Capstone Project
The Battle of Neighborhoods

By - Jerry Jing

Overview

- New residents coming to NYC often have a hard time deciding on the locations that they prefer to stay.

- By Clustering the neighborhoods, we should be able to find out some relationship between safeness, properties selling price and type of places people more likely to visit.

- New residents moving to NYC will be able to know not only the safeness of the neighborhoods and the price range of properties but also what kind of places they could visit in different neighborhoods.

Dataset

Crimes dataset consists of 24 columns, 7 of them were pre-selected for this project:

- **CMPLNT_NUM**: The unique ID of the criminal offence.
- **CMPLNT_FR_DT**: Date of crimes reported.
- **OFNS_DESC**: Offence Descriptions.
- **LAW_CAT_CD**: Offence Level.
- **BORO_NM**: Borough, where crime was committed.
- **Latitude**: Latitude of the crime scene.
- **Longitude**: Longitude of the crime scene.

There are 22 columns in the NYC Property Sales dataset, 2 of them will be used for this study:

- **BOROUGH**: Borough, where the sales occurred.
- **SALE PRICE**: The price of the property sold.

Coordinates of NYC neighborhoods will be extracted from the 2014 NYC Neighborhoods Names geojson file.
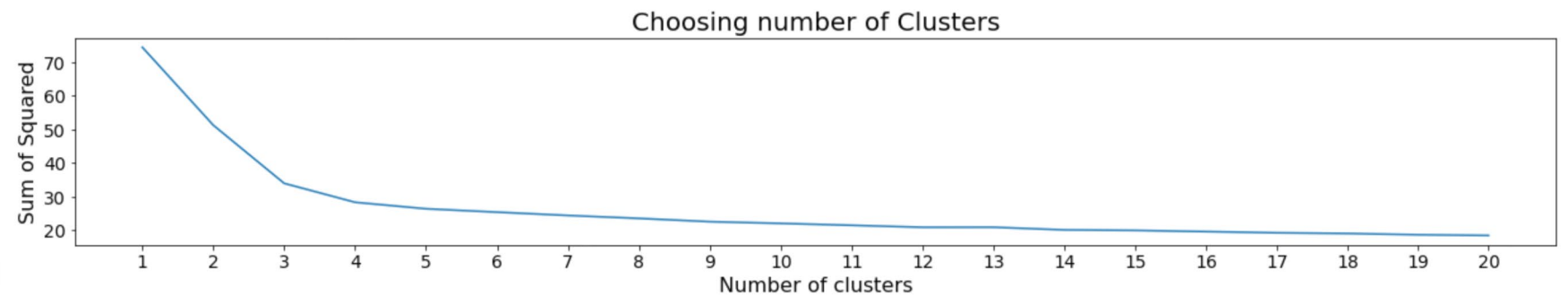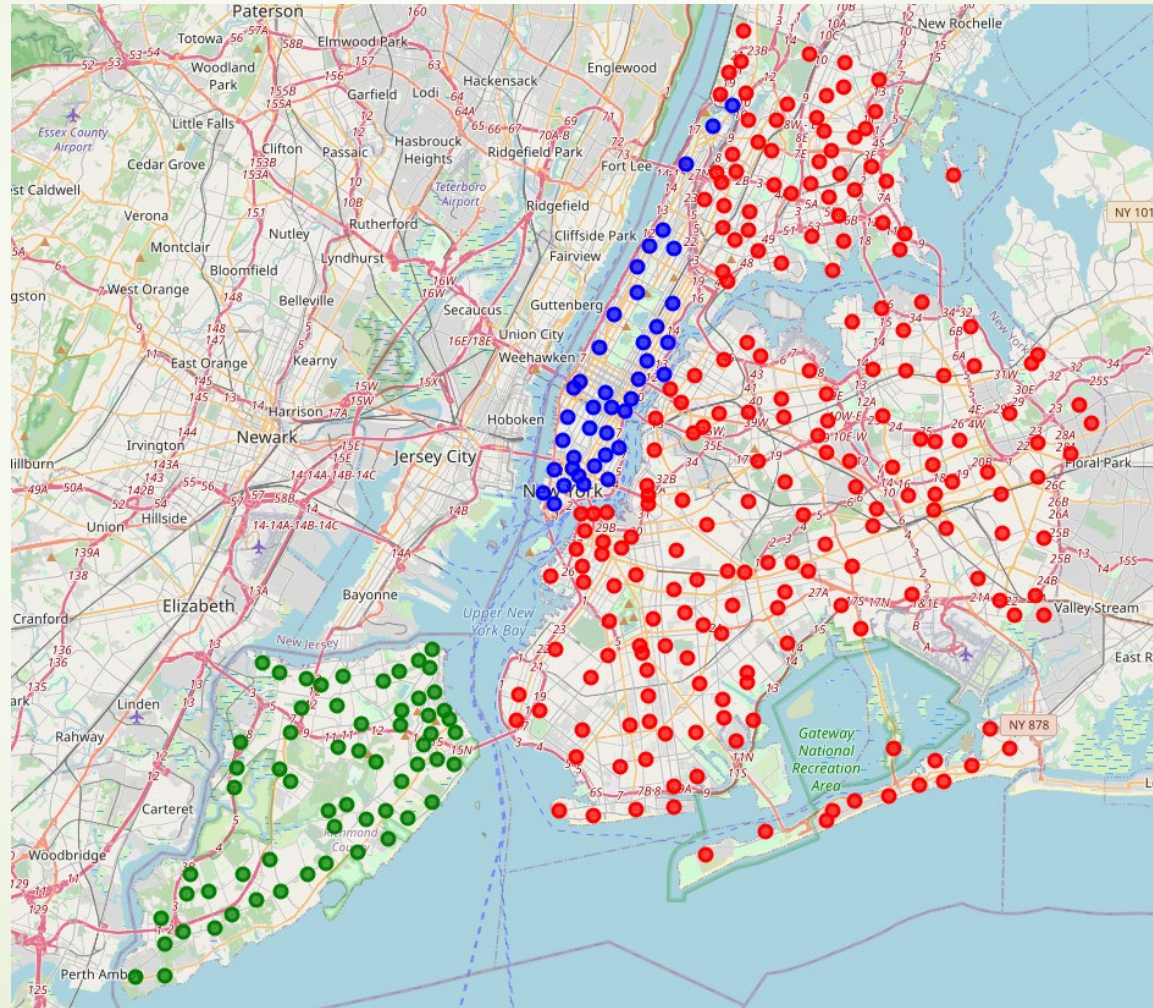
Methodology

1. Datasets processing and wrangling.

2. Coordinates from each neighborhood will be used to explore nearby popular venues visited by Foursquare users using Foursquare API.

3. Utilize the popular venues gotten from Foursquare, we can group the neighborhoods with their top 10 most commonly visited venues.

4. Determine the number of clusters using Sum of Squared Errors (SSE) and apply K-Means Clustering.

Methodology

	Borough	Safety Score	Safety Rank	Property Price	Property Price Rank	Neighborhood	Latitude	Longitude
0	Bronx	0.720667	3	0.243970	3	Wakefield - Bronx	40.894705	-73.847201
1	Bronx	0.720667	3	0.243970	3	Co-op City - Bronx	40.874294	-73.829939
2	Bronx	0.720667	3	0.243970	3	Eastchester - Bronx	40.887556	-73.827806
3	Bronx	0.720667	3	0.243970	3	Fieldston - Bronx	40.895437	-73.905643
4	Bronx	0.720667	3	0.243970	3	Riverdale - Bronx	40.890834	-73.912585
...	...	...	...	...	...	...	...	...
301	Manhattan	0.775386	2	1.000000	1	Hudson Yards - Manhattan	40.756658	-74.000111
302	Queens	0.671501	4	0.222888	4	Hammels - Queens	40.587338	-73.805530
303	Queens	0.671501	4	0.222888	4	Bayswater - Queens	40.611322	-73.765968
304	Queens							
305	Staten Island							

306 rows ×

Result & Discussion

Result & Discussion

Based on visualization on NYC map, the 3 clusters are:

1. Around Manhattan Area,
2. Around Bronx, Queens and Brooklyn Area,
3. Around Staten Island Areas.

Staten Island is relatively cheaper and safer compared to other boroughs. Also, the neighborhoods have many different types of restaurants and Bodegas which is pretty convenient for nearby residents.

Conclusion

This study explores the suitable location for new residents moving to NYC, each new resident will have different requirements for choosing a location to stay. Thus, this study can be utilized by new residents coming to NYC to have an idea of the neighborhoods here.