The “STOCK” Network: *NYSE Analysis*

Group Members: Sherry and Hiam

December 8th, 2021

Motivation

A gentle breeze, a bit of rain… We don’t mind this. Although, when it comes to a full downpour of rain or a bad snow storm, this inclement weather may devastate a life, damage property, or close off an internet connection. Did some climatic event affect your financial situation? Let’s investigate if daily weather has some effect on stock markets.

Abstract

For this project, we extracted daily weather data for the city of New York (NYC) between the beginning of 2018 until the end of month of October 2021, as well as stock information for 23 different stock exchanges where their headquarters are located within the city. The data was manipulated through the use of Pandas to populate clean datasets. Once transformed, the data was loaded onto MongoDB, where we further queried data to create our visualizations.

In the observed dynamic data, a trend does not seem to be apparent. When a regression analysis is performed however, the daily maximum temperature presents a strong possibility of a correlation with the stock close prices. As well, there is a moderate possibility of a correlation with the maximum temperature and volume.

Project Objective

Hypothesis: Maximum temperature correlates to the local Headquarters of the NYC stock exchanges between January 2018 and October 2021.

Questions to answer:

1. Is there a trend between the daily max temperature (°C) and stock volume over time?
2. Is there a trend between the daily max temperature (°C) and stock closing price over time?
3. Is there a relationship between the daily max temperature (°C) and stock volume?
4. Is there a relationship between the daily max temperature (°C) and stock closing price?

To answer these questions and to confirm our hypothesis, we extracted data for NYC’s daily weather and stock information for 23 stock markets within NYC between January 2018 and

October 2021. Ultimately, we had to map the daily max temperature with the daily stock volume and closing prices for each stock market on days the NYSE was open.

Methods

Data Sources Used:

- NYC Daily Weather Data: World Weather Online API
- Stock Information: Alphavantage API
- Geo Information: Google Maps and Geopy

Final Production Database: MongoDB

Javascript Libraries Used:

- Visualizations for Analysis and Correlation: Chart.js
- To get access to our routed data: **D3**
- To create our interactive webpage: Bootstrap 5.0 template

Web Page Breakdown

Home Page: Our landing page is where you will find our motivation, hypothesis, and steps taken to gather needed data for the following pages. In addition, it displays the results we obtained from the data analysis and correlation performed.

Analysis Page: In here, we find two multi y-axis graphs that demonstrate the trend of stock volumes, stock closing prices, and max temperature over time (between January 2018 - October 2021).

- There is a dropdown menu which allows you to switch between 23 stock markets within NYC, and see their trend respectively.
- Both graphs are interactive, and you can get the exact values of each point by just clicking on it.
- There is also the option to only display one of the two y variables on each graph. For instance, if you only want the stock volume to appear on the first graph, you simply click on the Max Temperature line in the legend, and it will automatically cross it out and remove temperature data from the graph.

Correlation Page: In this page, we find an interactive scatter plot which can be manipulated by two things: the stock symbol selected through the dropdown, and two radio buttons to display the volume or price for the stock chosen. The result is a scatter plot of Max Temp vs Close Price or Max Temp vs Volume. Additionally, the Pearson r^2 value is calculated, to help determine the fit of the correlation for the chart.

Map Page: The map page shows an interactive map displaying the stock HQs of 23 stock markets within NYC. There is a ‘zoom in’ feature which allows you to look into the map areas. When clicking on a particular marker, a popup shows up displaying the following about the HQ marker chosen:

- Stock Symbol and Full Company Name

- HQ Address
- Sector
- Industry

Code Highlight

ETL - Weather Data

NYC Weather Daily Weather Data:

We used a pre-programed package that allowed the retrieval and transformation of historical data from World Weather Online into a pandas dataframe and CSV file. We specifically retrieved daily weather data for NYC, from the beginning of 2014 until mid of November 2021. Once our API key was inputted, a whole list of weather variables are outputted and saved in a CSV file within our specified folder. Cleaning was performed to gather information needed to load to a Mongo database.

Lessons Learned: When we first used the World Weather History Package for data retrieval, we had to discover what was getting retrieved, as we didn't necessarily need everything and anything to do with NYC Weather Data:

- We had to refine our search to specific dates. We extracted from January 2018 to Mid November 2021, as we didn't yet know the time frame we wanted to analyze specifically
- We also refined the weather variables that were outputted. There was a long list of weather factors that could have been extracted, but we reduced it to a few, again not knowing which ones we would analyze specifically yet.
- The naming convention of the city was confirmed through many trials and errors:
    - For NYC specifically: **location_list** = ['New+York,ny']

ETL - Stock Exchange Data

Alphavantage Stock History Data:

The website documentation had various grouped categories for their api's to use according to your needs. Extraction of daily stock data, one can use an api and receive information in a csv or json format. As well, there is a library/wrapper that can be used to retrieve the data that was acceptable to pandas. For this application, a call was made to the api's, using a key, and getting a json back. Once the calls were made, the files were opened and cleaned and saved as a csv file. Cleaning was performed to gather information needed to load to a Mongo database.

Lessons Learned: Requests for Alpha Vantage can be done every 5 minutes. But often, a request may not be processed, due to the site being busy.

QUERIES IN FLASK APP

Please refer back to app.py in order to see the queries done for each route created in detail.

We imported all needed dependencies, and created our connection to our mongo Database. In order to make requests from the local server and avoid errors, we had to use the CORS import. We then created our app routes for our interactive webpage.

```
from flask import Flask, render_template, redirect
from flask_pymongo import PyMongo
from flask import jsonify

from flask_cors import CORS

from scipy.stats.stats import pearsonr

app = Flask(__name__)

app.config["MONGO_URI"] = "mongodb://localhost:27017/NYSE_Weather_db"
mongodb_client = PyMongo(app)

# To request from local server and avoid CORS error
CORS(app, support_credentials=True)
```

Lessons Learned: When we first created our app routes, we were merging both our html and javascript paths within the same route. This led to an error where the server couldn't read both html and javascript paths together. We had to separate the html and javascript paths by creating separate routes for each.

Step 1: Create routes where our .html files were being rendered to create our web pages.

Step 2: Create routes for data retrieval, which will be linked into our .js files.

Example: Here we have the route for the Map Data. We rendered through our mongoDB collection where the Stock Exchange Geo Information is located.

```
#Map Data Route
@app.route("/data")
def getData():
    # loop through here and append to list
    all_map_data = mongodb_client.db.StockGeoInfo.find()
    myData=[]
    for each in all_map_data:
        del each['_id']
        myData.append(each)
    #print(myData)
    #print("hello map")
    return (jsonify(myData))
```

Lessons Learned: When we first rendered through this, the collection was not being read properly due to the '_id' document, because it was being seen as an object within the collection. So this field was removed from the data returned. The same logic was applied to all the other routes where data was being retrieved from MongoDB.

In the map.js file, we created a constant which stores the flask APP route that we want, and then using d3 we extracted the data.

```
// Load in data from url
const url = "/data";
console.log(url);

d3.json(url).then(function(response) {…
});
```

Findings

Please refer back to our [Github Repo] to run the application in order to render through all Stock Markets. Here we are only looking at IBM Analysis and Correlation as an example.

Using the Max Temperature Data and the Stock Volume Data, we plotted the Stock Volume and Max Temperature between January 2018 to October 2021

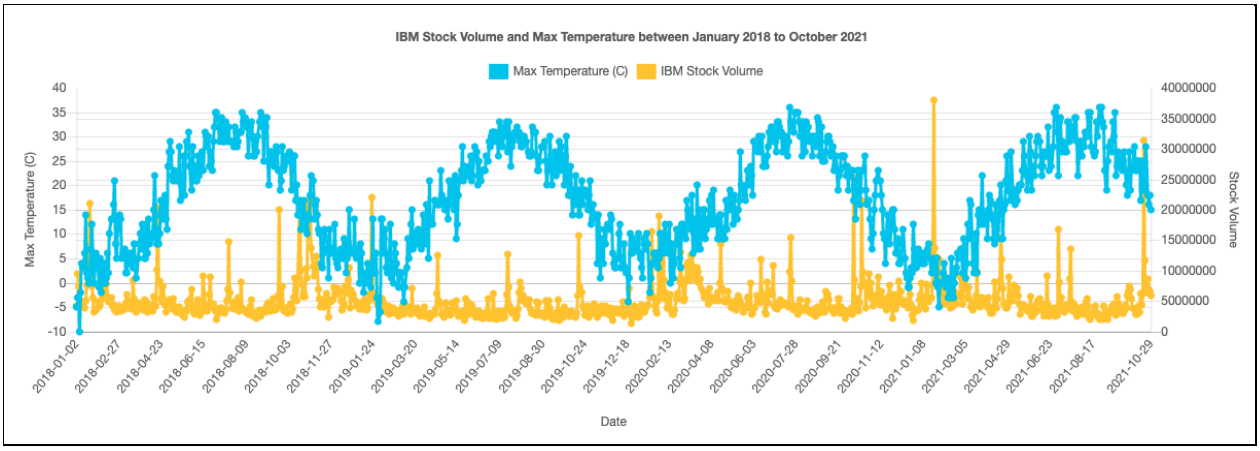

Figure 1.0: Max Temperature in Celsius in New York City and IBM's Stock Volume between January 2018 to October 2021

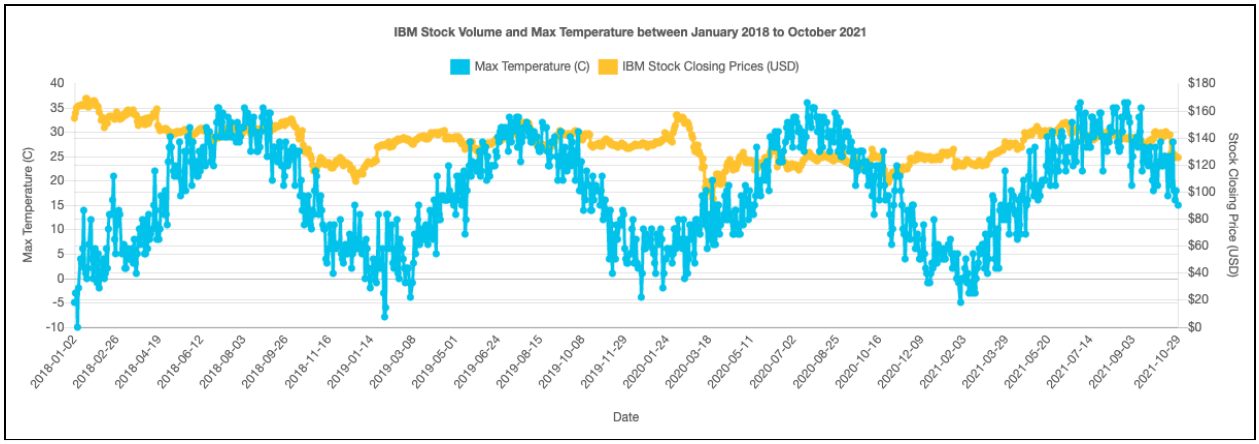

Figure 1.1: Max Temperature in Celsius in New York City and IBM's Stock Closing Prices between January 2018 to October 2021

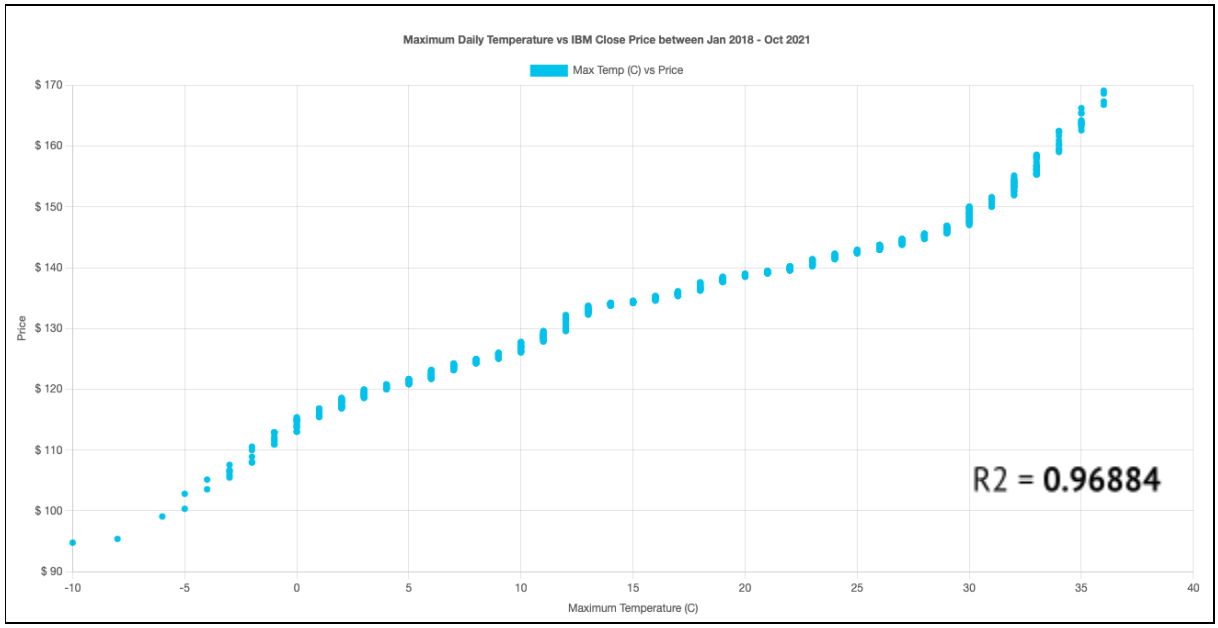

Figure 1.2: Max Temperature in Celsius in New York City vs IBM's Stock Closing Prices for January 2018 to October 2021.

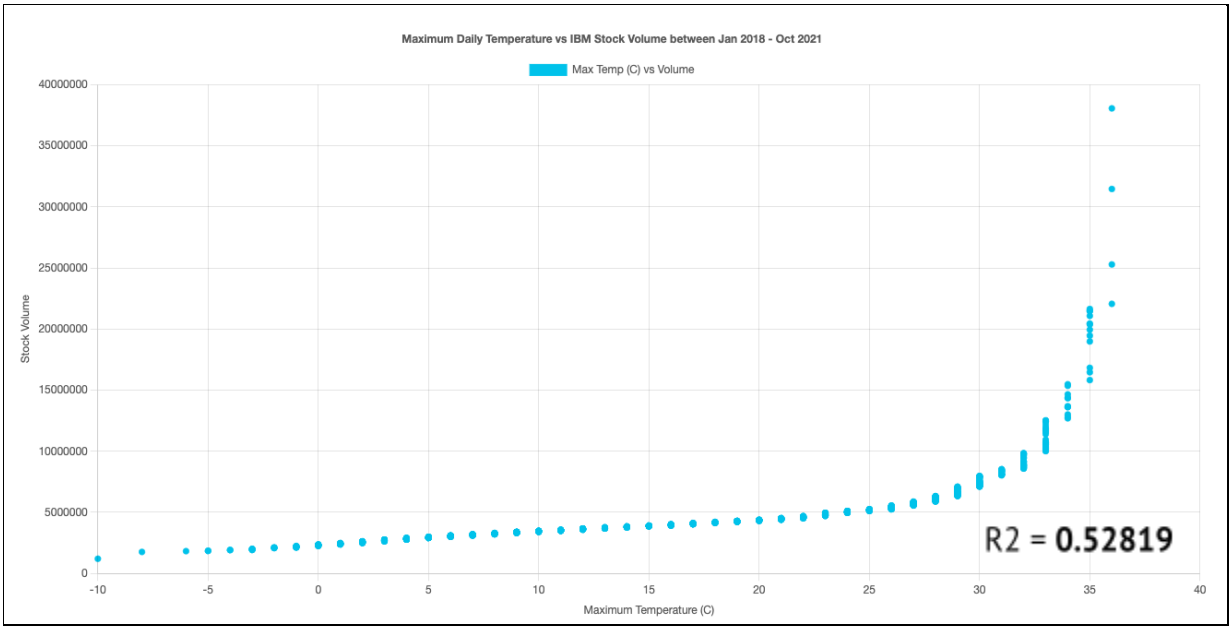

Figure 1.3: Max Temperature in Celsius in New York City vs IBM's Stock Volume for January 2018 to October 2021.

Our analysis focused on three main variables: daily max temperature, stock volume, and stock closing price.

- The stock closing price is the last price at which a financial investment with monetary value (or better known as security) is traded during the regular trading day. A security's

closing price is the standard benchmark used by investors to track its performance over time.[1]
- Stock volume on the other hand is the number of shares of a security traded during a given period of time. Generally, securities with a higher daily volume are more liquid than those without given that they are more active. Volume can help indicate relative significance of a market move.[2]

In ***Figures 1.0 and 1.1***, we can see a sinusoidal trend for the max temperature across time, which is justified by the change in seasons. As for the stock volume, there seems to be a linear trend that is mostly consistent over time. Between the max temperature and stock volume, there does not appear to be a particular trend. This same observation is applicable to the second figure.

In order to confirm this further, we created a regression model for each stock variable being analyzed vs the max temperature, to see if there is any relationship between the two (***Figures 1.2*** and ***1.3***).

The strength of a correlation between the Maximum Temperature and a Stock Closing Price or Transaction Volume data is measured using the Pearson r^2 values:

- Weak relationship: $r^2 < 0.5$
- Moderate relationship: $0.5 < r^2 < 0.7$
- Strong relationship: $r^2 > 0.7$

As one reviews the plots, over the expanse of the approximated three year period, one can notice some correlations by viewing the Pearson 'r-squared' value. The daily maximum temperature presents a strong possibility of a correlation with the stock close prices, given that the r^2 value is greater than 0.9. As well, there is a moderate possibility of a correlation with the maximum temperature and volume, with an r^2 value greater than 0.5. However, these correlations may adjust slightly depending on the stock that is chosen for the production of the charts.

We do have to note that Pearson's r-squared correlation uses a linear regression model. Some of the plots for the data appear non-linear and an alternative correlation method could be used for accuracy.

Conclusions

Based on the findings obtained, we do not reject the null hypothesis declared of Weather (Max Temp) correlates to the local HQs of NYC stock exchanges during January 2018 to October 2021.

When a regression analysis is performed, the daily maximum temperature presents a strong possibility of a correlation with the stock close prices (for IBM the r^2=0.96884). As well, there is a moderate possibility of a correlation with the maximum temperature and volume, with an r^2 value of greater than 0.52819 for IBM.

Notes on Conclusions

Obviously maximum temperature does not dictate the stock closing price or transaction volume. There are many other factors that affect the stock market, and this result can be very controversial. The analysis observed is just showing the trends observed.

Limitations

- Single Weather Factor:
  - Looking strictly at the maximum temperature effect on stock volume and closing prices.
- Limiting number of HeadQuarters:
  - We only retrieved data for 23 HQs in NYC.
- Time Frame Scope:
  - Only 3 years worth of data were used for analysis and correlation
- Linear Regression Model.

Future Considerations

- Multiple Weather Factors: Humidity, Snow Fall, Precipitation, etc.
- Increase the number of HQs for analysis.
- Expanding out Time Frame Scope.
- Have a date filter where we could look into specific timeframes for hidden trends
- Looking at alternative models to fit the non-linear trends.

References

(1) https://www.investopedia.com/terms/c/closingprice.asp

(2) https://www.investopedia.com/terms/v/volume.asp