent management and employee engagement–a meta-analysis of their impact on talent retention. Industrial and Commercial Training.

151. [151] DeSimone, J. A., Köhler, T., & Schoen, J. L. (2019). If it were only that easy: The use of meta-analytic research by organizational scholars. Organizational Research Methods, 22(4), 867-891.

152. [152] Laws, J. L. (1975). The psychology of tokenism: An analysis. Sex roles, 1(1).

153. [153] Crane, A., & Matten, D. (2010). Managing corporate citizenship and sustainability in the age of globalization. Business Ethics, 20-24.

154. [154] Christian, M. S., Garza, A. S., & Slaughter, J. E. (2011). Work engagement: A quantitative review and test of its relations with task and contextual performance. Personnel psychology, 64(1), 89-136.

155. [155] Taneja, S., Sewell, S. S., & Odom, R. Y. (2015). A culture of employee engagement: A strategic perspective for global managers. Journal of Business Strategy, 36(3), 46-56.

156. [156] Park, T. Y., & Shaw, J. D. (2013). Turnover rates and organizational performance: a meta-analysis. Journal of applied psychology, 98(2), 268.

157. [157] Edmondson, A. C., & Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. Annu. Rev. Organ. Psychol. Organ. Behav., 1(1), 23-43.

158. [162] Sarasvathy, S. D. (2008). Effectuation: Elements of entrepreneurial expertise. In Effectuation. Edward Elgar Publishing.

159. [158] Naman, J. L., & Slevin, D. P. (1993). Entrepreneurship and the concept of fit: A model and empirical tests. Strategic management journal, 14(2), 137-153.

160. [159] Wilke, C. O. (2019). Fundamentals of data visualization: a primer on making informative and compelling figures. O'Reilly Media.

161. [160] Wang, F., Li, M., Mei, Y., & Li, W. (2020). Time series data mining: a case study with big data analytics approach. IEEE Access, 8, 14322-14328.

162. [161] Kelleher, J. D., Mac Namee, B., & D'arcy, A. (2020). Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies. MIT press.

163. [163] Wiltbank, R., Read, S., Dew, N., & Sarasvathy, S. D. (2009). Prediction and control under uncertainty: Outcomes in angel investing. Journal of business venturing, 24(2), 116-133.

164. [164] Miller, D., Steier, L., & Le Breton–Miller, I. (2016). What can scholars of entrepreneurship learn from sound family businesses?. Entrepreneurship Theory and Practice, 40(3), 445-455.

165. [165] Roach, D. C., Ryman, J. A., & Makani, J. (2016). Effectuation, innovation and performance in SMEs: an empirical study. European Journal of Innovation Management.

166. [166] Fisher, G. (2020). The complexities of new venture legitimacy. Organization Theory, 1(2), 2631787720913881.

167. [167] Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2021). When to use and how to report the results of PLS-SEM. European Business Review, 33(1), 2-24.

168. [168] Ode, E., & Ayavoo, R. (2020). The mediating role of knowledge application in the relationship between knowledge management practices and firm innovation. *Journal of Innovation & Knowledge, 5(3)*, 210-218.

169. [169] Sarstedt, M., Hair, J. F., Nitzl, C., Ringle, C. M., & Howard, M. C. (2020). Beyond a tandem analysis of SEM and PROCESS: Use of PLS-SEM for mediation analyses! *International Journal of Market Research, 62(3)*, 288-299.

170. [170] Henseler, J., Ringle, C. M., & Sarstedt, M. (2021). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43(1)*, 115-135.

171. [171] Chih-Pei, H. U., & Chang, Y. Y. (2017). John W. Creswell, research design: Qualitative, quantitative, and mixed methods approaches. *Journal of Social and Administrative Sciences, 4(2)*, 205-207.

172. [172] Müller, O., Fay, M., & Vom Brocke, J. (2018). The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. *Journal of Management Information Systems, 35(2)*, 488-509.

173. [173] Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (Eds.). (1998). Applied regression analysis: a research tool. New York, NY: Springer New York.

174. [174] Meng, F., Wang, X., Chen, H., Zhang, J., Yang, W., Wang, J., & Zheng, Q. (2016). The influence of organizational culture on talent management: A case study of a real estate company. *Journal of Chinese Human Resource Management, 7(2)*, 129-146.

175. [175] Dey, N., Das, H., Naik, B., & Behera, H. S. (Eds.). (2019). Big data analytics for intelligent healthcare management. Academic Press.

176. [176] Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.

177. [177] Kelleher, J. D., Mac Namee, B., & D'arcy, A. (2020). Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies. MIT press.

178. [178] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T.P., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide.

179. [179] Zhu, Y., Hernandez, L. M., Mueller, P., Dong, Y., & Forman, M. R. (2013). Data acquisition and preprocessing in studies on humans: what is not taught in statistics classes?. *The American Statistician, 67(4)*, 235-241.

180. [180] Nemeth, M., & Peterkova, A. (2018, June). Proposal of data acquisition method for industrial processes in automotive industry for data analysis according to Industry 4.0. In 2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES) (pp. 000157-000162). IEEE.

181. [181] Bell, E., Bryman, A., & Harley, B. (2022). Business research methods. Oxford university press.

182. [182] Mitchell, R. (2018). Web scraping with Python: Collecting more data from the modern web. " O'Reilly Media, Inc.".

183. [183] Fujita, H., Selamat, A., & Omatu, S. (Eds.). (2017). New Trends in Intelligent Software Methodologies, Tools and Techniques: Proceedings of the 16th International Conference SoMeT_17 (Vol. 297). IOS Press.

184. [184] Creswell, J. W., & Poth, C. N. (2016). Qualitative inquiry and research design: Choosing among five approaches. Sage publications.

185. [185] Silvola, R., Harkonen, J., Vilppola, O., Kropsu-Vehkapera, H., & Haapasalo, H. (2016). Data quality assessment and improvement. *International Journal of Business Information Systems, 22(1)*, 62-81.

186. [186] Armina, R., Zain, A. M., Ali, N. A., & Sallehuddin, R. (2017, September). A review on missing value estimation using imputation algorithm. In Journal of Physics: Conference Series (Vol. 892, No. 1, p. 012004). IOP Publishing.

187. [187] Little, R. J., & Rubin, D. B. (2019). Statistical analysis with missing data (Vol. 793). John Wiley & Sons.

188. [188] Aggarwal, C. C., & Reddy, C. K. (Eds.). (2020). Data mining: The textbook. Springer.

189. [189] Morgenthaler, S. (2009). Exploratory data analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 1(1), 33-44.

190. [190] Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. (Eds.). (2010). Handbook of educational data mining. CRC press.

191. [191] Kelleher, J. D., Mac Namee, B., & D'arcy, A. (2020). Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies. MIT press.

192. [192] Dipanjan, S., Raghav, B., & Tushar, S. (2017). Practical Machine Learning with Python: A Problem-Solver's Guide to Building Real-World Intelligent Systems.

193. [193] Mahesh, B. (2020). Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet], 9, 381-386.

194. [194] Zhang, F., & O'Donnell, L. J. (2020). Support vector regression. In Machine learning (pp. 123-140). Academic Press.

195. [195] Segal, M. R. (2004). Machine learning benchmarks and random forest regression.

196. [196] Schwenker, F. (2013). Ensemble methods: Foundations and algorithms [book review]. IEEE Computational Intelligence Magazine, 8(1), 77-79.

197. [197] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

198. [198] Meade, N., & Islam, T. (2006). Modelling and forecasting the diffusion of innovation– A 25-year review. International Journal of forecasting, 22(3), 519-545.

199. [199] Zotteri, G., Kalchschmidt, M., & Caniato, F. (2005). The impact of aggregation level on forecasting performance. International Journal of Production Economics, 93, 479-491.

200. [200] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. International Journal of Forecasting, 34(4), 802-808.

201. [201] Yan, B. N., Lee, T. S., & Lee, T. P. (2015). Analysis of research papers on E-commerce (2000–2013): based on a text mining approach. Scientometrics, 105, 403-417.

202. [202] Madni, H. A., Anwar, Z., & Shah, M. A. (2017, September). Data mining techniques and applications—A decade review. In 2017 23rd international conference on automation and computing (ICAC) (pp. 1-7). IEEE.

203. [203] Dol, S. M., & Jawandhiya, P. M. (2023). Classification Technique and its Combination with Clustering and Association Rule Mining in Educational Data Mining—A survey. Engineering Applications of Artificial Intelligence, 122, 106071.

204. [204] Zheng, A., & Casari, A. (2018). Feature engineering for machine learning: principles and techniques for data scientists. " O'Reilly Media, Inc.".

205. [205] Ghahramani, A., Pantelic, J., Lindberg, C., Mehl, M., Srinivasan, K., Gilligan, B., ... & Wellbuilt for Wellbeing Project Team. (2018). Learning occupants' workplace interactions from wearable and stationary ambient sensing systems. Applied energy, 230, 42-51.

206. [206] Centofanti, F., Fontana, M., Lepore, A., & Vantini, S. (2022). Smooth lasso estimator for the function-on-function linear regression model. Computational Statistics & Data Analysis, 176, 107556.

207. [207] Petković, M., Ceci, M., Pio, G., Škrlj, B., Kersting, K., & Džeroski, S. (2022). Relational tree ensembles and feature rankings. Knowledge-Based Systems, 251, 109254.

208. [208] Singer, J. D., Willett, J. B., & Willett, J. B. (2003). Applied longitudinal data analysis: Modeling change and event occurrence. Oxford university press.

209. [209] Liamputtong, P. (2019). *Handbook of research methods in health social sciences.*

210. [210] Raschka, S. (2018). *Model evaluation, model selection, and algorithm selection in machine learning. arXiv preprint arXiv:1811.12808.*

211. [211] Hutter, F., Kotthoff, L., & Vanschoren, J. (Eds.). (2019). *Automated Machine Learning: Methods, Systems, Challenges. Springer Nature.*

212. [212] Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., ... & Maier, H. R. (2021). *The future of sensitivity analysis: An essential discipline for systems modeling and policy support. Environmental Modelling & Software, 137, 104954.*