REPORT OF BATTLE OF NEIGHBORHOODS

INDIAN RESTAURANT IN NY

Sakthivel P

31-JAN-2020

The Battle of the Neighborhoods - Week 2 Report

Table of contents

1. Introduction & Business Problem:

a. Business Problem:

In this project we will try to find a optimal location for opening a Indian restaurant in New York. This report will be targeted towards people who are interested in opening a new Indian restaurant in New York.

Since there are lots of restaurants in New York we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no Indian restaurants in vicinity. We would also prefer locations as close to city center as possible along with the above conditions

We will use the data science technique learned over the Certification program to arrive at an optimal solution for the stakeholders wishing to open an Indian restaurant in New York

b. Background:

The City of New York is famous for its array of international cuisines influenced by the city's immigrant history. The Indian population in New York City is over 650,000. Indians, the second largest Asian group in the city, comprised 20.5 percent of the Asian population. By borough, 65 percent of Indian New Yorkers lived in Queens, 12 percent in Manhattan, 12 percent in Brooklyn, 6.5 percent in the Bronx, and 5 percent in Staten Island. Neighborhoods with large Indian populations included South Ozone Park, Richmond Hill, Jackson Heights, Jamaica, Bellerose, Queens Village, Glen Oaks-Floral Park-New Hyde Park, Briarwood - Jamaica Hill, Elmhurst, and Flushing in Queens

Indians in New York City had higher median household and median family incomes than the general population. Indian median household income of USD 62,775 was higher than the citywide figure of USD 50,331. Median family income for Indians of USD 62,911 was above the citywide median family income of USD 55,434. Indian per capita income of USD 29,540 was close to the citywide figure of USD 30,717.

Given this background it is really an interesting proposition to open an Indian restaurant in New York, as it would make business sense and value

c. Target Audience:

To recommend the correct location to the stakeholders which neighborhood of New York city will be best choice to start a restaurant. The stakeholders also expect to understand the rationale of the recommendations made.
This would interest anyone who wants to start a new restaurant in New York city.

d. **Success Criteria:**

The success criteria of the project will be a good recommendation of borough/Neighborhood choice based on lack of such restaurants in that location

2. Data:

Based on our Business problem, the following factors will influence the decision making:
  a. All existing Indian restaurants in the neighborhood
  b. Ratings of the current Indian restaurants in various neighborhoods
  c. NYC Indian population concentration across boroughs

Following data sources will be needed to extract/generate the above required information:
  a. Demographics of NYC Data source: https://en.wikipedia.org/wiki/Demographics_of_New_York_City. Web scraping techniques will be used to get NYC's population demographics from Wikipedia. The following is the web scraped data

```
[{'jurisdiction': 'Queens', 'population_census': '2,229,379', '%_white': 44.1, '%_black_or_african_amercia
n': 20.0, '%_Asian': 17.6, '%_other': 12.3, '%_mixed_race': 6.1, '%_hispanic_latino_of_other_race': 25.0,
'%_catholic': 37.0, '%_jewish': 5.0}, {'jurisdiction': 'Manhattan', 'population_census': '1,537,195', '%_w
hite': 54.4, '%_black_or_african_amercian': 17.4, '%_Asian': 9.4, '%_other': 14.7, '%_mixed_race': 4.1, '%
_hispanic_latino_of_other_race': 27.2, '%_catholic': 11.0, '%_jewish': 9.0}, {'jurisdiction': 'Bronx', 'po
pulation_census': '1,332,650', '%_white': 29.9, '%_black_or_african_amercian': 35.6, '%_Asian': 3.0, '%_ot
her': 25.7, '%_mixed_race': 5.8, '%_hispanic_latino_of_other_race': 48.4, '%_catholic': 14.0, '%_jewish':
5.0}, {'jurisdiction': 'Staten Island', 'population_census': '443,728', '%_white': 77.6, '%_black_or_afric
an_amercian': 9.7, '%_Asian': 5.7, '%_other': 4.3, '%_mixed_race': 2.7, '%_hispanic_latino_of_other_race':
12.1, '%_catholic': 11.0, '%_jewish': 5.0}, {'jurisdiction': 'NYC Total', 'population_census': '8,008,27
8', '%_white': 44.7, '%_black_or_african_amercian': 26.6, '%_Asian': 9.8, '%_other': 14.0, '%_mixed_race':
4.9, '%_hispanic_latino_of_other_race': 27.0, '%_catholic': 17.0, '%_jewish': 6.0}]
```

  b. Ratings and Number of existing Indian restaurants, their type and location in every neighborhood will be obtained using Foursquare API and https://cocl.us/new_york_dataset. The following is the sample of the data obtained

	Borough	Neighborhood	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

3. Methodology

a. Business Understanding:

Our main goal is to get optimum location for new restaurant business in New York City for concerned stakeholders

b. Analytic Approach:

New York city neighborhood has a total of 5 boroughs and 306 neighborhoods. In this project first part is slicing and dicing the Indian restaurants across neighborhoods and borough. Then extracting a list that contains 'Likes', 'Ratings' and 'Tips' for the restaurants from Foursquare. And second part is grouping neighborhood based on average ratings and identifying neighborhood that has least rating. This is done to compare with the demographics data and arrive at suitable location for opening the restaurant

c. Exploratory data analysis

    i. New York Demographics data
        a. Beautiful soup (html parser) was used to parse the Wikipedia page on demographics of NYC. Data wrangling is performed on the dataset for converting the raw data. The processed data is shown below

	jurisdiction	%_white	%_black_or_african_amercian	%_Asian	%_other	%_mixed_race	%_hispanic_latino_of_other_race
0	Queens	44.1	20.0	17.6	12.3	6.1	25.0
1	Manhattan	54.4	17.4	9.4	14.7	4.1	27.2
2	Bronx	29.9	35.6	3.0	25.7	5.8	48.4
3	Staten Island	77.6	9.7	5.7	4.3	2.7	12.1
4	NYC Total	44.7	26.6	9.8	14.0	4.9	27.0

    ii. Slicing and Dicing of Indian restaurants data
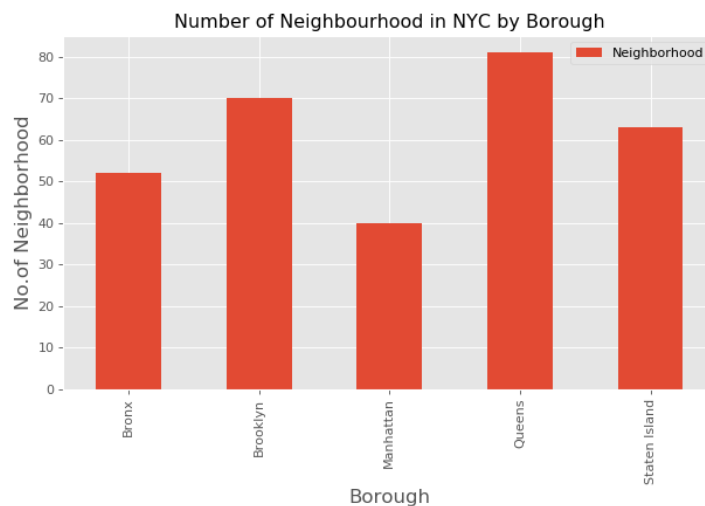
Number of Indian Restaurants in NYC by Borough
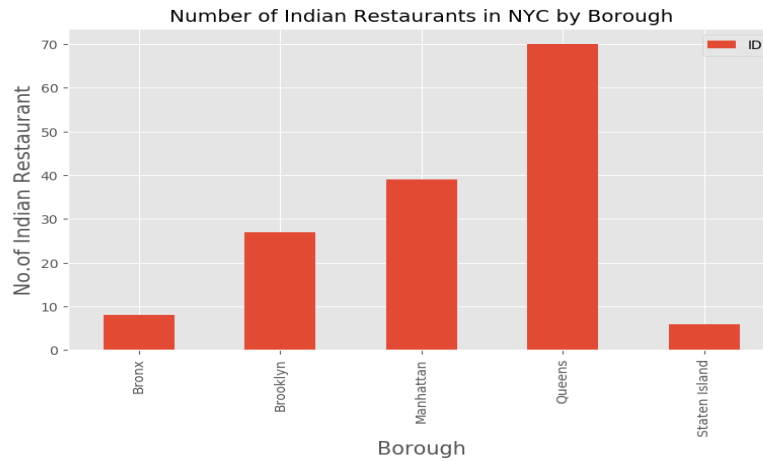
To start with, the analysis started with identifying number of Indian restaurants across boroughs. The distribution was compared with the distribution of Indian population across

Boroughs	Spread of Indian population	% of Indian restaurants
Bronx	6%	5.3%
Brooklyn	12%	18%
Manhattan	12%	26%
Queens	65%	46.7%
Staten Island	5%	4%

Inference 1: All Boroughs have a nearly equal % of Indian restaurants corresponding to the spread of Indian population except for Queens borough which has significantly lower % of Indian restaurants

To understand the inference further the number of neighborhoods by Borough was analyzed. If the neighborhoods in NYC are lower than others, then Inference 1 might be nullified

Number of Neighbourhood in NYC by Borough

Inference 2: The lower % of Indian restaurants corresponding to the spread of Indian population in Queensborough is significant since Queens has the highest number of neighborhoods by borough

Next the dataframe was appended with Likes, Ratings and Tips for each Indian restaurant using the foursquare API for analysis. The dataframe after appending data looks like the one below:

In [15]: indian_rest_stats_ny.head()

Out[15]:

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
0	Bronx	Woodlawn	4c0448d9310fc9b6bf1dc761	Curry Spot	5	7.6	10
1	Bronx	Norwood	0	0	0	0.0	0
2	Bronx	Williamsbridge	0	0	0	0.0	0
3	Bronx	Parkchester	4c194631838020a13e78e561	Melanies Roti Bar And Grill	3	6.2	2
4	Bronx	Spuyten Duyvil	4c04544df423a593ac83d116	Cumin Indian Cuisine	13	5.9	9

Then the Indian restaurants were grouped into Neighborhood based on average ratings. This will be required to narrow down to the neighborhoods we would be required to focus based on the rating combined with the inferences above.

The resultant table is like the one shown below:

	Neighborhood	Average Rating
70	Tribeca	9.00
76	West Village	8.85
45	Murray Hill	8.80
25	Fort Greene	8.70
11	Chelsea	8.70
54	Prospect Lefferts Gardens	8.60
48	North Side	8.50
63	South Side	8.50
2	Bay Ridge	8.40
62	Roosevelt Island	8.40

As explained in the analytic approach section, our area of focus would be the neighborhood with the least average rating. There are two hypotheses on which the analysis is based on (These Hypothesis are not tested and are assumed to be true in this project)

Hypothesis 1: Average rating of 0 could mean either the restaurant was started newly, or the ratings were very low

Hypothesis 2: A newly started restaurant with low or no ratings need not be considered as competition

Based on these Hypothesis, the Neighborhoods are filtered based on ratings below 5. The data is then appended with the Latitude and Longitude coordinate of the Boroughs. The resultant table will look like the one below

	Borough	Neighborhood	Latitude	Longitude	Average Rating
26	Manhattan	Manhattan Valley	40.797307	-73.964286	3.925
37	Queens	Rego Park	40.728974	-73.857827	0.000
27	Brooklyn	Midwood	40.625596	-73.957595	0.000
28	Manhattan	Morningside Heights	40.808000	-73.963896	0.000
29	Staten Island	New Brighton	40.640615	-74.087017	0.000
30	Staten Island	New Dorp	40.572572	-74.116479	0.000
31	Bronx	Norwood	40.877224	-73.879391	0.000
32	Queens	Ozone Park	40.680708	-73.843203	0.000
33	Staten Island	Park Hill	40.609190	-74.080157	0.000
34	Queens	Queensboro Hill	40.744572	-73.825809	0.000

Visualization: The data is then plotted in the map and visualized for the neighborhoods with the least ratings
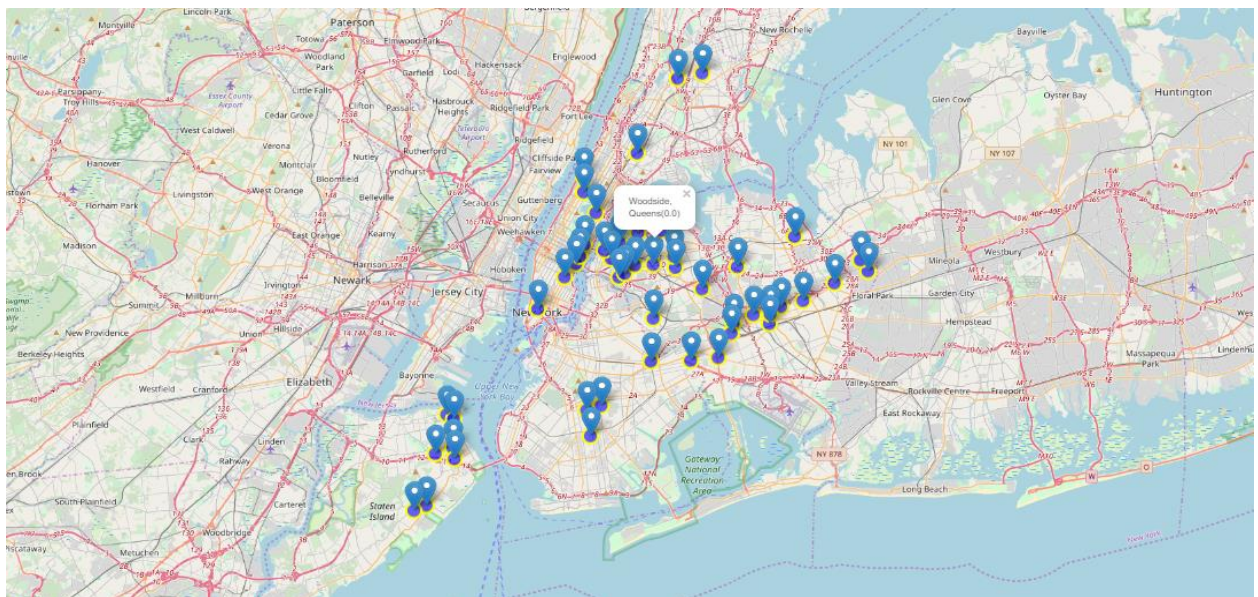
Inference 3: Queensborough has the most number of neighborhoods with the least restaurant ratings

4. Results

Let's have the summary of the inferences to derive the results

Inf 1: All Boroughs have a nearly equal % of Indian restaurants corresponding to the spread of Indian population except for Queens borough which has significantly lower % of Indian restaurants

Inf 2: The lower % of Indian restaurants corresponding to the spread of Indian population in Queens borough is significant since Queens has the highest number of neighborhoods by borough

Inf 3: Queens borough has the most number of neighborhoods with the least restaurant ratings

From the inferences above it is evident that Queens borough will be the best place to open an Indian restaurant as it has lower % of Indian restaurants with respect to population spread, highest number of neighborhoods and the least average ratings for Indian restaurants

5. Discussions:

Murray Hill, Tribeca, Midtown in Manhattan are some of the best neighborhoods for Indian cuisine currently. Hence opening another restaurant is not advisable since the competition will be hectic. Queens has the lowest rated Indian Restaurants in NYC and it has 65.0% of the Indian population in NYC

Recommendation based on result:

Hence Queens would be the best choice to start a restaurant given that
1. it is the most populous (Indian population) urban area in New York City (NYC)
2. most of the restaurants are new or have a very poor ratings in the locality
3. it has the highest Asian ethnic minority population in NYC

6. Conclusion

Further fine tuning of the model can be done to identify neighborhood level data if the hypothesis mentioned in the model are tested with primary research

This analysis is performed based on the basic premise the new restaurant opened will be a good value for money. If that goes for a toss, the new restaurant will add one more data point to restaurant with least ratings. Hence, the major contribution towards success is the value proposition that the new restaurant will offer.