

GROUP PROJECT OUTLINE



College of Professional Studies

Northeastern University, Boston

Presented By: Group 6

ALY6980: Capstone Project

Prof. Roy Wada

Date of Submission: May 15, 2023

INTRODUCTION

The composition of a company's board of directors is a critical factor that can impact on the company's performance. Directors bring diverse skills, experiences, and perspectives to the decision-making process, which can influence the overall direction and success of the company. In recent years, there has been a growing interest in understanding the characteristics of board members and how these characteristics impact company performance.

This analysis is based on data provided by FreeFloatMedia, which includes information about the directors of multiple companies between 2018 and 2023 and their influence on key performance indicators (KPIs) during their tenure. The purpose of this individual project is to explore the relationship between director characteristics and company performance, with a focus on age and gender.

The analysis will begin with a literature review to justify the research structure and methodology. Then, the project will describe the data source and the software platform used for the analysis. Next, the preliminary results of the analysis will be presented, including histograms of director age, the distribution of age by gender, and summary statistics for age by gender.

The findings of this analysis will have important implications for companies in terms of how they select and appoint directors to their boards. By understanding the factors that impact director characteristics and company performance, companies can make informed decisions that lead to better outcomes.

ANALYSIS

The techniques used to explore the data include descriptive statistics, data visualization, and regression analysis. Descriptive statistics were used to summarize the age and gender characteristics of the board of directors, while data visualization techniques such as histograms were used to visualize the distribution of age and age by gender. Regression analysis was used to investigate the relationship between director age and KPIs. While one regression method is checking the relationship between performance indices, we also have other regression methods to analyze the relationship between the directors and the factors that are responsible for producing higher influence than others in the board.

To tackle the sponsor's business question of exploring the relationship between director characteristics and company performance, the analysis focused on age, gender and listed influence factors or drivers as key characteristics. As a team, we have break-down the mentioned factors as our individual research project and try to analyze the data in respective manner to have a holistic view of the data analysis in this provided dataset. Initially, the analysis used descriptive statistics and data visualization techniques to explore the age and gender composition of the board of directors, and regression analysis using Linear Regression paired with statistical correlation to investigate the relationship among the variables. These techniques were used to provide insights into the factors that impact company performance and guide decision-making around board appointments.

During our preliminary analysis, we performed the basic data cleaning process over the data to bring the dataset to a state where we can explore it more in a meaningful way and drafted visual charts to understand it more. In our first analysis, we have tried to investigate the relationship between the influence scores of directors and their respective gender. Here, we have tried to understand if gender has any specific implications over the influence of directors among the organization. In our second analysis, we have tried to analyze the data variables to understand the relationship between factors that are influencing the board members to be effective in their respective organizations by studying their influence scores recorded throughout these past 6 years since 2018. And finally, we have performed analysis to understand the performance of current directors using their recorded Win-Rates in following categories such as EBITDA, TSR, Carbon Intensity, and Controversy.

The research structure and analytical plan were justified based on a literature review of existing studies on the relationship between director characteristics and company performance. The literature review revealed that age and gender are important factors that impact the composition of the board of directors and have implications for company performance. Additionally, regression analysis was identified as an appropriate technique to investigate the relationship between director characteristics and KPIs. The analytical plan was designed to provide a comprehensive understanding of the factors that impact company performance and guide decision-making around board appointments.

PRELIMINARY RESULTS

Gender Analysis

The primary step to start the analysis was by checking significant variations in the age and gender distribution among the board of directors. Initially, the data were categorized into two data frames taking gender into consideration i.e., male and female from ‘DIRECTOR_COREMETRICS’ table. The first table (Table1) provides the summary statistics when gender is set as males and the second table (Table 2) when it is female.

The summary statistics for age by gender show that the mean age for male directors is 60.38 years with a standard deviation of 9.72, while the mean age for female directors is 57.53 years with a standard deviation of 8.62. The summary statistics suggest that there is a slightly lower age range for female directors compared to male directors.

Table 1: Summary statistics of Board of Directors who are Male.

	DIRECTOR: AGE	COMPANY: ACTIVE DIRECTORS	COMPANY: ALL DIRECTORS	INFLUENCE:20 23	INFLUENCE:20 22	INFLUENCE:20 21
count	59678.00	66006.00	66006.00	66006.00	59365.00	50615.00
mean	60.38	9.94	13.29	12.12	12.44	12.83
std	9.72	3.03	4.70	13.74	13.89	14.40
min	21.00	3.00	3.00	0.00	0.00	0.00
25%	54.00	8.00	10.00	4.00	4.00	4.00
50%	61.00	9.00	13.00	7.00	8.00	8.00
75%	67.00	12.00	16.00	15.00	15.00	16.00
max	100.00	29.00	39.00	100.00	100.00	100.00

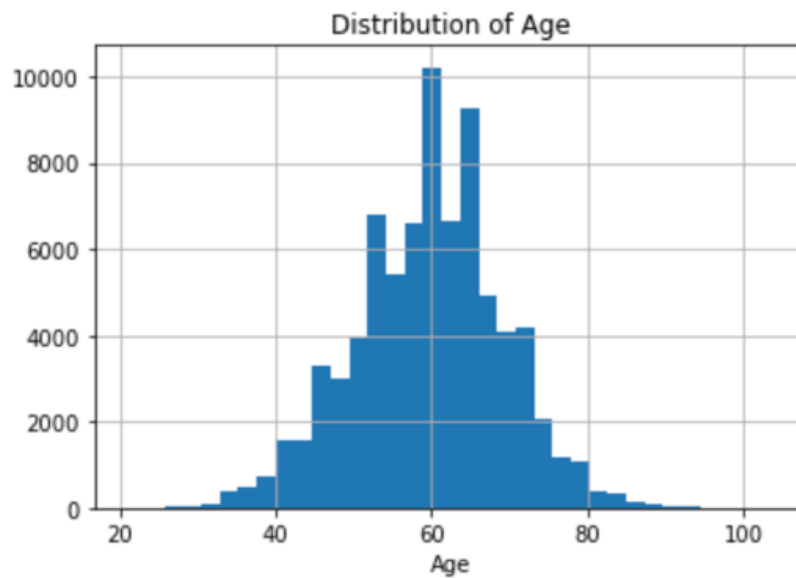
The age of the board members ranged from 21 years to 100 years, with the concentration majorly from 54 – 67 years. The company’s growth can be understood by its establishment in the market and product/service quality. Factors such as the influence of the board members are evaluated, with mean decreasing yearly.

Table 2: Summary statistics of Board of Directors who are Female.

	DIRECTOR: AGE	COMPANY: ACTIVE DIRECTORS	COMPANY: ALL DIRECTORS	INFLUENCE: 2023	INFLUENCE: 2022	INFLUENCE: 2021
count	18877.00	20525.00	20525.00	20525.00	17678.00	14004.00
mean	57.53	10.05	13.55	7.60	7.95	8.20
std	8.62	2.94	4.55	8.18	8.32	8.69
min	24.00	3.00	3.00	0.00	0.00	0.00
25%	52.00	8.00	10.00	3.00	3.00	3.00
50%	58.00	10.00	13.00	6.00	6.00	6.00
75%	63.00	12.00	16.00	9.00	10.00	10.00
max	104.00	29.00	39.00	100.00	100.00	100.00

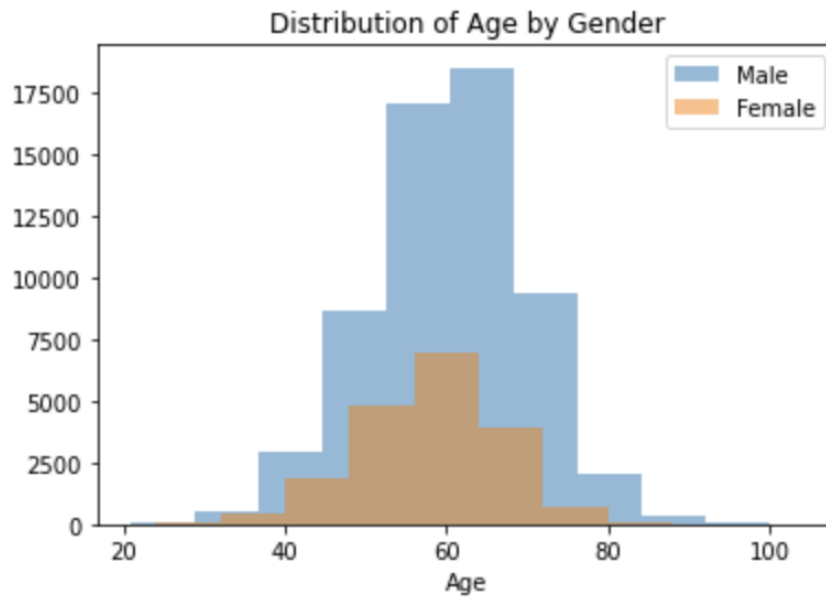
The age of the board members ranged from 24 years to 104 years, with the concentration majorly from 52 – 63 years. The influence of the board members when gender is considered female shows same result as male, where the mean of the influence is decreasing.

Figure 1: Histogram distribution among the Age of Board of Directors



Following is the data presentation as histogram of age shows a roughly normal distribution with a peak around 60 years old (Figure 1). However, there is a significant tail to the right indicating some directors are in their seventies and mid-seventies.

Figure 2: Histogram distribution of the KDE graph of Age by Gender



The second graph (Figure 2) shows the distribution of age by gender, with female directors being slightly younger on average compared to male directors. The kernel density estimate plot for females is skewed to the right, indicating a smaller proportion of female directors are in the upper age brackets compared to male directors.

Influence-Factor Analysis

After performing the initial gender analysis, we moved forward to analyze the variable relationship of influence factors with the director's average influence in the past 6 years. To observe the statistics of the data using “*describe*” function that provided me with the below information in Table 3:

Table 3: Statistical distribution of Director and Company information

	DIRECTOR: NAME	DIRECTOR: GENDER	COMPANY: NAME	COMPANY: TICKER	COMPANY: DOMICILE	COMPANY: SECTOR
count	86531	86531	86531	86468	86531	86025
unique	70156	2	9529	8956	67	11
top	Wei Zhang	Male	Banco Comercial Portugues, S.A.	SOL	US	Industrials
freq	13	66006	29	44	25040	13350

Here, from the above table, we can get an initial idea about the distribution of categorical data that points us regarding the data uniqueness for each participating column. Here, we are dealing with only two genders namely Male or Female. Most of the companies have US domiciles and belong to the Industrial sector. If we observe carefully, almost 3/4th of the data records belongs to the male directors than female ones which show a brief distribution of gender participation in company board.

As we are here highly interested in understanding the influence factor distribution and perform my analysis over that, we tried to focus my analysis on the Influence drivers that have been listed in the data. To understand the distribution of the data, we prepared a rudimentary frequency table (Table 4) where we can clearly observe that majority of the observations are not having the listed influences factors, however, it will be interesting to observe out of these factors which factors are still dominant in building a good director's influence in the team.

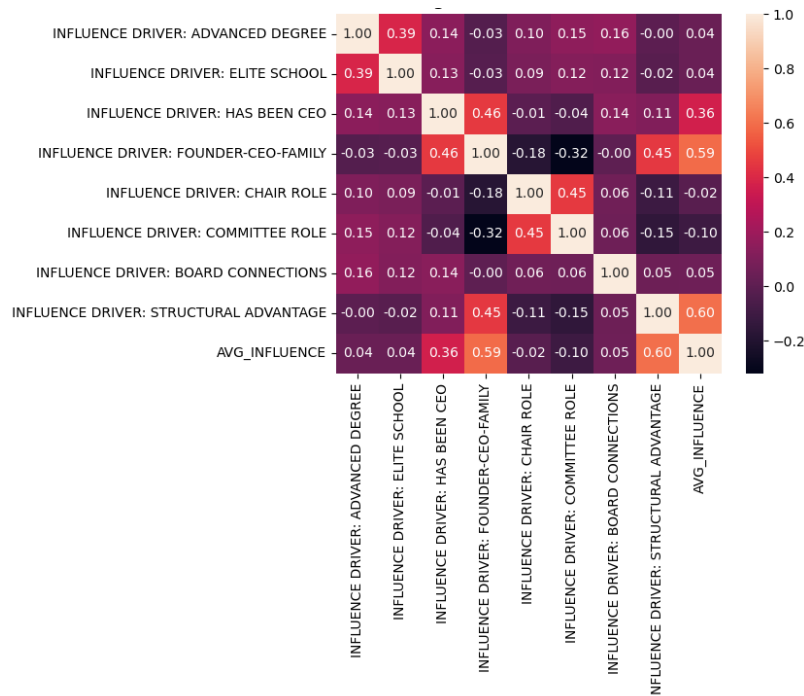
Table 4: Count of Influence Drivers/Factors throughout the data

	INFLUEN CE DRIVER: ADVANC ED DEGREE	INFLUEN CE DRIVER: ELITE SCHOOL	INFLUEN CE DRIVER: HAS BEEN CEO	INFLUEN CE DRIVER: FOUNDER -CEO- FAMILY	INFLUEN CE DRIVER: CHAIR ROLE	INFLUENC E DRIVER: COMMITT EE ROLE	INFLUENCE DRIVER: BOARD CONNECTION S	INFLUENCE DRIVER: STRUCTUR AL ADVANTAG E
No	50564	71936	56174	72391	64606	32762	77784	79069
Yes	35967	14595	30357	14140	21925	53769	8747	7462

After analyzing the above frequency table, I tried to interpret these influence drivers/factors with influences that each director is having from the past 6 years since 2018 to the recent year. To get the influence data, we extracted the average from all the years of influence for each director and created a new variable, ‘AVG_INFLUENCE’ which is a means of all the influence. The idea behind this is to observe which factor is leading to a higher average influence number for directors across these years. To observe that, we tried to build a heatmap with the correlation values among the variable to the target variable.

* (Placeholder) A chart for Influence factor distribution across gender of Directors

Fig. 3: Correlation Matrix of Influence Factors



Here, we can observe that factors such as “Structural Advantage”, “Founder-CEO-Family” are having higher correlation with the average influence than other factors. This indicates that the board of directors who are insiders are having much higher influential capacity than members appointed through other means.

Fig. 4: Comparison of Two factors driving average Influence.

The figure consists of two side-by-side box plots. The left plot is titled 'INFLUENCE DRIVER: STRUCTURAL ADVANTAGE' and the right plot is titled 'INFLUENCE DRIVER: COMMITTEE ROLE'. Both plots have 'AVG_INFLUENCE' on the y-axis, ranging from 0 to 100. The x-axis for both plots has two categories: 0 and 1. In the left plot, category 0 is represented by a blue box and category 1 by an orange box. In the right plot, category 0 is represented by a blue box and category 1 by an orange box. Both plots show a significant increase in average influence for category 1 compared to category 0, with category 1 reaching nearly 100% influence in both cases.

Figure 5: Model Summary to observe significant variables.

OLS Regression Results							
=====							
Dep. Variable:	AVG_INFLUENCE	R-squared:	0.519				
Model:	OLS	Adj. R-squared:	0.519				
Method:	Least Squares	F-statistic:	1.169e+04				
Date:	Mon, 08 May 2023	Prob (F-statistic):	0.00				
Time:	18:14:16	Log-Likelihood:	-3.1034e+05				
No. Observations:	86531	AIC:	6.207e+05				
Df Residuals:	86522	BIC:	6.208e+05				
Df Model:	8						
Covariance Type:	nonrobust						
=====							
		coef	std err	t	P> t	[0.025	0.975]

const		4.2289	0.059	71.175	0.000	4.112	4.345
INFLUENCE DRIVER: ADVANCED DEGREE		0.0825	0.067	1.235	0.217	-0.048	0.213
INFLUENCE DRIVER: ELITE SCHOOL		0.7942	0.087	9.140	0.000	0.624	0.964
INFLUENCE DRIVER: HAS BEEN CEO		4.0289	0.073	55.299	0.000	3.886	4.172
INFLUENCE DRIVER: FOUNDER-CEO-FAMILY		11.9902	0.107	112.013	0.000	11.780	12.200
INFLUENCE DRIVER: CHAIR ROLE		2.1312	0.077	27.703	0.000	1.980	2.282
INFLUENCE DRIVER: COMMITTEE ROLE		1.1920	0.072	16.495	0.000	1.050	1.334
INFLUENCE DRIVER: BOARD CONNECTIONS		0.0372	0.101	0.368	0.713	-0.161	0.235
INFLUENCE DRIVER: STRUCTURAL ADVANTAGE		19.6503	0.120	163.823	0.000	19.415	19.885
=====							
Omnibus:	29258.127	Durbin-Watson:	1.706				
...							
=====							

*(Placeholder) A Model summary after performing variable selection using gender and influence drivers combined.

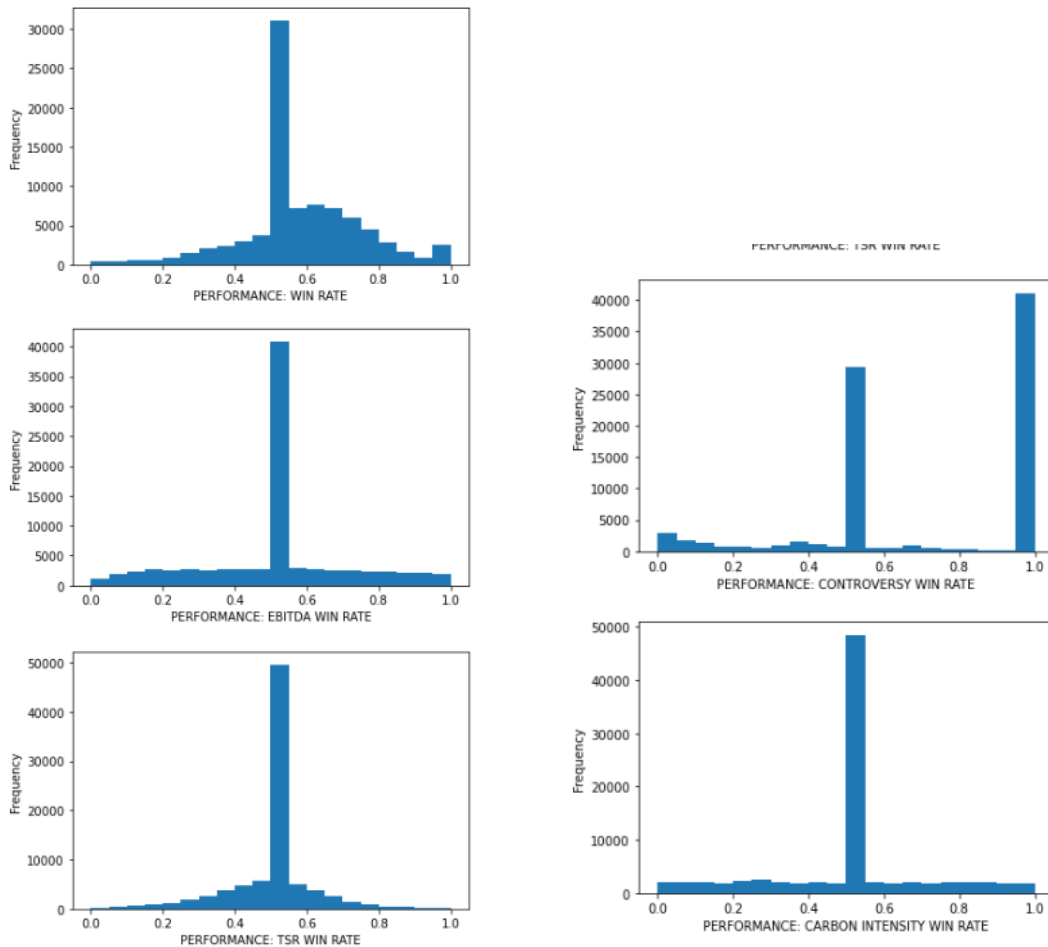
As we have observed the gender and influence factor above responsible for understanding the director influence standing across the board, we are now interested in understanding the performance of directors across multiple approaches including EBITDA, TSR, Carbon Intensity and Controversy. These are the win methods, which are responsible for explaining how successfully a director can influence the members of an organization.

The chart displays the performance win rate for ten different sectors across five key metrics. The y-axis represents the 'Performance Win Rate' from 0.0 to 0.8. The x-axis lists the metrics: PERFORMANCE_WIN_RATE, PERFORMANCE_EBITDA_WIN_RATE, PERFORMANCE_TSR_WIN_RATE, PERFORMANCE_CARBON_INTENSITY_WIN_RATE, and PERFORMANCE_CONTROVERSY_WIN_RATE. A legend on the right identifies the sectors by color.

Metric	Sector 1	Sector 9	Total Win Rate
PERFORMANCE_WIN_RATE	~0.55	~0.05	~0.60
PERFORMANCE_EBITDA_WIN_RATE	~0.50	~0.00	~0.50
PERFORMANCE_TSR_WIN_RATE	~0.49	~0.01	~0.50
PERFORMANCE_CARBON_INTENSITY_WIN_RATE	~0.49	~0.01	~0.50
PERFORMANCE_CONTROVERSY_WIN_RATE	~0.62	~0.18	~0.80

From the above graph we can observe that the performance win rate is higher for the performance categories of PERFORMANCE_WIN_RATE, PERFORMANCE_EBITDA_WIN_RATE, and PERFORMANCE_TSR_WIN_RATE. The Information Technology sector has the highest win rate for most of the performance categories. The Real Estate sector has the lowest win rate for most of the performance categories.

Figure 7: Histogram for the Performance Metrics.



The histograms show the distribution of each of the five performance metrics columns, with the x-axis representing the value range and the y-axis representing the frequency count. Each column has a range of values from 0 to 1, with higher values indicating better performance. For the performance win rate metric, the highest frequency count is observed for values ranging from 0.45 to 0.5. This suggests that a significant number of companies have a win rate that falls within this range, and that there is a clustering of companies around this range. Similarly, for the performance: controversy win rate metric, the highest frequency count is observed for a value of 1.0, indicating that a significant number of companies have a perfect score on this metric. The second highest frequency count is observed for a value of 0.5, indicating that there is a clustering

of companies with a score of 0.5 on this metric. For many of the columns, the frequency count is highest for values ranging from 0.45 to 0.5, indicating that there is a clustering of companies around this range for most performance metrics. While we can even observe that the graph is distributed over the entire value range.

Table 5: Descriptive Table for the Win Rates

	PERFORMANCE: WIN RATE	
count	86772	
unique	996	
top	Unrated	
freq	25782	
	PERFORMANCE: EBITDA WIN RATE	
count	86772	
unique	1002	
top	Unrated	
freq	38275	
	PERFORMANCE: TSR WIN RATE	
count	86772	
unique	989	
top	Unrated	
freq	43679	
	PERFORMANCE: CARBON WIN RATE	
count	86772	
unique	1002	
top	0.5	
freq	46326	
	PERFORMANCE: CONTROVERSY WIN RATE	
count	86772	
unique	932	
top	1	
freq	41105	

The table provided contains performance metrics for different categories, specifically focusing on win rates. The dataset consists of 86,772 entries. In the "Performance: Win Rate" category, there are 996 unique values, with the most frequent value being "Unrated" appearing 25,782 times. Similarly, in the "Performance: EBITDA Win Rate" category, there are 1,002 unique values, with "Unrated" being the most common value occurring 38,275 times. The "Performance: TSR Win Rate" category has 989 unique values, with "Unrated" appearing 43,679 times. In the "Performance: Carbon Win Rate" category, there are 1,002 unique values, and the most frequent value is "0.5," which occurs 46,326 times. Finally, the "Performance: Controversy Win Rate" category has 932 unique values, with "1" being the top value appearing 41,105 times. These metrics provide insights into the win rates across different performance areas, with "Unrated" indicating missing or unavailable data for those specific performance metrics.

CONCLUSION

At the conclusion of the Capstone project, we would like to suggest that so far, we have performed rudimentary analysis to observe the data provided and understand the variables and their relationship with the performance and factors involved. We continuously dig deeper with our analysis and try to extract more insights from it. We are looking for methods to optimize the dataset to a point which can be clearly used with the dashboarding tools to showcase analytical results.

Our team aims to deliver insightful key outcomes in this project that include the finalized Tableau dashboard, a comprehensive analysis of the impact of director age and influence metrics on performance. A comparison of director performance metrics across different countries and regions, and recommendations for stakeholders on how to use the dashboard and predictive models for better-informed decision-making regarding director performance and corporate governance. These deliverables will provide valuable insights and tools for stakeholders to improve governance practices and make informed decisions.

To share the work on LinkedIn or an online portfolio while honoring the terms of a non-disclosure agreement (NDA), it is crucial to ensure that no confidential or sensitive information is disclosed. Will focus on highlighting the skills, methodologies, and tools used in the project without revealing specific proprietary information. We will emphasize our contributions to the data analysis process, the insights gained, and the impact of the project on improving corporate governance. We will Avoid mentioning any confidential data, company names, or specific details that may violate the NDA. Instead, we will present our work in a general and informative manner that showcases our expertise and proficiency in data analysis and visualization.

REFERENCES

- The Harvard Law School Forum on Corporate Governance. (2019, June 28). *Dual-Class Shares: Governance Risks and Company Performance*. <https://corpgov.law.harvard.edu/2019/06/28/dual-class-shares-governance-risks-and-company-performance/>
- Judge, W. Q., & Zeithaml, C. P. (1992). Institutional and strategic choice perspectives on board involvement in the strategic decision process. *Academy of Management Journal*, 35(4), 766–794. <https://doi.org/10.2307/256315>
- Hermalin, B. E., & Weisbach, M. S. (2012). Information Disclosure and Corporate Governance. *Journal of Finance*, 67(1), 195–233. <https://doi.org/10.1111/j.1540-6261.2011.01710.x>
- Ngo, H. (2022, August 9). *How to Clean Your Data in Python*. Medium. <https://towardsdatascience.com/how-to-clean-your-data-in-python-8f178638b98d>
- *How to calculate summary statistics — pandas 2.0.1 documentation*. (n.d.). How to Calculate Summary Statistics — Pandas 2.0.1 Documentation. https://pandas.pydata.org/docs/getting_started/intro_tutorials/06_calculate_statistics.html