**Research Questions**

These are the three questions that guide EDA process:

1. How do Economic and ESG (Environmental, Social, Governance) scores correlate with Stock Prices?
2. What are the trends in stock prices and trading volumes over time across various sectors?
3. Which sector are most sensitive to specific economic indicators such as employment rates, the Human Development Index (HDI), and the Producer Price Index (PPI)?

**Dataset Description**

The dataset contains 946889 records/observations and fifteen variables. These variables can be categorized into 1 Date variable, 8 float variables, 1 integer variable & 5 string variables. Following is a data dictionary containing variable descriptions and types as below:

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Data Type** | **Description** |
| Date | Date | Date when the stock data was recorded. |
| Open | Float | Stock Price at the beginning of the trading day |
| High | Float | Stock's highest price during the trading day |
| Low | Float | Stock lowest price during the trading day |
| Close | Float | Stock price at the end of the trading day |
| Adj Close | Float | After adjustments of all applicable splits & dividend distributions the stock’s closing price |
| Volume | Integer | During a given period (monthly considered), the number of shares or contracts traded in the entire market |
| Environment Score | Float | Score which reflects the company’s environmental impact and sustainability practices. |
| Social Score | Float | Score which reflects the company’s social responsibility and actions affecting stakeholders. |
| Governance Score | Float | Score reflects how well the company is managed. |
| Company Name | String | Full name of company |
| Industry Group | String | Category describing the company’s primary business activity |
| Primary Sector | String | Sector to which company business belongs. |
| Country | String | Location of Company headquarters |
| Ticker | String | Unique stock ticker symbol of the company for trading purposes. |
| Average Hourly Earnings | Float | Laborers’ average earnings per hours |
| Producer Price Index (PPI) | Float | Wholesale prices measure. Producers require for their goods & services |
| Employment Cost Index (ECI) | Float | It is the cost of employee wages and benefits measured for the employer |
| Human Development Index (HDI) | Float | Health, Education, and income terms measure a country's well-being. |
| Labor Productivity | Index | It is a measure where the amount of work done |
| Payroll Employment | Integer | Per employer, it is the total number of paid employees |
| Rolling Means | Float | Measure of average stocker price over the specific time period. |
| Stock Price Data | Float | The current trading price of the company’s share in the market |
| Unemployment Rates | Index | Measure of percentage of skilled individuals with no job. |
| Volumes of Trades | Integer | Number of company shares bought and sold in the given period. |

The above dataset is created and scrupulously assembled from many authoritative sources, each providing valuable insights into stock market dynamics, HDI (human development index), labor statistics, and economic indicators. The complete dataset consists of approximately 950,000 records, containing fifteen distinct variables, and as mentioned earlier values ranging from stock prices and trading volumes to environmental, social, and governance scores, along with crucial labor & economic indices.

Starting with stock market data we have scraped historical stock price data for US companies using ***the*** Python **‘*yfinance*’**library. The data includes opening, high, low, and closing prices of stocks, along with adjusted closing prices and trading volumes. We also used Bureau of Labor ([BLS GOV](https://www.bls.gov)) statistics, our primary source for labor-related data, which includes average hourly earnings, ECI (employment cost index), payroll employment, labor productivity, and unemployment rates. The BLS data is available in Excel format, and we have integrated it into our dataset. The ESG scores (Environment, Social & Governance) scores are scrapped from the Yahoo Finance website ([Yahoo Finance]((https:/finance.yahoo.com))) using web-scraping for all the companies initially pulled from the ‘**yfinance**’ library.

Further, the interest rate over the years, the economic indicator mentioned earlier Producer Price Index (PPI) is retrieved from the Federal Reserve Economic Data [(FRED)](https://fred.stlouisfed.org/) database where our focus is to offer insights into wholesale price movements and economic trends. Lastly, the data of the Human Development Index (HDI) is obtained from [(ourworldindata org)]((https:/ourworldindata.org/human-development-index)) Our World in Data website.

While compiling the data, to confirm the dataset’s integrity and relevance we have done various meticulous verification of data sources and formats. This dataset has various types of data such as dates, floats, integers & strings, which offers a holistic view of the financial and economic landscape. Our research is supported as this dataset includes stock market fluctuations, labor market dynamics, and broad economic and social indicators to analyze trends and how they impact different sectors and times.