Analysis for best place to start a new restaurant in New York

Introduction

New York is one of biggest cities of the world. The city is financial capital of USA and an extremely expensive city to live and start a new business venture. People from every corner of the world are settled here for doing business and jobs. The core of this project is to suggest an optimal location to start a new restaurant. Although the success of the restaurant is highly dependent on quality and taste of the food, but accessibility is one of the main factor too. For analysis following factors need to be considered.

1. How New York is divided into different segments.
2. Population density of the areas.
3. Competitors.
4. Choices of people in a particular area.

In order to find the solution, we will use data of New York city and apply machine learning techniques to estimate some best areas.

Role

I will prepare this report as a data scientist and the report will help the client to make an initial decision for top few locations. He will have to decide it by himself that he should go to an extremely competitive area with a lot of existing restaurants and potential customers or he will opt to open it a place that has very low number of restaurants.

Data

One city will be analyzed in this project : New York City. We will be using the below datasets for analyzing New York city

Data 1 :

Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order 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 the latitude and longitude coordinates of each neighborhood. This dataset exists for free on the web. Link to the dataset is: https://geo.nyu.edu/catalog/nyu_2451_34572

Data 2 :

For the below analysis we will get data from wikipedia as given below :

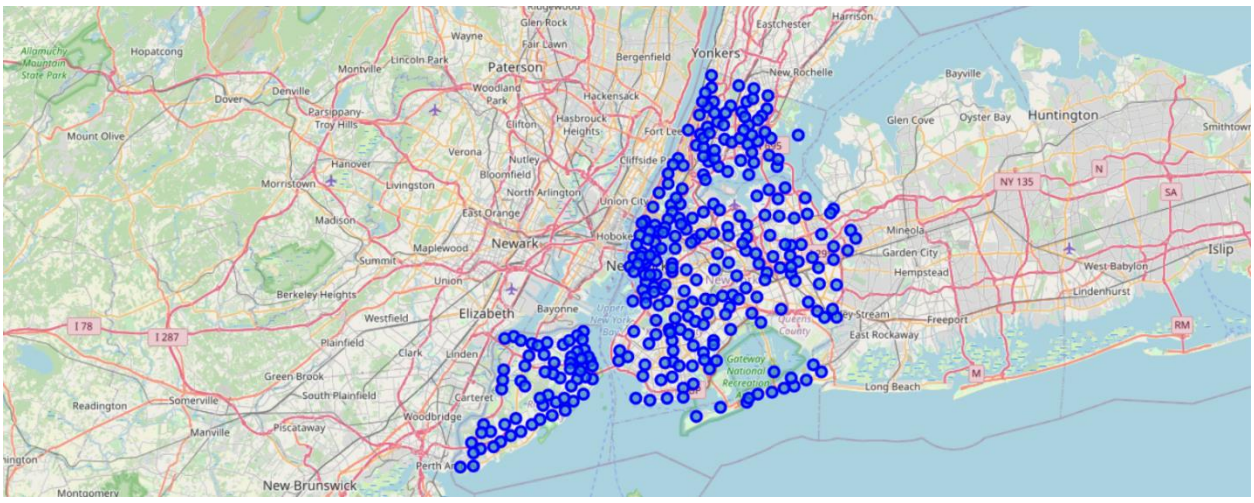
1. New York Population
2. Cuisine of New York city

[1] https://en.wikipedia.org/wiki/New_York_City

[2] https://en.wikipedia.org/wiki/Cuisine_of_New_York_City

Data 3 :

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.



Methodology 1

New York has 5 boroughs; each has different density of population and the choices of the people differ in what they like. Using the data [2] in the link, I plotted the taste of each borough using word count. Most popular food in New York city are following

1. Italian
2. Mexican
3. Indian
4. Jewish
5. Puerto Rican

Figure 1: Word Cloud of food for NEW YORK

The Bronx

Figure 2: The Bronx - Food Preferences

The Bronx is third most densely populated area of New York. It is south of Westchester County; northeast and east of Manhattan, across the Harlem River; and north of Queens, across the East River. The Bronx has a land area of 42 square miles (109 km2) and a population of 1,418,207 in 2019. Of the five boroughs, it has the fourth-largest area, fourth-highest population, and third-highest population density. Most popular food choice of the people are following.

1. Italian
2. Puerto Rican
3. Dominican
4. Alabanian

Queens

Queens is a borough of New York City, coterminous with Queens County, in the U.S. state of New York. It is the largest borough geographically and is adjacent to the borough of Brooklyn, at the western end of Long Island and Nassau County to its east. Queens also shares water borders with the

Figure 3: Queens - Food Preferences

boroughs of Manhattan, the Bronx, and Staten Island (via the Rockaways). The borough of Queens is the second largest in population. Among its residents following cuisines are mostly preferred.

1. Italian
2. Indian
3. Pakistani
4. Bangladeshi

Manhattan

Figure 4: Manhattan - Food Preferences

Manhattan, often referred to by residents of the New York City area as the City, is the most densely populated of the five boroughs of New York City, and coextensive with the County of New York, one of the original counties of the U.S. state of New York. Manhattan serves as the city's economic and administrative center, cultural identifier and historical birthplace. The borough consists mostly of Manhattan Island, bounded by the Hudson, East, and Harlem rivers; as well as several small adjacent islands. Most popular cuisines of Manhattan are as following.

1. Italian
2. Chinese
3. Puerto Rican
4. Jewish

Staten Island

Staten Island is a borough of New York City, coterminous with Richmond County, in the U.S. state of New York. Located in the southwest portion of the city, the borough is separated from New Jersey by the Arthur Kill and the Kill Van Kull and from the rest of New York by New York Bay. With an estimated population of 476,143. Staten Island is the least populated of the boroughs.

Figure 5: Staten Island - Food Preferences

1. Italian
2. Indian
3. Sri Lankan , Mexican, Polish

Brooklyn

Brooklyn is a borough of New York City, coterminous with Kings County, located in the U.S. state of New York. It is the most populous county in the state, the second-most densely populated county in the United States and New York City's most populous borough, with an estimated 2,559,903 residents. Named after the Dutch village of Breukelen, it shares a land border with the borough of Queens at the western end of Long Island. Brooklyn has several bridge and tunnel connections to the borough of Manhattan across the East River, and the Verrazzano-Narrows Bridge connects it with Staten Island. People mostly like following cuisines.

Figure 6: Brooklyn - Food Preferences

1. Italian
2. Puerto Rican
3. Dominican
4. Albanian.

Methodology 2

In order to handle the data better, the boroughs are divided into two groups.

1. Group A [Manhattan, Brooklyn]
2. Group B [The Bronx, Queens and Staten Island

The venues of restaurants are retrieved using foursquare API and then using k-means method of sklearn library, the data is divided in two clusters. Silhouette Coefficient is highest for two clusters that's the reason of using two clusters.

Results

Group A

For Manhattan and Brooklyn chase, the number of venues retrieved are 9,708 of 397 different types. Using the clustering approach for which the code is available in repository we divide the areas in two clusters that show similar in together

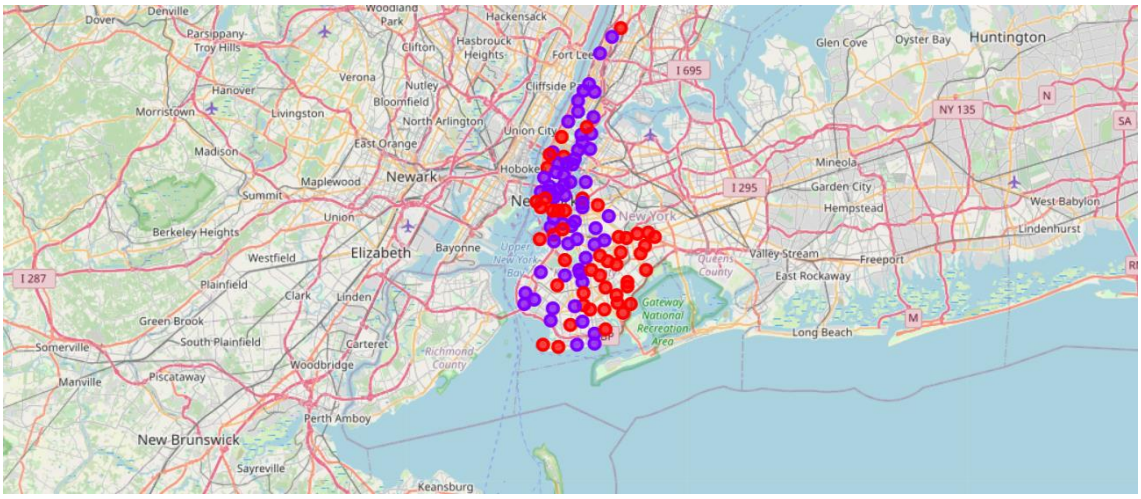


Figure 7: Clusters division for Brooklyn and Manhattan

Top 10 most saturate locations of Group A.

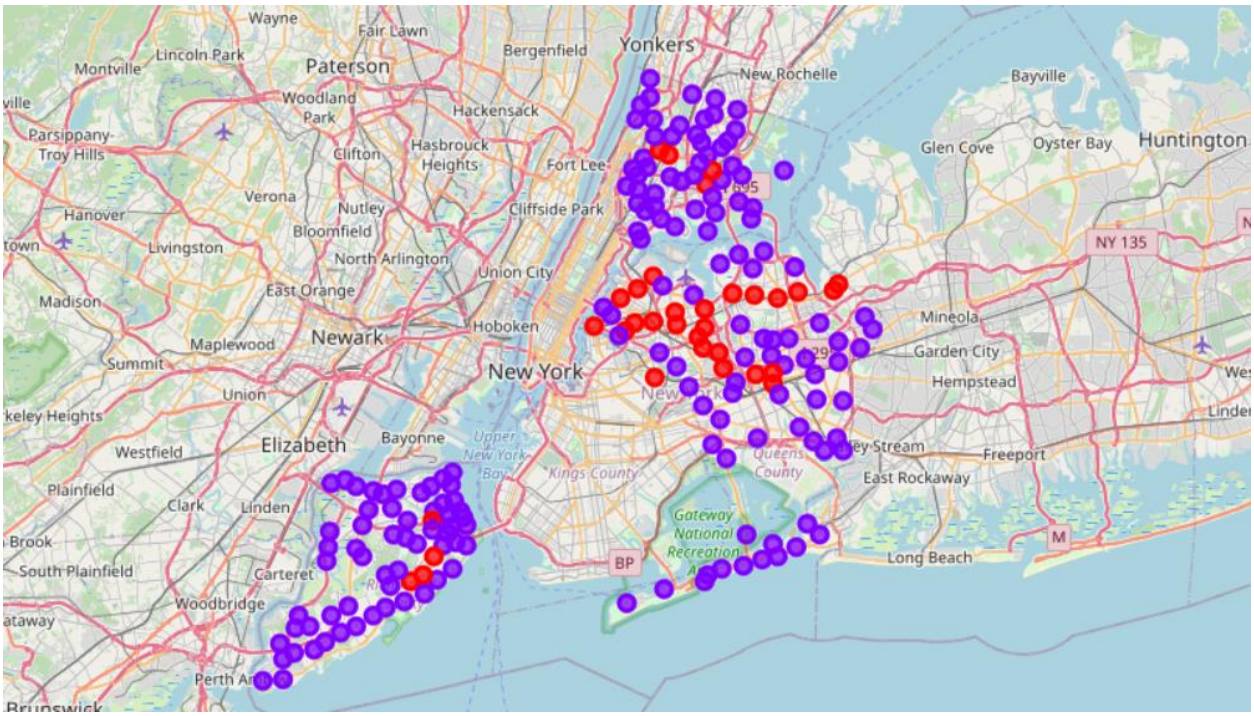
	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Brooklyn	Sunset Park	40.645103	-74.010316	39	0
1	Manhattan	Central Harlem	40.815976	-73.943211	36	0
2	Brooklyn	Prospect Park South	40.647009	-73.962613	36	0
3	Brooklyn	Bay Ridge	40.625801	-74.030621	36	0
4	Brooklyn	Ditmas Park	40.643675	-73.961013	35	0
5	Manhattan	Washington Heights	40.851903	-73.936900	34	0
6	Manhattan	Inwood	40.867684	-73.921210	34	0
7	Brooklyn	Bath Beach	40.599519	-73.998752	33	0
8	Brooklyn	Flatbush	40.636326	-73.958401	33	0
9	Manhattan	Manhattanville	40.816934	-73.957385	33	0

10 Least saturated locations of Group A.

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Brooklyn	Sea Gate	40.576375	-74.007873	2	1
1	Brooklyn	Bergen Beach	40.615150	-73.898556	3	1
2	Brooklyn	Mill Island	40.606336	-73.908186	3	1
3	Brooklyn	East New York	40.669926	-73.880699	4	1
4	Brooklyn	East Flatbush	40.641718	-73.936103	6	1
5	Brooklyn	Paerdegat Basin	40.631318	-73.902335	7	1
6	Brooklyn	Highland Park	40.681999	-73.890346	7	1
7	Brooklyn	Starrett City	40.647589	-73.879370	7	1
8	Brooklyn	Borough Park	40.633131	-73.990498	7	1
9	Manhattan	Battery Park City	40.711932	-74.016869	8	1

Group B

The group consists of The Bronx, Queens and Staten Island. The number of optimal clusters is also 2 here and then the distribution of clusters is shown as following.



Top 10 most populated areas of the Group B

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Queens	Elmhurst	40.744049	-73.881656	53	0
1	Queens	Jackson Heights	40.751981	-73.882821	50	0
2	Queens	Murray Hill	40.764126	-73.812763	45	0
3	Queens	Flushing	40.764454	-73.831773	44	0
4	Queens	Woodside	40.746349	-73.901842	43	0
5	Queens	Sunnyside Gardens	40.745652	-73.918193	39	0
6	Queens	Ravenswood	40.761705	-73.931575	37	0
7	Queens	Bayside	40.766041	-73.774274	35	0
8	Staten Island	Sunnyside	40.612760	-74.097126	33	0
9	Queens	Auburndale	40.761730	-73.791762	33	0

10 Least populated areas of the Group B

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Staten Island	Todt Hill	40.597069	-74.111329	0	1
1	Bronx	Clason Point	40.806551	-73.854144	0	1
2	Staten Island	Bloomfield	40.605779	-74.187256	0	1
3	Staten Island	Rossville	40.549404	-74.215729	0	1
4	Staten Island	Butler Manor	40.506082	-74.229504	0	1
5	Staten Island	Port Ivory	40.639683	-74.174645	0	1
6	Queens	Breezy Point	40.557401	-73.925512	1	1
7	Staten Island	Howland Hook	40.638433	-74.186223	1	1
8	Queens	Bayswater	40.611322	-73.765968	1	1
9	Queens	Roxbury	40.567376	-73.892138	1	1

Discussion

As we can see from the results, there is no area in Brooklyn, Manhattan and Queens where we don't have a restaurant. Elmhurst neighborhood in Queens has the highest number (53) of restaurants. In fact, there more than 7 neighborhoods in Queens, 4 neighborhoods in Brooklyn and 1 in Manhattan which have more than 35 restaurants.

There are 5 neighborhoods (Todt Hill, Bloomfield, Rossville, Butler Manor and Port Ivory) in Staten Island which don't have any restaurant. Clason Point in The Bronx is also in the same condition so that makes these areas extremely favorable to start a restaurant.

Conclusion

The analysis briefly covers the restaurants in the New York and give a complete idea to the investor that which areas he can target. In methodology section the word clouds also give an idea about the choice of the people in boroughs. Some further manual research should be done for economic analysis and data validation.