CAPSTONE FINAL PROJECT

College of Professional Studies

Northeastern University, Boston

Presented By: Group 6

Amryta Panda

Sneha Prabhu

Durga Bhanu Nayak

ALY6980: Capstone Project

Prof. Roy Wada

Date of Submission: Jun 28, 2023

INTRODUCTION

Corporate governance plays a crucial role in ensuring transparency, accountability, and responsible decision-making within organizations. At the heart of effective governance lies the Board of Directors, a group of individuals entrusted with overseeing the strategic direction and performance of a company. Given their pivotal role, analyzing the influence, structure, and performances of boards is essential for understanding the dynamics that drive corporate success or failure. This project aims to delve into the realm of corporate governance by conducting a comprehensive analysis of the Board of Directors across multiple companies. By examining the composition, diversity, expertise, and influence-factors of these boards, we seek to identify the key factors that contribute to effective governance and ultimately enhance the overall performance and sustainability of organizations.

After conducting an extensive literature review on the importance of timely evaluation of regular Board members and the factors that contribute to an effective board, it has become evident that certain key elements significantly influence a board's success. Our project, sponsored by FreeFloatMedia, involves collecting demographic data, influence factors, and performance ratings of board members from various organizations. In line with the findings discussed in Cossin and Caballero's article "The Four Pillars of Board Effectiveness" (2022), the skillset of individuals plays a pivotal role in maintaining an effective board. Analyzing the data provided by our sponsor, we have chosen to focus on these factors and investigate the essential skillsets required for board directors who consistently demonstrate a high average influence metric over the years. This research aims to employ data analytics principles to explore the relationships among these factors.

To gain a deeper understanding, we have utilized summary statistics and implemented a Linear Regression model to examine the significance of various variables and their ability to predict a director's higher average influence. Additionally, we utilized the factors identified over our final deliverable as a Tableau Dashboard. Here, we have integrated our individual analysis into the main group project. By collaborating as a team, we tried to present meaningful insights extracted from the provided data, offering valuable guidance for businesses to implement in their processes.

The final dashboard is prepared to include three main aspects of the project as per our understandings. These are overall diversity, influence factors and performance of the board members. Hence, on each of the story boards we have tried to present these aspects in terms of overall statistics, company-wise statistics, and individual statistics. This is an interactive dashboard that will take input from the user to select the company or director name and present their corresponding stats accordingly.

METHODS

The project will utilize a combination of quantitative and qualitative research methodologies. For conducting the factor analysis, Python programming language will be employed as it offers a wide range of functionalities for reading and analyzing data from various sources, including flat files, relational databases, and web scraping. Python libraries provide comprehensive tools for data access, analysis, visualization, and statistical modeling. The analysis process will adhere to standard data analytics procedures, encompassing data loading, cleansing, exploratory data analysis, visualization, and extraction of relevant information.

To handle the provided data in Excel format, Visual Studio Code software will be utilized to run the Jupyter notebook. Leveraging the integrated Jupyter environment, the ‘pandas’ library will be employed to read the Excel file using the ‘read_excel’ command. This approach enables the data to be read into a pandas DataFrame object, facilitating efficient data manipulation and analysis. For visualizations, two Python libraries, namely ‘matplotlib’ and ‘seaborn’, will be utilized. These libraries offer a comprehensive set of functions for creating visually appealing charts and graphs.

Drawing inspiration from W. Bruce's article, "Six ways Owners can Keep their Boards Fresh" (2023), which emphasizes the importance of clear communication to maintain team cohesion, I aim to prioritize clarity and effective data communication throughout the project. This will ensure that my ideas and insights are conveyed clearly to team members and stakeholders. In the literature, it has been noted that histograms and KDE (Kernel Density Estimation) plots are commonly used for visualizing variable distributions (Navarro, 2015). Additionally, descriptive statistics such as mean and standard deviation are useful for assessing central tendency and variability of the data (Field, 2013).

To further explore age disparities between men and women, we will employ the describe() method to generate summary statistics separately for each gender. This will provide insights into the age variable's range, mean, standard deviation, and other statistical indicators for male and female directors. The descriptive data obtained from the describe() function will enhance our understanding of the age distribution among male and female directors.

Building upon the factors studied, we will try to build the final deliverables in the form of a dashboard which we have prepared using Tableau platform. To accomplish this, we have utilized Tableau Desktop (version 2023.1) software package. This gives us the opportunity to build and prepare dashboards by extracting data from multiple sources such as excel, text, database, etc. Additionally, it also empowers users to deploy the dashboard as a packaged format to public server. Furthermore, utilizing Tableau as a platform will greatly enhance the visualization of insights by presenting them through a variety of graphs, reports, and interactive dashboards. The upcoming section will feature visually compelling figures that convey narratives on the overall performance of the Board of Directors, their influence, and key performance indicators (KPIs) that encapsulate critical statistics.

Overall, this project aims to explore the relationship between director traits and firm performance, utilizing Python and Tableau as the software platform. The analysis will employ a combination of descriptive and inferential statistics, along with appropriate visualization tools, to gain valuable insights from the data. These chosen methodologies are grounded in existing literature and ensure a rigorous and systematic investigation of the research question.

Expected Impact:

This project's findings will contribute to the broader discourse on corporate governance and provide practical recommendations for enhancing board effectiveness across diverse industries. By identifying the factors that contribute to successful governance, we aim to empower companies to strengthen their boards, improve decision-making processes, and drive sustainable growth.

PRELIMINARY RESULTS

Upon importing the data from the Excel sheet into the Python environment, we obtained preliminary insights from the various sheets. The first sheet contains the Directors' core metrics data, consisting of **86,772** records and comprising **35** attributes. This dataset provides essential information about the directors, including their effectiveness metrics such as influence details and win rate information across multiple platforms. The second sheet contains company-related data, including details about board members and performance metrics specific to each company. This dataset offers insights into the company's governance structure and performance indicators. The third sheet combines the director-level data with their corresponding company data, providing a consolidated view.

For this project, we will focus on analyzing all the datasets to gain a comprehensive overview of the entire dataset. This approach will enable us to establish connections between the directors' data and the associated company data. Due to our strong interest in understanding the distribution of influence factors and conducting an in-depth analysis, we focused our attention on the listed Influence drivers within the dataset. To gain insights into the data distribution, we created a basic frequency table (Table 1) which revealed that most of the observations did not possess the listed influence factors. However, it remains intriguing to determine which factors still play a dominant role in shaping a director's influence within a team.

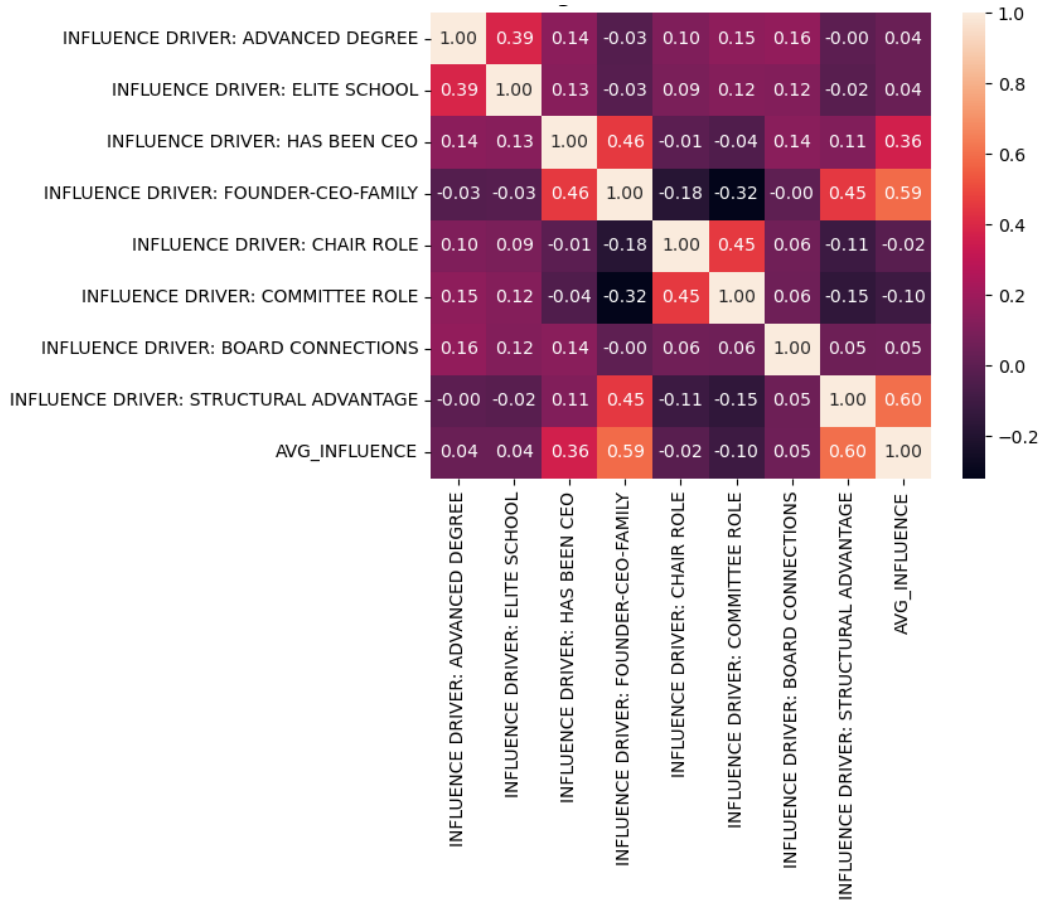
Table 1: Count of Influence Drivers/Factors throughout the data

	INFLUENCE DRIVER: ADVANCED DEGREE	INFLUENCE DRIVER: ELITE SCHOOL	INFLUENCE DRIVER: HAS BEEN CEO	INFLUENCE DRIVER: FOUNDER-CEO-FAMILY	INFLUENCE DRIVER: CHAIR ROLE	INFLUENCE DRIVER: COMMITTEE ROLE	INFLUENCE DRIVER: BOARD CONNECTIONS	INFLUENCE DRIVER: STRUCTURAL ADVANTAGE
No	50564	71936	56174	72391	64606	32762	77784	79069
Yes	35967	14595	30357	14140	21925	53769	8747	7462

Building upon the analysis of the performance level tables and graphs mentioned earlier, we proceeded to interpret the influence drivers or factors associated with each director's influence from 2018 to the present year. To obtain the influence data, we calculated the average influence for each director across the past six years and generated a new variable named 'AVG_INFLUENCE.' This variable represents the mean influence value across all years. The purpose of this calculation is to identify the factors that contribute to a higher average influence score for directors over this period. To visualize these relationships, we created a heatmap displaying the correlation values between the variables and the target variable. By examining the correlation heatmap, we aim to uncover significant relationships between the variables and the

average influence of directors, shedding light on the factors that have a strong impact on their overall influence within the team.

Fig. 1: Correlation Matrix



Upon analysis, it is evident that factors such as "Structural Advantage" and "Founder-CEO-Family" exhibit a stronger correlation with the average influence compared to other factors. This finding suggests that board members who possess insider status, such as being part of the company's structure or having familial ties with the founder or CEO, tend to exert a significantly higher level of influence within the board. This underscores the notion that individuals appointed through these channels hold greater influential capacity compared to those selected through alternative means.

Analyzing Factors against EBITDA Win Rate

Upon analyzing the regression summary for influence factors, I proceeded to examine the overall average influence of the directors and its relationship with the EBITDA win rate. The analysis revealed a clear and gradual increase in the average influence of directors, leading to a positive impact on their EBITDA performance, specifically in terms of the win rate. This observation is supported by the regression analysis, which demonstrates a distinct positive correlation between these two variables.

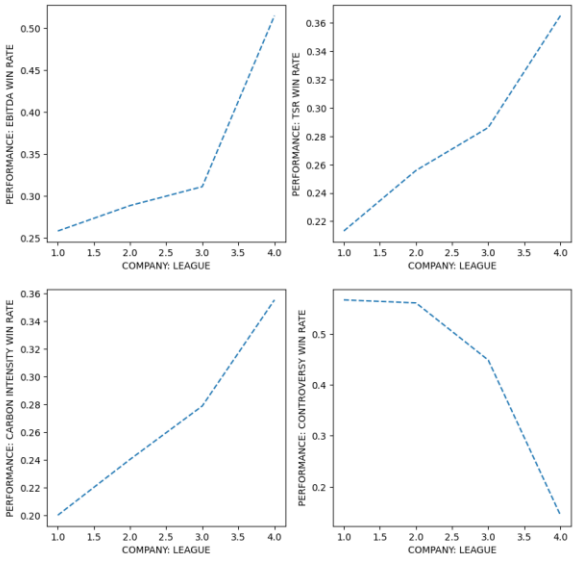
During our exploration of the data, we became curious about comparing the performance rates at a company level. To gain insights, we aggregated the average performance win rate values for each company league. The leagues are categorized based on the company's market capitalization, with higher league numbers indicating higher market capitalization. The following grouped data was generated to identify any patterns or extract meaningful insights at the company level.

Table 2: Average Performance Categories across each Company Leagues

COMPANY: LEAGUE	EBITDA WIN RATE	TSR WIN RATE	CARBON INTENSITY WIN RATE	CONTROVERSY WIN RATE
1	0.258	0.213	0.200	0.567
2	0.289	0.256	0.240	0.561
3	0.311	0.286	0.279	0.449
4	0.515	0.365	0.355	0.144

To gain a deeper understanding of the data presented in the table above, I decided to visually represent it in a chart (Fig. 2) that facilitates a clear comparison among the leagues. The line graph below demonstrates the average win rate performance across different categories, namely EBITDA, TSR, Carbon Intensity, and Controversial. The graph reveals an interesting pattern: companies in higher leagues exhibit higher average win rate performance in the EBITDA, TSR, and Carbon Intensity categories. Conversely, lower league companies demonstrate the highest performance in terms of Controversial win rate. This observation suggests that as companies ascend in leagues, they are less likely to engage in controversial activities that could potentially impact their overall performance in other crucial areas. By visualizing the data in this manner, we can easily grasp the comparative performance of companies across different leagues and discern potential correlations between league ranking and performance metrics.

Fig. 2: Higher League Companies are having lesser Controversial Win Rate than other categores



Linear Regression Model

Let's understand the significance of each variable involved in the process using a Linear Regression statistic.

Fig. 3: Model Summary to observe significant variables.

OLS Regression Results

Dep. Variable:	PERFORMANCE: EBITDA WIN RATE	R-squared:	0.049
Model:	OLS	Adj. R-squared:	0.049
Method:	Least Squares	F-statistic:	638.1
Date:	Mon, 19 Jun 2023	Prob (F-statistic):	0.00
Time:	21:34:04	Log-Likelihood:	-22065.
No. Observations:	86531	AIC:	4.415e+04
Df Residuals:	86523	BIC:	4.422e+04
Df Model:	7		
Covariance Type:	nonrobust		

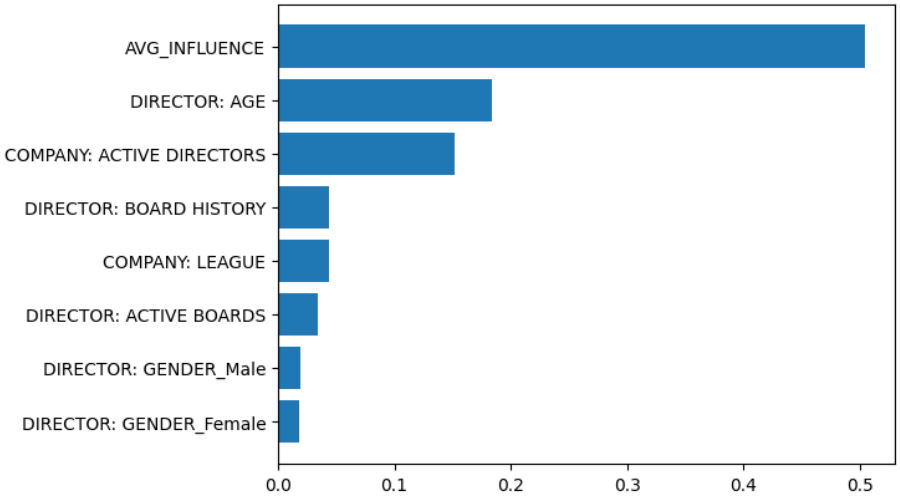
	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.1205	0.005	-22.806	0.000	-0.131	-0.110
DIRECTOR: AGE	0.0058	0.000	48.496	0.000	0.006	0.006
DIRECTOR: ACTIVE BOARDS	0.0342	0.003	10.221	0.000	0.028	0.041
DIRECTOR: BOARD HISTORY	-0.0279	0.003	-10.126	0.000	-0.033	-0.022
COMPANY: LEAGUE	0.0317	0.001	21.303	0.000	0.029	0.035
COMPANY: ACTIVE DIRECTORS	0.0018	0.000	4.690	0.000	0.001	0.003
AVG_INFLUENCE	0.0029	8.79e-05	33.343	0.000	0.003	0.003
DIRECTOR: GENDER_Female	-0.0685	0.003	-23.531	0.000	-0.074	-0.063
DIRECTOR: GENDER_Male	-0.0521	0.003	-17.532	0.000	-0.058	-0.046

Here, from the above summary statistics of the Linear Regression model, we can see that all the variables are statistically significant and are having either positive or negative impact on the EBITDA performance win rate.

Random Forest Model

Using this information, I would like to employ the Random Forest model to extract the effective variables that are highly affecting the EBITDA performance win rate of the board members.

Fig. 4: Random Forest Feature Importance.



Based on the feature importance plot (Fig. 4) generated using the Random Forest model, we can clearly observe that the variable 'AVG_INFLUENCE' holds the highest importance in determining the EBITDA performance win rate value. As calculated earlier in our analysis, 'AVG_INFLUENCE' represents the average influence value of a director over a six-year period (from 2018 to 2023). Additionally, the variables 'DIRECTOR: AGE' and 'ACTIVE_DIRECTORS' are also found to significantly impact the performance win rate. A higher age of board members suggests a potentially higher level of quality experience, leading to increased influence among the members. Similarly, having a greater number of influential active directors within an organization can result in improved market performance, consequently enhancing individual performance. Furthermore, it is worth noting that among all the variables, the gender of directors, particularly female gender, has the lowest impact on the model, aligning with our expectations.

PROJECT DELIVERABLES

In this project, as mentioned above in Introduction section, we aim to provide an interactive platform/dashboard for users to dynamically check and observe the company-wise and director-wise statistics and study the factors that are making them influential and letting them perform better than others.

It is an organized way to present data which is meaningful and easier to understand than simple text. The aggregation and presentation of data enables us to relate the insights to the context that we are hoping to gather. Here in this project, we are presenting the below deliverables as the final completed project.

List of deliverables:

1. Jupyter Notebook containing Factor Analysis in Python.
2. A comprehensive and interactive Tableau Dashboard.
3. Project report comprising the analysis and dashboard.

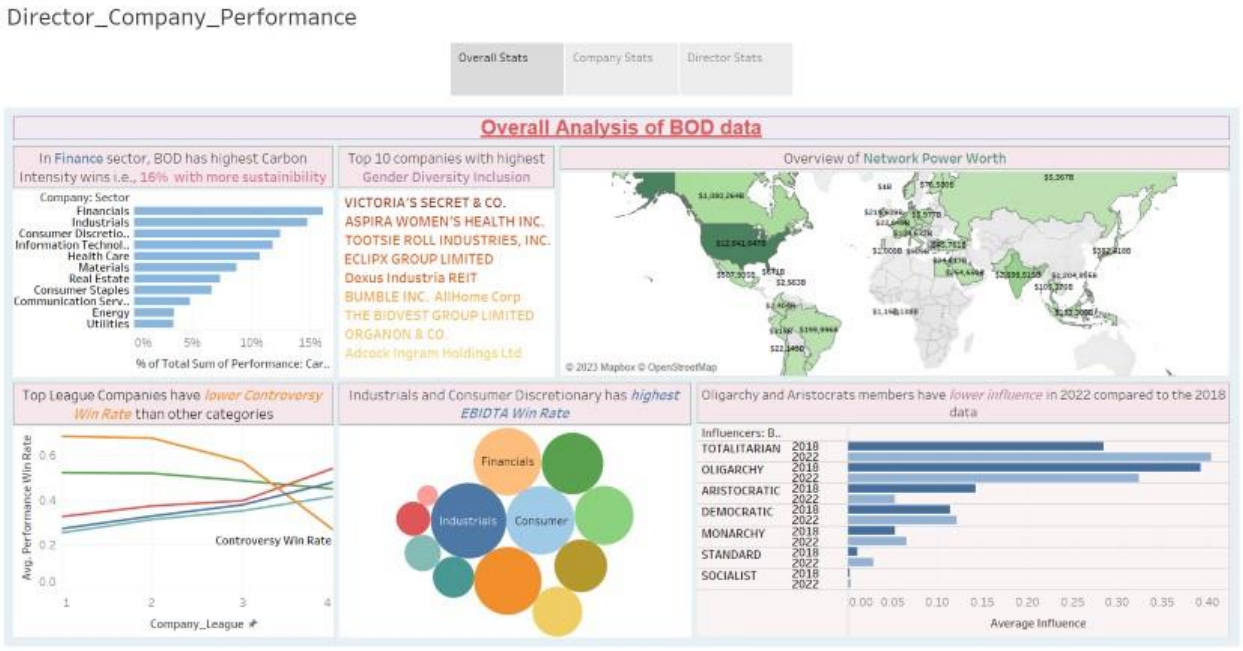
We initiated our project using the Jupyter Notebook to understand the complete dataset and to figure out the hidden patterns and factors to be focused on. As this is a director level data, here we performed the diversity analysis, factor analysis and performance analysis of the directors as per our individual contribution to the project respectively. As a result, we have successfully integrated the individual efforts to gain combined insight into each aspect such as diversity, influence factors, and performance. This is eventually utilized to build the final dashboard.

As the final dashboard deliverable, we have prepared a packaged Tableau dashboard that we would like to walk through each of the story sections. Here, we have prepared three story dashboards that include overall statistics of the director-company data, company-wise interactive dashboard, and director-wise interactive dashboard respectively to observe the data.

OVERALL DATA STATS

The dashboard created for the Board of Directors provides a comprehensive overview of major variables and key insights related to corporate governance. It offers a range of visualizations and graphs that shed light on crucial aspects of board dynamics and performance.

Through six distinct graphs, it provides valuable insights into key aspects of the dataset. Carbon intensity being a major focus in recent years, the bar graph provides the industry-wise carbon intensity win rate across different sectors. Next, a text graph shows the highest female gender inclusion among the companies.

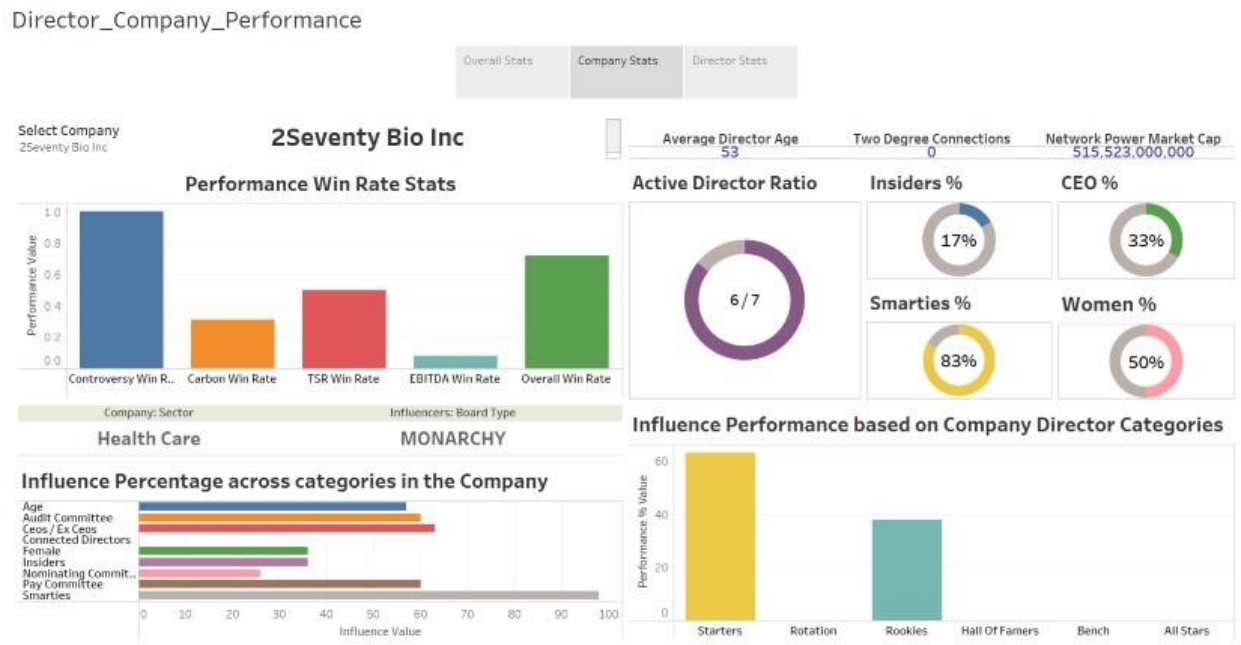


Being a board member also comes with profiting the company and bringing more business, as shown in the geographical graph presenting network power. The 4 performance win rates (EBIDTA, carbon intensity, controversy, and TSR) in the four company leagues. Moreover, EBIDTA win rates in various industries shown in the bubble chart. Then the influence of different board types of structures is shown through bar graph, demonstrating the influence in 2018 and how it changed through the course in 2022.

INDIVIDUAL COMPANY STATS

In this dashboard, we have included insights on a company level. Using this interactive dashboard, we can provide any company name through the drop-down menu and observe its composition. Below we can see that as we select the company name from top-left drop-down menu, all the stats get updated instantly and provides multiple graphical information such as performance categories standings, influential factors among the directors, composition of performance group, etc.

Additionally, we have also included information such as which sector the company belongs to, and what regime is the board type which can be found just below the Performance win rate stats bar graph. By including the average director's age and market capitalization network power information provides additional insights on what age group the company usually comprises of its board members and how well it is performing in the market.



Along with the performance and influence insights of the directors, we can observe the composition of the board members clearly by observing the dough-nut charts such as how many directors are currently active corresponding to the total number of directors of the company, the number of directors who are insiders (belongs to CEO family members), how many directors are having the previous work experience as a CEO, what percentage of directors are have went to prestigious schools and are considered as Smart, and how well the company promotes the diversity among its board.

INDIVIDUAL DIRECTOR STATS

In addition to the overall and company-wise outlook, we have developed an innovative and interactive dynamic dashboard that offers a comprehensive overview of various essential aspects of directors' profiles. The dashboard encompasses performance metrics, active directorship in companies, age, gender, board history, influence characteristics, and percentage of influence over the years, including specific years. With the inclusion of interactive filters for director names and IDs, users can effortlessly explore individual directors' information.

Through intuitive data visualization techniques, the dashboard provides valuable insights into the performance and influence patterns of directors over time. Users can analyze directors' holistic performance trends as well as delve into specific years of interest. This functionality enables effective decision-making and evaluation of director performance, empowering organizations, and stakeholders with a powerful resource.

The user-friendly interface and rich visualization capabilities of the dashboard make it accessible and easy to understand for all users. It enhances transparency in understanding the dynamics of directors and their contributions. By providing a clear and comprehensive representation of director-related data, the dashboard significantly contributes to improved governance and facilitates informed decision-making processes.

Overall, our dynamic dashboard offers a valuable tool for organizations and stakeholders seeking to gain actionable insights into director performance, influence dynamics, and other critical factors. It revolutionizes the way director information is accessed, analyzed, and utilized, fostering improved governance practices and informed decision-making across various industries.

CONCLUSION

Hence, we can state that the comprehensive Tableau dashboard provides a holistic view of the company's investment portfolio, offering valuable insights into the performance of individual directors and the standings of different companies in the current market. The analysis presented in the dashboard showcases overall, company-wise, and director-wise performance. To gain a deeper understanding of the portfolio, it is recognized that more thorough factor analysis is needed. This analysis should focus on attributes such as diversity, skillsets, and performance traits. These factors play a significant role in driving higher EBITDA performance win rates.

Additionally, it is observed that higher company league standings are associated with higher performance in EBITDA, TSR (Total Shareholder Return), and a lower carbon intensity win rate. This suggests that companies performing well in these areas can expect better overall performance and success.

Furthermore, the analysis reveals that board members with a structural advantage, CEO experience, or belonging to a family with insider connections exert higher influence over others. This indicates the importance of leadership and industry expertise in driving success within the company. These findings highlight the importance of average influence, director age, and active directors in determining EBITDA performance win rates. Understanding these dynamics can contribute to informed decision-making and strategic initiatives aimed at optimizing board composition and enhancing organizational performance.

Lastly, the average influence of totalitarian regimes has shown significant growth over the past six years. This finding highlights the impact of political and regulatory environments on business operations and performance.

Overall, the Tableau dashboard and the insights derived from it provide a comprehensive overview of the company's investment portfolio, highlighting key factors that drive performance and suggesting areas for further analysis and improvement.

REFERENCES

1. *Free Float Media*. (n.d.). Free Float Media. https://www.freefloatmedia.com/
2. Scgn, & Scgn. (2023, May 3). Examining the balance between performance metrics and remuneration for the board of directors. *Nairametrics*. https://nairametrics.com/2023/05/02/examining-the-balance-between-performance-metrics-and-remuneration-for-the-board-of-directors/
3. Dalton, D.R., Daily, C.M., Johnson, J.L. and Ellstrand, A.E. (1999) Number of Directors and Financial Performance A Meta-Analysis. The Academy of Management Journal, 42, 674-686. - References - Scientific Research Publishing.(n.d.).https://www.scirp.org/reference/ReferencesPapers.aspx?ReferenceID=1460248
4. 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
5. Predum, R. (2020, May 26). Customer Segmentation Analysis with Python. Medium. https://towardsdatascience.com/customer-segmentation-analysis-with-python-6afa16a38d9e
6. Mulani, S. (2020, October 13). 2 Easy Ways to Standardize Data in Python for Machine Learning - AskPython. AskPython. https://www.askpython.com/python/examples/standardize-data-in-python
7. Cossin, D., & Caballero, J. (2022). The Four Pillars of Board Effectiveness. *IMD Business School for Management and Leadership Courses*. https://www.imd.org/research-knowledge/articles/the-four-pillars-of-board-effectiveness/
8. Werner, B. (2023, April 13). Six Ways Owners Can Keep Their Boards Fresh. *Forbes*. https://www.forbes.com/sites/forbesbusinesscouncil/2023/04/13/six-ways-owners-can-keep-their-boards-fresh/?sh=5cd42b8a4d77
9. Stutzman, M., Thornton, L., & Nestler, M. (2022, June 13). *Boards Are Tying Goals to ESG Metrics*. The Harvard Law School Forum on Corporate Governance. https://corpgov.law.harvard.edu/2022/06/13/boards-are-tying-goals-to-esg-metrics/
10. *Linear Regression in Python using Statsmodels*. (n.d.). https://datatofish.com/statsmodels-linear-regression/