

Battle Of the Neighborhoods

An analysis of Manhattan Restaurant Distribution

Mohamed Naseef
Data Science Aspirant

Start a Restaurant in New York?

New York City is the most populous city in the United States and the center of the New York metropolitan area.

As of 2019, there were 27,043 restaurants in the city, up from 24,865 in 2017.

Due to the ever-increasing competition, it is inferred that to open up a new restaurant in New York is quite a challenge.

Why Not!

If we are able to open up a restaurant in a neighborhood with lesser no of high-rated restaurants, the chance to prosper would be fairly higher.

Using data science, we can identify these neighborhoods that are least risky for a new restaurant business.

Data Acquisition and Cleansing

A collection of the New York City Neighborhoods along with the coordinate information is available for free in NYU Spatial Repository as a geojson file

The relevant information geojson file is extracted into a pandas dataframe

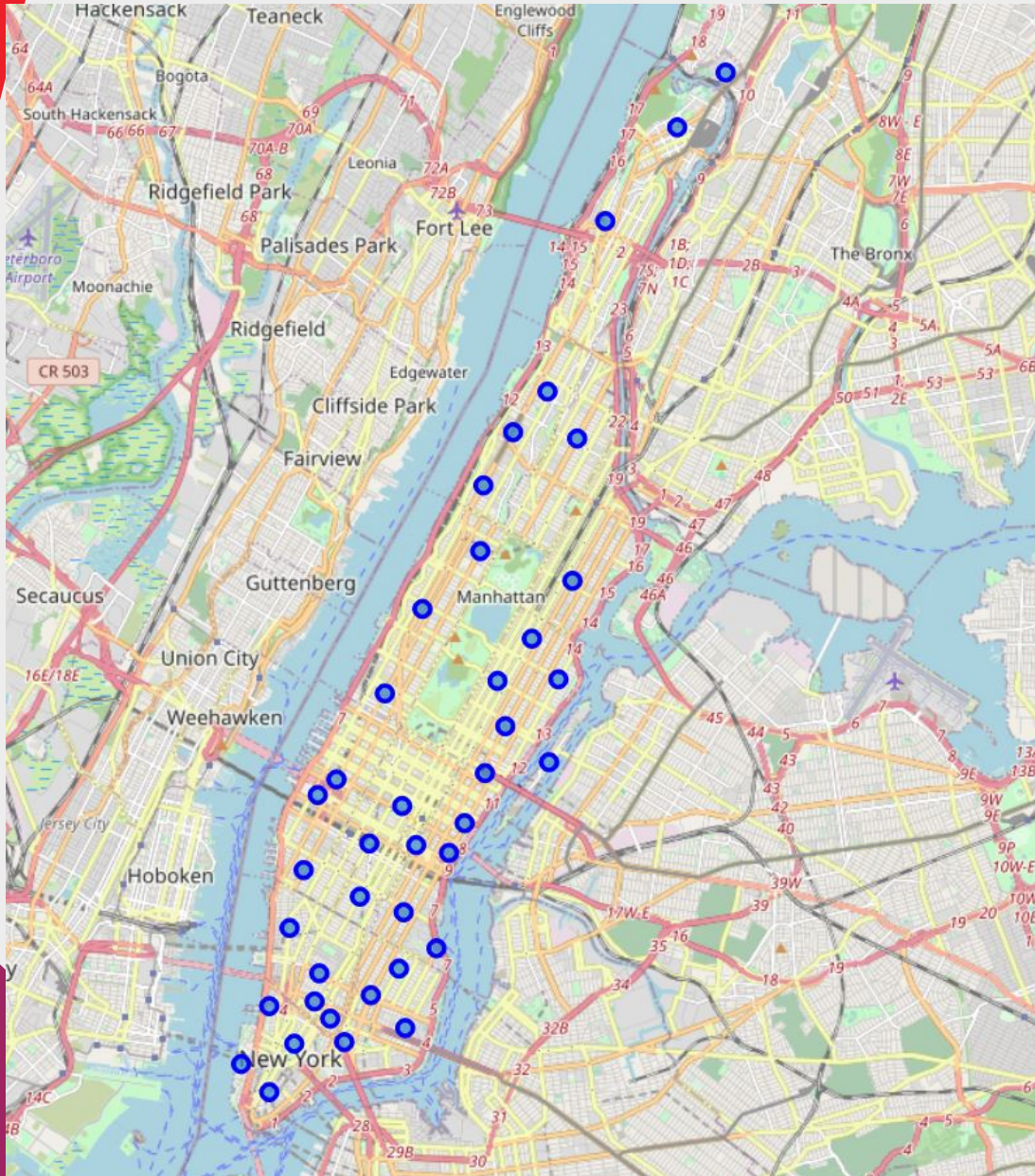
The dataframe is then filtered to contain only Manhattan neighborhoods

For the restaurants and their rating information for each neighborhood, we rely on the Foursquare API

Using this API, we define a function to return the number of restaurants, the average rating of the top-rated 5 restaurants and the average rating of the least-rated 5 restaurants

This information is added to the dataframe

Manhattan Neighborhoods



Manhattan Borough consists of 40 neighborhoods

Majority of neighborhoods are present at Southern Manhattan

Some of the neighborhoods are coastal, while some inland

K-means Clustering

K-means clustering was used to group the neighborhoods into clusters based on three features:

- Number of neighborhoods
- Average high rating
- Average low rating

The neighborhoods are grouped into 3 clusters

Cluster 1

Neighborhood	No of Restaurants	Average Top Rating	Average Low Rating
Chinatown	30	7.84	6.24
Washington Heights	30	8.04	6.92
Hamilton Heights	30	7.06	5.94
East Harlem	30	7.47	6.5

Truncated

Cluster-1, which contains majority of the neighborhoods, already have a large number of restaurants with a fairly high or average rating.

These neighborhood would prove highly risky for a new business.

Cluster 2

Neighborhood	No of Restaurants	Average Top Rating	Average Low Rating
Marble Hill	14	7.82	7.38
Roosevelt Island	0	0	0
Morningside Heights	12	7.23	6.3
Stuyvesant Town	5	6.9	6.9

Cluster-2, which includes 4 neighborhoods -Marble Hill, Roosevelt Island, Morningside Heights, Stuyvesant Town - have very few restaurants and average rating.

These neighborhoods are ideal for starting a new restaurant.

Cluster 3

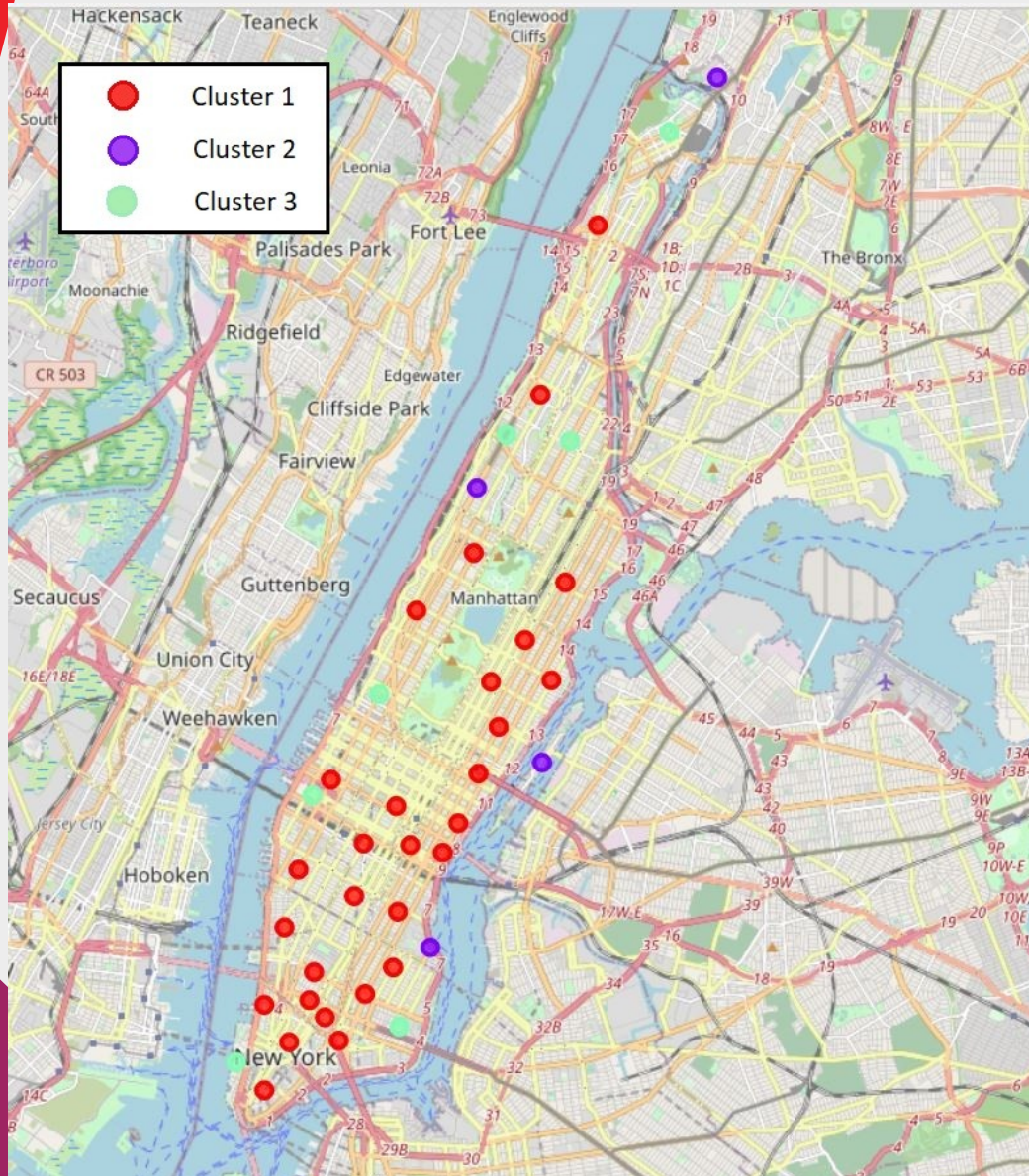
Neighborhood	No of Restaurants	Average Top Rating	Average Low Rating
Inwood	26	7.41	6.75
Manhattanville	21	7.47	6.67
Central Harlem	24	7.96	6.5
Lincoln Square	25	8.38	6.12

Truncated

Cluster-3, which contain a few neighborhoods including Manhattanville and Hudson Yards, have an average number of restaurants and a fair rating of restaurants

These are medium risk neighborhoods for starting up a restaurant.

Manhattan Neighborhood Clustered



Cluster 1
High Risk Neighborhoods

Cluster 2
Low Risk Neighborhoods

Cluster 3
Medium Risk Neighborhoods

Conclusion

A large number of neighborhoods, especially in lower Manhattan, already have a sufficient number of restaurants. Starting a new restaurant here would prove to be a high-risk.

There are a few neighborhoods with an average number of restaurants with some having low rated restaurants as well. These neighborhoods, mostly coastal, are of comparatively low risk.

However the best neighborhoods in Manhattan that have fewer restaurants and an average or low rating would be the most ideal neighborhoods to start a new restaurant. These were found to be:

- Marble Hill
- Roosevelt Island
- Morningside Heights
- Stuyvesant Town

Future Directions

Grouping into a higher number of clusters, can help identify those neighborhoods, despite having an average number of restaurants, has very low rated restaurants. These can also be considered since the presence of low rated restaurants implies room for new better restaurants.

Incorporating the cuisine of the restaurants as a feature will help in identifying what restaurant cuisine has better chance of survival in any neighborhood.

Including the population and demographic in the analysis can improve the results.