COVID-19 Spread Assessment in New York Counties

Prepared by: Tasneem Yousuf

Problem Statement

- With the spread of corona virus around the world, a massive surge in the number of cases is also being reported in the state of New York.
- New York is the most populous city in the United States of America (USA).
- The state of New York consists of 62 counties.
- There are different sorts of places and venues situated and operational in the NY state with various eateries, cafes, shopping malls, night life, and transportation stations.
- The virus is bound to spread among the people socializing and continuing with their daily dine out, transportation, and night life activities.

Background

- Corona virus disease (COVID-19) is an infectious disease caused by a new virus.
- The disease causes respiratory illness (like the flu) with symptoms such as a cough, fever, and in more severe cases, leads to pneumonia and breathing difficulties.
- This disease spreads through contact with an infected person, or a surface or object that has the virus on it.

Motivation and Project Scope

- This study analyses the lifestyle with regards to choice of restaurants, travel, and nightlife preferences of people impacted by the disease.
- This project aims to find how night life and different travel and transportation venues across counties is linked to the spread of COVID-19 in the state of New York.
- In addition, it also aims to analyse the extent to which different kinds of food joints are linked to the number of deaths in the New York.

Significance

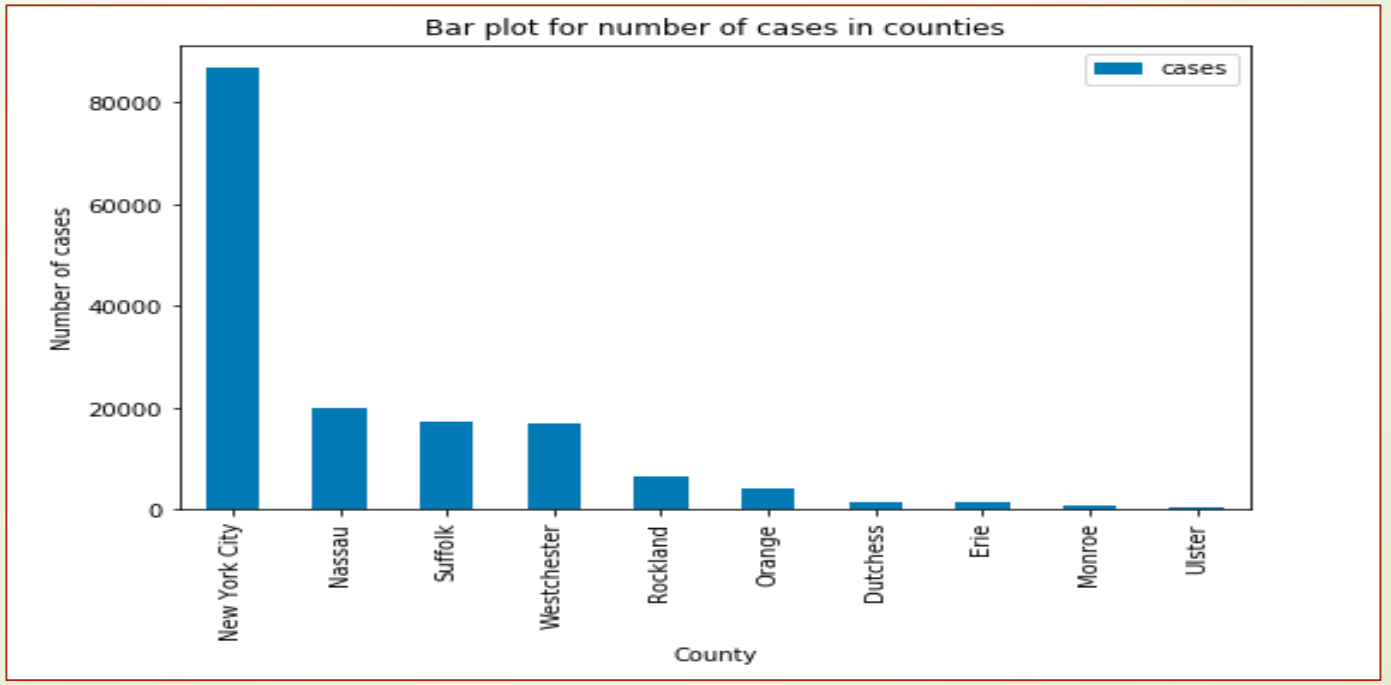
- A large number of COVID-19 data are being collected around the world in these times.
- So, it is important to make sense of this data at deeper levels so to enable people to make better lifestyle choices.
- This kind of study aims at promoting healthy choices when hopefully the world comes out of this pandemic.

Data

- COVID-19
  - https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv
  - Attributes: date, county, state, fips, cases, deaths
- 2020 population data of New York state counties
  - https://worldpopulationreview.com/us-counties/ny/
  - Attributes: County name, population in 2020, Growth since 2010

Extract, Transform, and Load (ETL)

- The dataframe was filtered to contain only counties of New York.
- When sorted, it was found that the following counties had the most number of cases as reported on 9th April 2020.



Normalization

- The number of deaths and cases in a particular county were taken as a percentage as follows:

$$Percentage\ of\ Deaths = \frac{Number\ of\ deaths}{Number\ of\ cases} * 100$$

$$Percentage\ of\ Cases = \frac{Number\ of\ cases}{2020\ Population\ of\ county} * 100$$

Methodology

- The FourSquare endpoint ***search*** was used to formulate two URLs to access venues in the categrory of food, and travel and night life.

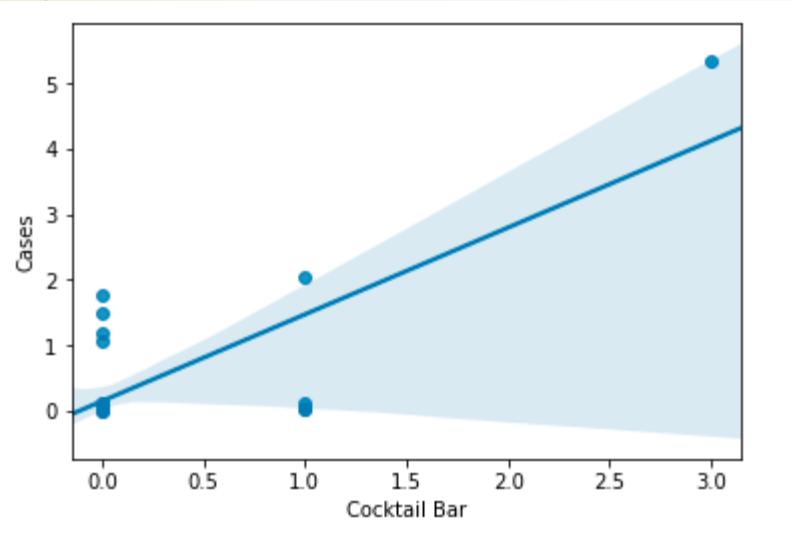
URL parameter	Value
categoryIdFood	'4d4b7105d754a06374d81259'
categoryIdTravel	'4d4b7105d754a06379d81259'
categoryIdNightLife	'4d4b7105d754a06376d81259'
LIMIT	50
Radius	750
Version	'20200412'
Near	<County name, NY>
Intent	'browse'

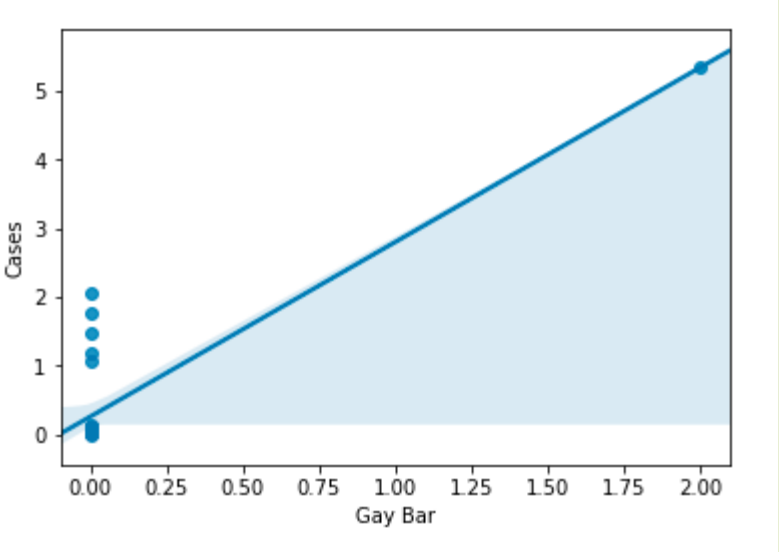
Methodology

- The resulting JSON format was studied and all venues were extracted in a dataframe which was then grouped for each county to contain the sum of venue categories in each county. The final transformed dataframe is shown below:

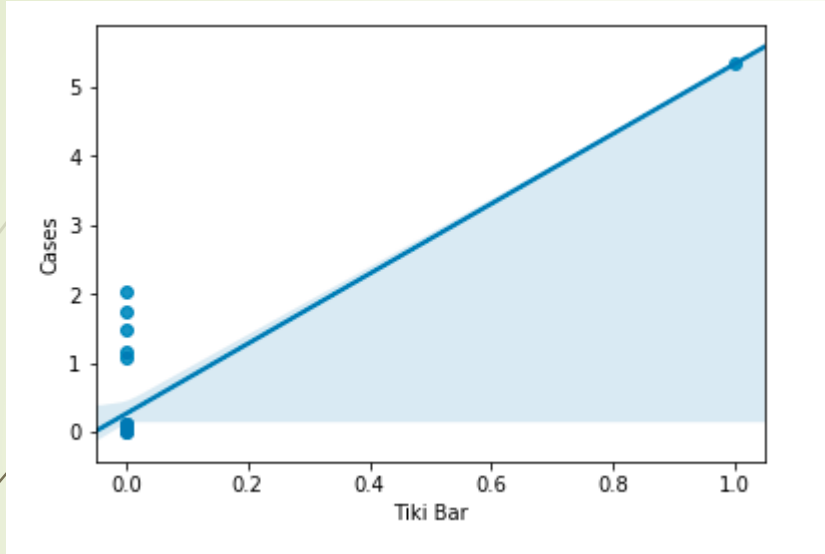
	County	Cases	Deaths	American Restaurant	Argentinian Restaurant	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bar
0	Albany	0.123406	3.166227	4	0	0	0	0	0	0
1	Allegany	0.047383	0.000000	2	0	0	0	0	1	1
2	Cattaraugus	0.022124	0.000000	1	0	0	0	0	0	0
3	Chautauqua	0.014069	5.555556	2	0	0	0	0	0	0
4	Clinton	0.049569	0.000000	2	0	0	0	0	0	0

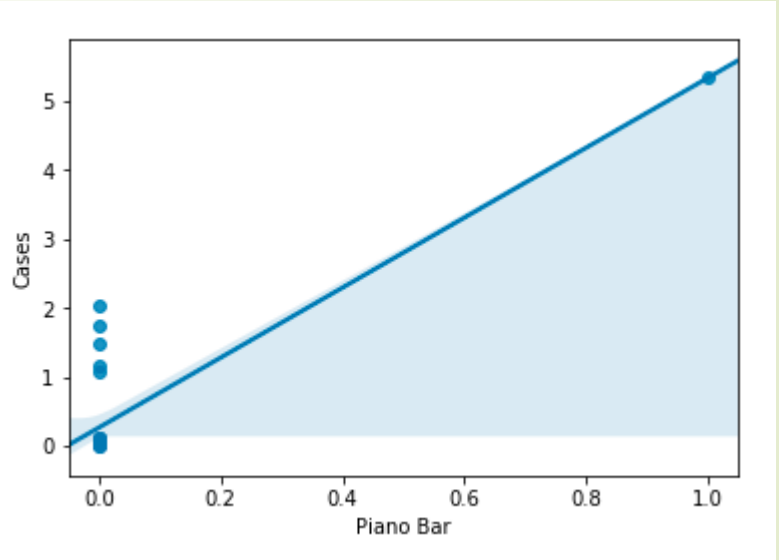
Results

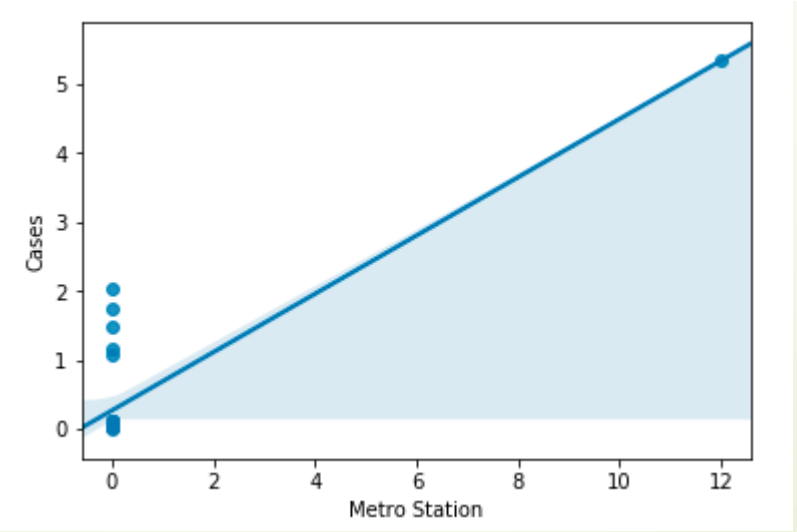

Cases
Cocktail Bar
0
1
2
3
4
5
0.0
0.5
1.0
1.5
2.0
2.5
3.0


Cases
Gay Bar
0
1
2
3
4
5
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

Results


Cases
Tiki Bar
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0


Cases
Piano Bar
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0


Cases
Metro Station
0
1
2
3
4
5
0
2
4
6
8
10
12

Observation

- The numbers of Cocktail bar, Gay Bar, Metro Station, Piano Bar, and Tiki Bar were found to be positively correlated with the number of COVID-19 cases in the respective county according to the obtained Pearson coefficient and p-values.

Conclusion

- In this project, number and abundance of various kinds of venues in different counties of New York was studied in relationship to the number of cases and deaths due to COVID-19.
- The exploratory data analysis technique of correlation was performed for this research and the resulting relationships were visualized by Scatter plots, and determined by Pearson coefficients and p-values.
- Some popular night life venues and Metro Station transportation venues were observed to be the cause of a plethora of cases in the state of New York.