Impacts of Environment on US Stock Price

An exploratory data analysis

Team members: Nat Pham, Deep Desai, Rishabh Prasad, Benson

TAMU Datathon

Goldman Sachs

Problem

- Stock markets increase both economic activities and energy consumption across countries
  - Hence, stock markets are expected to have a significant effect on environment indicators such as CO2 emissions, greenhouse gas emissions etc.
- Many studies find that weather has a close relationship with human's mood and behavior
  - Hence, weather factors might also have an impact on stock return and trading volume
- We are interested in exploring how the environment and stock markets are correlated and tied together
- Our results can enable us to build models to predict risk events and discover investment opportunities based on environmental variations

Dataset Usage/ Integration

- Both datasets given provided a unique challenge to overcome. While the US Stocks Daily dataset had more than 37 million entries, the Environment dataset was split over multiple views.
- We created two subteams
  - One parsed through different views in the Environment database for features that seemed most impactful overall.
  - Another tried to figure out a proper way to subset and aggregate the entries while maintaining coherency of the data.
- Once we had completed that, we used a combination of SQL Queries and the Snowflake connector for Python to load the result set into local .csv files.

Data Normalization

Blending multiple datasets from different sources:

- A major hurdle was the specific daily entries in US Stocks database, but the yearly entries of the Environment database
- To counter this disparity, we created two different datasets
  - An aggregate dataset of average values over all the years from 1970 to present (to match gas plots)
  - A subset of Stock History from 2018 to present (to match temperature plots)

Data-cleaning

- To ensure that data cleanliness is maintained, we made sure to import only relevant data from Snowflake and later used Pandas and NumPy to filter columns for specific values.
- We tried to maintain a fine line between too much data and too little data.

Data-transformation

- During aggregation some features got abnormally large, and thus we used a log transformation in the visualizations to maintain coherency of the data.
- This allowed us to see the general trend relatively undisturbed by outliers altering the scale but still providing a clear full view of the data itself.

Final Product

- Demo of an interactive program that lets users choose:
  - Environmental indicators (CO2 Emission, GHG Emission, Temperature (H), Precipitation Level (inches) and
  - Sectors ('Consumer Cyclical', 'Financial', 'Technology', 'Pharmaceuticals', 'Financial Services', 'Industrials', 'Basic Materials', 'Conglomerates', 'Services', 'Industrial Goods', 'Consumer Defensive', 'Communication Services', 'Real Estate', 'Utilities', 'Media', 'Energy', 'Healthcare', 'Building')

- And output visualizations of changes in the environment indicator and stock price by sector over time are created using boxplots, scatterplots, and line graphs in the Matplotlib and Seaborn packages.

- Visualizations and insights are also presented in the following slides

Insights

In a cursory glance, we can see that there is a rather large spike in volume of stocks traded for the Financial Services, Energy, and Real Estate industry. This might be based on the stock exchange as particular US Stock Exchanges in other sectors also have a similar spike.

Corresponding to the rise in trading, there is a slight dip in average prices in various industries.

Future Work

- Add further compatibility with dynamic plotting in the user interface
- Generate specific correlation matrix metrics to provide insights to users
- Account for larger set of environmental factors, like precipitation, humidity, and PM2.5

Limitations

- Lack of data documentation in the data sources made it difficult to understand the variables and interpret the analysis results
  - We missed a few instances of data cleaning that were quite obvious to us late during the project. For example, there were some instances where the “High” in a single day was in the millions. Cross checking with the history online showed this to be obviously false.
  - Similarly, when comparing to some of the indicators, some features like “Price”are averaged over entire years. This is a gross oversimplification of a sensitive market such as the stock market, but this is helpful in seeing a general trend.

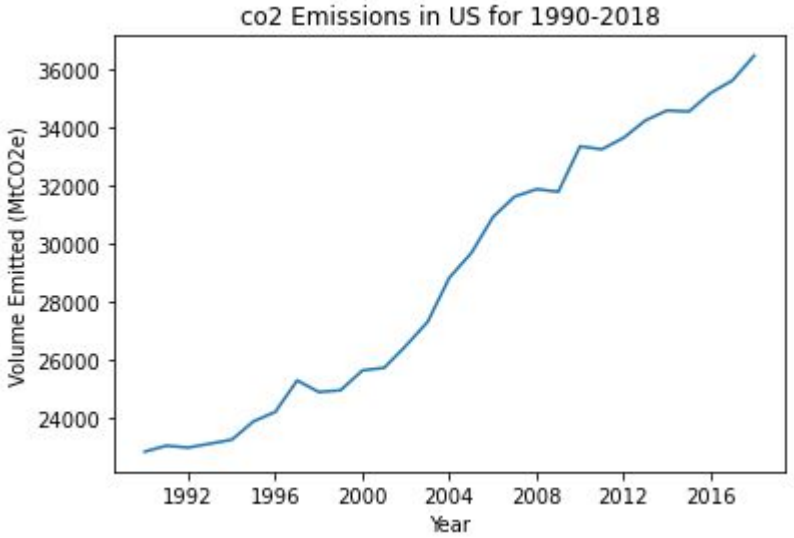
Limitations (cont.)

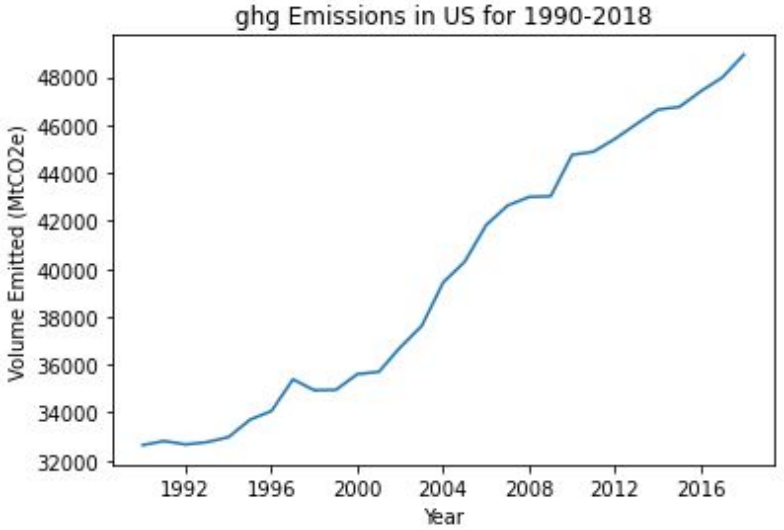
- Environment indicators data were only available on a yearly estimates, while stock prices are volatile by day/hours. This makes it more challenging to correctly identify the correlation of these two factors.

CO2 & GHG Emissions and US Stock Prices

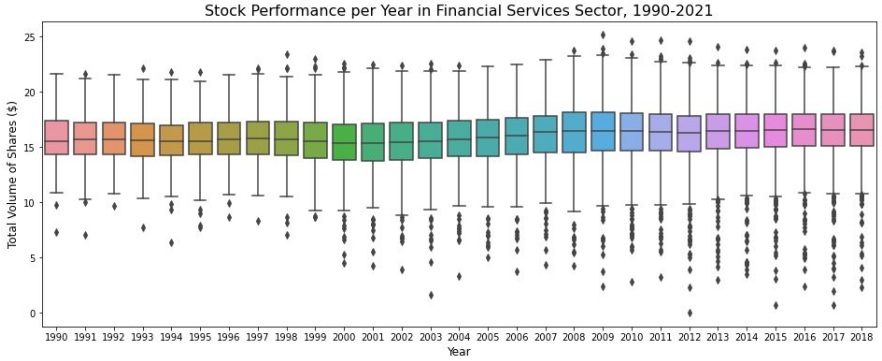
We look at average volume emitted in the US, and distribution of US stock prices by sector by year

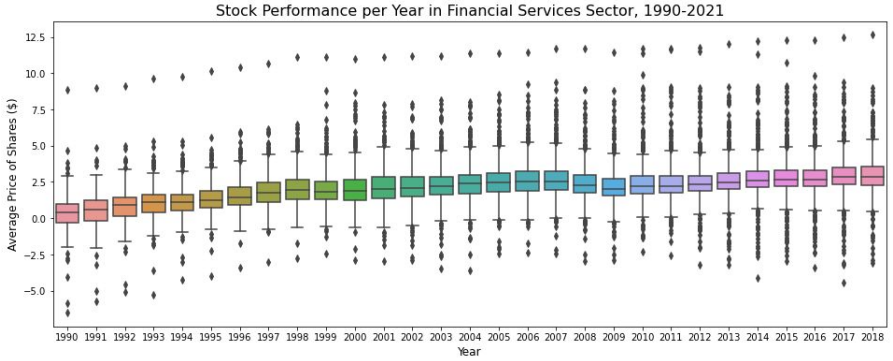
CO2 and Greenhouse Gas Emissions Increased Over Time

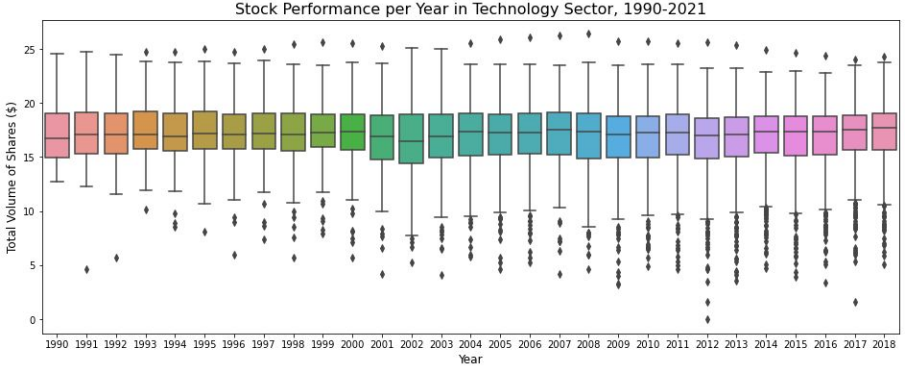


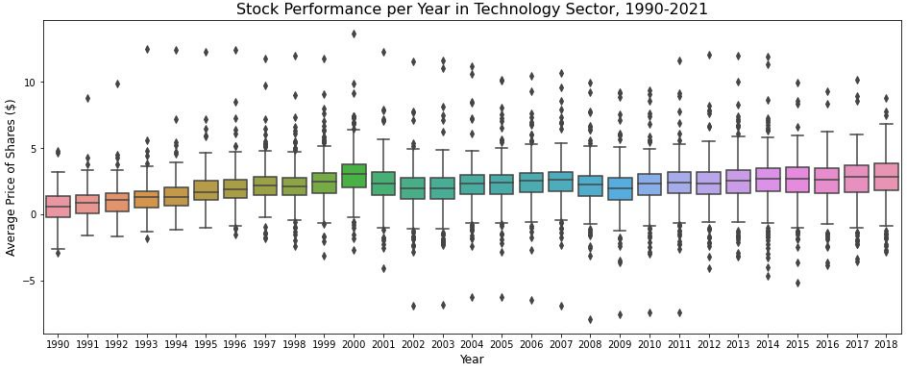


Stock Share Performance by Sector and Value

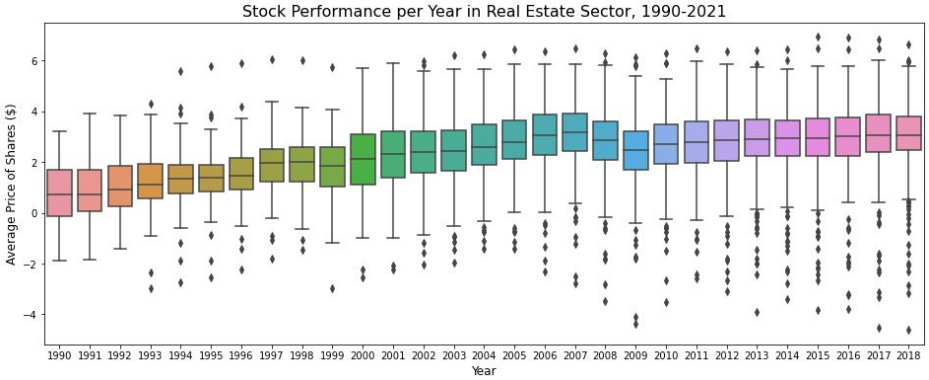


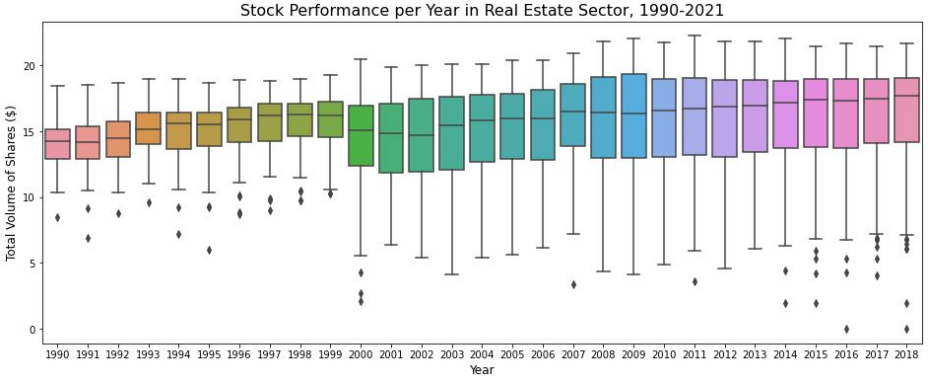


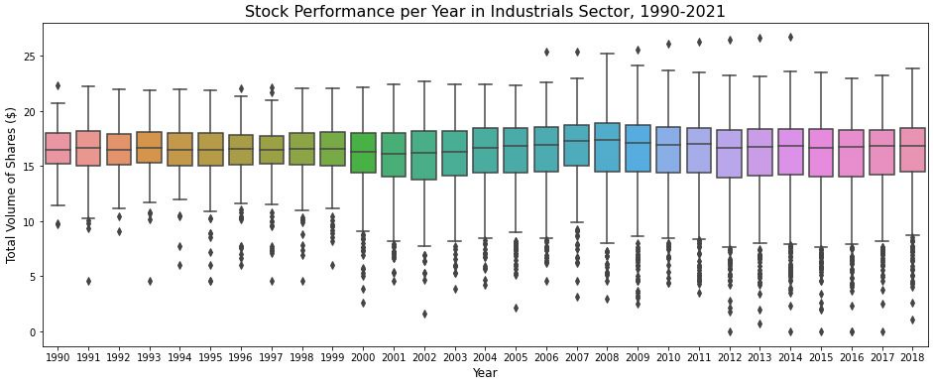


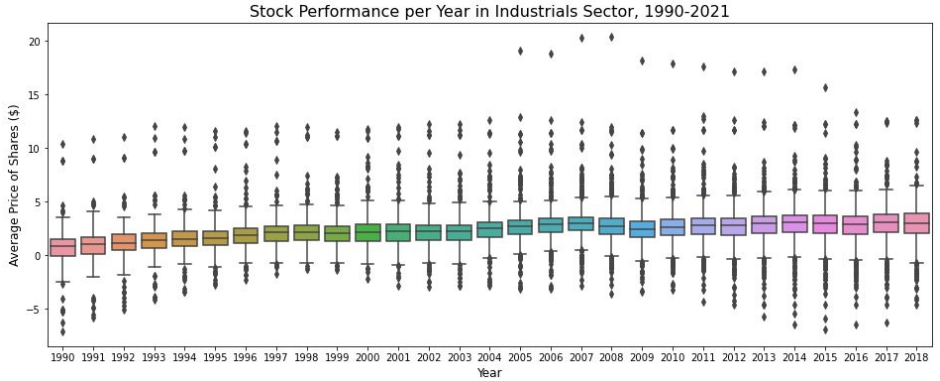


Real Estate and Industrials

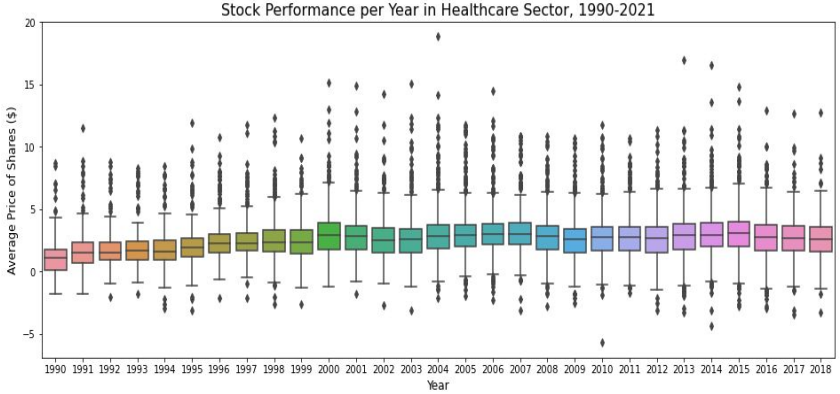


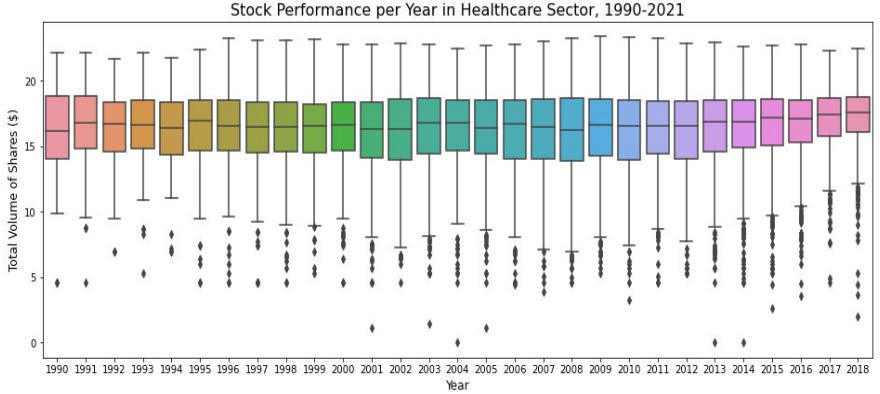


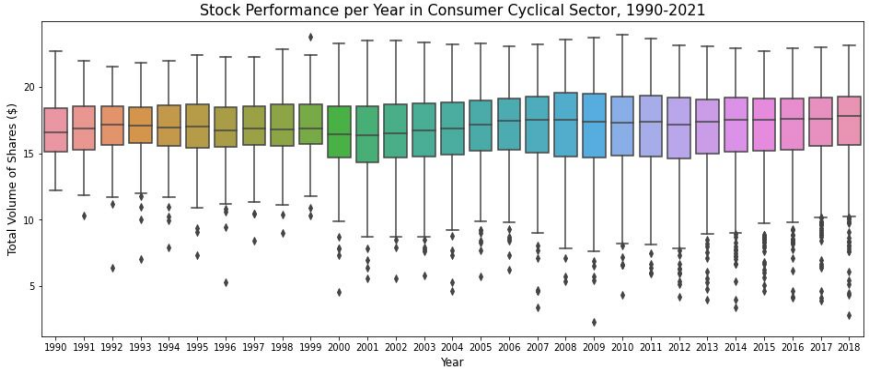


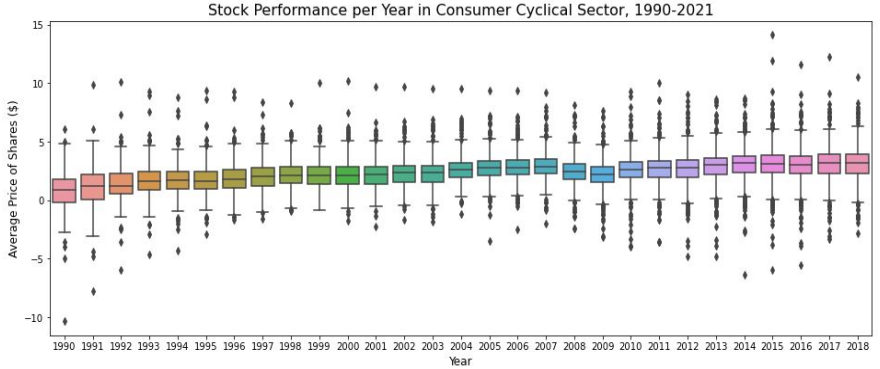


Healthcare and Consumer Cyclical

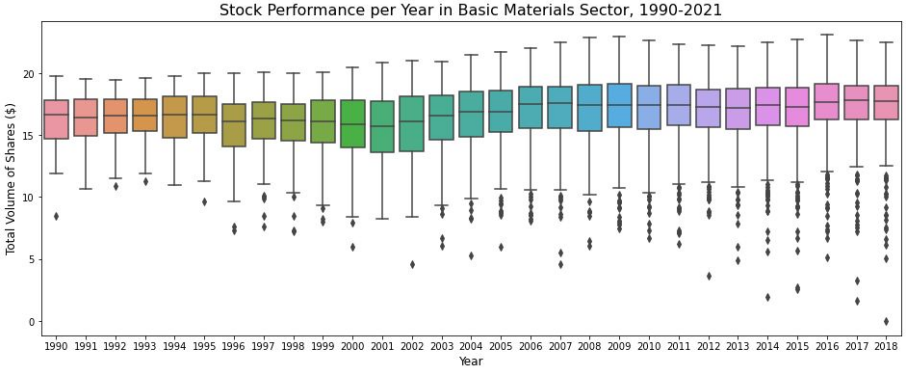


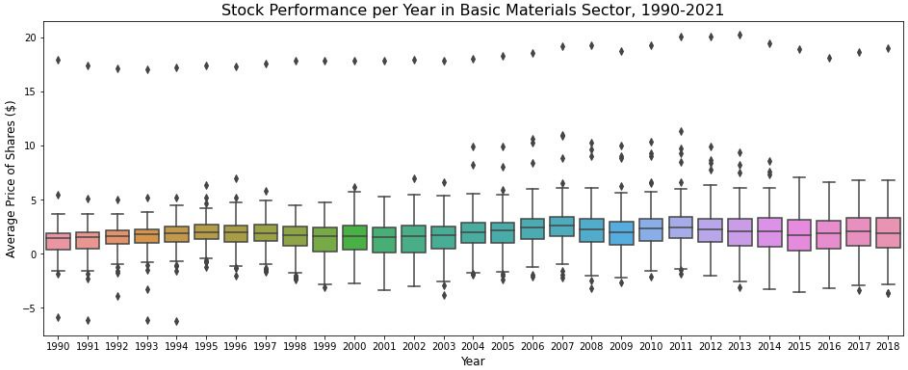


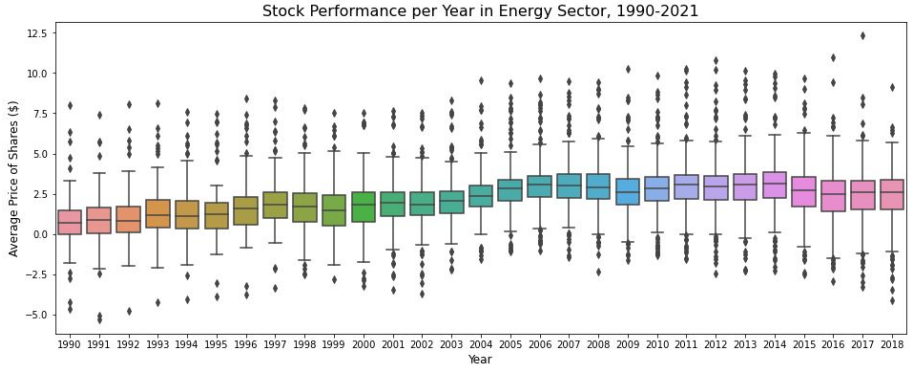


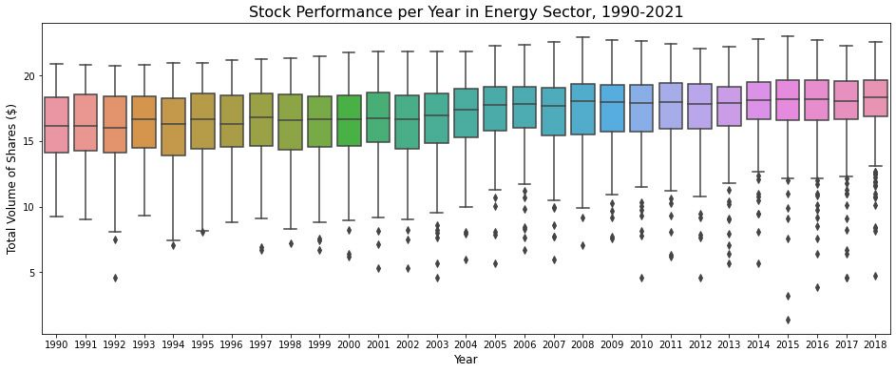


Basic Materials and Energy







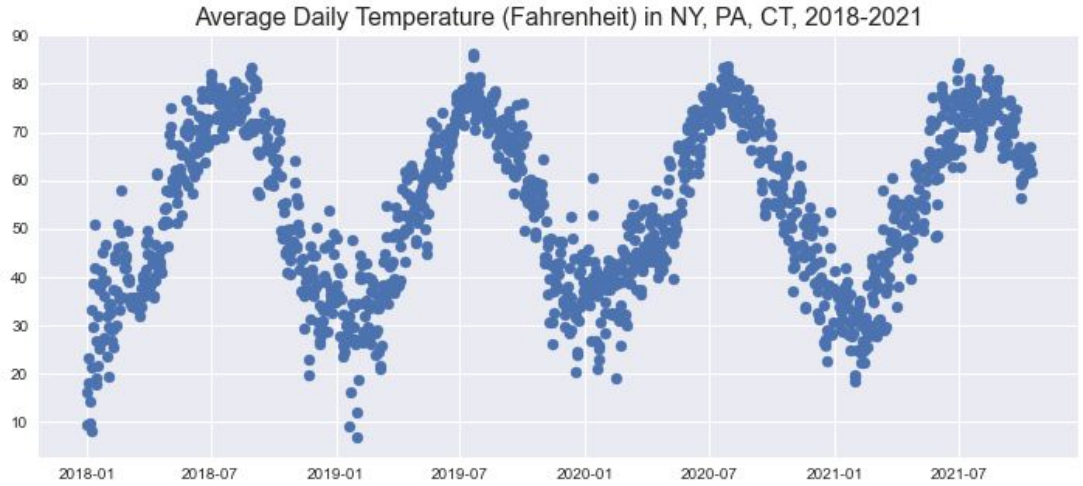


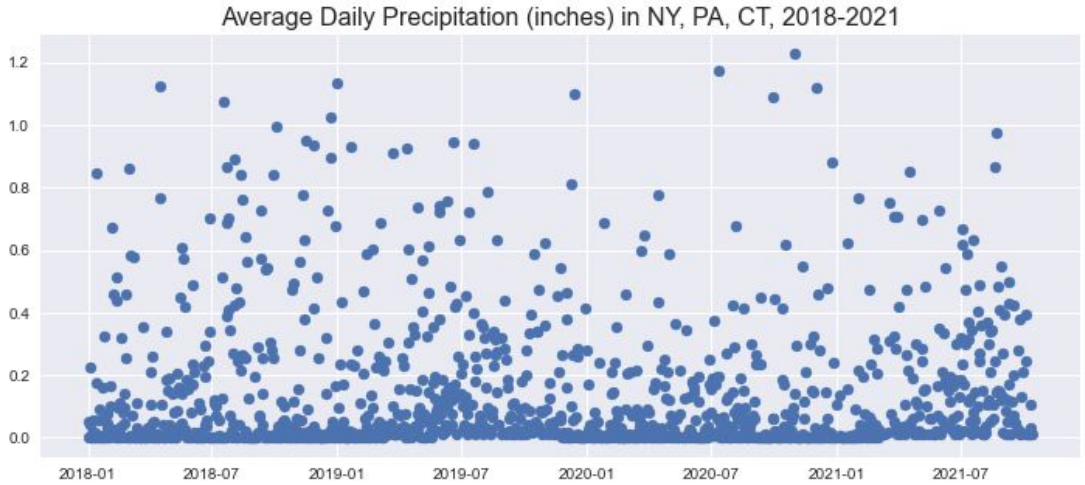
Daily Temperature, Precipitation Levels and Stock Prices

We look at daily temperature and precipitation levels in 3 Northeastern states (NY, CT, PA), and average daily low/high price, and total stock volume traded by sector

Daily Temperature and Precipitation Levels

Daily weather data is available from 2018-01-01 to 2021-07-20

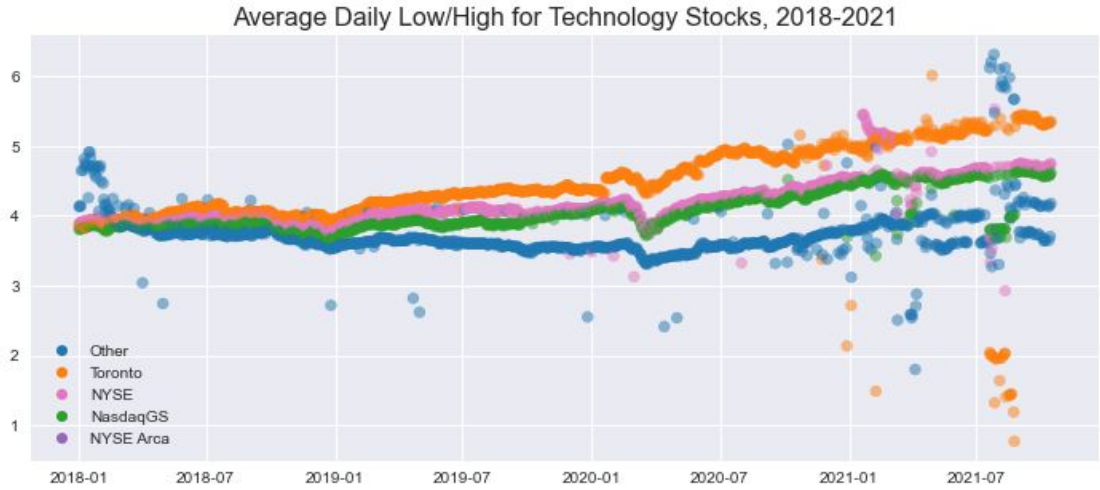


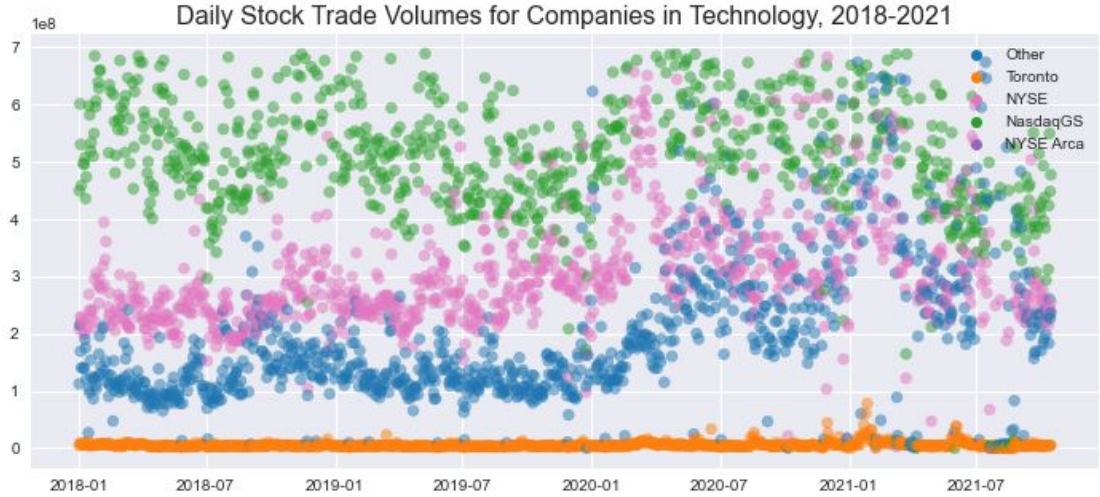


Daily Stock Share Performance by Sector

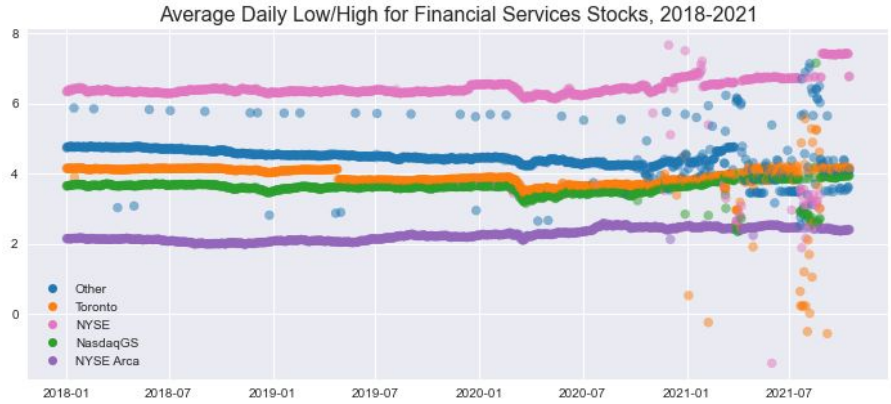
We analyzed data for 24 sectors, and presented results for the 8 biggest ones here.

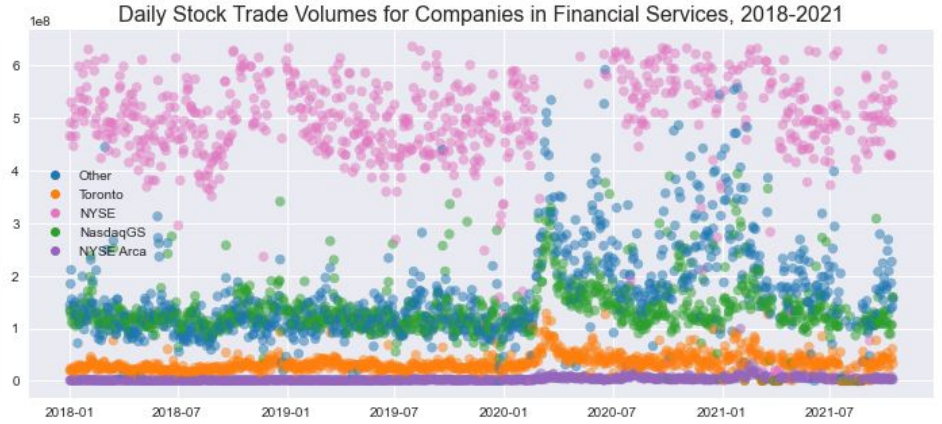
Results are stratified by 4 Stock Exchange with highest presence in the dataset, and Other exchanges

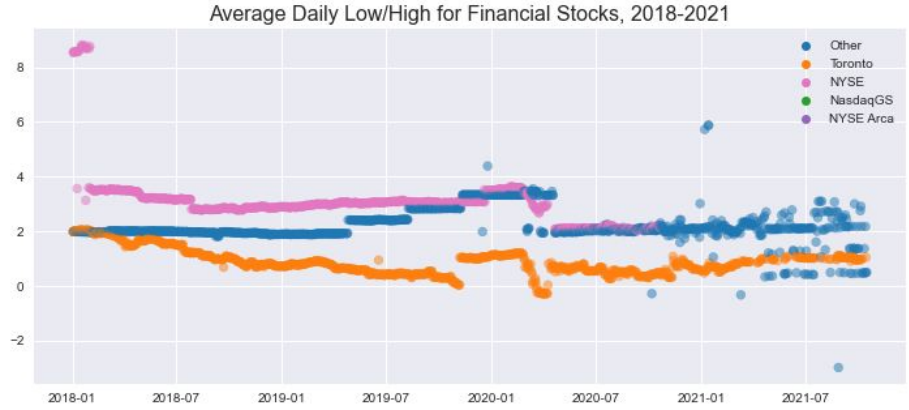


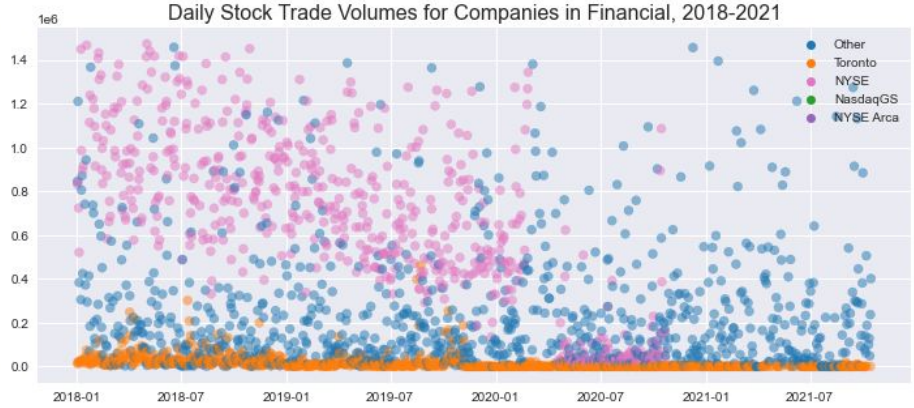


Financial Services and Financial

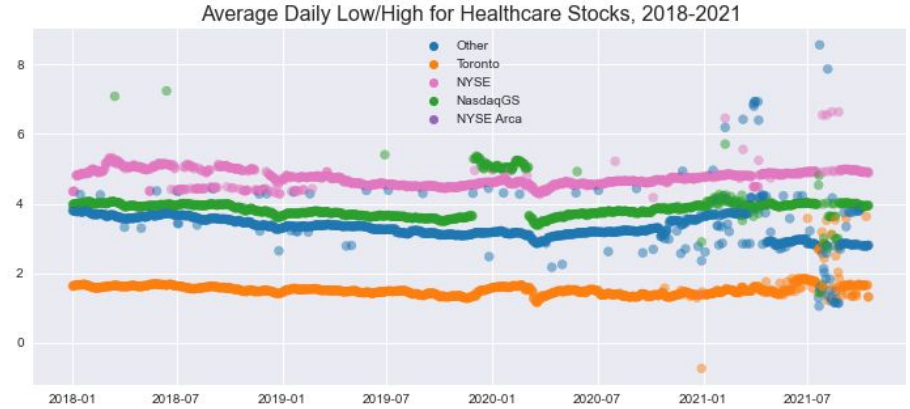


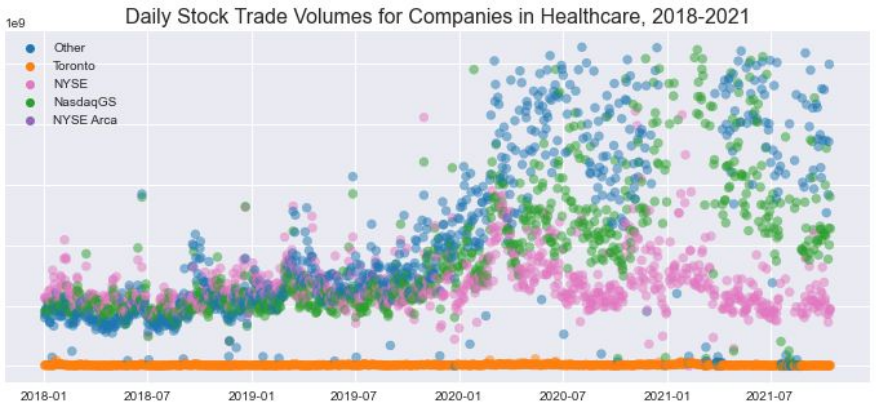


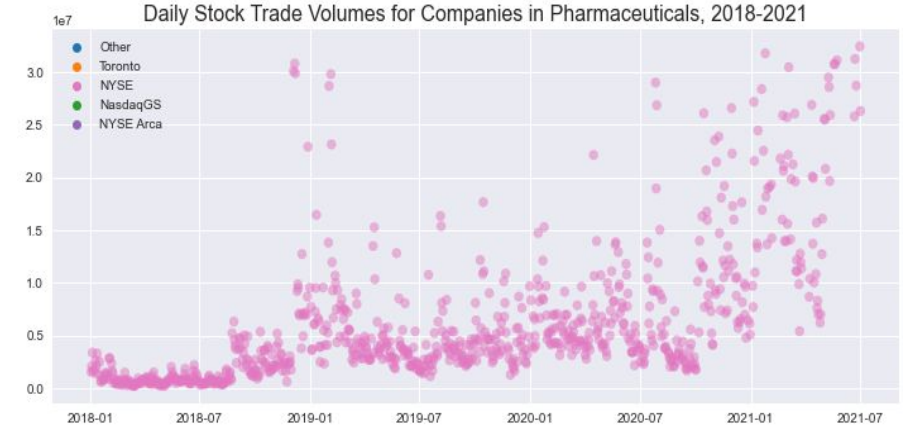


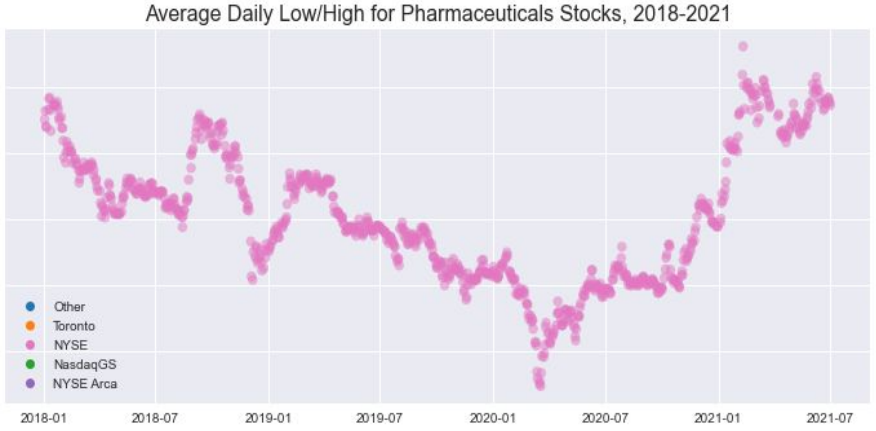


Healthcare & Pharmaceuticals

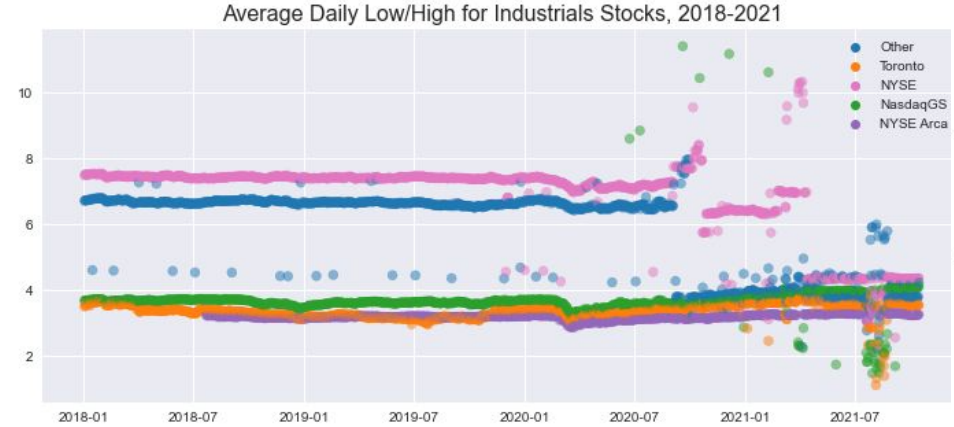


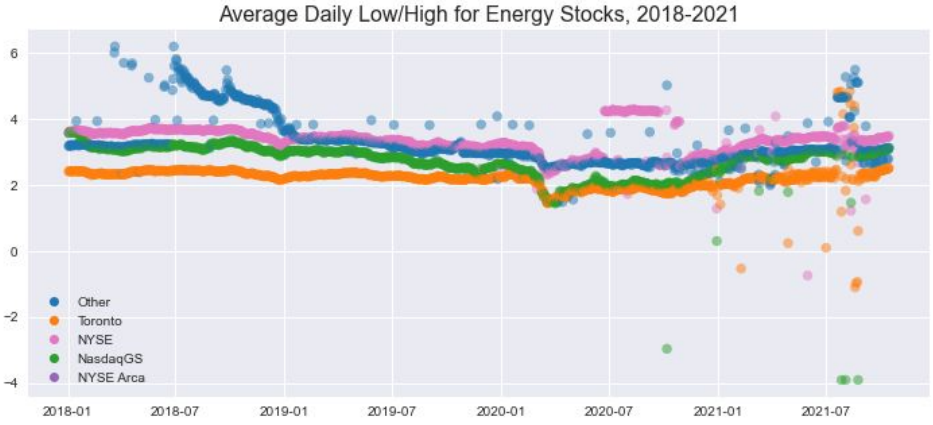


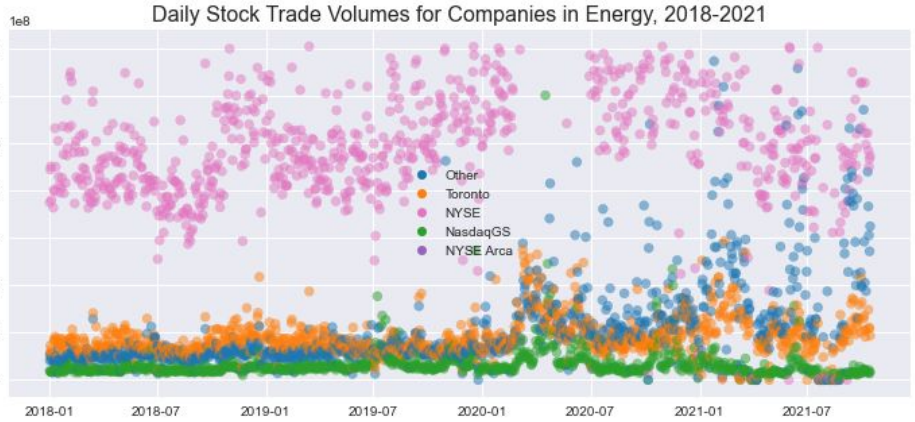




Energy & Industrials







Real Estate & Consumer Cyclical