Finding the optimal location for opening a Restaurant Supply Store in New York

PRANAV NARAYANAN

25th January, 2021

1. Introduction

The City of New York is famous for its excellent cuisine. Its food culture includes an array of international cuisines influenced by the city's immigrant history.

Due to this influence, many restaurants are operational in the city catering to a wide variety of audience. This also means that there is a huge demand for the supply of items needed to run these restaurants.

1.1 Problem

In this analysis, we will discuss the optimal place to start a restaurant supply chain shop which can cater to maximum restaurants and reduce the operational costs incurred in running the shop.

1.2 Target Audience

The target audience for this analysis would be any person or company who is interested in opening a restaurant supply store in New York.

After the analysis they would obtain the optimum position to open the store which will minimise the cost of transport as well as generate sufficient income due to its prime location.

2. Data Acquisition and Cleaning

2.1 Data Sources

Data related to the city of New York such as the Borough and Neighbourhood information can be obtained from here.

2.2 Data Cleaning

The data obtained from above link contains lot of information out of which we are particularly interested in details such as neighbourhood, boroughs, the latitude and longitude for each neighbourhood.

3. Exploratory Data Analysis

3.1 Obtaining the list of venues related to food for each neighbourhood using Foursquare API

We use the Foursquare API to retrieve the list of venues upto a maximum of 100 venues for each neighbourhood. Since we would like to restrict the venues to only food related ones, we pass a parameter called as “CategoryId” within the request to Foursquare API.

Once the relevant venues are retrieved from Foursquare API, the dataset looks something like this -

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Margaritaville Restaurant and Lounge	40.898564	-73.837016	Caribbean Restaurant
2	Wakefield	40.894705	-73.847201	Cooler Runnings Jamaican Restaurant Inc	40.898083	-73.850259	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Essa Deli	40.892181	-73.854427	Deli / Bodega
4	Wakefield	40.894705	-73.847201	Big Daddy's Caribbean Taste Restaurant	40.899767	-73.857135	Caribbean Restaurant

3.2 Data Wrangling

Once the data has been obtained from Foursquare API, we apply multiple data wrangling techniques such as one-hot encoding and grouping to obtain insights regarding the data.

Below is the result of one-hot encoding-

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Asian Restaurant	Australian Restaurant	Austrian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bangladeshi Restaurant	Bar	Bath House	Beach	Beach Bar	Beer Bar
0	Wakefield	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Wakefield	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Wakefield	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Wakefield	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Wakefield	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Once the encoding is done, we group the Neighbourhoods and obtain the mean for every venue within get neighbourhood.

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Asian Restaurant	Australian Restaurant	Austrian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bangladeshi Restaurant	Bar	Bath House	Beach	Beac Ba
0	Allerton	0.0	0.0	0.020000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.060000	0.0	0.0	0.0	0.0	0
1	Annadale	0.0	0.0	0.107143	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.071429	0.071429	0.0	0.0	0.0	0.0	0
2	Arden Heights	0.0	0.0	0.096774	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.032258	0.032258	0.032258	0.0	0.0	0.0	0.0	0
3	Arlington	0.0	0.0	0.055556	0.0	0.0	0.0	0.0	0.027778	0.0	0.0	0.000000	0.027778	0.027778	0.0	0.0	0.0	0.0	0
4	Arrochar	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.066667	0.000000	0.0	0.0	0.0	0.0	0

3.3 Classification Model

Now that we have data in the required format, we need to apply the K-Means grouping algorithm to group the data into similar clusters. This helps us in identifying similarities between the Neighbourhoods.

The main parameter if K-Means algorithm is the cluster size. To find the best cluster size, we use silhouette_score.

Below is the result of silhouette_score for different cluster sizes.

```
silhouette_score results :  {2: 0.17743357998254314, 3:
0.14066439991387775, 4: 0.12914863106062802, 5:
0.1328577503914053, 6: 0.11736892508853244, 7:
0.12027984810897324, 8: 0.08326902182481445, 9:
0.07694941825109619}
```

The highlighted value shows that Cluster size of 2 is best suited for K-Means algorithm.

3.4 After Classification

After the classification model was successfully run on the dataset, we obtain the following result by further wrangling the data. We find out the most common venues within the neighbourhoods and then apply the Cluster Labels(0 and 1) to each of the entries.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bronx	Wakefield	40.894705	-73.847201	0	Caribbean Restaurant	Deli / Bodega	Chinese Restaurant	Fast Food Restaurant	Donut Shop	Food	Asian Restaurant	Southern / Soul Food Restaurant	Steakhouse	Juice Bar
1	Bronx	Co-op City	40.874294	-73.829939	0	Bakery	Chinese Restaurant	Deli / Bodega	Fast Food Restaurant	Donut Shop	Caribbean Restaurant	Fried Chicken Joint	Mexican Restaurant	Seafood Restaurant	American Restaurant
2	Bronx	Eastchester	40.887556	-73.827806	0	Pizza Place	Deli / Bodega	Caribbean Restaurant	Fast Food Restaurant	Bakery	Donut Shop	Chinese Restaurant	Burger Joint	Café	Asian Restaurant
3	Bronx	Fieldston	40.895437	-73.905643	0	Pizza Place	Mexican Restaurant	Deli / Bodega	Fast Food Restaurant	Bagel Shop	Donut Shop	Sushi Restaurant	Bakery	Sandwich Place	Diner
4	Bronx	Riverdale	40.890834	-73.912585	0	Pizza Place	Diner	Bagel Shop	Deli / Bodega	Fast Food Restaurant	Café	Mexican Restaurant	Food Truck	Bakery	Sandwich Place

4. Analysis

After the data has been clustered, we analyse each cluster to find which cluster contains more relevant data by finding its shape.

The cluster 0 has 233 rows while Cluster 1 has 73 rows.

Hence we use Cluster 0 as our dataset for further evaluation.

4.1 Finding Centroid of Cluster 0

The next step is to find the centroid of Cluster 0. This is in fact the optimal location for opening our restaurant supply store.

We calculate the centroid of geographical coordinates by using the averaging methodology. A detailed explanation is given here.

We obtain the centroids as **Latitude : 40.69394045210022, Longitude : -73.93700185136717.**

4.2 Reverse lookup of Geographic Coordinates

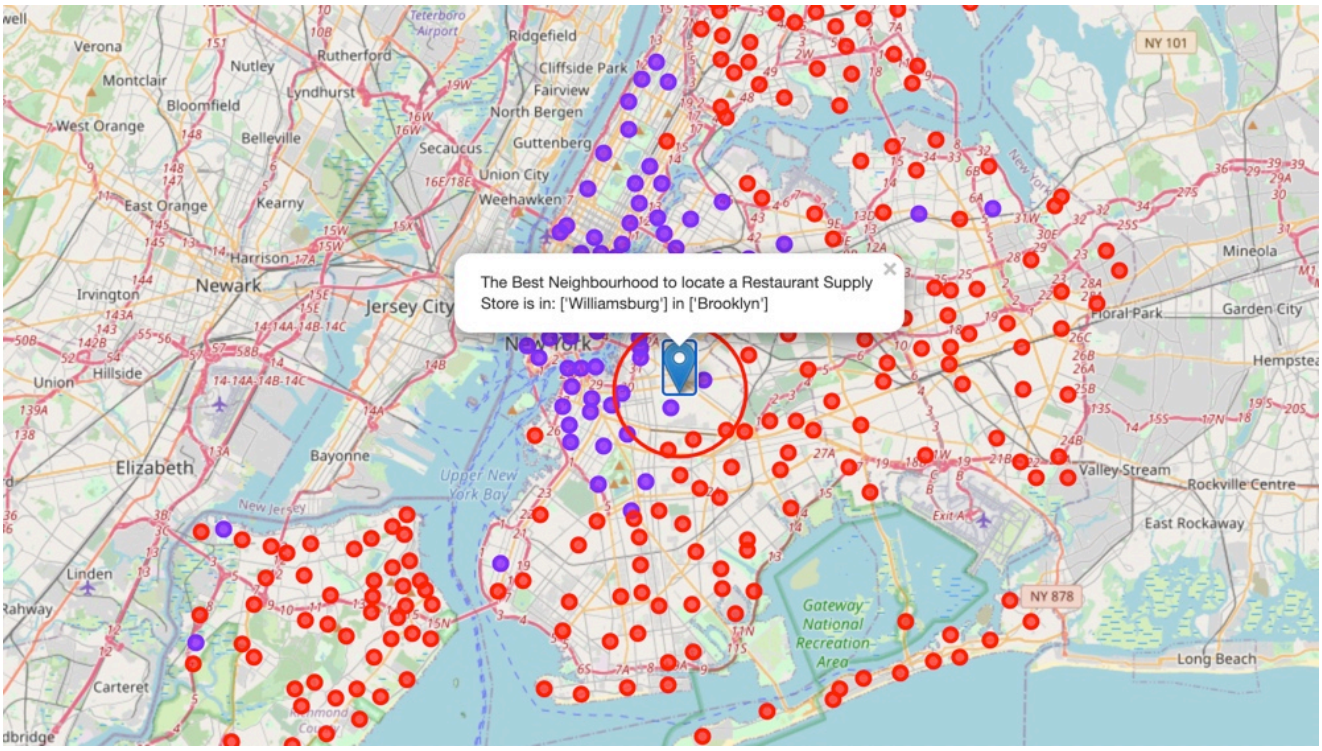
We use OpenCageGeocode to reverse lookup the coordinates and obtain the correct details such as Neighbourhood or Borough. This is used to locate the data on the map using folium.

The reverse lookup yields the following data -

```
'components': {'ISO_3166-1_alpha-2': 'US',
                'ISO_3166-1_alpha-3': 'USA',
                '_category': 'building',
                '_type': 'building',
                'building': 'Building 7',
                'city': 'New York',
                'city_district': 'Kings County',
                'continent': 'North America',
                'country': 'United States',
                'country_code': 'us',
                'house_number': '93',
                'postcode': '11206',
                'residential': 'BEDFORD STUYVESANT/
EXPANDED STUYVEVSANT '
                               'HEIGHTS HISTORIC
DISTRICT',
                'road': 'Lewis Avenue',
                'state': 'New York',
                'state_code': 'NY',
                'suburb': 'Brooklyn'},
  'confidence': 10,
  'formatted': 'Building 7, 93 Lewis Avenue, New York, NY
11206, United States '
               'of America',
  'geometry': {'lat': 40.6941276, 'lng': -73.9369924}}
```

This gives us the precise location within the city of New York to open our restaurant supply store.

A map with the cluster and the location marked is shown below -



5. Conclusion

I feel confident with the recommendation I have given my client as it is backed up with demonstrated data analysis. While nothing can ever be 100% certain they will certainly be better informed than they were prior to this comprehensive analysis.

Much more inference can be obtained with more work. A potential side business for my client might be assisting new restaurant owners where they might locate a new restaurant, who their competition is and who their clientele might be.