Capstone Project

The Battle of Neighborhoods

Name: Jerry Jing

Table of Contents

Introduction

Background

New York City (NYC) is one of the most important financial centers in the world, it's the most populous city in the United States (U.S.) with an estimated population of 8,398,748 in the year 2018 and covers 784 km².

NYC cosmists of 5 boroughs and its core is at Manhattan where all the world's major commercial, financial and cultural centers are located.

Problem

This study will rank Boroughs according to its total crime rates. Using the top 10 visited categories of venues (using Foursquare API) from each neighborhood, safety score and sales price of properties to find clusters using K-Means Clustering.

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.

Target Audience

New residents moving to NYC would benefit from the results from this study.

New residents moving to NYC seeking to stay in Manhattan 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. On top of that, it could also be a deciding factor for them to choose the area they wish to stay.

Dataset

Data Source

The New York City Crimes dataset can be obtained from Kaggle. The dataset contains all criminal offences recorded by the NYPD from 2014 to 2015.

NYC Property Sales dataset can also be found in Kaggle. It contains data on Properties sold in New York City over a 12-month period from September 2016 to September 2017.

2014 New York City Neighborhood Names geojson file from NYU Libraries. Dataset contains locational information for 306 neighborhoods from NYC.

Data Wrangling

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

	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	0.671501	4	0.222888	4	Queensbridge - Queens	40.756091	-73.945631
305	Staten Island	0.154422	5	0.164314	5	Fox Hills - Staten Island	40.617311	-74.081740

306 rows × 8 columns

Figure 1

Datasets will be processed and combined into 1 dataframe with only useful features for this study, as shown by Figure 1.

(9826, 12)

	Neighborhoods	Safety Score	Safety Rank	Property Price	Property Price Rank	Borough	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield - Bronx	0.720667	3	0.24397	3	Bronx	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield - Bronx	0.720667	3	0.24397	3	Bronx	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
2	Wakefield - Bronx	0.720667	3	0.24397	3	Bronx	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
3	Wakefield - Bronx	0.720667	3	0.24397	3	Bronx	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
4	Wakefield - Bronx	0.720667	3	0.24397	3	Bronx	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop

Figure 2

Coordinates from each neighborhood will be used to explore nearby popular venues visited by Foursquare users using Foursquare API. Figure 2 shows the sample results obtained from Foursquare.

	Neighborhoods	Safety Score	Safety Rank	Property Price	Property Price Rank	No.1 Most visited Venue	No.2 Most visited Venue	No.3 Most visited Venue	No.4 Most visited Venue	No.5 Most visited Venue	No.6 Most visited Venue	No.7 Most visited Venue	No.8 Most visited Venue	No.9 Most visited Venue	No.10 Most visited Venue
0	Allerton - Bronx	0.720667	3.0	0.243970	3.0	Pizza Place	Deli / Bodega	Chinese Restaurant	Supermarket	Fast Food Restaurant	Bakery	Check Cashing Service	Mexican Restaurant	Grocery Store	Gas Station
1	Annadale - Staten Island	0.154422	5.0	0.164314	5.0	American Restaurant	Sports Bar	Deli / Bodega	Pizza Place	Diner	Pharmacy	Restaurant	Train Station	Cosmetics Shop	Pub
2	Arden Heights - Staten Island	0.154422	5.0	0.164314	5.0	Pharmacy	Coffee Shop	Pizza Place	Bus Stop	Field	Ethiopian Restaurant	Event Service	Event Space	Exhibit	Factory
3	Arlington - Staten Island	0.154422	5.0	0.164314	5.0	Deli / Bodega	American Restaurant	Boat or Ferry	Bus Stop	Grocery Store	Yoga Studio	Fish & Chips Shop	Exhibit	Factory	Falafel Restaurant
4	Arrochar - Staten Island	0.154422	5.0	0.164314	5.0	Bus Stop	Italian Restaurant	Deli / Bodega	Pizza Place	Athletics & Sports	Middle Eastern Restaurant	Food Truck	Bagel Shop	Outdoors & Recreation	Sandwich Place
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
299	Woodhaven - Queens	0.671501	4.0	0.222888	4.0	Deli / Bodega	Park	Bank	Pharmacy	Nail Salon	Dive Bar	Metro Station	Thai Restaurant	Bagel Shop	Sandwich Place
300	Woodlawn - Bronx	0.720667	3.0	0.243970	3.0	Deli / Bodega	Pizza Place	Playground	Pub	Grocery Store	Food Truck	Bakery	Liquor Store	Trail	Donut Shop
301	Woodrow - Staten Island	0.154422	5.0	0.164314	5.0	Pharmacy	Mexican Restaurant	Cosmetics Shop	Bakery	Pizza Place	Bank	Donut Shop	Coffee Shop	Liquor Store	Grocery Store
302	Woodside - Queens	0.671501	4.0	0.222888	4.0	Grocery Store	Thai Restaurant	Latin American Restaurant	Bakery	Filipino Restaurant	Donut Shop	Pub	Pizza Place	Bar	American Restaurant
303	Yorkville - Manhattan	0.775386	2.0	1.000000	1.0	Coffee Shop	Italian Restaurant	Gym	Deli / Bodega	Bar	Wine Shop	Mexican Restaurant	Sushi Restaurant	Japanese Restaurant	Diner

304 rows × 15 columns

Figure 3

Then, we utilize the popular venues gotten from Foursquare, we can group the neighborhoods with their top 10 most commonly visited venues. As shown in Figure 3.
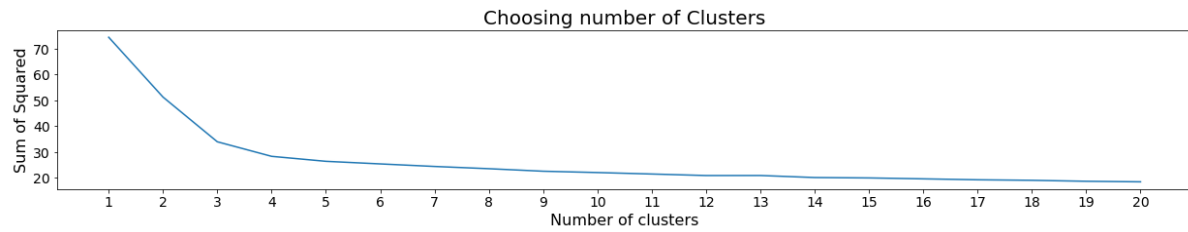
Figure 4

Next, we will determine the number of clusters using Sum of Squared Errors (SSE) and apply K-Means Clustering. Figure 4 shows a plot of SSE for choosing the number of clusters.

And finally, we will be examining the results of this study.
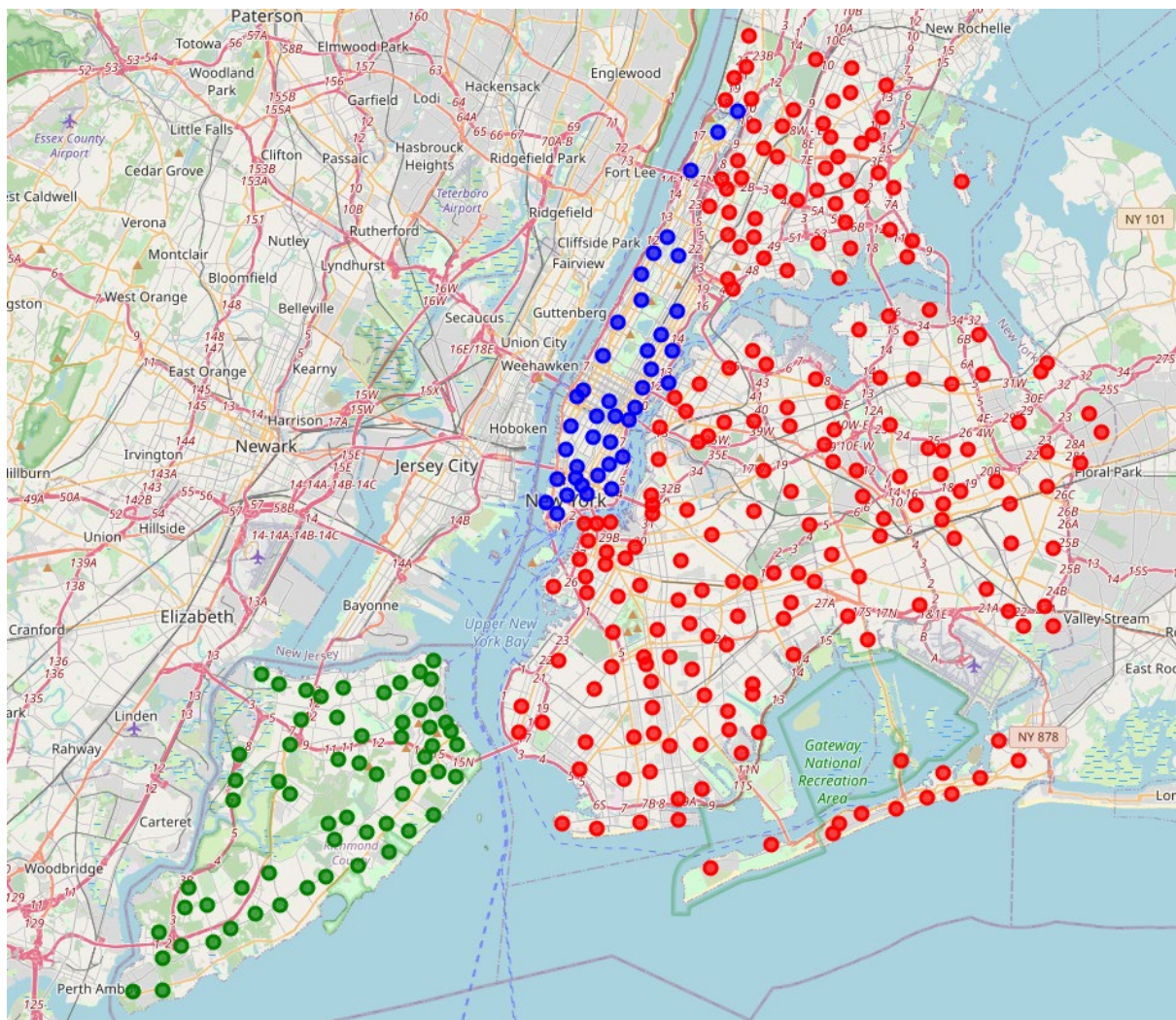
Result & Discussion

Figure 5

By visualizing the clusters of neighborhoods on an actual map with Folium, we see that 3 clusters are:

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

People in cluster 1 loves pizza, Deli/Bodega and bars are often visited, there might have high numbers of Chinese communities, also might have moderate numbers of Italian and Caribbean communities. Properties are moderate in price, higher crime rates

People in Cluster 2 loves coffee, there might be high numbers of Italian Communities, there might also be small numbers of Mexican, Chinese and Korean communities. Properties are most expensive here, also higher crime rate

Cluster 3 might have high numbers of Italian communities, people here like to visit cafes, also might have a small number of Mexican and Chinese Communities. Properties least expensive here, crime rate is lowest

For new residents coming to NYC, I would recommend them to pick their new stay in Staten Island, as it 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.