

Battle of the Neighbourhoods

LOCATION RECOMMENDATION FOR A NEW INDIAN RESTAURANT IN NYC

- Ishan Iyer

1. Introduction and Business Problem

Problem background and description

New York City is famous for its excellent cuisine and robust restaurant culture influenced heavily by the city's immigrant history. It's basically a big brewing pot with Jewish, Italian, Asian, Middle Eastern etc. restaurants dotted all across the city. They serve up everything from pizzas and pastrami, to falafel and tandoori chicken, basically you name it and it's very likely to be found in New York. The city is also very famous for its numerous Michelin starred fine dining restaurants. So, it's amply clear that to survive in such a competitive market, it is crucial to strategically plan the location of a new restaurant.

Chef QWERTY has been working on opening his own Indian casual dining restaurant. Once he has conceptualized his menu his focus shifts to a very important part of any restaurant, or establishment for that matter; the location. He decides he'll take some help from his old friend XYZ's who just completed a data science course online. XYZ is delighted to take on this project and assures QWERTY that he has come to the right person because data can solve most problems and give optimal solutions. He tells QWERTY that it'll take him 2-3 days to program the model as various factors such as population, demographics, competitors, cuisines, nearby farmers/wholesale markets for freshness of ingredients, untapped areas etc. need to be considered in order to pick the perfect and most profitable location.

Target Audience

The target audience for this analysis is mainly Chef QWERTY, but this template can be used for anyone looking to open a new restaurant in New York City.

Goal

To recommend the optimal neighbourhood/borough to open a new Indian restaurant in New York City, and basing this recommendation on predominantly 2 factors; frequency of Indian restaurants in the area, and distance to the nearest market.

2. Data

The following datasets related to New York City will be used to conduct this analysis:

Neighbourhood data

This will consist of the borough name, neighbourhood name, and coordinates for each neighbourhood. This CSV file can be accessed for free via the following link:

<https://data.cityofnewyork.us/City-Government/Neighborhood-Names-GIS/99bc-9p23>

| | Borough | Neighbourhood | Latitude | Longitude |
|---|---------|---------------|-----------|------------|
| 0 | Bronx | Wakefield | 40.894705 | -73.847201 |
| 1 | Bronx | Co-op City | 40.874294 | -73.829939 |
| 2 | Bronx | Eastchester | 40.887556 | -73.827806 |
| 3 | Bronx | Fieldston | 40.895437 | -73.905643 |
| 4 | Bronx | Riverdale | 40.890834 | -73.912585 |

Table 1: Neighbourhood Dataframe

Population, demographic and cuisine related data

These datasets will consist of population distributions, economic distributions, types of cuisines and restaurants, location of Michelin starred restaurants and other demographic data. These can be easily found on Wikipedia; the following are the links for the same:

https://en.wikipedia.org/wiki/New_York_City

https://en.wikipedia.org/wiki/Portal:New_York_City

https://en.wikipedia.org/wiki/Economy_of_New_York_City

https://en.wikipedia.org/wiki/Cuisine_of_New_York_City

| | Borough | Indian Americans | (Indian Americans)/sq_mi | % of Indian Americans |
|------|---------------|------------------|--------------------------|-----------------------|
| Rank | | | | |
| 1 | Queens | 144,896 | 1,326.5 | 6.2 |
| 2 | Brooklyn | 25,270 | 357.9 | 1.0 |
| 3 | Manhattan | 24,359 | 1,060.9 | 1.5 |
| 4 | The Bronx | 16,748 | 398.6 | 1.2 |
| 5 | Staten Island | 6,646 | 113.6 | 1.4 |

Table 2: NYC Demographics Data

| | Borough | County | Population | Area sq_mi | Area sq_km | Persons/sq_mi | Persons/sq_km |
|---|---------------|-------------------|------------|------------|------------|---------------|---------------|
| 0 | The Bronx | Bronx | 1,471,160 | 42.10 | 109.04 | 34,653 | 13,231 |
| 1 | Brooklyn | Kings | 2,648,771 | 70.82 | 183.42 | 37,137 | 14,649 |
| 2 | Manhattan | New York | 1,664,727 | 22.83 | 59.13 | 72,033 | 27,826 |
| 3 | Queens | Queens | 2,358,582 | 108.53 | 281.09 | 21,460 | 8,354 |
| 4 | Staten Island | Richmond | 479,458 | 58.37 | 151.18 | 8,112 | 3,132 |
| 5 | | City of New York | 8,622,698 | 806.863 | 783.83 | 28,188 | 10,947 |
| 6 | | State of New York | 19,849,399 | 1,547.116 | 122,284 | 416.4 | 159 |

Table 3: NYC Population Data

| | Borough | Neighbourhood | Cuisine |
|---|-----------|---------------|--|
| 0 | The Bronx | Bedford Park | Mexican, Puerto Rican, Dominican, Korean |
| 1 | The Bronx | Belmont | Italian, Albanian |
| 2 | The Bronx | City Island | Italian, Seafood |
| 3 | The Bronx | Morris Park | Italian, Albanian |
| 4 | The Bronx | Norwood | Filipino |
| 5 | The Bronx | Riverdale | Jewish |
| 6 | The Bronx | South Bronx | Puerto Rican, Dominican |
| 7 | The Bronx | Wakefield | Jamaican, West Indian |
| 8 | The Bronx | Woodlawn | Irish |
| 9 | Queens | Astoria | Greek, Italian, Eastern European, Brazilian, ... |

Table 4: NYC Cuisine Data

Markets and ingredient related data

This dataset will contain the location, operating hours, season dates of farmers markets across NYC. This can also be accessed for free via the following link:

<https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2>

| | Borough | Market Name | Street Address | Latitude | Longitude | Days of Operation | Hours of Operations | Season Dates | Accepts EBT | Open Year-Round | Stellar Cooking Demonstrations | Food Activities for Kids | Location Point |
|---|---------------|----------------------------------|---|-----------|------------|--------------------|---------------------|-----------------------|-------------|-----------------|--------------------------------|--------------------------|-------------------------|
| 0 | Brooklyn | Urban Oasis Farmers Market | 681 Clarkson Ave | 40.656255 | -73.936608 | Wednesday | 2 - 5:30 p.m. | 06/26/2019-11/06/2019 | No | No | No | No | (40.656255, -73.936608) |
| 1 | Staten Island | Staten Island Mall Greenmarket | Marsh Ave & Ring Rd | 40.583804 | -74.161245 | Saturday | 8 a.m. - 3 p.m. | Year-Round | Yes | Yes | No | No | (40.583804, -74.161245) |
| 2 | Manhattan | Mount Sinai Hospital Greenmarket | E 99th St bet Madison & Park Aves | 40.789169 | -73.952743 | Wednesday | 8 a.m. - 5 p.m. | 06/12/19-11/27/19 | Yes | No | No | No | (40.789169, -73.952743) |
| 3 | Bronx | 170 Farm Stand | E 170th St & Townsend Ave | 40.839882 | -73.916783 | Wednesday | 2:30 - 6:30 p.m. | 07/10/2019-11/27/2019 | Yes | No | No | Yes | (40.839882, -73.916783) |
| 4 | Manhattan | Grass Roots Farmers Market | W 145th St bet Edgecombe & Bradhurst Aves, at ... | 40.823647 | -73.943844 | Tuesday & Saturday | 9 a.m. - 4 p.m. | 07/11/2019-11/21/2019 | Yes | No | No | No | (40.823647, -73.943844) |

Table 5: NYC Farmers Market Data

Foursquare data

The NYC geo coordinates will be used by the Foursquare API to provide information on venues and types of establishments in each neighbourhood. This will have sections for venue type, venue location etc.

| | Neighbourhood | Neighbourhood Latitude | Neighbourhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---------------|------------------------|-------------------------|-----------------------------|----------------|-----------------|----------------------|
| 0 | Wakefield | 40.894705 | -73.847201 | Lollipop Gelato | 40.894123 | -73.845892 | Dessert Shop |
| 1 | Wakefield | 40.894705 | -73.847201 | Ripe Kitchen & Bar | 40.898152 | -73.838875 | Caribbean Restaurant |
| 2 | Wakefield | 40.894705 | -73.847201 | Rite Aid | 40.896649 | -73.844846 | Pharmacy |
| 3 | Wakefield | 40.894705 | -73.847201 | Jackie's West Indian Bakery | 40.889283 | -73.843310 | Caribbean Restaurant |
| 4 | Wakefield | 40.894705 | -73.847201 | Ali's Roti Shop | 40.894036 | -73.856935 | Caribbean Restaurant |

Table 6: Foursquare API Data

3. Methodology

Analytic Approach:

New York City has a total of 5 Boroughs and 306 Neighbourhoods. This project will make a few assumptions basis the exploratory data analysis, which will help us narrow down on a Borough after which clustering and classification techniques will be applied on the chosen Borough.

Exploratory Data Analysis:

Data 1: NYC Geo data

- In this step we load and transform the geo dictionaries into python data frames.

- This dataset contains location data regarding each neighbourhood, and will be used to obtain Foursquare data for the same.
- Geopy and Folium libraries are used to superimpose these coordinates onto a map of NYC.

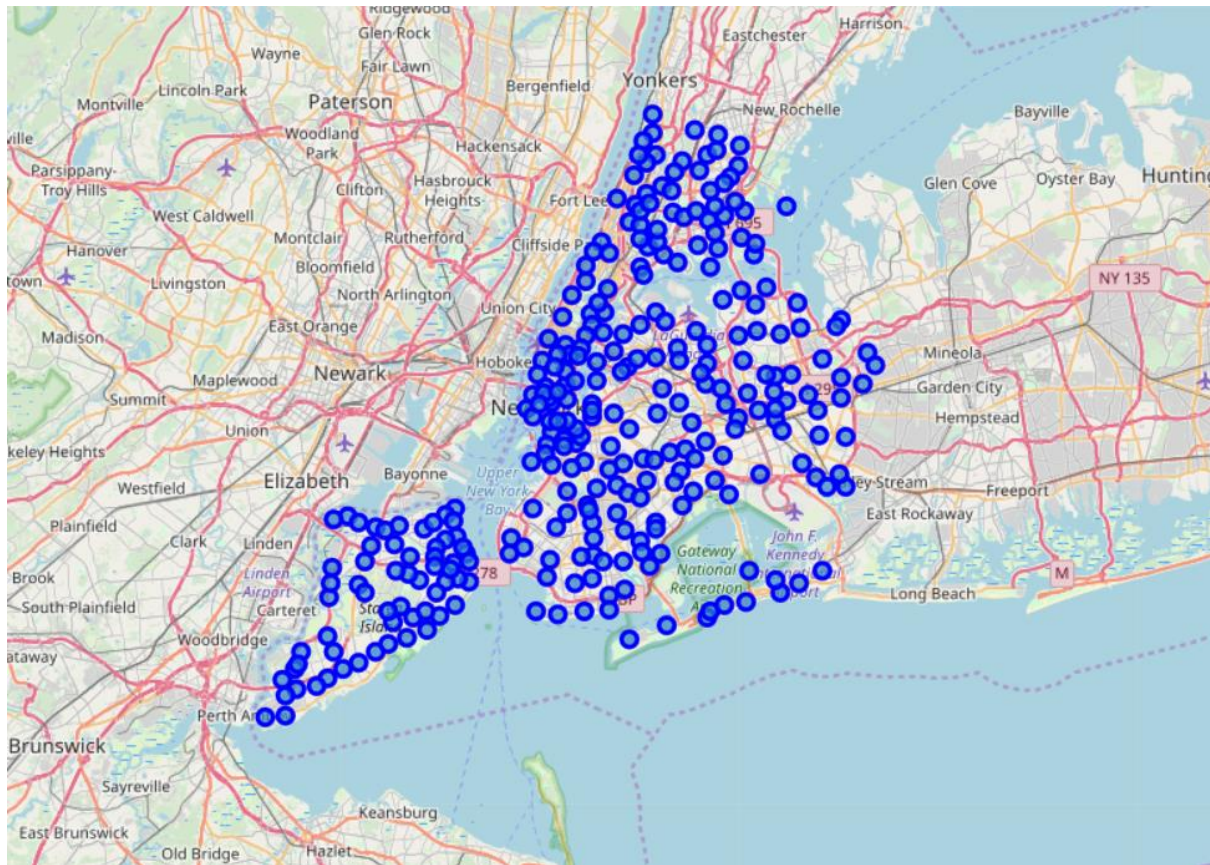


Figure 1: Map of NYC neighbourhoods

Data 2: NYC Farmers Market data

- This dataset contains information wrt the location and availability of all farmers markets and food boxes around NYC.
- A total of 139 farmers markets are present in New York, with the highest concentration in Brooklyn and Manhattan.
- Geopy and Folium libraries are used to superimpose the location of these farmers markets onto a map of NYC.

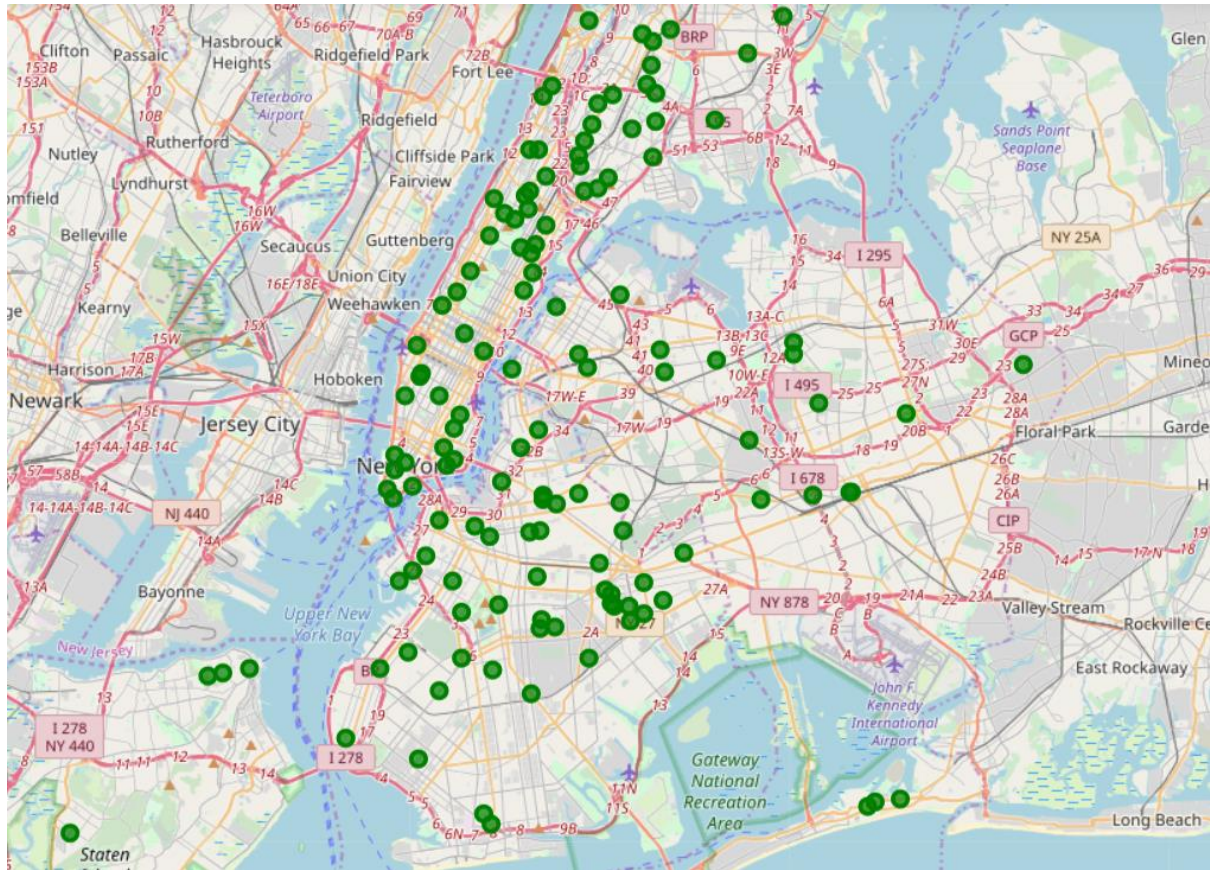


Figure 2: Map of NYC Farmers Market locations

Data 3: NYC population and demographic data

To analyze the population and demographic data we first scraped data from Wikipedia pages using a python library called BeautifulSoup, which is used for parsing HTML and XML documents.

Insights:

- Manhattan is the geographically smallest, and most densely populated borough.
- Brooklyn is the city's most populated Borough.
- Queens is geographically the largest Borough

| | Borough | County | Population | Area sq_mi | Area sq_km | Persons/sq_mi | Persons/sq_km |
|---|---------------|-------------------|------------|------------|------------|---------------|---------------|
| 0 | The Bronx | Bronx | 1,471,160 | 42.10 | 109.04 | 34,653 | 13,231 |
| 1 | Brooklyn | Kings | 2,648,771 | 70.82 | 183.42 | 37,137 | 14,649 |
| 2 | Manhattan | New York | 1,664,727 | 22.83 | 59.13 | 72,033 | 27,826 |
| 3 | Queens | Queens | 2,358,582 | 108.53 | 281.09 | 21,460 | 8,354 |
| 4 | Staten Island | Richmond | 479,458 | 58.37 | 151.18 | 8,112 | 3,132 |
| 5 | | City of New York | 8,622,698 | 806.863 | 783.83 | 28,188 | 10,947 |
| 6 | | State of New York | 19,849,399 | 1,547.116 | 122,284 | 416.4 | 159 |

- Queens and Manhattan have the highest number of Indians per square mile, with queens having the highest number of Indian-Americans overall.
- This makes Queens and Manhattan prime targets basis both size and percentage of Indian Americans present.

Data 4: NYC cuisine data

This data was scraped using BeautifulSoup and will be used to find out the most preferred cuisines in each borough. https://en.wikipedia.org/wiki/Cuisine_of_New_York_City

Most preferred cuisines (in order of preference):

Brooklyn – Italian, Mexican, Puerto Rican

Manhattan – Italian, American, Indian

Queens – Indian, Irish, Pakistani, Mexican

Bronx – Italian, Puerto Rican, Albanian

Basis this, a well-informed choice would be to pick Queens due to the preference and % of Indian-Americans.

Even though Manhattan has similar preferences, it is a very saturated market with a lot of footfall. It is normally harder to break into such markets, and factors such as cost of inventory and property are higher as well.

Data 5: Foursquare location data

NYC geo data acquired previously can be used as an input for Venue related data via the Foursquare API. We will use this Foursquare data to explore neighbourhoods in and around Queens.

| | Neighbourhood | Neighbourhood Latitude | Neighbourhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---------------|------------------------|-------------------------|-----------------------------|----------------|-----------------|---------------------------|
| 0 | Astoria | 40.768509 | -73.915654 | Favela Grill | 40.767348 | -73.917897 | Brazilian Restaurant |
| 1 | Astoria | 40.768509 | -73.915654 | Titan Foods Inc. | 40.769198 | -73.919253 | Gourmet Shop |
| 2 | Astoria | 40.768509 | -73.915654 | CrossFit Queens | 40.769404 | -73.918977 | Gym |
| 3 | Astoria | 40.768509 | -73.915654 | Sitan Muay Thai | 40.766108 | -73.913224 | Martial Arts Dojo |
| 4 | Astoria | 40.768509 | -73.915654 | Al-sham Sweets and Pastries | 40.768077 | -73.911561 | Middle Eastern Restaurant |

Table 7: Queens Foursquare Data

This is then filtered to show only Restaurants (in Venue Category), and then further filtered to show only Indian Restaurants. This Indian Restaurant data is then used for clustering and classification.

Clustering:

First, we used one hot encoding to assign low bits to all values, besides one high bit value ie. 1. Then, we used K-Means clustering to form clusters wrt to the frequency of Indian restaurants in the area. To find the optimal number of clusters, we used the Silhouette Score method. This method runs the clustering for different values of n_clusters and returns the accuracy of each value. Using this we can pick the optimal n_clusters, which in this case was 6 or 7.

```
For n_clusters=2, The Silhouette Coefficient is 0.8416450996729327
For n_clusters=3, The Silhouette Coefficient is 0.827920816163456
For n_clusters=4, The Silhouette Coefficient is 0.9199402865337593
For n_clusters=5, The Silhouette Coefficient is 0.912491122974994
For n_clusters=6, The Silhouette Coefficient is 0.931135531135531
For n_clusters=7, The Silhouette Coefficient is 0.9392551892551894
For n_clusters=8, The Silhouette Coefficient is 0.9230769230769231
```

Using n_clusters=6, we carried on with the clustering and found out that cluster1 has the lowest total sum value and thus represents the most unsaturated areas.

As you can see below, cluster1 are all those in blue.

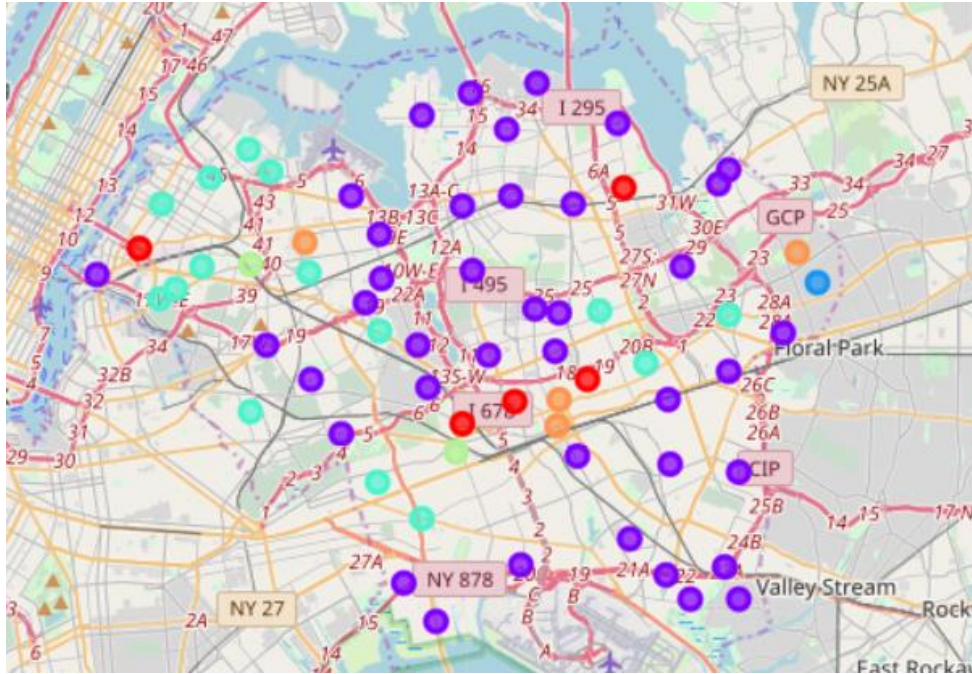


Figure 3: Cluster labels (6) overlaid on a map of Queens

Results

We can find the comparatively unsaturated market areas by creating a separate data frame for all cluster1 values, as shown below. This gives us the separated unsaturated clusters which we can use in tandem with the farmers market data for a final result.

| | Borough | Neighbourhood | Latitude | Longitude | Total | Cluster_Labels |
|----|---------|-------------------|-----------|------------|-------|----------------|
| 0 | Queens | Howard Beach | 40.654225 | -73.838138 | 0 | 1 |
| 1 | Queens | South Corona | 40.742382 | -73.856825 | 0 | 1 |
| 2 | Queens | Forest Hills | 40.725264 | -73.844475 | 0 | 1 |
| 3 | Queens | Downtown Flushing | 40.761164 | -73.829368 | 0 | 1 |
| 4 | Queens | East Elmhurst | 40.764073 | -73.867041 | 0 | 1 |
| 5 | Queens | Maspeth | 40.725427 | -73.896217 | 0 | 1 |
| 6 | Queens | Glendale | 40.702762 | -73.870742 | 0 | 1 |
| 7 | Queens | South Ozone Park | 40.668550 | -73.809865 | 0 | 1 |
| 8 | Queens | College Point | 40.784903 | -73.843045 | 0 | 1 |
| 9 | Queens | Whitestone | 40.781291 | -73.814202 | 0 | 1 |
| 10 | Queens | Auburndale | 40.761730 | -73.791762 | 0 | 1 |
| 11 | Queens | Little Neck | 40.770826 | -73.738898 | 0 | 1 |
| 12 | Queens | Douglaston | 40.766846 | -73.742498 | 0 | 1 |
| 13 | Queens | Bellerose | 40.728573 | -73.720128 | 0 | 1 |
| 14 | Queens | Kew Gardens Hills | 40.722578 | -73.820878 | 0 | 1 |

Table 8: Unsaturated clusters in Queens

On superimposing the farmers markets markers onto the map with unsaturated clusters, we can easily see the distance and convenience factor for each cluster. The clusters are shown in blue with the farmers markets overlaid in green.

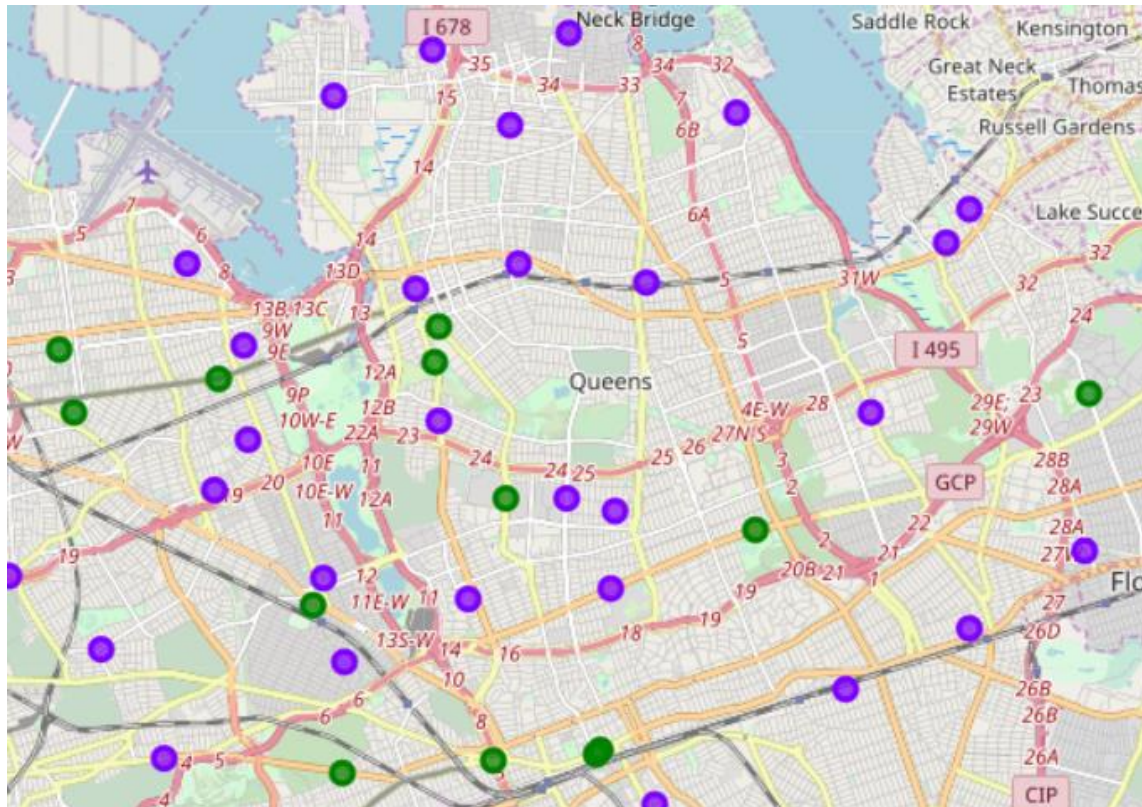


Figure 4: Unsaturated clusters + Farmers Markets in Queens

Discussion

- Manhattan and Brooklyn consist of many diverse restaurants, and have high footfall. Thus, to break into these areas a very strong menu and ambience is required.
- Bronx, Queens and Staten Island contain some mildly unsaturated markets, whereas the rest are completely saturated.

Conclusion

Queens was picked as this borough consists of the highest % of Indian-Americans, highest number of Indians/sq.mile and the food of choice in most neighbourhoods is Indian. Even though Manhattan is a faster moving market, with a large number of Indian-Americans, it's a very competitive space. XYZ feels the restaurant will benefit from first opening a branch in Queens, with the option of branching out to Manhattan/Brooklyn in the near future. At the end of the day it's about picking the option with the lowest risk and maximizing returns.

XYZ carefully analyses the cluster selection and quality & distance to farmer markets, and decides Chef QWERTY should open his first restaurant in Downtown Flushing, Queens.

With the data collected from basic internet sources, XYZ says this is the best outcome he can come up with.

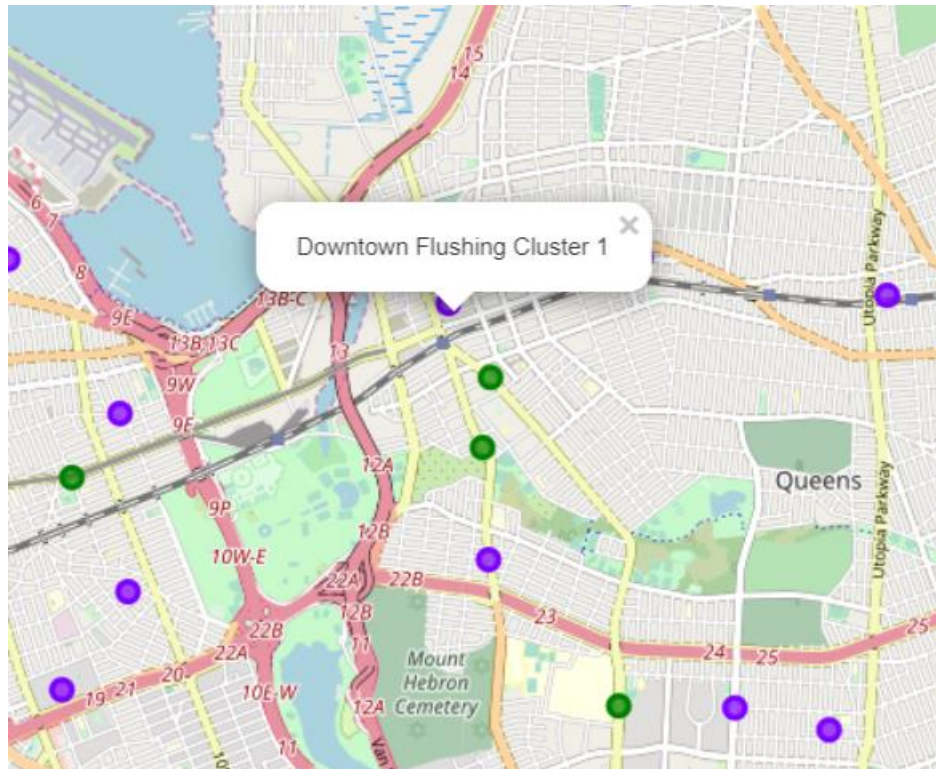


Figure 5: Ideal cluster picked by XYZ