Isabella Valentino
FA595 Final Project: Sunshine Anomaly

Overview and Summary of Project

My project aims to assess the reliability of the stock market Sunshine Anomaly, a hypothesis proposing a correlation between weather conditions and stock market performance. According to this anomaly, when the weather at a stock market exchange location is sunny, the market tends to be up, and conversely, when the weather is not sunny, the market tends to be down. To enhance the scope of my investigation, I am utilizing the New York Stock Exchange (NYSE) data obtained from Yahoo Finance, along with weather data from four major cities: New York (NY), Los Angeles (LA), Chicago, and Houston, sourced from Visual Crossing.

The program is designed to read information from five distinct files: the stock market data file, "NYSE.csv," and weather data files for each city: "weather_NY.csv," "weather_LA.csv," "weather_Chicago.csv," and "weather_Houston.csv." Subsequently, it creates new files for each city, extracting only the columns essential for testing purposes. To visualize the relationship between weather conditions and stock market trends, I utilize these newly generated files to plot graphs using the matPlotLib library.

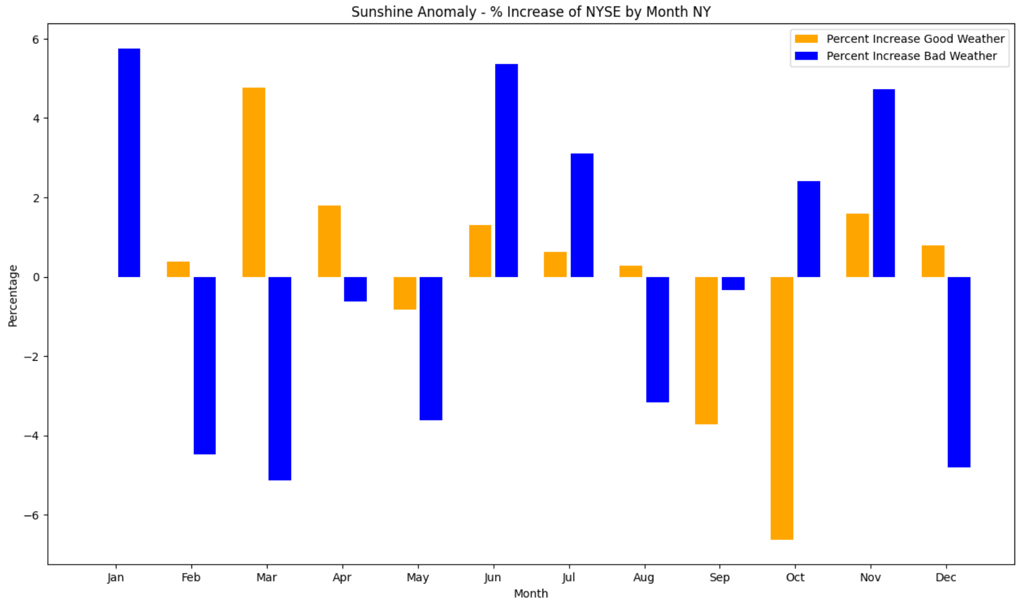
Snapshot of first three lines of NYSE.csv downloaded from Yahoo Finance

Date	Open	High	Low	Close	Adj Close	Volume
10/31/2022	14795.62988	14825.01953	14698.82031	14747.03027	14747.03027	4820620000
11/1/2022	14747.03027	14921.48047	14725.08984	14790.70996	14790.70996	4481210000
11/2/2022	14790.70996	14929.37988	14496.33008	14497.84961	14497.84961	4899000000

Snapshot of first three lines of newFile_NY.csv created by the program

Date	Up or Down	Weather
10/31/2022	-48.6	Rain, Partially cloudy
11/1/2022	43.68	Rain, Partially cloudy
11/2/2022	-292.86	Clear

Snapshot of the plot showing percent increase by weather for each month in NY



In March, there is a notable positive correlation between Good Weather (Clear) and a positive market performance. The total points for the market during clear weather conditions are 704.84, representing a 4.76% increase from the initial opening value. This suggests that during March, when the weather is clear, the market tends to perform well. Conversely, during Bad Weather (non-clear conditions), the market shows a negative correlation. The total points during non-clear weather conditions are -758.91, indicating a 5.13% decrease from the initial opening value. This implies that in March, bad weather conditions are associated with a decline in the market.

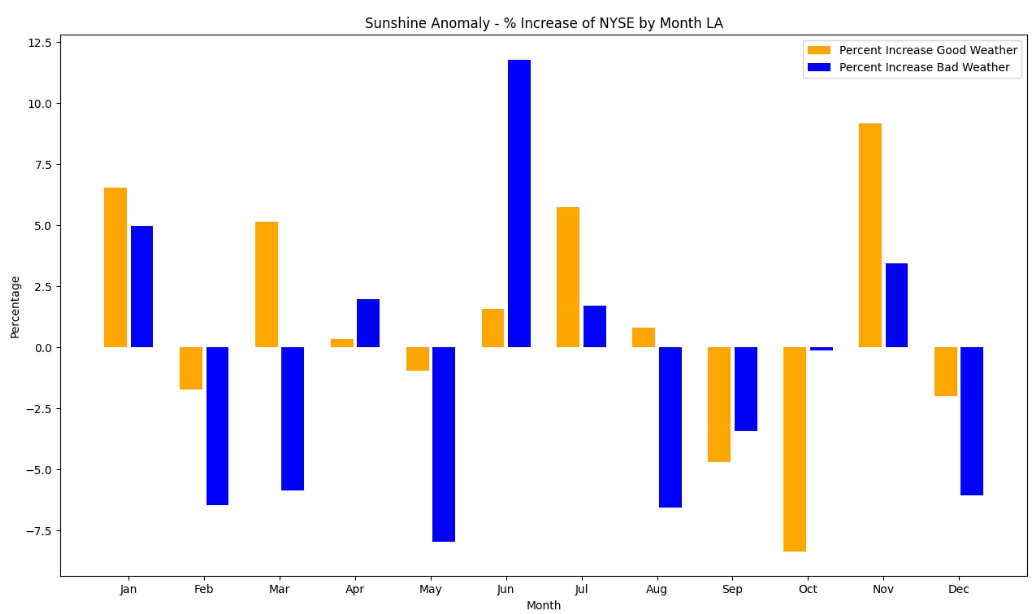
In April, the correlation between weather conditions and market performance is not as strong as in March, but there are still discernible patterns. During days characterized by Good Weather, specifically Clear conditions, the market demonstrates a positive response. The total points for April in these weather conditions amount to 264.82, representing a 1.79% increase from the initial opening value. This implies that the market tends to fare better during clear weather days in April. Conversely, on days marked by Bad Weather, denoted by non-clear conditions, the market experiences a decrease of -93.84 total points, reflecting a -0.63% change. This negative correlation suggests that, in April, adverse weather conditions are associated with a slight downturn in the market.

Overall, in both March and April, there is a consistent pattern where Good Weather tends to be associated with positive market performance, while Bad Weather is linked to a decline in the market. However, it's essential to note that this observed pattern does not apply to all months, as evidenced by the analysis of other months in the dataset. When considering the entire year, the correlation between weather conditions and market performance appears to vary, and the Sunshine Anomaly may not hold consistently throughout.

Snapshot of first three lines of newFile_LA.csv created by the program

Date	Up or Down	Weather
10/31/2022	-48.6	Clear
11/1/2022	43.68	Partially cloudy
11/2/2022	-292.86	Rain, Partially cloudy

Snapshot of the plot showing percent increase by weather for each month in LA

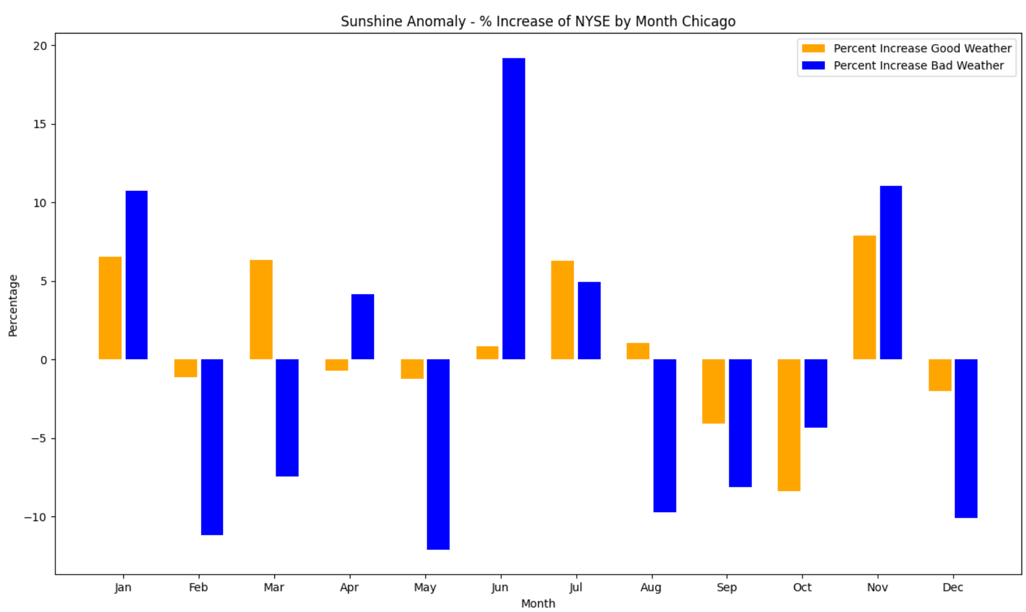


In March, the Sunshine Anomaly hypothesis holds true for Los Angeles (LA). On days characterized by Clear Weather, the market demonstrates a positive response with a total of 758.67 points, resulting in a 5.13% increase from the initial opening value. Conversely, during days with Bad Weather, marked by non-clear conditions, the market experiences a decrease of -866.81 total points, reflecting a -5.86% change. This alignment with the Sunshine Anomaly suggests that, similar to New York, LA shows a consistent pattern of positive market performance during clear weather conditions in March. However, this pattern does not extend across all months. The analysis of other months in the dataset reveals varying trends, indicating that the Sunshine Anomaly might not consistently apply throughout the entire year.

Snapshot of first three lines of newFile_Chicago.csv created by the program

Date	Up or Down	Weather
10/31/2022	-48.6	Rain, Partially cloudy
11/1/2022	43.68	Clear
11/2/2022	-292.86	Clear

Snapshot of the plot showing percent increase by weather for each month in Chicago

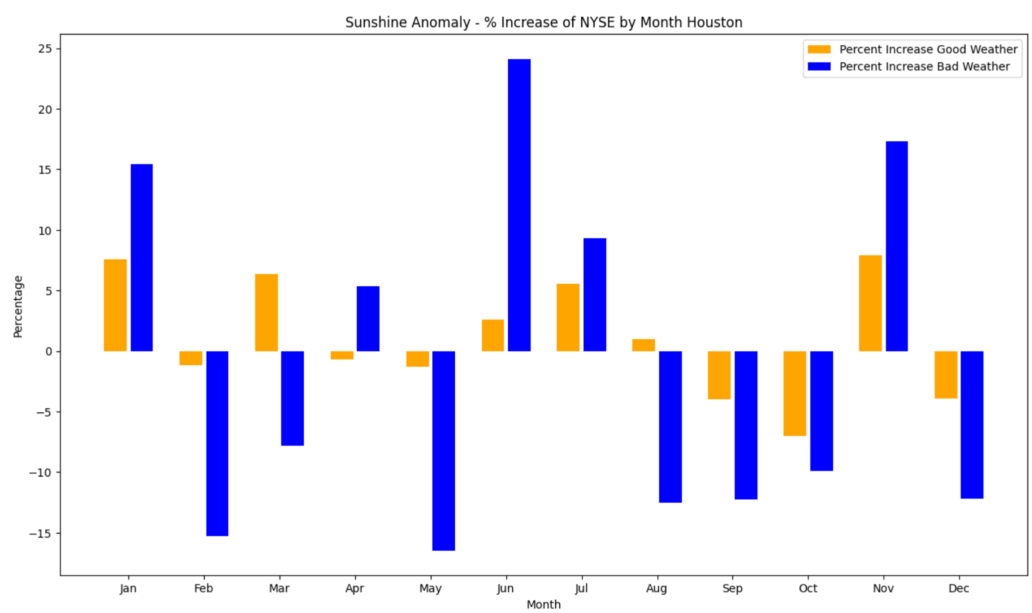


In March, the Sunshine Anomaly analysis for Chicago aligns with the observed trend in New York and LA. Clear weather conditions are associated with a positive market performance, reflected in a total points increase of 939.04, equivalent to a 6.35% rise from the initial opening value. Conversely, days characterized by Bad Weather show a negative correlation, with a decrease of -1101.25 total points, representing a -7.44% change. These findings support the notion that favorable weather conditions tend to coincide with positive market outcomes in March. However, this pattern does not extend across all months. The analysis of other months in the dataset reveals varying trends, indicating that the Sunshine Anomaly might not consistently apply throughout the entire year.

Snapshot of first three lines of newFile_Houston.csv created by the program

Date	Up or Down	Weather
10/31/2022	-48.6	Partially cloudy
11/1/2022	43.68	Partially cloudy
11/2/2022	-292.86	Partially cloudy

Snapshot of the plot showing percent increase by weather for each month in Houston



In March, the Sunshine Anomaly hypothesis appears to hold true for Houston. On days characterized by Clear Weather, the market demonstrates a positive response with a total of 939.04 points, resulting in a 6.35% increase from the initial opening value. Conversely, during days with Bad Weather, marked by non-clear conditions, the market experiences a decrease of -1155.32 total points, reflecting a -7.81% change. This alignment with the Sunshine Anomaly suggests that, similar to other cities, Houston shows a consistent pattern of positive market performance during clear weather conditions in March. However, it's crucial to emphasize that this observed pattern doesn't extend uniformly across all months. The analysis of other months in the dataset reveals varying trends, indicating that the Sunshine Anomaly might not consistently apply throughout the entire year.

Target Audience

My program is tailored for individuals intrigued by the interplay between the stock market and weather conditions. This audience may include both investors seeking insights into market behavior and researchers evaluating the accuracy of the Sunshine Anomaly. In the future, I aspire to enhance the program's functionality, allowing users to compare various market sectors over time. This expansion aims to uncover patterns and discern which sectors perform better in specific months. Such advancements would not only benefit investors in making informed decisions but also broaden the program's appeal to those seeking a comprehensive understanding of market dynamics and trends.

Specific Programming Techniques Used

The program categorizes outcomes on a monthly basis, utilizing four distinct lists designed to accommodate twelve items, each representing a specific month. These lists serve as repositories for consolidating the total points reflecting market fluctuations during periods of good and bad weather. Additionally, the lists play a pivotal role in computing the percentage increase in the market's exchange for both good and bad weather conditions.

Example of lists in program to tally points gained or lost by month and store percent by month

```
totalPointsGoodWeather = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

totalPointsBadWeather = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

percentIncreaseGoodWeather = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

percentIncreaseBadWeather = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
```

To mitigate the reliance on global variables, the code employs a modular approach with small, focused functions handling specific segments. For example, the `createDataFile()` function processes the five downloaded data files, including NYSE.csv and the four city weather files, and generates new files for each city. Another function, `plotGraph()`, is dedicated solely to plotting a bar graph based on the provided data. A total of six functions were employed, utilizing for loops to iterate through the data. Conditional statements, such as if and elif, were employed for value comparisons. The matPlotLib library was imported and utilized by the `plotGraph()` function for drawing the bar graph. The data, sourced from Yahoo Finance and Visual Crossing, was downloaded in .csv format (Comma Separated Values).

Challenges

One of the major challenges I had in designing my program was dealing with the downloaded data files from Yahoo Finance and the Visual Crossing site. Since the stock market is not open on the weekends the data from the two files did not match up line for line. In my first iteration of the program, I actually went through and removed all weekends from the weather.csv file so each line from both files would match with the same date. After I had the program running, it did not sit right with me that one should have to manually remove the weekends from the weather file especially if more testing is to be done on other years and different stock exchanges. Besides the obvious labor involved, this method would be prone to errors. I decided to loop through each line in the stock data file and each line in the weather data file comparing dates and only writing to the new output file if the dates matched. Another issue that arose was that the dates downloaded for each file were not in the same format. I handled this using the split command and then comparing the dates with individual variables for month, day and year.

Code that loops through both csv files creating a new file only if the dates from both files match (Lines 38-58 in FA595_Final_Project.py)

```
s = 0  # s is index for line number in stockData
for w in range(lineCountWeatherData):  # w is index for line number in weatherData
  if s < lineCountStockData:
      stockDataDate = stockData['Date'][s]
      # Since the date in the weather file is in a different format than the date in the stockData file
      # I split the values into individual variables for testing later
      stockDataYear, stockDataMonth, stockDataDay = stockDataDate.split('-')
      weatherYear, weatherMonth, weatherDay = weatherData['datetime'][w].split('-')
      # Only write to newFile.csv if the two dates actually match from the
      # weather file weather.csv file and the stockData file NYSE.csv
```

```
    if int(stockDataDay) == int(weatherDay) and int(stockDataMonth) == int(weatherMonth)
and int(stockDataYear) == int(weatherYear):
        weather = weatherData['conditions'][w]
        # upOrDown contains the total points gained or lost for that day on the NYSE from open
        # to close
        upOrDown = round(stockData['Close'][s] - stockData['Open'][s], 2)
        # Data Rows of newFile.csv file
        rows = [stockDataDate, upOrDown, weather]
        # writing the data rows to newFile.csv
        csvwriter.writerow(rows)
        # if the two dates match move to the next line of stockData by increasing its index
        s += 1
```

Lastly, I faced a challenge when striving to construct the program without utilizing global variables. From the beginning, I set a design objective to avoid global variables due to their potential to create complications. This goal was successfully achieved by structuring my code into discrete functions, each dedicated to fulfilling specific tasks. In total, six functions were created to address various aspects of the program's functionality.

Future Extensions

Upon completing this project, I find myself grappling with the complexity of monitoring numerous variables to accurately correlate the validity of the Sunshine Anomaly assumption. It becomes apparent that obtaining data exclusively from individuals who purchase stocks within the city associated with the monitored stock exchange would offer a more refined perspective. For instance, my current data retrieval is centered on stock data from the New York Stock Exchange (NYSE) and corresponding weather data for New York, Los Angeles, Chicago, and Houston. However, this approach overlooks the fact that stock purchasers on the NYSE may be located in other states or even different countries, experiencing vastly different weather conditions. To enhance accuracy, a more focused comparison could involve analyzing purchases and sales exclusively from individuals and institutions situated in the city of the exchange.

The current method lacks the capability to recognize whether market movements result from a substantial purchase by a hedge fund or institution, potentially unrelated to the local weather. A more reasonable approach might involve comparing individual purchases to sales of investors located in the exchange area. Although obtaining data in such fine detail may pose challenges, it could shed light on whether investors are more likely to buy stocks on sunny days than to sell. This nuanced analysis, considering individual behavior, appears more insightful than simply correlating market movements with weather conditions.

In future versions, I aim to incorporate features enabling the program to compare different market sectors over time and identify sectors that perform better in specific months. Additionally, I plan to integrate file dialog boxes that empower users to select the files they wish to analyze. This flexibility allows users to download data spanning multiple years and from different exchanges without the need to modify hard coded file names in the code.