DSCI 510 Final Report
Nuo Chen

Name of the project: How do various factors shape companies' ESG risk ratings and performance? – ESG impact analysis

Description: Nowadays, ESG (Environmental, Social, and Governance) has garnered widespread popularity, addressing concerns that span from environmental conservation to financial investment. By scrutinizing a company's environmental footprint, governance practices, and societal impact, stakeholders gain a profound understanding of its resilience and vulnerabilities. Through an exploration of our dataset, which exclusively features companies from the renowned S&P 500 index, the project aims to investigate the various factors influencing the ESG ratings of companies and seek to gain insights into the relationships and patterns that emerge within the ESG landscape.

Data Sources: Initially, I intended to gather three datasets for analysis, However, I discovered that the address information in the S&P 500 ESG Risk Ratings dataset only includes state abbreviation, lacking the full state names. This presented a challenge as I couldn't directly correlate company address information with states' GDP data, which exclusively includes the full state names. Thus, I decided to incorporate an additional data source - U.S. states and territories abbreviations, to subsequently execute an inner join to link all 3 datasets altogether. Below are the datasets I included in my analysis.

1. S&P 500 ESG Risk Ratings, csv format dataset from Kaggle: https://www.kaggle.com/datasets/pritish509/s-and-p-500-esg-risk-ratings/data
This dataset consists of 503 rows and 15 columns of data. The columns in the dataset can be divides into 2 groups: the initial 7 columns comprise company information, while the subsequent 8 columns contain ESG-related data. The details are as follows.

	Symbol	Name	Address	Sector	Industry	Full Time Employees	Description	Total ESG Risk score	Environment Risk Score	Governance Risk Score	Social Risk Score	Controversy Level	Controversy Score	ESG Risk Percentile	ESG Risk Level
0	A	Agilent Technologies Inc	5301 Stevens Creek Boulevard\nSanta Clara, CA ...	Healthcare	Diagnostics & Research	18,000	Agilent Technologies, Inc. provides applicatio...	15.0	0.3	6.3	8.6	Low	1.0	11th percentile	Low
1	AAL	American Airlines Group Inc	1 Skyview Drive\nFort Worth, TX 76155\nUnited ...	Industrials	Airlines	132,500	American Airlines Group Inc., through its subs...	29.0	12.0	5.0	12.0	Moderate	2.0	62nd percentile	NaN
2	AAP	Advance Auto Parts Inc	4200 Six Forks Road\nRaleigh, NC 27609\nUnited...	Consumer Cyclical	Specialty Retail	40,000	Advance Auto Parts, Inc. provides automotive r...	12.0	0.0	3.0	8.0	Moderate	2.0	4th percentile	Negligible

Table 1 S&P 500 ESG Risk Ratings dataset

2. U.S. states and territories by GDP, link to perform web scraping from Wikipedia: https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_GDP
Using the request library and beautiful soup package, I have accessed the website and create gdp_raw.csv file. This dataset consists of 54 rows and 3 columns – state, gdp_2022, gdp_2023. Few rows of the dataset are as follows:

state	gdp_2022	gdp_2023
California *	3,598,103	3,755,487

Texas *	2,355,960	2,436,346
New York *	2,053,180	2,135,672

Table 2 U.S. states by GDP dataset

3. U.S. states and territories abbreviations, link to perform web scraping from Wikipedia:
https://en.wikipedia.org/wiki/List_of_U.S._state_and_territory_abbreviations
Using the request library and beautiful soup package, I have accessed the website and create state_name_raw.csv file. This dataset consists of 78 rows and 2 columns – state, state_abbreviation. Few rows of the dataset are as follows:

state	state_abbreviation
Alabama	US-AL
Alaska	US-AK
Arizona	US-AZ

Table 3 U.S. states abbreviations dataset

4. Financial data of related companies, including market value, latest revenue, and stock volatility from Yahoo: https://finance.yahoo.com.
Using the yfinance library in Python, I have extracted the 503 rows and 3 columns of data and store them in fin_raw.csv file. According to Yahoo Finance, the data comprises the most recent information disclosed in the latest financial report. Few rows of the dataset are as follow:

Symbol	market_value	lastest_revenue	stock_volatility
A	33349068800	6.848000e+09	0.016883
AAL	8103907840	4.897100e+10	0.023060
AAP	3153115392	1.115472e+10	0.031097

Table 4 Yahoo Finance dataset

Analysis Performed:
1. Understanding the ESG scores: In ESG dataset, the most important columns are: Total ESG Risk Score, Environment Risk Score, Governance Risk Score, Social Risk Score. Based on definition, $Total\ ESG\ Risk\ Score = Environment\ Risk\ Score + Governance\ Risk\ Score + Social\ Risk\ Score$. Companies with lower ESG risk score might be perceived as safer options from an ESG perspective, whereas those with higher score may warrant a more meticulous review by investors.

Environmental Risk Score	Quantifies a company's ecological responsibility and impact, reflecting its commitment to safeguarding the planet.
Governance Risk Score	Evaluates the transparency, ethics, and quality of corporate governance, influencing long-term viability.
Social Risk Score	Measures a company's societal contributions, spanning employee welfare, diversity, and community engagement.

Table 5 ESG Score Explanation, extracted from kaggle

After employing .describe() method for analysis and leveraging the seaborn package for visualization, it is evident that the governance risk score exhibited a higher density and showcased more stable scores among the target companies.

	Total ESG Risk score	Environment Risk Score	Governance Risk Score	Social Risk Score
count	432.000000	432.000000	432.000000	432.000000
mean	21.393519	5.675463	6.669676	9.023148
std	7.247546	5.345653	2.201841	3.790888
min	7.000000	0.000000	3.000000	1.100000
25%	16.000000	1.500000	5.000000	6.600000
50%	21.000000	3.800000	6.000000	8.700000
75%	26.000000	8.925000	7.700000	11.525000
max	46.000000	25.000000	15.500000	21.000000

Table 6 description of ESG dataset
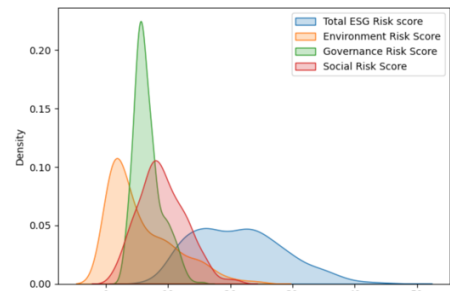
Figure 1 Distribution of ESG Risk Scores

2. Relation between geography (location) and ESG rating: First I group the data by state and compute the average ESG risk score for each state. The findings reveal that Louisiana (LA) has the highest average ESG risk score, whereas Utah (UT) exhibits the lowest score. Moreover, the analysis highlights that the Southern region of the U.S. tends to possess more states with high ESG risk score in comparison to other regions.

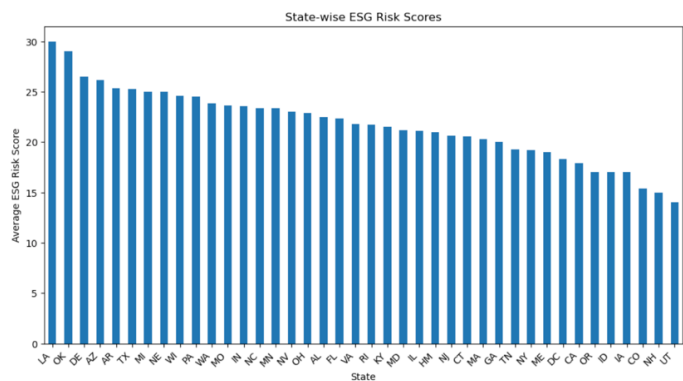|	state_abbreviation	average ESG scores
0	LA	30.000000
1	OK	29.000000
2	DE	26.500000
3	AZ	26.166667
4	AR	25.333333
5	TX	25.282051
6	MI	25.000000
7	NE	25.000000
8	WI	24.600000
9	PA	24.533333

Table 7 State-wise ESG risks scores

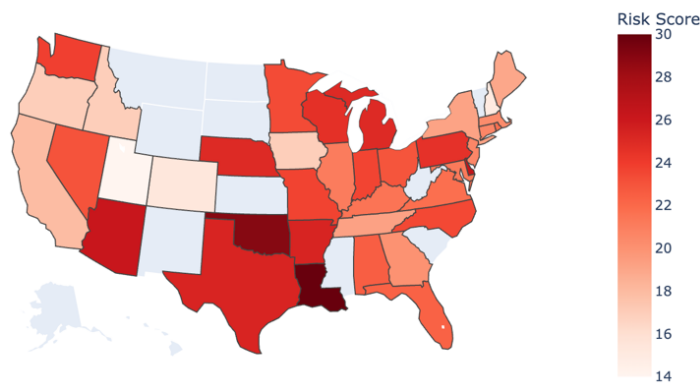*Figure 2 State-wise ESG risks scores (barplot)*

Figure 3 State-wise ESG risks scores (map)

Next, dive deeper into the economy landscape of each state by integrating the states' GDP data, aiming to better understand the potential relationship between the average ESG risk score and the

state's economy condition. Additionally, I calculate the GDP percentage changed from 2022 to 2023 and create a new column named "change_percentage" using the formula $\frac{gdp_{2023}-gdp_{2022}}{gdp_{2022}}$. I then calculate the correlation coefficient between gdp_{2023} & $average\ ESG\ risk\ scores$, as well as between $change\_percentage$ & $average\ ESG\ risk\ scores$. Result shows no correlation between GDP and ESG risk score. However, a negative correlation is observed between the percentage changed in GDP and ESG risk score. This suggests that in states experiencing economic growth, companies within are more likely to have lower ESG risk.
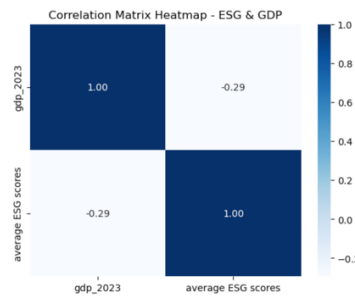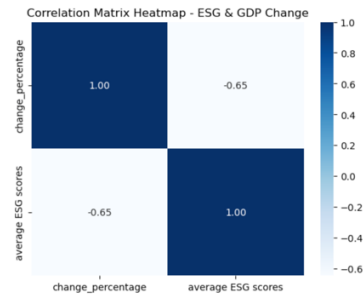

Figure 4 Heatmap – ESG&GDP

Figure 5 Heatmap – ESG&GDP Growth

3. Relation between Sector/Industry and ESG rating: In this ESG dataset, there are a total of 11 sectors. I conduct a grouping based on companies' sectors and calculate their average ESG risk scores, using a bar plot for visualization. The top 3 sectors with the highest ESG risk score are: Energy (33.15), Utilities (27.75) and Basic Materials (27.42). Conversely, the top 3 sectors with the lowest ESG risk score are: Real Estate (13.1), Technology (16.85) and Consumer Cyclical (18.79). This result suggests that Real Estate sector exhibits lower ESG risk, potentially positioning it as a safer choice for ESG-conscious investment, while the Energy sector presents higher instability.
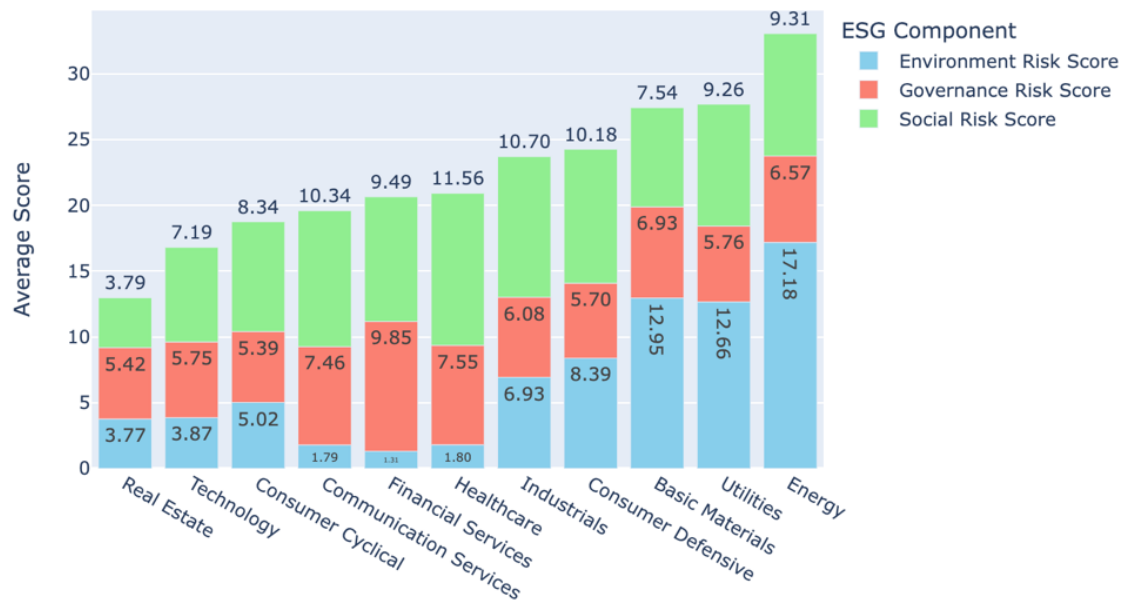

Figure 6 Sector-wise ESG risks scores

Figure 7 Sector-wise ESG risks scores (specific)

To further explore specific industries within sectors. I focus on the top three sectors with the lowest total ESG risk scores. I first use .isin() method in Python to select the specific sector groups and subsequently group the data by industry. Then, I calculate the mean value for each score group. The findings reveal that 4 out of 5 Real Estate Investment Trust (REIT) related industries are considered as the safest ESG investment options.
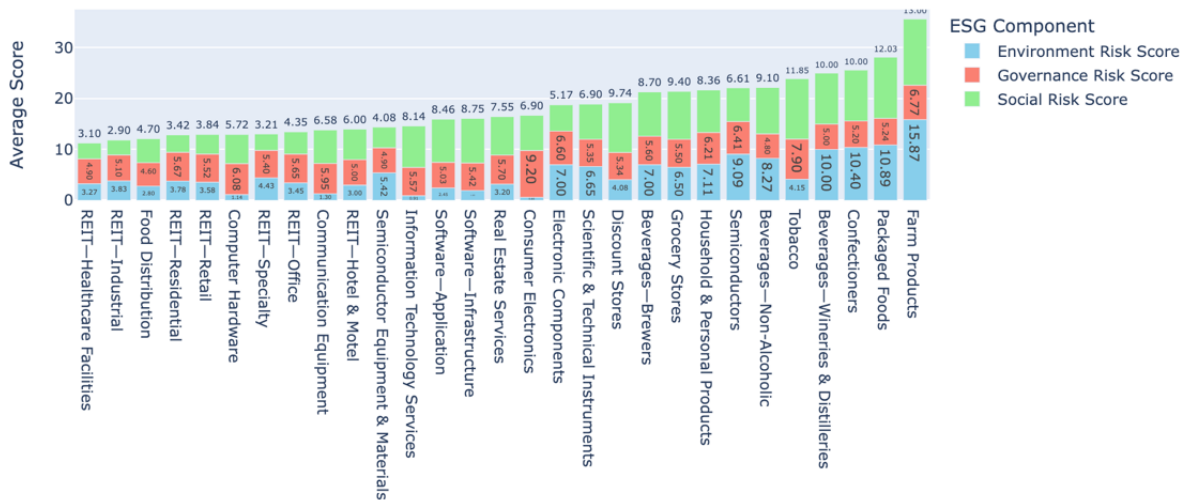

Figure 8 Industry-wise ESG risks scores

4. Relation between Company and ESG rating: The earlier analysis primarily focuses on the macro perspective to examine the factors influencing companies' ESG ratings. Shifting to a micro perspective, I aim to analyze whether the company's size, financial performance, or financial risk management affects its ESG rating. Specifically, I select market value, latest revenue, and stock volatility as financial indices from the yfinance package for coefficient analysis. Surprisingly, the findings reveal almost no correlation between these factors.

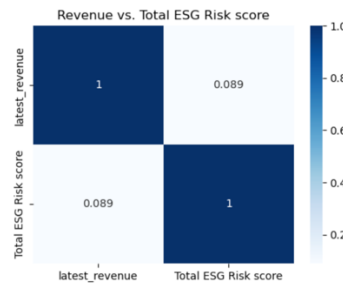Figure 9 Heatmap – ESG & Market Value

Figure 10 Heatmap – ESG & Revenue

Figure 11 Heatmap – ESG & Stock Volatility

At a more detailed level, I organize the total ESG risk scores in the dataset and extract the top 10 companies with the lowest ESG risk scores. I then create a bar plot to visualize this information. Investors might find it beneficial to focus on companies such as Cbre Group and HASBRO Inc., which stand out with scores as low as 7, indicating a strong commitment to environmental, social, and governance (ESG) factors. Similarly, companies like Keysight Technologies and CDW Corp in the Technology sector also demonstrate low ESG risks, with scores of 9, making them noteworthy for investor attention.
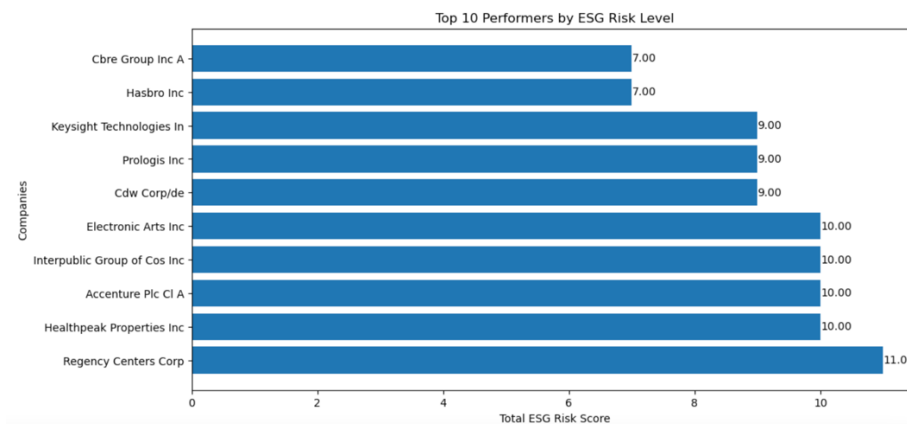

Figure 12 Top 10 Companies with lowest ESG risk score

Future work: Given more time, I aim to employ A/B testing and construct related machine learning models to explore the factors influencing the ESG risk score of companies. Additionally, I plan to conduct further research to investigate the financial factors that are most likely to affect the ESG risk level. Moreover, within the ESG dataset, there are intriguing columns labeled "controversy level" and "controversy score." According to definitions, these columns aid in identifying companies involved in incidents or events that might negatively impact stakeholders, the environment, or the company's operations. I'm also interested in exploring the relationship between controversies and various factors such as geography and economics.