Final Project: Applied Data Science Capstone

Subject: Optimal location for an Indian restaurant in New York City.

Introduction

As part of the final project we try to find the optimal location for opening a new Indian Restaurant in New York City.

We would try to find locations with low occurrences of an Indian restaurant in NYC. Also it would be ideal to have these locations in neighbourhoods with high demands for restaurants.

In order to find these locations we look at the number of existing restaurants as a proxy for demand.

Data

We would need the following information:

1. Neighbourhood information: The latitudes and longitudes of the various neighbourhoods in NYC.

2. Total number of restaurants in the various neighbourhoods.

3. Number of Indian restaurants in the various neighbourhoods.

We can get the neighbourhood information from the below json files:

https://cocl.us/new_york_dataset

https://raw.githubusercontent.com/veltman/snd3/master/data/nyc-neighborhoods.geo.json

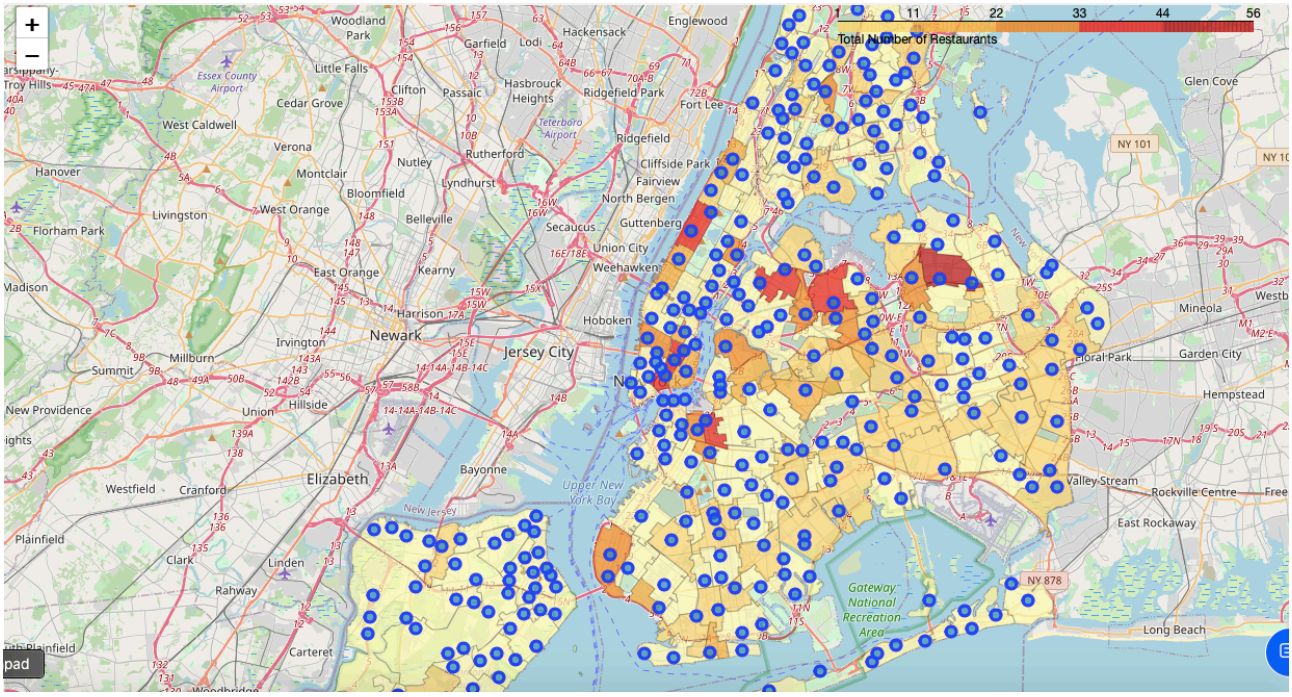
For the other two, we can use the Fourstone API calls to get the relevant information.

Methodology

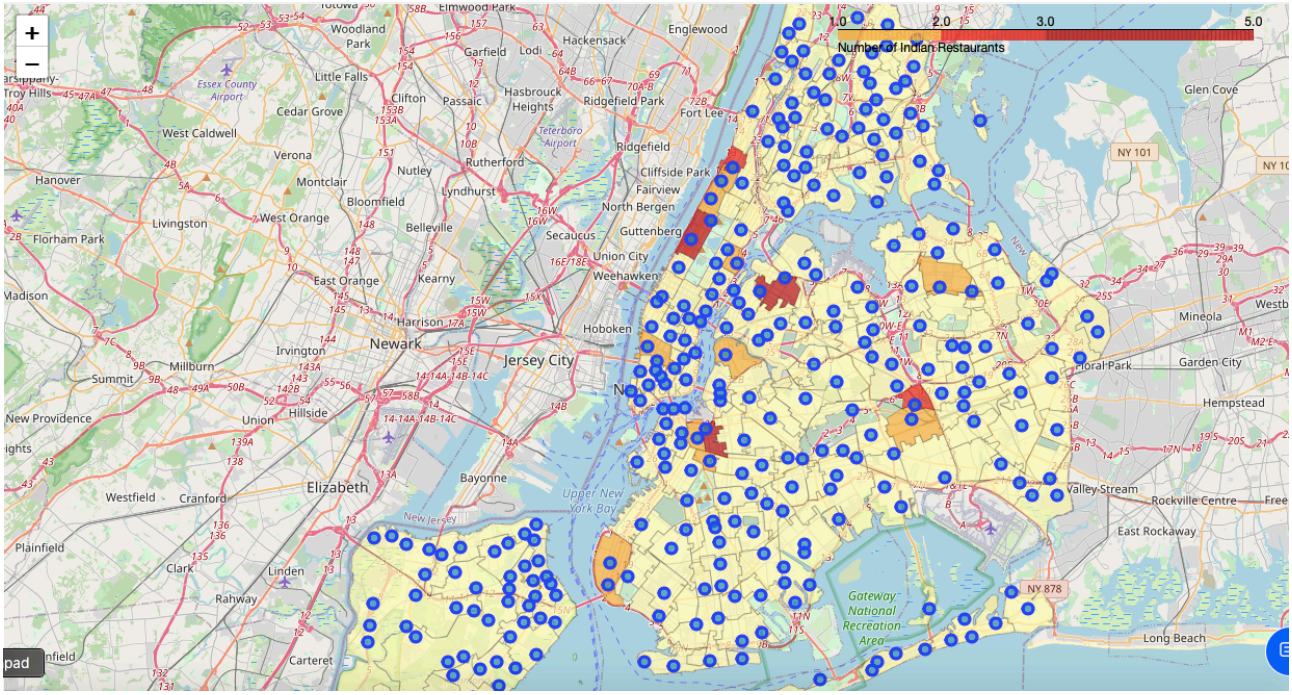
- We start by getting the neighbourhood data for the city of NYC, i.e., the coordinates of the various neighbourhoods along with the boroughs which contain them.
- We then get the venue data from Foresquare for all the neighbourhoods in the city.
- We isolate the data about the restaurants from this venue data in order to further study it.
- We further create a new data frame for just the Indian restaurants in each neighbourhood.
- We create a Choropleth map of the city using the neighbourhood boundaries json and the count of the total number of restaurants in each neighbourhood.
- We create a similar Chloropleth map for the Indian restaurants in each neighborhood.
- In order to find the locations with the least number of Indian restaurants but the most number of total restaurants, we create a new attribute as the ratio of the number of Indian restaurants to that of total restaurants.
- We create a Chloropleth map for this ratio as well.
- We segregate the locations with the least number of Indian restaurants.
- We segregate the locations with the most number of total restaurants from the above list.
- We shortlist the top five from the above list and look at the other venues in the neighbourhood.
- We select for the most neutral venue in terms of specialisation as the optimal location.

Results

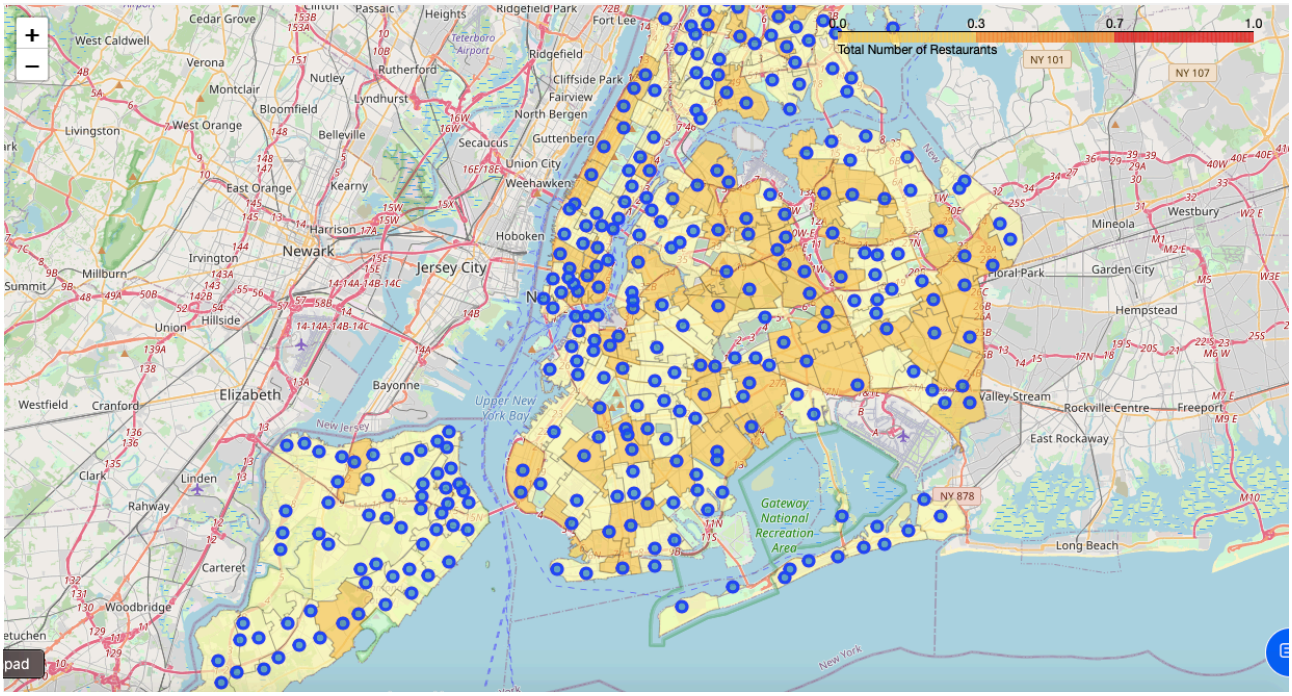
Choropleth map of the count of the total number of restaurants in each neighbourhood:



Chloropleth map for the Indian restaurants in each neighborhood:



We see a lot of neighbourhoods that have a low number of Indian restaurants also have a low number of total restaurants. These areas might not be the ones which have a high demand for restaurants and thus are not important to us.
In order to get highlight the areas with low number of Indian but high number of total restaurants, we make a new column called 'Ratio'
Ratio is defined as the number of Indian restaurants in the area / number of total restaurants in the area.
Chloropleth map for this ratio:



Looks like a lot of areas lie in the lower ranges of the ratio column. One way to go ahead is to look at only the areas with the lowest number of Indian Restaurants, i.e. 0
In order to shortlist the number of neighbourhoods from the above, we can build on the idea that we should concentrate on the areas which have a high number of total restaurants but a low number of Indian restaurants (in this case, zero). This would ensure that there is enough demand in the area for food in general but no competitors for Indian food.

Shortlist:

	Neighborhood	Count Indian	Count Total	Ratio
144	Jackson Heights	0.0	39	0.0
77	Chinatown	0.0	37	0.0
101	East Village	0.0	35	0.0
156	Little Italy	0.0	32	0.0
230	South Side	0.0	30	0.0

Final analysis on these 5 shortlisted venues:

One reason for no occurence of an Indian restaurant in these locations could be because these might be popular for a certain cuisine and thus the demand for food is also majorly for that one specialised cuisine.

It would be best for us to find the neighborhood which is the most neutral in terms of the frequency of occurence of different types of venues. In order to determine the most neutral location, let's look at the venue with the highest frequency of occurence as a percentage.

1. Chinatown: Chinese restaurants comprise 8% of total venues in the neighborhood.
2. East Village: Bars comprise 7% of total venues in the neighborhood.
3. Jackson Heights: Latin American restaurants comprise 11% of total venues in the neighborhood.
4. Little Italy: Bakeries comprise 6% of total venues in the neighborhood.
5. South side: Bars comprise 8% of total venues in the neighborhood.

Conclusion

Thus, we can conclude that **Little Italy** and **East Village** are the optimal neighborhoods to open an Indian Restaurant