Team: Aimee Zheng, Marina Sun, Sage Pei

API's, SQL's, and Visualizations
Github: https://github.com/aimee-z/SI-206-Project.git

Original Goals:
Our original goal was to utilize the GeoKeo and Abstract IP Geolocation APIs, and EPA.gov website. We wanted to use the GeoKeo database to find the current average population and income in Michigan cities by zip code, and then use the Abstract IP Geolocation API to see how many individuals have searched for the EPA.gov website and where they are from based on zip code. The aim was to understand the relationship between population, income, and environmental initiative of individuals residing within Michigan cities. Overall though, our main goal was to develop our programming knowledge through hands-on utilization of API's and SQL (i.e. use three APIs and create three visualizations).

Achieved Goals:
As we began work on our project, we ended up redirecting the project scope due to limitations of our API's. While we thought that GeoKeo would be able to find county locations, further investigation revealed that the API was missing key locations that we needed to continue our project. As a result, we decided to instead look at COVID-19 data for the U.S.A as a whole, just for New York, and then Bitcoin data in order to see relations between COVID-19 rates and Bitcoin stock price. The APIs we used for this shift are: The COVID Tracking Project API and Coin Paprika API. We stored all data using 25 rows at a time, resulting in 100 rows each for our three tables and created 5 visualizations, but we did not use three APIs.

Problems Faced:
We had difficulty in correlating the dates with each other. For overall COVID-19 data, the dates began in January, but New York COVID-19 data only began after March. However, we solved this by including code that would make the two equal, starting with a column of sequential days beginning at 49. Additionally, we faced problems in using a third API that would allow us to use a time range to make valuable insights for our analyses. We tried to use Polygon's stock API, but this only offered stock data on one day or stock data across an undefined amount of time. The other APIs we found had paywalls or were discontinued before the pandemic, resulting in time ranges before the one we needed. Ultimately, we decided to move forward using the COVID-19 and Bitcoin API's, as they both provided useful information we needed during specific dates.

Calculations:
1. Data calculated from finding the percentage of COVID-19 cases nationally that were in NY on a given day:

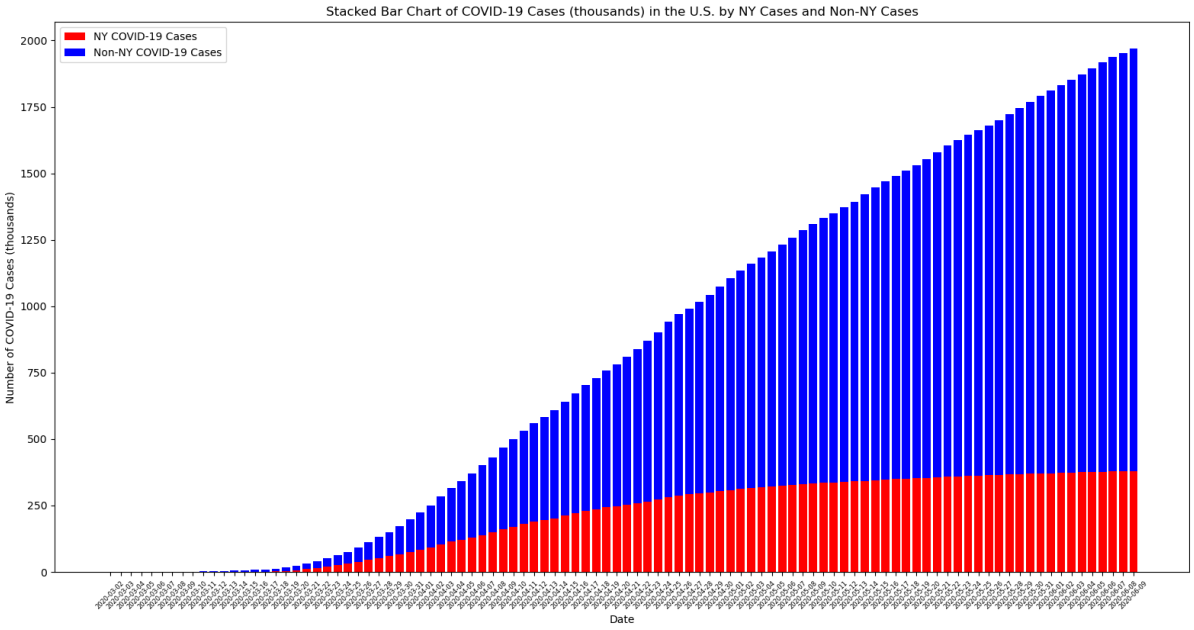
```
Percentage of national COVID-19 cases that were identified in New York on a give
================================================================================

On 2020-03-02 the percent of national COVID-19 cases in New York was 0.0%.
On 2020-03-03 the percent of national COVID-19 cases in New York was 0.88%.
On 2020-03-04 the percent of national COVID-19 cases in New York was 0.41%.
On 2020-03-05 the percent of national COVID-19 cases in New York was 0.98%.
On 2020-03-06 the percent of national COVID-19 cases in New York was 5.62%.
On 2020-03-07 the percent of national COVID-19 cases in New York was 6.13%.
On 2020-03-08 the percent of national COVID-19 cases in New York was 7.93%.
On 2020-03-09 the percent of national COVID-19 cases in New York was 8.29%.
On 2020-03-10 the percent of national COVID-19 cases in New York was 10.09%.
On 2020-03-11 the percent of national COVID-19 cases in New York was 10.18%.
On 2020-03-12 the percent of national COVID-19 cases in New York was 9.68%.
On 2020-03-13 the percent of national COVID-19 cases in New York was 10.23%.
On 2020-03-14 the percent of national COVID-19 cases in New York was 11.55%.
On 2020-03-15 the percent of national COVID-19 cases in New York was 11.29%.
On 2020-03-16 the percent of national COVID-19 cases in New York was 12.65%.
On 2020-03-17 the percent of national COVID-19 cases in New York was 14.35%.
On 2020-03-18 the percent of national COVID-19 cases in New York was 18.42%.
On 2020-03-19 the percent of national COVID-19 cases in New York was 23.67%.
On 2020-03-20 the percent of national COVID-19 cases in New York was 30.04%.
On 2020-03-21 the percent of national COVID-19 cases in New York was 33.87%.
On 2020-03-22 the percent of national COVID-19 cases in New York was 38.1%.
On 2020-03-23 the percent of national COVID-19 cases in New York was 40.93%.
On 2020-03-24 the percent of national COVID-19 cases in New York was 41.47%.
On 2020-03-25 the percent of national COVID-19 cases in New York was 41.33%.
On 2020-03-26 the percent of national COVID-19 cases in New York was 40.43%.
On 2020-03-27 the percent of national COVID-19 cases in New York was 40.08%.
On 2020-03-28 the percent of national COVID-19 cases in New York was 39.89%.
On 2020-03-29 the percent of national COVID-19 cases in New York was 39.46%.
On 2020-03-30 the percent of national COVID-19 cases in New York was 38.67%.
On 2020-03-31 the percent of national COVID-19 cases in New York was 38.48%.
On 2020-04-01 the percent of national COVID-19 cases in New York was 37.51%.
On 2020-04-02 the percent of national COVID-19 cases in New York was 36.76%.
On 2020-04-03 the percent of national COVID-19 cases in New York was 36.33%.
On 2020-04-04 the percent of national COVID-19 cases in New York was 35.95%.
On 2020-04-05 the percent of national COVID-19 cases in New York was 35.67%.
On 2020-04-06 the percent of national COVID-19 cases in New York was 35.29%.
On 2020-04-07 the percent of national COVID-19 cases in New York was 34.65%.
On 2020-04-08 the percent of national COVID-19 cases in New York was 34.58%.
On 2020-04-09 the percent of national COVID-19 cases in New York was 34.27%.
```

2. Data from the calculations of the difference between NY COVID-19 cases compared to national cases on a given day:

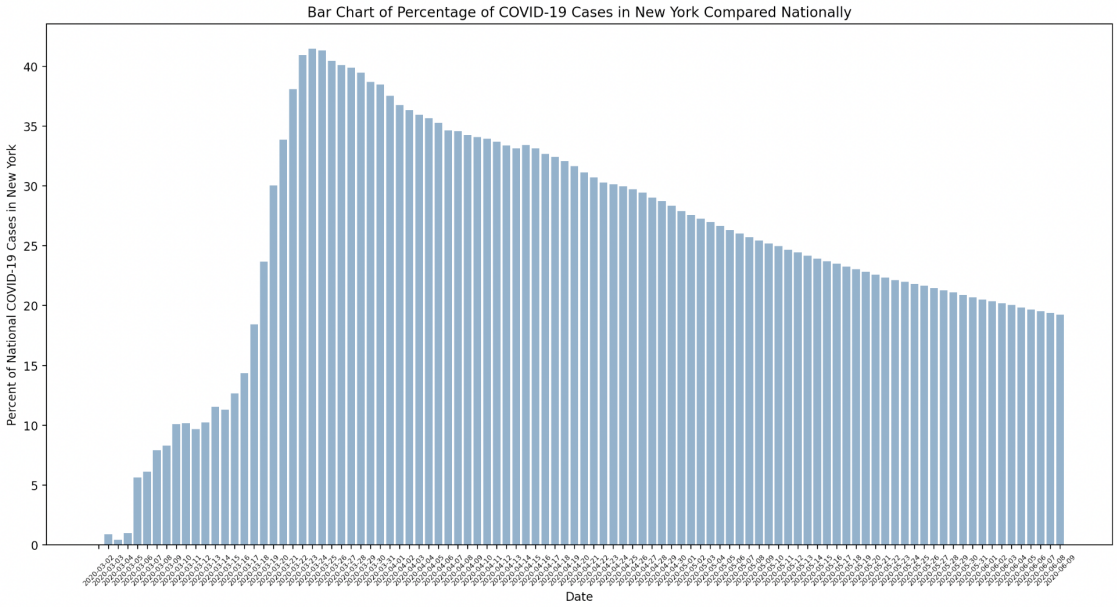
```
Difference of NY COVID-19 cases compared to national cases on a given day:
==========================================================================

On 2020-03-02 the difference of NY COVID-19 cases compared to national cases was 72 cases.
On 2020-03-03 the difference of NY COVID-19 cases compared to national cases was 113 cases.
On 2020-03-04 the difference of NY COVID-19 cases compared to national cases was 241 cases.
On 2020-03-05 the difference of NY COVID-19 cases compared to national cases was 304 cases.
On 2020-03-06 the difference of NY COVID-19 cases compared to national cases was 420 cases.
On 2020-03-07 the difference of NY COVID-19 cases compared to national cases was 551 cases.
On 2020-03-08 the difference of NY COVID-19 cases compared to national cases was 697 cases.
On 2020-03-09 the difference of NY COVID-19 cases compared to national cases was 973 cases.
On 2020-03-10 the difference of NY COVID-19 cases compared to national cases was 1346 cases.
On 2020-03-11 the difference of NY COVID-19 cases compared to national cases was 1720 cases.
On 2020-03-12 the difference of NY COVID-19 cases compared to national cases was 2341 cases.
On 2020-03-13 the difference of NY COVID-19 cases compared to national cases was 3097 cases.
On 2020-03-14 the difference of NY COVID-19 cases compared to national cases was 3959 cases.
On 2020-03-15 the difference of NY COVID-19 cases compared to national cases was 5090 cases.
On 2020-03-16 the difference of NY COVID-19 cases compared to national cases was 6507 cases.
On 2020-03-17 the difference of NY COVID-19 cases compared to national cases was 8203 cases.
On 2020-03-18 the difference of NY COVID-19 cases compared to national cases was 10551 cases.
On 2020-03-19 the difference of NY COVID-19 cases compared to national cases was 13388 cases.
On 2020-03-20 the difference of NY COVID-19 cases compared to national cases was 16538 cases.
On 2020-03-21 the difference of NY COVID-19 cases compared to national cases was 20224 cases.
On 2020-03-22 the difference of NY COVID-19 cases compared to national cases was 24640 cases.
On 2020-03-23 the difference of NY COVID-19 cases compared to national cases was 30123 cases.
On 2020-03-24 the difference of NY COVID-19 cases compared to national cases was 36229 cases.
On 2020-03-25 the difference of NY COVID-19 cases compared to national cases was 43741 cases.
On 2020-03-26 the difference of NY COVID-19 cases compared to national cases was 54885 cases.
On 2020-03-27 the difference of NY COVID-19 cases compared to national cases was 66721 cases.
On 2020-03-28 the difference of NY COVID-19 cases compared to national cases was 78825 cases.
On 2020-03-29 the difference of NY COVID-19 cases compared to national cases was 91313 cases.
On 2020-03-30 the difference of NY COVID-19 cases compared to national cases was 105468 cases.
On 2020-03-31 the difference of NY COVID-19 cases compared to national cases was 121170 cases.
On 2020-04-01 the difference of NY COVID-19 cases compared to national cases was 139457 cases.
On 2020-04-02 the difference of NY COVID-19 cases compared to national cases was 158923 cases.
On 2020-04-03 the difference of NY COVID-19 cases compared to national cases was 180276 cases.
On 2020-04-04 the difference of NY COVID-19 cases compared to national cases was 202557 cases.
On 2020-04-05 the difference of NY COVID-19 cases compared to national cases was 220073 cases.
On 2020-04-06 the difference of NY COVID-19 cases compared to national cases was 239675 cases.
On 2020-04-07 the difference of NY COVID-19 cases compared to national cases was 261917 cases.
On 2020-04-08 the difference of NY COVID-19 cases compared to national cases was 282428 cases.
```

Visualizations:

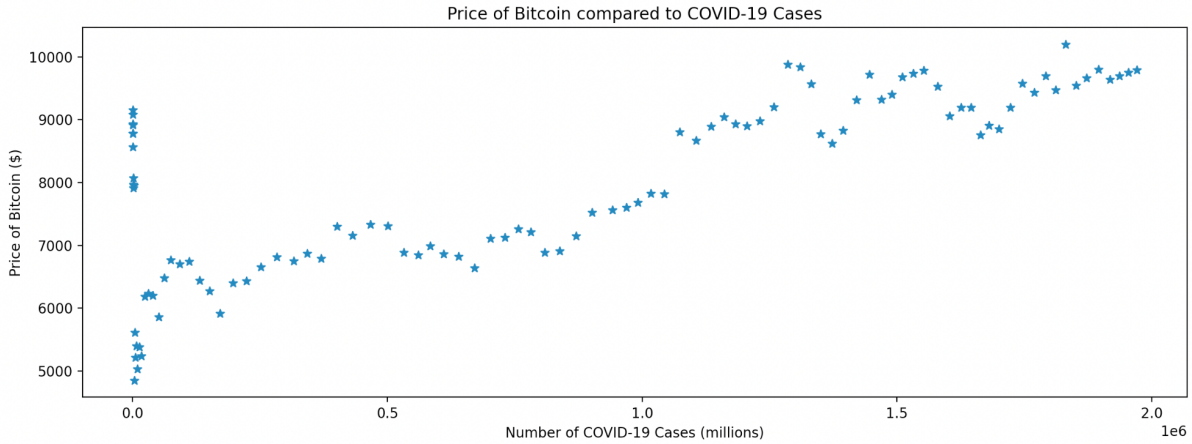
1. Stacked Bar Chart of NY Cases in comparison to non-NY cases



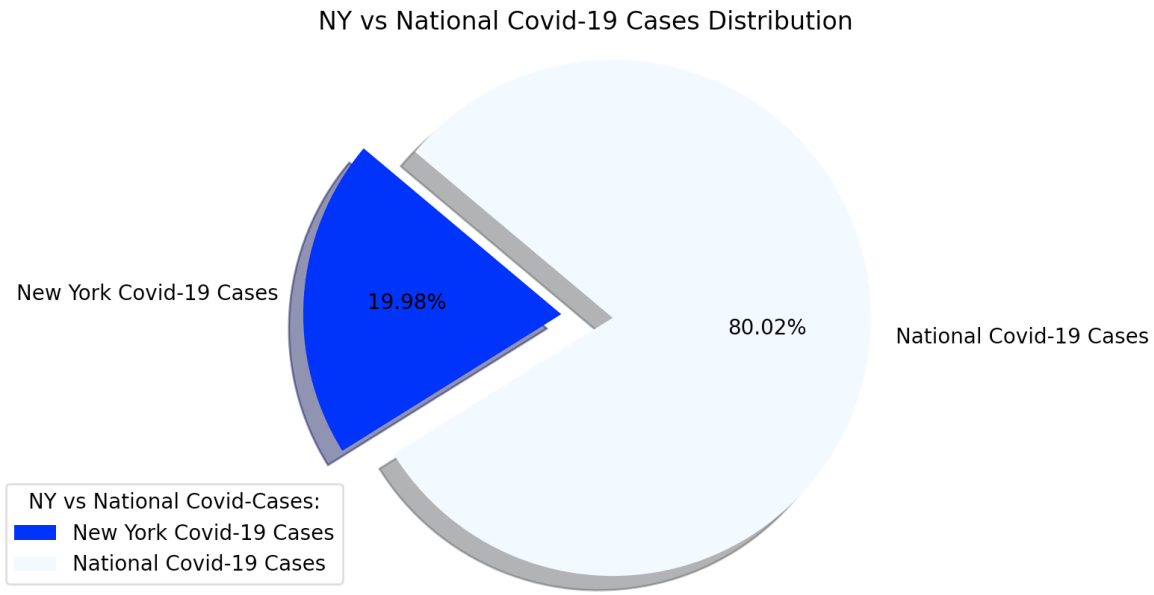
2. Bar Chart of percent of NY cases in comparison to entire nation



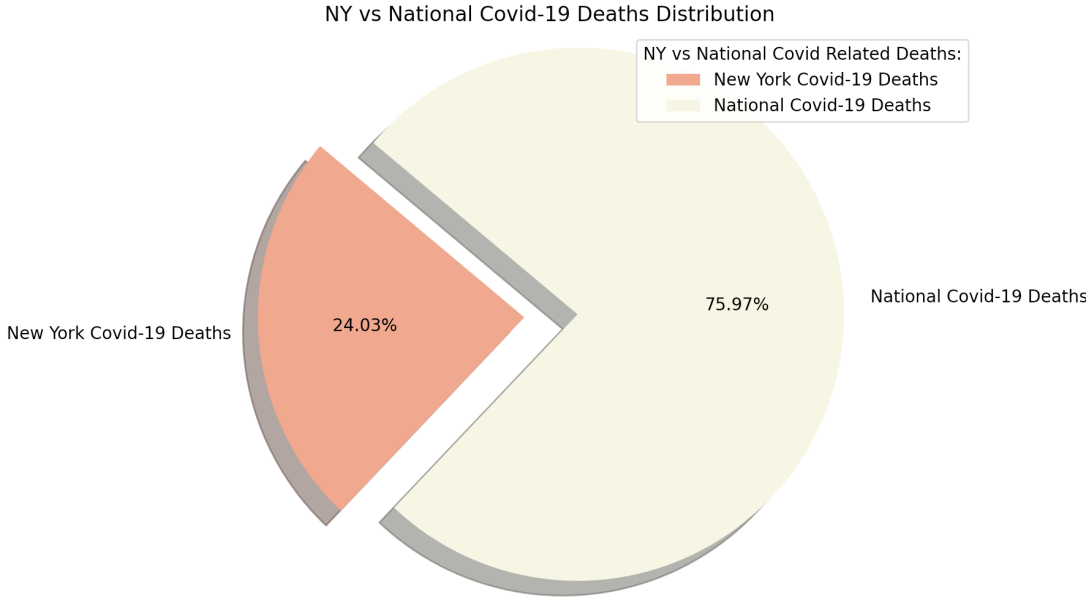
3. Scatterplot comparing total national cases (millions) to Bitcoin price ($)



4. Pie Chart of Average Number of Covid-19 Cases in NY to the entire country



5. Pie Chart of Average Number of Covid-19 related Deaths in NY to the entire country



Instructions for Running Code:
To run our code, please just make sure you have some form of SQL reader installed on your computer so you are able to see the tables that we were able to create.

1) Please run through the main.py file 4 times in order to fill in the database (we compiled our code into one file, making it easier for you).
2) You will be able to see our visualizations and csv files with all calculations included.

Function Documentation: (also included in the main.py file)
National Covid-19 Data:

a) def covid_api():
   i) Takes in no inputs. Uses request module in order to get historical daily Covid-19 statistics from Covid-19 API.
   ii) It returns Covid-19 data for the date range that we specified in JSON formatting
b) def covid_table(data, cur, conn):
   i) Takes in COVID-19 JSON formatted data from covid_api(), database cursor and connector as inputs.
   ii) Returns nothing. Creates a new table called COVID_Data.
c) def set_up_covid(data, cur, conn, start):

i) Takes in COVID-19 JSON formatted data from covid_api(), database cursor and connector, and start value (integer) as inputs. Returns nothing.
ii) Indexes date, total cases, percentage of the population with positive cases, hospitalizations, and deaths from COVID-19 data into COVID_Data table.
iii) Uses start value to limit amount of data to 25 collected/stored at a time

NY Covid-19 Data:

d) def ny_covid_api():
    i) Takes in no inputs. Uses requests module to get historical daily New York COVID-19 statistics from COVID-19 API.
    ii) Returns COVID-19 data for date range in JSON formatting.
e) def ny_covid_table(data, cur, conn):
    i) Takes in COVID-19 JSON formatted data from ny_covid_api(), database cursor and connector as inputs.
    ii) Returns nothing. Creates a new table called NY_COVID_Data
f) def set_up_ny_covid(data, cur, conn, start):
    i) Takes in COVID-19 JSON formatted data from ny_covid_api(), database cursor and connector, and start value (integer) as inputs. Returns nothing.
    ii) Indexes date, total cases from New York COVID-19 data into NY_COVID_Data table.
    iii) Uses start value to limit amount of data to 25 collected/stored at a time
g) def join_tables(cur,conn):
    i) Takes in database cursor and connector as inputs.
    ii) Uses JOIN to return a list of tuples in the format (NY cases, date, National cases) where date is the same.
h) def calculate_ny_nat_cases(lst_of_tups):
    i) Takes in a list of tuples from join_tables() and calculates the percentage of national COVID-19 cases that were identified in New York on a given day.
i) def calculate_national_avg (cur,conn):
    i) Takes in database cursor and connector as inputs. Uses SELECT avg to calculate the average number of cases from Covid_Data and returns the national average

j) def calculate_national_deaths(cur,conn):
    i) Takes in database cursor and connector as inputs. Uses SELECT avg to calculate the average amount of deaths from Covid_Data and returns the national average
k) def calculate_ny_avg (cur,conn):
    i) Takes in database cursor and connector as inputs. Uses SELECT avg to calculate the average number of cases from NY_COVID_DATA
    ii) Returns the NY average
l) def calculate_ny_deaths(cur,conn):
    i) Takes in database cursor and connector as inputs. Uses SELECT avg to calculate the average amount of deaths from NY_COVID_DATA.
    ii) Returns the NY average
m) def write_diff_to_file(filename, lst_of_tups):
    i) 'Takes in filename (string) and a list of tuples from join_tables().
    ii) Returns text file ('difference.txt') that writes the difference value of NY/National COVID-19 cases for specified 100 days.'"
n) def write_calculation_to_file(filename, lst_of_tups, percent_lst):
    i) Takes in filename (string), list of tuples from join_tables(), and list of calculated percentages from calculate_ny_nat_cases().
    ii) Returns a text file ('calculations.txt') that writes the percentage value of NY/National COVID-19 cases for specified 100 days.

Bitcoin Data:

o) def bitcoin_api():
    i) Takes in no inputs. Uses requests module to get historical daily Bitcoin data from CoinPaprika API.
    ii) Returns Bitcoin data for date range in JSON formatting.
p) def bitcoin_table(data2, cur, conn):
    i) Takes in Bitcoin JSON formatted data from bitcoin_api(), database cursor and connector as inputs.
    ii) Returns nothing. Creates a new table called Bitcoin_Data.
q) def set_up_bitcoin(data2, cur, conn, start):

i) Takes in Bitcoin JSON formatted data from bitcoin_api(), database cursor and connector, and a start value (integer) as inputs. Returns nothing.
ii) Indexes date, open, high, low, and closing values from Bitcoin data into Bitcoin_Data table.
iii) Uses start value to limit amount of data to 25 collected/stored at a time

Visualizations/Graphs:

r) def bitcoin_graph(lst_of_tups, cur, conn):
   i) Takes in a list of tuples containing the number of NY Covid-19 cases, date, and National Covid-19 cases, and the database cursor and connector as inputs
   ii) Uses matplotlib to return a scatterplot of the Bitcoin prices ($) to National Covid cases (millions) across the specified date range.

s) def create_percent_bar(lst_of_tups, percent_lst):
   i) Takes in a list of tuples containing the number of NY Covid-19 cases, date, and National Covid-19 cases, a list of the percentage of NY Covid cases that made up National Covid case and the database cursor and connector as inputs
   ii) Returns a bar chart comparing the % of NY Covid-19 cases to National Covid-19 cases

t) def create_stacked_bar(lst_of_tups):
   i) Takes in a list of tuples containing the number of NY Covid-19 cases, date, and National Covid-19 cases, and the database cursor and connector as inputs.
   ii) Returns a stacked bar chart comparing raw numbers of NY Covid-19 cases (thousands) to Non-NY Covid-19 cases

u) def create_pie_chart_avg():
   i) Takes in no inputs.
   ii) Uses average number of NY and National Covid-19 cases calculated in calculate_national_avg() and calculate_ny_avg() to return a pie chart comparing the average cases in NY and nationally over the selected time frame.

v) def create_pie_chart_deaths():
   i) Takes in no inputs.
   ii) Uses average number of NY and National Covid-19 related deaths calculated in calculate_national_deaths()and calculate_ny_deaths() to return a pie chart

comparing the average death rates in NY and nationally over the selected time frame.

Main:

w) def main ():

i) Takes in no inputs and returns nothing

ii) Calls the covid_api(), covid_table(), set_up_covid(), ny_covid_api(), ny_covid_table(),set_up_ny_covid(), join_tables(), bitcoin_api(), bitcoin_table(), and set_up_bitcoin(), set_up_calculations(),write_calculation_to_file(),write_diff_to_file(), create_percent_bar(), create_stacked_bar(),bitcoin_graph(), calculate_national_avg(),calculate_national_deaths(),calculate_ny_avg (), calculate_ny_deaths(),create_pie_chart_avg(),create_pie_chart_deaths(),functions .

iii) Limits the amount of data collected/stored to 25 at a time

Resources Used:

Date	**Issue Description**	**Location of Resource**	**Result (did it resolve the issue?)**
12/2	How to use COVID-19 API	https://covidtracking.com/data/api/version-2	Yes, allowed us to utilize this API to get historic national COVID-19 numbers as well as NY specific numbers.
12/2	How to use Bitcoin API	https://api.coinpaprika.com/#tag/Coins/paths/~1coins~1{coin_id}~1ohlcv~1historical/get	Yes, allowed us to utilize this API to get Bitcoin prices across the same date range that we used in the COVID-19 API, which helped us make our calculations and visualizations.
12/3	Wasn't sure how to limit data rows to 25 without doing a loop	https://piazza.com/class/ksgl012i3yw3lz?cid=331	Yes, a student replied with an idea of how to get started (i.e. set

	in the function rather than executing the code as whole 4 times		an else-if for the number of current rows in the database, and Prof. Ericson told us to use INSERT OR IGNORE, which helped us significantly in implementing the function.
12/4	Unsure how to use JOIN function to grab information from multiple databases	https://www.w3schools.com/sql/sql_join_left.asp	Yes, implemented code for the LEFT JOIN function to grab dates, national COVID cases, and NY cases from their respective databases.
12/5	Unsure how to write sentences out to csv file for each of the 100 calculations	https://www.w3schools.com/python/ref_string_format.asp	Yes, used the format() method to insert date and calculations into string formatting.
12/5	Numbers written out to csv files included many decimal places. We wanted to reduce these to make the numbers look cleaner	https://realpython.com/python-rounding/	Yes, used the round() function to reduce decimal places to 2.
12/6	Wanted to get more ideas on what visualizations to use with Matplotlib	https://matplotlib.org/	Yes, gave us all the possible graphs to make with Matplotlib, which gave us a general scope to work towards
12/8	Didn't know how to create scatterplot with ticker ranges	https://stackoverflow.com/questions/12608788/changing-the-tick-frequency-on-x-or-y-axis-in-matplotlib/12608937	Yes, we implemented code that would reduce large numbers to understandable graphs.
12/9	Didn't know how to	https://www.geeksfor	Yes, helped format a

	create a stacked bar chart using matplotlib	geeks.org/create-a-stacked-bar-plot-in-matplotlib/	bar chart to place NY COVID cases at the bottom of the stack, and non-NY cases at the top.
12/9	X-axis labels (dates) on stacked bar plot overlapped with each other, making it hard to interpret the plot	https://stackoverflow.com/questions/10998621/rotate-axis-text-in-python-matplotlib	Yes, implemented the code to rotate the x-axis labels 45 degrees.
12/9	Rotated X-axis labels were easier to read, but we were still looking to make the text smaller	https://www.delftstack.com/howto/matplotlib/how-to-set-tick-labels-font-size-in-matplotlib/#plt.xticksfontsize-to-set-matplotlib-tick-labels-font-size	Yes, implemented the code to decrease the font size to 6.
12/11	Didn't know how to create pie charts in general with the data collected	https://matplotlib.org/stable/gallery/pie_and_polar_charts/pie_features.html	Yes, implemented select aspects of the code to create a pie chart
12/11	Didn't know how to display legend + % directly onto pie chart	https://www.w3schools.com/python/matplotlib_pie_charts.asp	Yes, implemented code to use auto% function and legend function
12/12	Didn't know how to make the pie charts more visually appealing or to have more clarity in presenting data	https://www.w3schools.com/colors/colors_names.asp	Yes, implemented selected colors to create pie charts with emphasized data