

# GICS Sector Performance

2023-01-07

## INTRODUCTION

Data exploration and transformation will be conducted to determine the relationship between the stock market and revenue for S&P 500 companies and their GICS Sectors.

## DATA SOURCES

### Flat File (CSV): Fortune 1000 Company Revenue

Fortune ranks the top 1000 companies based on employability and revenue generated from U.S. incorporated companies. It contains information regarding revenue, profit, assets, and market value. The dataset was scraped from the 2021 Fortune 500 companies' magazine and posted on Kaggle.

<https://www.kaggle.com/datasets/ramjasmaurya/fortune-1000-companieslatest?select=fortune+1000+companies+in+dec2021.csv>

### Website: S&P 500 Companies

The S&P 500 stock market index are issued based on the company's float-adjusted market capitalization and is broken down by Global Industry Classification Standard (GICS) sector and industry. Additional information is provided for each company including the headquarters location and founded year. This information is posted on Wikipedia.

[https://en.wikipedia.org/wiki/List\\_of\\_S%26P\\_500\\_companies#:~:text=S%26P%20500%20component%20stocks%20%20%20Symbol,%20%20%202014%20more%20rows%20](https://en.wikipedia.org/wiki/List_of_S%26P_500_companies#:~:text=S%26P%20500%20component%20stocks%20%20%20Symbol,%20%20%202014%20more%20rows%20)

### API: Historical Data

Marketstack provides a REST API to obtain historical market stock data for the past 30+ years. By passing a stock ticker to the API, it retrieves information such as open, high, low, close, and volume of the company stock for any date or date range. The overall stock market data is licensed and sourced from multiple high-authority market data providers, and the US exchange data is sourced from Tiingo, Inc.

<https://marketstack.com/documentation>

## RELATIONSHIPS

**CSV File:** Contains a list of Fortune 1000 company names, record number, rank, revenue, % change in revenue compared to previous year, profits in millions, % change in profits compared to previous year, assets, market value, and change in rank compared to last year.

**Website:** Contains a list of SP 500 company names and stock ticker, SEC filing reports, GICS Sector, GICS Sub-Industry, Headquarters Location, Date first added, CIK, and founded year.

**API:** Contains market stock data by stock ticker including date, exchange, open, high, low, close, volume, adjusted open, adjusted high, adjusted low, adjusted close, and adjusted volume.

The relationships between the data sources are connected based on the company name and associated stock ticker. The CSV file containing Fortune 1000 companies' revenue has a one-to-one relationship with the Wikipedia webpage that contains the S&P 500 companies by company name. An inner join will need to be made to filter only the S&P 500 companies that have revenue listed in the Fortune 1000 list. The result of this relationship will have a one-to-many relationship with the API containing historical stock data by stock ticker.

## PROJECT PLAN

To determine the relationship between the stock market and revenue of S&P 500 companies and their associated GICS Sectors, three tables will be created with associated visualizations to find key information. Each dataset will be cleaned and formatted in each milestone using techniques such as removing characters from monetary amounts, checking for duplicate or missing data from each record, and replacing header names. The CSV file and website data will be cleaned and formatted first allowing matching to be done early to determine what next steps will need to be done before visualizations. The resulting stock tickers will be fed to the historical stock data API.

The visualizations that will be the outcome of the project will group each GICS sectors and their associated sub-industry groups to see if any group outperforms another significantly in revenue or stock price. This will provide insight to which sectors have the highest revenue and stock price.

## ETHICAL IMPLICATIONS

The separate data sources were collected to demonstrate the financial standing of companies that met the requirements to be on the S&P 500 or Fortune 1000 list. Outside factors that have an impact on revenue and stock data are not considered in this research. Each data source had their data outsourced by another entity and provided it for analysis.

## CHALLENGES

Based on the variables described, the main challenge will be coming up with a merged list between the S&P 500 Companies and Fortune 1000 Company Revenue. All companies listed on S&P 500 may not also be on the Fortune 1000 list. Alternative measures may be put in place to notate which did not have revenue associate or a relationship may need to be created. Another challenge will be that the csv file that contains the Fortune 1000 companies has a lot of formatting that will need to be removed to be in a useful format. The CSV file variables will also need to be converted from objects to text, integers, or floats as required to perform analysis.