THE BATTLE OF NEIGHBORHOODS

APPLIED DATA SCIENCE CAPSTONE
PROJECT PRESENTATION

INTRODUCTION

- PROBLEM STATEMENT: Stakeholders of a new restaurant chain want to study the region of New York City, specifically the region of Manhattan to open a new North Indian restaurant.
- Since the Indian population around the region has increased in the recent years as well as the fact that travelling tourists also like Indian cuisine, opening an Indian restaurant in this region can be profitable.

TARGET AUDIENCE & LOCATION PREFERENCES

- Since we are dealing with the Manhattan region, we are obviously dealing with affluent businessmen who can afford to open a large-scale restaurant chain in such areas.
- Also, they need to be in areas where maximum people are satisfied.
- So, the targeted audience is the Indian tourists, workers also and anyone who loves North Indian cuisine.
- We need to see that the restaurant opened has a substantial Indian population living or working around in the neighborhood or it is opened in tourist places which attract many Indians or people loving Indian food

DATA

- As mentioned in the report, we use two basic data sources for the project like the ones used in the week three lab session. They are:
  - JSON file from the appropriate website providing location data for creating maps in our python notebook. The link is provided here - https://geo.nyu.edu/catalog/nyu_2451_34572
  - We also use the Foursquare Venues list to extract the category data pertaining to the Indian restaurants in the region. The details are as follows:

A) From Foursquare Venues Categories - https://developer.foursquare.com/docs/resources/categories

B) Indian Restaurant ID - 4bf58dd8d48988d10f941735

METHODOLOGY

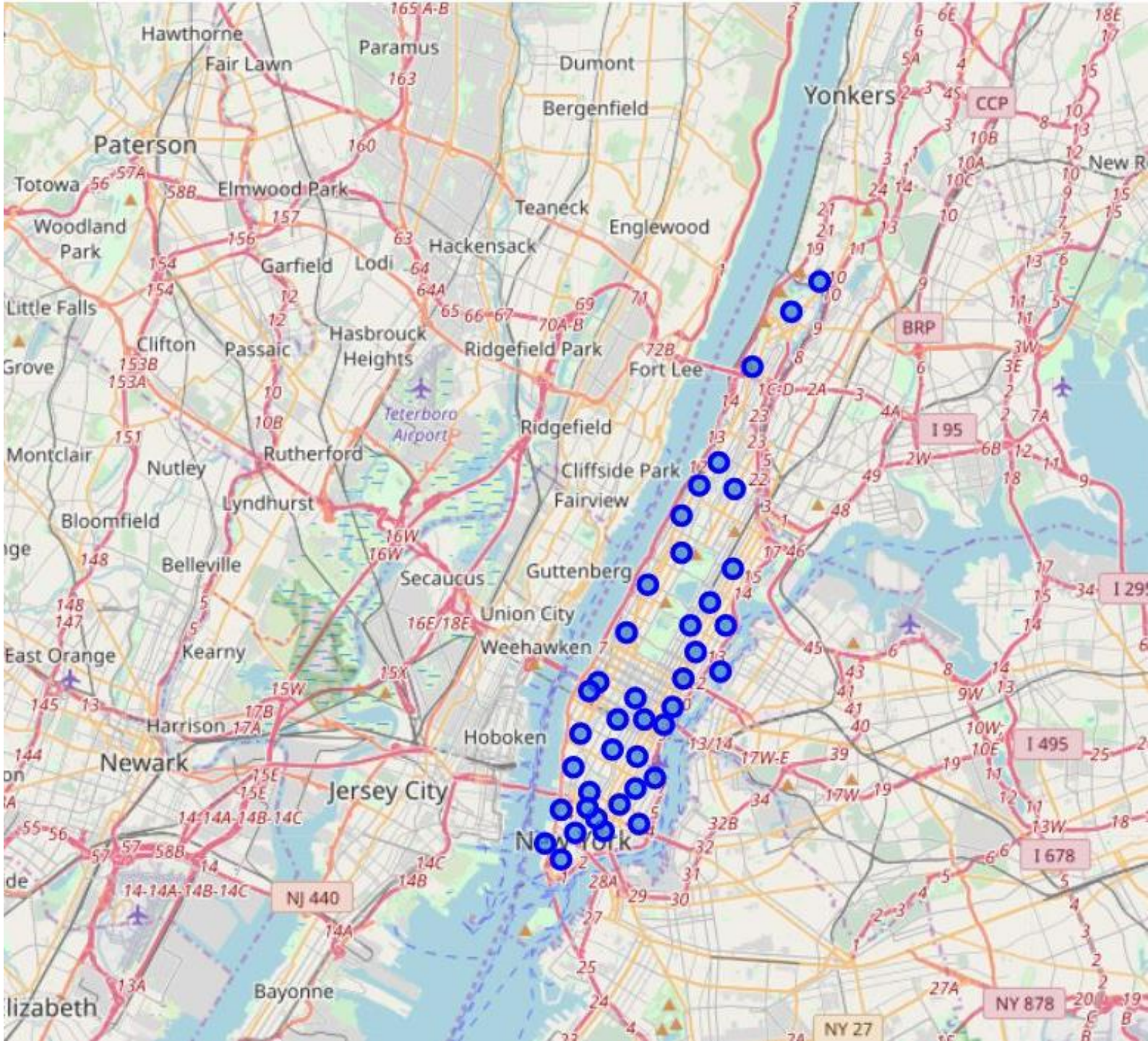
- The first step in the analysis is to load the initial data from the JSON file to a pandas data frame. On doing that we get a tabular format of data displaying each neighborhood in New York City along with its location co-ordinates i.e. longitudes and latitudes.

```
neighborhoods.head()
```

	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

METHODOLOGY

- The next step is to plot the map of Manhattan just to explain the geography of the region to the management team. This gives a good idea, about the exact location of the neighborhoods and which locations would have a higher land cost.



METHODOLOGY

- Since we are interested only in the Manhattan region, we can use the appropriate co-ordinates to focus our attention on that region. This can be done by using either a GeoCoder or manually by entering the latitude and longitude of the place. The neighborhoods are then captured in a pandas data frame.

```
man_data = neighborhoods[neighborhoods['Borough'] == 'Manhattan'].reset_index(drop=True)
man_data.head()
```

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

METHODOLOGY

- The next step is one of the most important step as we use the Foursquare credentials and API database in order to extract the data of existing Indian restaurants in the Manhattan region. We insert the appropriate 'Category Id' in the code to get the names of the Indian restaurants in the pandas data frame.

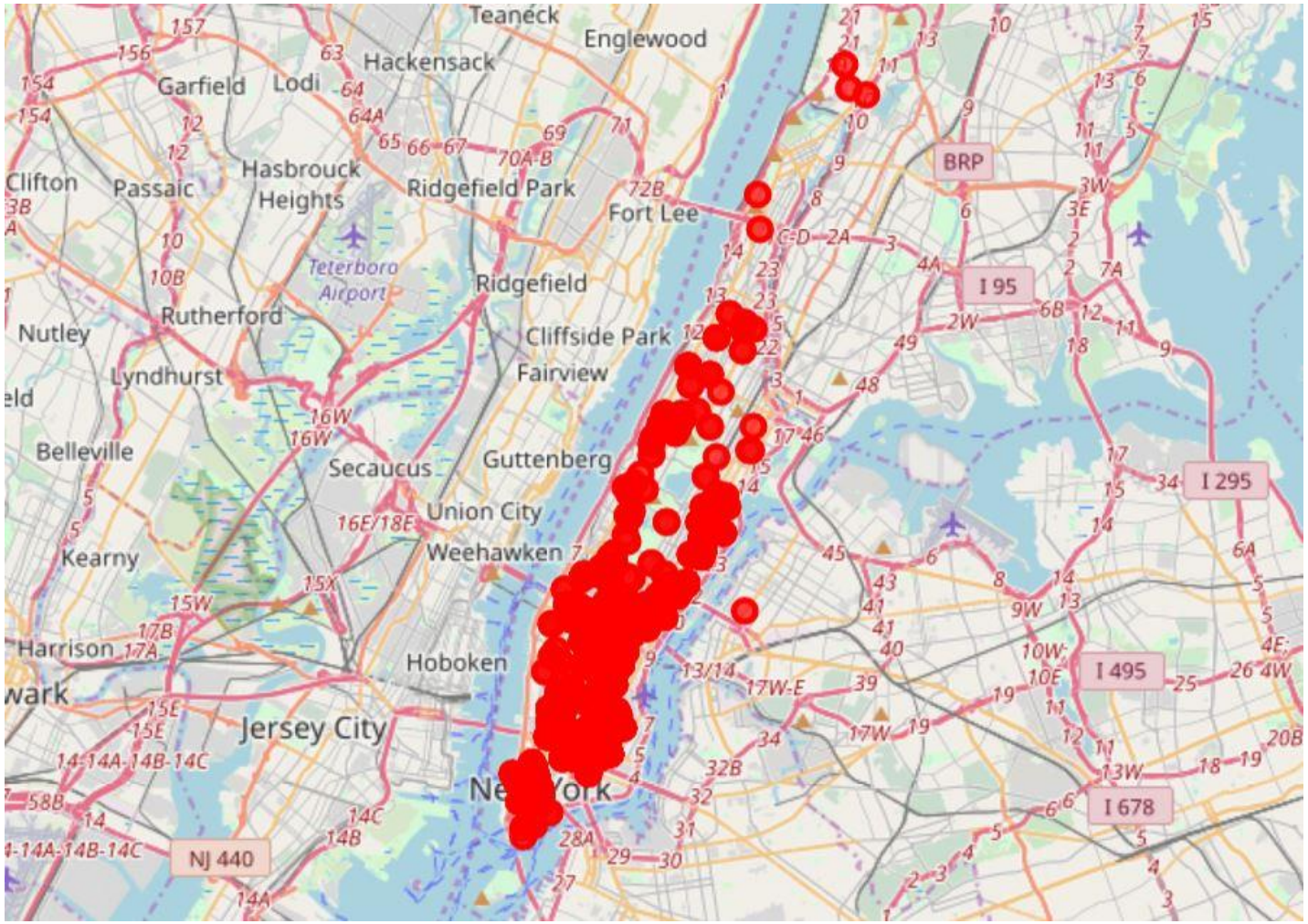
[11]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.910660	Riverdale Indian Cuisine	40.880886	-73.908800	Indian Restaurant
1	Marble Hill	40.876551	-73.910660	Tazmohol Indian Restaurant	40.879331	-73.903192	Indian Restaurant
2	Marble Hill	40.876551	-73.910660	Cumin Indian Cuisine	40.886459	-73.909816	Indian Restaurant
3	Chinatown	40.715618	-73.994279	Kabab Bites	40.720094	-73.995819	Indian Restaurant
4	Chinatown	40.715618	-73.994279	indi thai	40.719830	-73.990350	Indian Restaurant

[12]:
```
newyork_venues_ind.shape
```

[12]:
```
(1090, 7)
```

LOCATIONS OF 1090 INDIAN RESTAURANTS IN MANHATTAN



METHODOLOGY

- In the next step, we perform one-hot encoding as performed in the week 3 lab and then carry out frequency analysis which gives us neighborhood-wise results describing which are the most visited restaurants in the region.

```
----Battery Park City----
                 venue  freq
0  Indian Restaurant  0.84
1          Food Truck  0.08
2   Asian Restaurant  0.04
3   Tapas Restaurant  0.04
4      Burrito Place  0.00
```

```
----Carnegie Hill----
                       venue  freq
0         Indian Restaurant  0.92
1  North Indian Restaurant  0.08
2          Asian Restaurant  0.00
3             Burrito Place  0.00
4               Chaat Place  0.00
```

METHODOLOGY

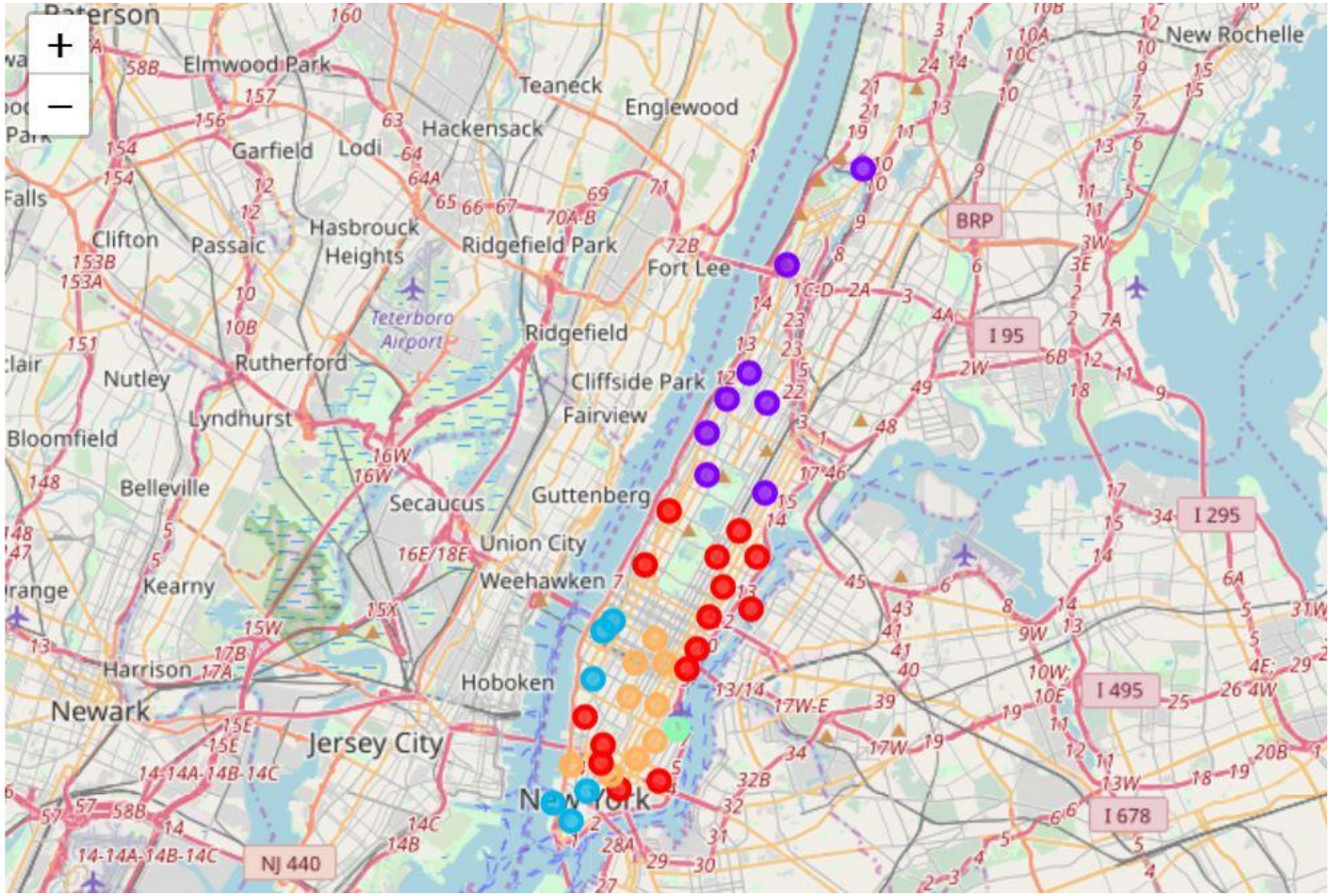
- The previous data can be grouped for each neighborhood and represented as a pandas data frame as follows:

```
manhattan_merged.head() # check the last columns!
```

	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
0	Manhattan	Marble Hill	40.876551	-73.910660	1.0	Indian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant
1	Manhattan	Chinatown	40.715618	-73.994279	0.0	Indian Restaurant	North Indian Restaurant	Dosa Place	Deli / Bodega	Vegetarian / Vegan Restaurant
2	Manhattan	Washington Heights	40.851903	-73.936900	1.0	Indian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant
3	Manhattan	Inwood	40.867684	-73.921210	NaN	NaN	NaN	NaN	NaN	NaN
4	Manhattan	Hamilton Heights	40.823604	-73.949688	1.0	Indian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant

METHODOLOGY

- The last step in the neighborhood analysis is the **k-means clustering** which is carried out to group certain neighborhoods and see which of them exhibit more demand for Indian cuisines in the most common venues tab. After performing k-means clustering we get the following visual results:



RESULTS

- The results are mainly based on the k-means clustering analysis and depict a group of neighborhoods. On individually assessing the clusters we get the following results:
- Cluster 1 consists of the following Neighborhoods:

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 0, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

	Neighborhood	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
1	Chinatown	Indian Restaurant	North Indian Restaurant	Dosa Place	Deli / Bodega	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	Indian Chinese Restaurant	Hookah Bar
7	Upper East Side	Indian Restaurant	North Indian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	Indian Chinese Restaurant	Hookah Bar	Himalayan Restaurant	Food Truck
8	Yorkville	Indian Restaurant	North Indian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	Indian Chinese Restaurant	Hookah Bar	Himalayan Restaurant	Food Truck
9	Lenox Hill	Indian Restaurant	North Indian Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	Indian Chinese Restaurant	Hookah Bar	Himalayan Restaurant
10	Roosevelt Island	Indian Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar	Himalayan Restaurant
11	Upper West Side	Indian Restaurant	South Indian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar	Himalayan Restaurant	Food Truck
12	Lincoln Square	Indian Restaurant	South Indian Restaurant	Food Truck	Vegetarian / Vegan Restaurant	Tapas Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar	Himalayan Restaurant
17	Greenwich Village	Indian Restaurant	North Indian Restaurant	Food Truck	Dosa Place	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	Indian Chinese Restaurant	Hookah Bar
19	Lower East Side	Indian Restaurant	Vegetarian / Vegan Restaurant	North Indian Restaurant	Deli / Bodega	Chaat Place	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	Indian Chinese Restaurant	Hookah Bar
22	Soho	Indian Restaurant	Dosa Place	North Indian Restaurant	Food Truck	Deli / Bodega	Asian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant
23	West Village	Indian Restaurant	Indian Chinese Restaurant	Food Truck	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Hookah Bar	Himalayan Restaurant
29	Carnegie Hill	Indian Restaurant	North Indian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	Indian Chinese Restaurant	Hookah Bar	Himalayan Restaurant	Food Truck
33	Sutton Place	Indian Restaurant	Food Truck	Asian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar
34	Turtle Bay	Indian Restaurant	Food Truck	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar	Himalayan Restaurant
35	Tudor City	Indian Restaurant	Food Truck	South Indian Restaurant	Chaat Place	Burrito Place	Vegetarian / Vegan Restaurant	Tapas Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant

RESULTS

- Cluster 2 consists of the following Neighborhoods:

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 1, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

	Neighborhood	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	Marble Hill	Indian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar	Himalayan Restaurant	Food Truck
2	Washington Heights	Indian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar	Himalayan Restaurant	Food Truck
3	Hamilton Heights	Indian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar	Himalayan Restaurant	Food Truck
4	Manhattanville	Indian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar	Himalayan Restaurant	Food Truck
5	Central Harlem	Indian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar	Himalayan Restaurant	Food Truck
6	East Harlem	Indian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar	Himalayan Restaurant	Food Truck
24	Manhattan Valley	Indian Restaurant	Himalayan Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar	Food Truck
25	Morningside Heights	Indian Restaurant	Himalayan Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar	Food Truck

RESULTS

- Cluster 3 consists of the following Neighborhoods:

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 2, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

	Neighborhood	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
13	Clinton	Indian Restaurant	Food Truck	South Indian Restaurant	Diner	Vegetarian / Vegan Restaurant	Tapas Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar
16	Chelsea	Indian Restaurant	Food Truck	Vegetarian / Vegan Restaurant	Indian Chinese Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Hookah Bar	Himalayan Restaurant
27	Battery Park City	Indian Restaurant	Food Truck	Tapas Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar
28	Financial District	Indian Restaurant	Food Truck	Tapas Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar
31	Civic Center	Indian Restaurant	Food Truck	Tapas Restaurant	Dosa Place	Asian Restaurant	Vegetarian / Vegan Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant
38	Hudson Yards	Indian Restaurant	Food Truck	South Indian Restaurant	Diner	Vegetarian / Vegan Restaurant	Tapas Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant	Hookah Bar

RESULTS

- Cluster 4 consists of the following Neighborhoods:

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 3, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

	Neighborhood	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
36	Stuyvesant Town	Indian Restaurant	Vegetarian / Vegan Restaurant	North Indian Restaurant	Dosa Place	Deli / Bodega	Chaat Place	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant	Indian Chinese Restaurant

- Cluster 5 consists of the following Neighborhoods:

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 4, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

	Neighborhood	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
14	Midtown	Indian Restaurant	Food Truck	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Diner	Vegetarian / Vegan Restaurant	Tapas Restaurant	Indian Chinese Restaurant	Hookah Bar
15	Murray Hill	Indian Restaurant	South Indian Restaurant	Food Truck	Vegetarian / Vegan Restaurant	North Indian Restaurant	Diner	Tapas Restaurant	Pakistani Restaurant	Indian Chinese Restaurant	Hookah Bar
18	East Village	Indian Restaurant	Dosa Place	Deli / Bodega	Vegetarian / Vegan Restaurant	North Indian Restaurant	Food Truck	Chaat Place	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant
20	Tribeca	Indian Restaurant	Food Truck	Tapas Restaurant	Dosa Place	Asian Restaurant	Vegetarian / Vegan Restaurant	South Indian Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant
21	Little Italy	Indian Restaurant	Dosa Place	North Indian Restaurant	Food Truck	Deli / Bodega	Asian Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant
26	Gramercy	Indian Restaurant	Vegetarian / Vegan Restaurant	South Indian Restaurant	North Indian Restaurant	Hookah Bar	Food Truck	Deli / Bodega	Chaat Place	Burrito Place	Tapas Restaurant
30	Noho	Indian Restaurant	Dosa Place	Vegetarian / Vegan Restaurant	North Indian Restaurant	Food Truck	Deli / Bodega	Chaat Place	Tapas Restaurant	South Indian Restaurant	Pakistani Restaurant
32	Midtown South	Indian Restaurant	South Indian Restaurant	Food Truck	Vegetarian / Vegan Restaurant	Diner	Burrito Place	Tapas Restaurant	Pakistani Restaurant	North Indian Restaurant	Indian Chinese Restaurant
37	Flatiron	Indian Restaurant	South Indian Restaurant	Food Truck	Vegetarian / Vegan Restaurant	North Indian Restaurant	Hookah Bar	Diner	Deli / Bodega	Burrito Place	Tapas Restaurant

DISCUSSIONS

- Now reading the tabular results for the clusters from the 'Results' section we see that in cluster 1, the 2nd most common venues is indeed North Indian Restaurants, the topic of our interest. On further inspection we see that this cluster is represented by the 'red' color on the map. These are heavily congested regions and naturally the prices of these locations will be on the higher side.
- However, if we look at the ***Upper West Side*** Neighborhood, we see that the relative density of Indian restaurants is low, yet the demand for Indian food is higher than other regions. Thus, this can be one of the potential sites for a North Indian Restaurant if adequate funds are available.
- The next region belongs to cluster 2 as there is relatively a smaller number of Indian restaurants as compared to the first cluster. Also, there is less congestion in this region which means that the land prices will be less as compared to cluster 1 neighborhoods.
- Thus, regions like ***Harlem, Upper Manhattan and Washington Heights*** are good neighborhoods to open a new Indian restaurant.

DISCUSSIONS

- There are 3 more clusters to be considered. Although cluster 3 and 5 exhibit good demand for Indian food, it is slightly risky to open a North Indian restaurant as they appear in the 4th or 5th most common venues in the list. Also, since these neighborhoods lie in the Central Manhattan region, the prices would be the highest. Thus, it would be a bad idea to open a new restaurant here.
- Cluster 4 doesn't display enough data or statistical evidence from which we can conclude that we should/shouldn't open a new restaurant her as it has only one data point.

CONCLUSION

- As far as this project is concerned, we were able to decide a rough or approximate region/neighborhood where we can open up a new North Indian restaurant by *analytically* studying the behavior and preferences of customers and *qualitatively* making decisions on the land price and competitions by looking and studying the cluster maps produced by our code.
- Although the goals of the project can be met on a higher level, like we have done in this project, this project can be refined further by providing more data to the algorithm like population data, ethnicity data, tourist information and eating preferences related to tourists as well as the working class travelling to Manhattan.