

Introduction

New York is located on the Atlantic coast in the southeast of New York state. It is the largest city and the largest port in the United States. The New York metropolitan area is one of the largest metropolitan areas in the world. It is rich in unique artistic and cultural characteristics, unmatched economic capacity and a free melting pot of all kinds of leisure and entertainment attractions. It is a dream practice place for many people.

The market competition here is fierce. Any new business venture or expansion needs careful analysis. The cost of doing business is very high. The insights from the analysis will contribute to a better understanding of the business environment and thus to strategic targeting of the market. This will help reduce the risk. And the return on investment will be reasonable.

New York's most attractive food is of course, New York's food can make you addicted to eat, eat once you will deeply love that taste, although we know a lot of New York's food:

1. Hamburg @ shakeshack
2. Hot dog @ Gray's papaya
3. Guantangbao @ lumingchun
4. Halal Food @ halal guys
5. Rougamo @ famous food in Xi'an
6. Chocolate biscuit @ levain bakery
7. Banana pudding @ Magnolia bakery
8. Jiebao @ Momofuku noodle bar
9. PASTRAMI @ KATZ'S DELICATESSEN
10. BAGEL @ RUSS & DAUGHTERS
11. Steak @ Peter Luger
12. EGGS BENEDICT @ DELMONICO'S
13. PANCAKES @ CLINTON ST. BAKING CO.
14. Green tea Melaleuca @ lady M
15. Lobster roll @ Luke's lobster

Business problem

It is very important to strategically plan in order to survive in such a competitive market. Various factors need to be studied in order to decide on the Location such as:

1. New York Population
2. New York City Demographics
3. Are there any Farmers Markets, Wholesale markets etc nearby so that the ingredients can be purchased fresh to maintain quality and cost?
4. Are there any venues like Gyms, Entertainment zones, Parks etc nearby where floating population is high etc
5. Who are the competitors in that location?
6. Cuisine served / Menu of the competitors
7. Segmentation of the Borough
8. Untapped markets
9. Saturated market

Target audience

This project would play a major role for owners who are willing to start a new restaurant and for tourists who are in trouble finding a good restaurant.

Data

To solve the problem, we will need the following data:

- List of neighborhoods in New York. This defines the scope of this project which is confined to the city of New York, the capital city of the country of America in Northern America.
- Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighborhoods.

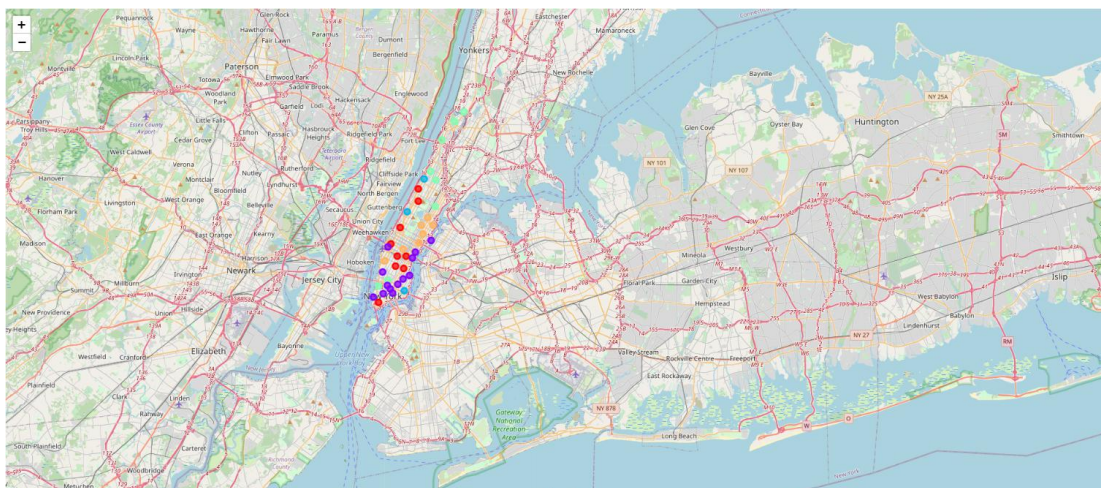
Methodology

Firstly, we need to get the list of neighborhoods in the city of New York. We will do web scraping using Python requests and BeautifulSoup packages to extract the list of neighbourhoods data. We need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the Geocoder. After gathering the data, we will populate the data into a pandas DataFrame and then visualize the neighbourhoods in a map using the Folium package. Next, we will use Foursquare API.

Foursquare will return the venue data in JSON format. Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. We will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. The results will allow us to identify which neighbourhoods have higher concentration of restaurants.

Results

The following image of the map shows the observed results.



Conclusion

So From the results we can infer that cluster 0 with color red has more number of places and it would be more suitable to open in any one of the p