The Battle of the Neighborhoods
Covid 19 Era

INTRODUCTION

- New York city review for the Covid 19 spread
- Location that require more Covid 19 testing

Business Problem

- It is evident that to survive and curb the various more testing must be done, and this study aids to shows the areas with more cases
- it is very important to strategically plan how to distribute testing kits to areas gravely infected

Success Criteria

- Boroughs that require more testing

NEW YORK-FACTS

- New York is the most populous city in the United States
- It is diverse and is the financial capital of USA
- It is a global hub of business and commerce
- In 2019, New York had its first case of coronavirus disease 2019 (Covid-19)
- In 2019, New York emerged as the United States major epicenter

FACTORS INFLUNCING THE TESTING QUANTITY

- Amount of infected citizens
- New York City Boroughs
- New York areas and zip codes
- Segmentation of the Borough

METHODOLOGY

The Data Science Workflow

- **Outline the initial data that is required:**
  - District data for New York including names, location data if available, and any other details required.
- **Obtain the Data:**
  - Research and find suitable sources for the district data for New York.
  - Access and explore the data to determine if it can be manipulated for our purposes.
- **Initial Data Wrangling and Cleaning:**
  - Clean the data and convert to a useable form as a data frame
- **Data Analysis and Location Data:**
  - Data manipulation and analysis to derive subsets of the initial data.
  - Identifying the high traffic areas using data visualisation and statistical analysis.
- **Visualization:**
  - Analysis and plotting visualizations.
  - Data visualization using various mapping libraries.

DATA DESCRIPTION

We will be using the below datasets for analysing New York city

- **Data 1:** Neighborhood has a total of 5 boroughs and 306 neighborhoods. To segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

- This dataset exists for free on the web. Link to the dataset is: https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm

	Borough	Neighborhood	ZIP Codes
0	Bronx	Central Bronx	10453, 10457, 10460
1	Bronx	Bronx Park and Fordham	10458, 10467, 10468
2	Bronx	High Bridge and Morrisania	10451, 10452, 10456
3	Bronx	Hunts Point and Mott Haven	10454, 10455, 10459, 10474
4	Bronx	Kingsbridge and Riverdale	10463, 10471

DATA DESCRIPTION

We will be using the below datasets for analysing New York city

- **Data 2:** New York city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in New York City. The below is image of the Foursquare API data.

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

DATA DESCRIPTION

We will be using the below datasets for analysing New York city

- ***Data 3:*** Second data which will be used is the data from NYC Health.

Website- https://github.com/nychealth/coronavirus-data/blob/master/boro.csv

BOROUGH_GROUP	COVID_CASE_COUNT	COVID_CASE_RATE
The Bronx	40148	2728.8
Brooklyn	46977	1732.54
Manhattan	21862	1164.06
Queens	54558	2183.62
Staten Island	12452	2477.13
Citywide	176086	

ANALYTIC APPROACH

- New York city neighbourhood has a total of 5 boroughs and 306 neighborhoods. In this project first part is clustering of Manhattan and Brooklyn. And second part is clustering of Bronx, Queens, and Staten Island. This is done because of the following Exploratory data analysis

RESULTS

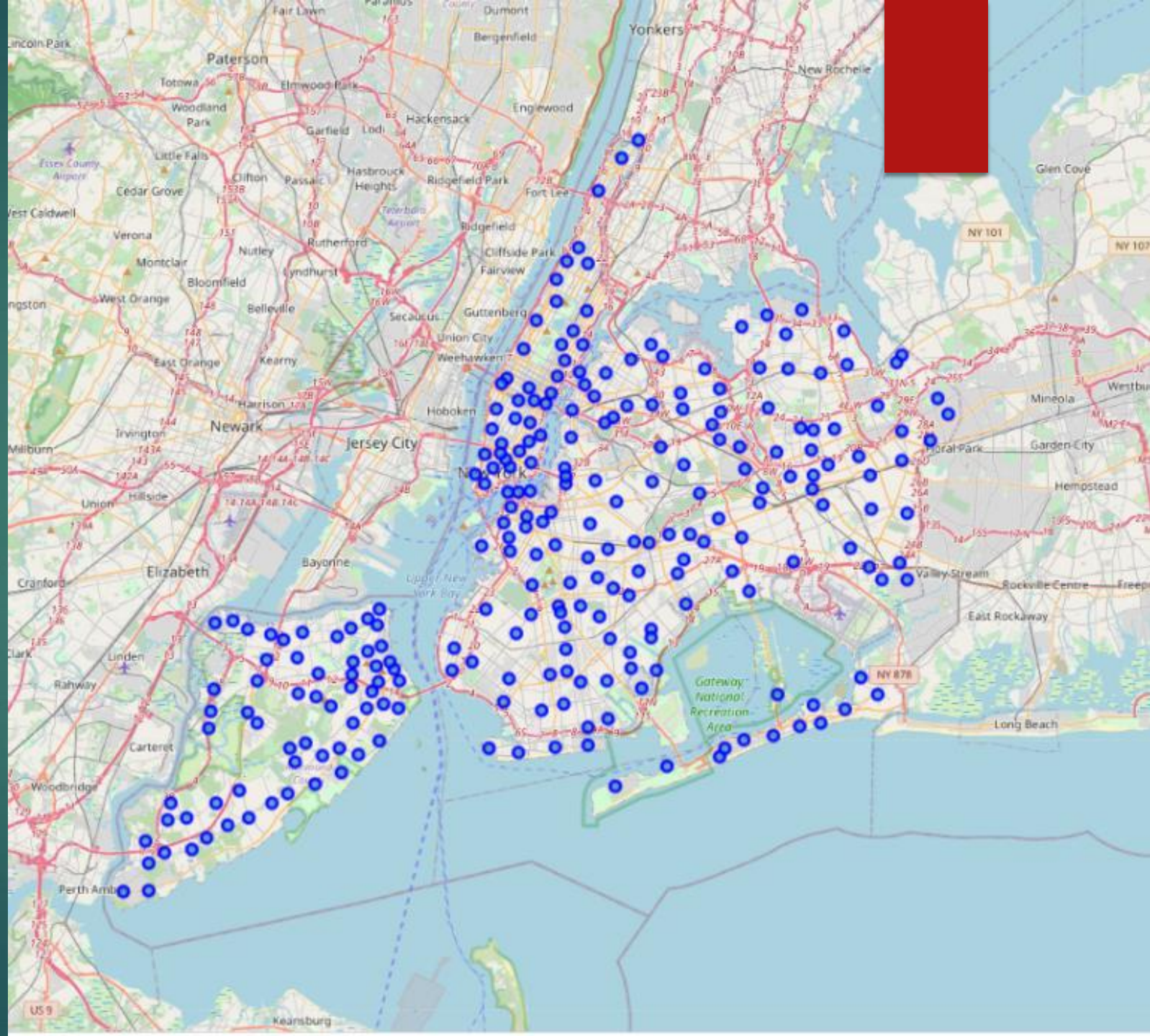
- From this data we filtered and used only the data that aligns with the Covid report eliminating the Zip codes clustering. As we focused only on areas with high volumes.

Neighborhood K-Means clustering based on mean occurrence of category

- To cluster the neighborhoods into two clusters we used the K-Means clustering Algorithm. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. It uses iterative refinement approach.

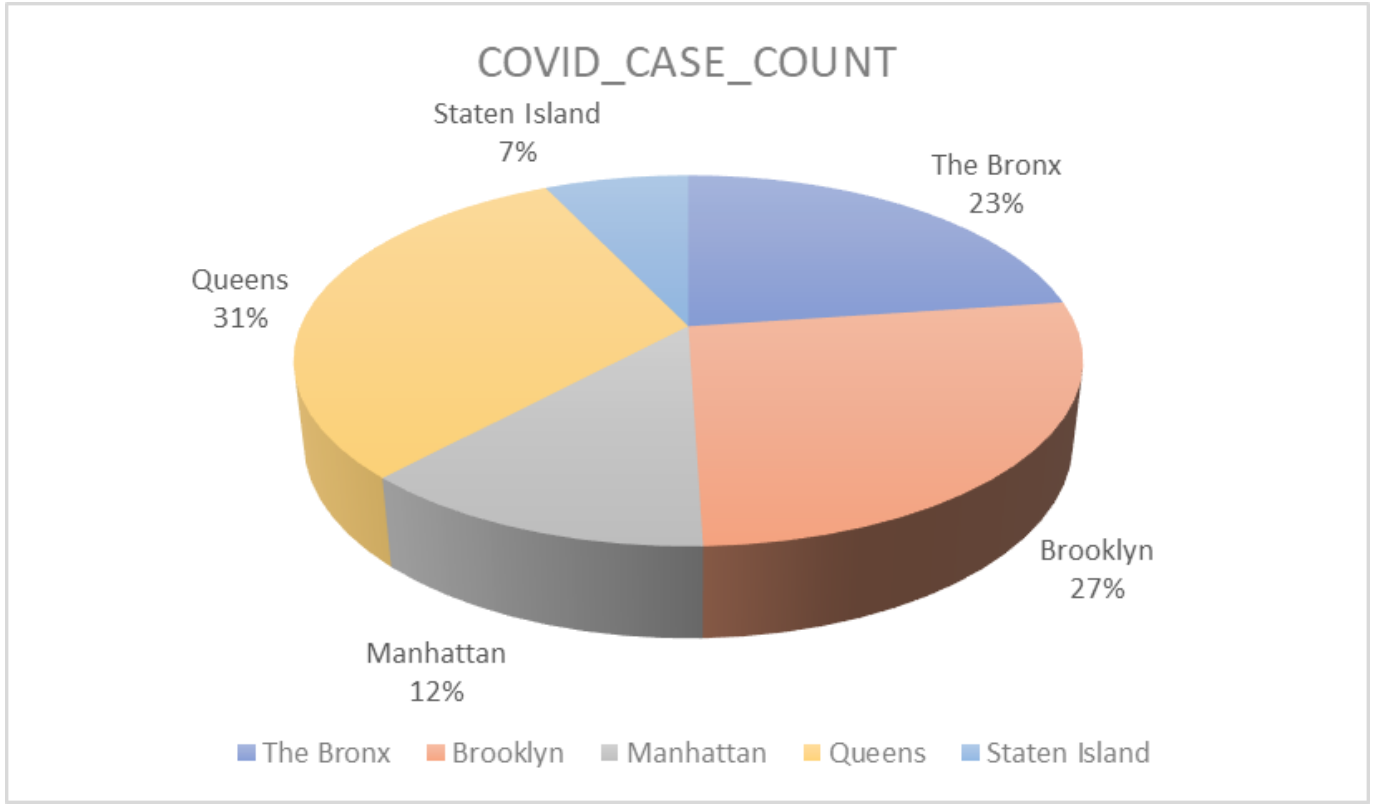
RESULTS-CONT'D

New York neighborhood visualization showing areas affected by Covid 19



RESULTS-CONT'D

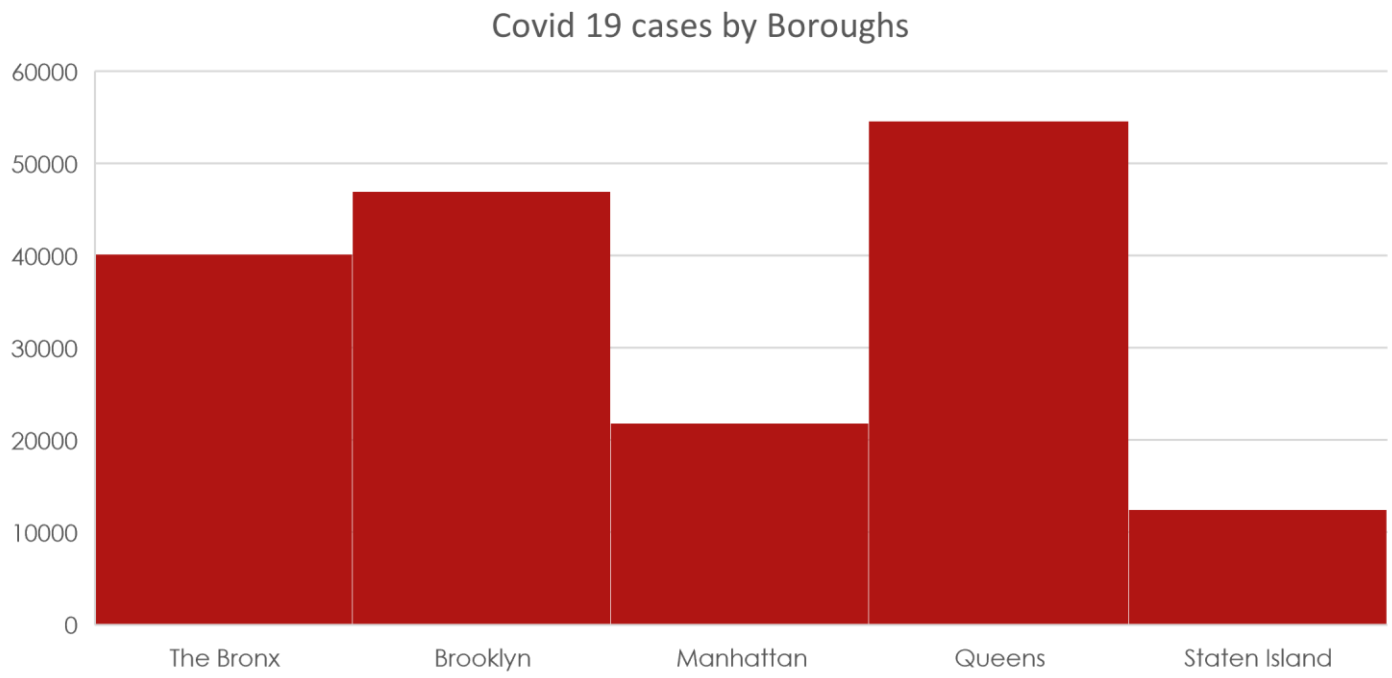
Visualization showing percentage affected by Covid 19



RESULTS-CON'D

Covid 19 cases by Boroughs

- This chart shows the number of positive cases COVID-19 on a daily basis since March 3



CONCLUSION

- Analysis is performed on limited data
- Re-run program with updated information
- Sufficient amount of data is required to come up with better results
- Queens, Brooklyn and The Bronx has high concentration of infected persons.
- Bronx, Queens and Staten Island also has had a substantial amount of testing but not as many as required