

Name: Antony Wu
Student ID: 0439 9524
Email: antonywu@umich.edu
Partner: Phillip Ripsam-Walsh

APIs, SQL, and Visualizations

Goals:

Initially, our group was interested in tracking weather in different countries across the world and comparing it to either depression rates in those countries as we believed there would be some correlation between the data.

After searching for APIs and websites to use for collecting this data, we found it difficult to find ones for weather and depression that worked for this project but in our search, we found APIs for Covid data and the stock market and decided it would be interesting to see how Covid-19 affected a certain stock in the year 2020. So we used API Ninja's Covid-19 API and gathered data on the new and total cases for each day from January 22, 2020, to December 31, 2020, as there was only data starting on January 22nd for that year. We also used Financial Modeling Prep's Historical Daily Price API and gathered data on the daily low, high, open, and close prices of Expedia stock as well as the daily price change, change percent, and change over time of the stock in the year 2020.

Problems:

This project definitely came with a lot of kinks and issues that we had to sort out. The first issue we came across was many APIs cost money and we weren't able to find any APIs that fit our original project plan so we had to scrap the whole idea and decide on another one. Another problem was that once we found an API for the stock data, it only had historical data dating back to April 2021, but we wanted to gather data for the year 2020 as it was the year that was most affected by Covid and we wanted the date to be our integer primary key that we were going to use to join the two datasets. Eventually, we were able to find an API that was free and had stock data for most of 2020 which gave us more than 100 data points to use for our stock data table.

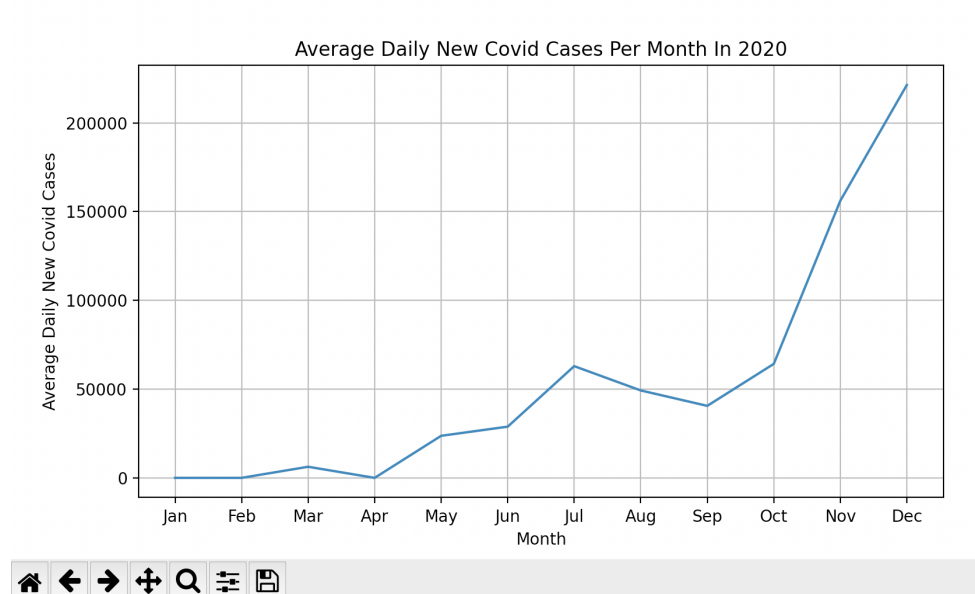
Calculations:

```

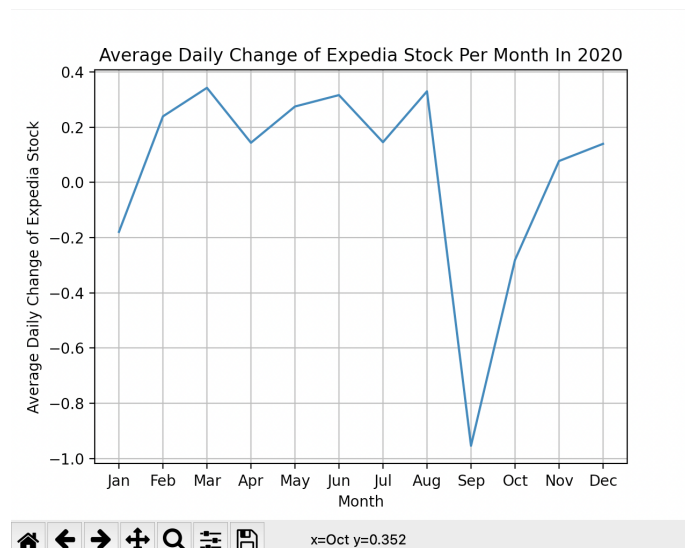
1  Average New Covid Cases and Expedia Price Changes in 2020:
2
3  January:
4  Average Cases Per Day: 0.5
5  Average Expedia Price Change Per Day: -0.179175
6
7  February:
8  Average Cases Per Day: 0.47368421052631576
9  Average Expedia Price Change Per Day: 0.23921052631578957
10
11 March:
12 Average Cases Per Day: 6253.045454545455
13 Average Expedia Price Change Per Day: 0.3426681818181817
14
15 + April:
16 Average Cases Per Day: 30031.238095238095
17 Average Expedia Price Change Per Day: 0.1436904761904762
18
19 May:
20 Average Cases Per Day: 23700.15
21 Average Expedia Price Change Per Day: 0.275
22
23 June:
24 Average Cases Per Day: 28851.045454545456
25 Average Expedia Price Change Per Day: 0.316359090909091
26
27 July:
28 Average Cases Per Day: 62941.045454545456
29 Average Expedia Price Change Per Day: 0.1456363636363636
30
31 August:
32 Average Cases Per Day: 49258.76190476191
33 Average Expedia Price Change Per Day: 0.33
34
35 September:
36 Average Cases Per Day: 40567.28571428572
37 Average Expedia Price Change Per Day: -0.9540476190476189
38
39 + October:
40 Average Cases Per Day: 64221.63636363636
41 Average Expedia Price Change Per Day: -0.2809090909090909
42
43 November:
44 Average Cases Per Day: 156122.6
45 Average Expedia Price Change Per Day: 0.07750000000000004
46
47 December:
48 Average Cases Per Day: 221347.36363636365
49 Average Expedia Price Change Per Day: 0.13977272727272733
50
51
```

From these images, we can see that for each month in the year 2020, we calculated the average number of new cases per day and the average price change per day of Expedia stock. We joined the Covid_data table with the Stock_data table based on both of them having the date which we converted to integers in the form (month)(day) so January 30th would be 130 and December 2nd would be 1202. From the covid table, we took the new cases per day and from the stock table we took the price change per day. From there, we iterated through each day (which excluded weekends because the stock market closes on the weekend) and calculated the average daily change in price and the average daily new cases for each month.

Visualizations:



This graph shows the average daily new covid cases in the U.S. per month. From this chart, we can see before the pandemic really started at the end of March, there weren't many new cases. However, throughout the month of April and onwards is when the new cases started to pick up and the countrywide lockdown was put in place.



This graph charts the average daily change of Expedia's stock price for each month, and from it, we can see that it started plummeting throughout the month of August. This was around the height of the pandemic so what we concluded from this data is because cases were so severe, people weren't able to travel very much and this had an impact on travel agencies such as Expedia. From the previous graph, we can see that the new cases were much higher in the winter months, and yet the daily change was much less compared to August. From our intuition, we believed the reason for this was due to the fact that people travel more in the summer and so the inability to travel during the lockdown definitely caused a hit to the stock price.

Code Instructions:

To run our code, the user must first run the covid.py file by doing **python3 covid.py** which gathers the data from both APIs, creates the tables, and inserts the data into the tables in the database. After doing this, the user must run the combined.py file and redirect the printed output into a txt file by doing **python3 combined.py > Data.txt** which joins the two tables for data that has equal dates and calculates the average daily new covid cases for each month and the average daily change in Expedia stock price for each month. This file also uses Matplotlib to create the 2 visualizations for each calculated dataset.

Functions:

In the covid.py file, the getJSON function takes in a URL and an API key and returns a dictionary representation of the JSON file that was requested from the API. The main function creates a dictionary for the covid data and a dictionary for the stock data, creates tables in the SQLite database, and inserts the data into the database. It creates the covid.db file which contains the database that was created. This function returns nothing.

In the combined.py file, the main function joins the two tables and from the list of data that is collected, calculates the average daily new covid cases for each month and the average daily change in Expedia stock price for each month and saves them in a dictionary with the keys being months as numbers 1-12. It then prints out each month's average daily new covid cases and average change in Expedia stock price. It then creates two line graphs, one for the covid cases data and one for the change in stock price data for each month. This function returns nothing.

Resources:

Date	Issue Description	Location of Resource	Result
04/09/23	Didn't know where to find free APIS	Github	After looking through the list of APIs in the GitHub repository provided in the instructions, we were able to find 2 that worked for us
04/11/23	Confused about how to access API we selected	APINinjas.com	Used API documentation and obtained a key to use it
04/12/23	Confused about how to join two tables	Runestone	After looking at some examples, we were able to figure out how

			to join tables together
04/16/23	Couldn't get our visualizations working	Runestone	After looking back at runestone readings for matplotlib, we were able to successfully plot our visualizations

GitHub Repository:

<https://github.com/PhillipRW03/Final-Project>