Prediction of a suitable location for opening a restaurant in Mumbai

By

Ice Bear

1.Introduction:

1.1 Background:

Mumbai also known as **Bombay** is the capital of the Indian state of Maharashtra. According to United Nations, as of 2018, Mumbai was the second city in India after Delhi and the seventh most populous city in the world with a population of roughly 20 million. As per Indian government population census of 2011, Mumbai was the most populous in India with an estimated city proper population of 12.5 million living under Municipal Corporation of Greater Mumbai . Mumbai is the centre of the Mumbai Metropolitan Region, the sixth most populous metropolitan area in the world with a population of over 23 million. Mumbai is also the Financial, Commercial, and the Entertainment capital of India generating about 6.16% of India's GDP alone. So it wouldn't be a surprise if a large corporate or even an individual would be interested in opening a restaurant or a restaurant chain here given the fact that there is a good sizable population that is providing a healthy contribution to this growing financial workhorse of India with a respectable purchasing power.

1.2 Problem:

With a total area of 603.4 km^2 the city has it's fair share of economic inequality and not all areas and corners are same in terms of economic and urban growth. This makes it quite a challenge to choose locations that may increase the success of the restaurant based solely on these locations. Therefore, This project aims to mitigate this problem by providing several locations and inferring a best possible location for the head start of the restaurant.

1.3 Interest:

Individuals that are interested in opening a restaurant or a large corporate that wants to start a restaurant chain would be interested in this project.

2.Data acquisition and cleaning:

1.1 Data Sources:

For the Project, we first need to have a data that shows all the Suburban and Urban areas in Mumbai along with their respective postal codes which can be found here. After this, by web-scraping using BeautifulSoup , the data was obtained in the notebook in the form of a list but it was not in a tabular form that I expected which was needed in order to carry out analysis.

```
Out[2]: ['Antop Hill\n400037',
         'B P T Colony\n400037',
         'Haffkin Institute\n400012',
         'Mazgaon Dock\n400010',
         'Lal Baug\n400012',
         'Parel Rly Work Shop\n400012',
         'Princess Dock\n400009',
         'Reay Road\n400033',
         'Cotton Exchange\n400033',
         'Dadar Colony\n400014',
         'V K Bhavan\n400010',
         'Wadala Rs\n400031',
         'Wadala Truck Terminal\n400037',
         'BEST STaff Quarters\n400012',
         'Dadar H.O\n400014',
         'Kalachowki\n400033',
         'Noor Baug\n400009',
         'Sewri\n400015',
         'V J B Udyan\n400027',
         'B.P.Lane\n400003',
         'Chinchbunder H.O\n400009',
         'Masjid\n400003',
         'Mazgaon Road\n400010',
         'Parel\n400012',
         'Tank Road\n400033',
         'L B S N E collage\n400033',
         'Mazgaon\n400010',
         'Chamarbaug\n400012',
         'Kidwai Nagar (Mumbai)\n400031',
         'Naigaon (Mumbai)\n400014',
         'Null Bazar\n400003',
         'Parel Naka\n400012',
         'Wadala\n400031',
         'C G S Colony\n400037',
         'Dockyard Road\n400010',
```

Fig 1:Format of all Areas/Neighborhoods of Mumbai in the list

1.2 Data Cleansing:

As the integers and strings were together in the form of a string in the list, the objective was to first separate them and represent them separately which was achieved by importing and using re library for splitting of strings and integers in the form of list and thus I was able to obtain a much more refined look.

```
Out[3]: [['Antop Hill\n', '400037', ''],
         ['B P T Colony\n', '400037', ''],
         ['Haffkin Institute\n', '400012', ''],
         ['Mazgaon Dock\n', '400010', ''],
         ['Lal Baug\n', '400012', ''],
         ['Parel Rly Work Shop\n', '400012', ''],
         ['Princess Dock\n', '400009', ''],
         ['Reay Road\n', '400033', ''],
         ['Cotton Exchange\n', '400033', ''],
         ['Dadar Colony\n', '400014', ''],
         ['V K Bhavan\n', '400010', ''],
         ['Wadala Rs\n', '400031', ''],
         ['Wadala Truck Terminal\n', '400037', ''],
         ['BEST STaff Quarters\n', '400012', ''],
         ['Dadar H.O\n', '400014', ''],
         ['Kalachowki\n', '400033', ''],
         ['Noor Baug\n', '400009', ''],
         ['Sewri\n', '400015', ''],
         ['V J B Udyan\n', '400027', ''],
         ['B.P.Lane\n', '400003', ''],
         ['Chinchbunder H.O\n', '400009', ''],
         ['Masjid\n', '400003', ''],
         ['Mazgaon Road\n', '400010', ''],
         ['Parel\n', '400012', ''],
         ['Tank Road\n', '400033', ''],
         ['L B S N E collage\n', '400033', ''],
         ['Mazgaon\n', '400010', ''],
         ['Chamarbaug\n', '400012', ''],
         ['Kidwai Nagar (Mumbai)\n', '400031', ''],
         ['Naigaon (Mumbai)\n', '400014', ''],
         ['Null Bazar\n', '400003', ''],
         ['Parel Naka\n', '400012', ''],
         ['Wadala\n', '400031', ''],
         ['C G S Colony\n', '400037', ''],
         ['Dockyard Road\n', '400010', ''],
```

Fig 2:List within a List

From here on, using For Loop I was able to form a separate list for both string and integers i.e a list for Postal Code and a list for Neighborhood.

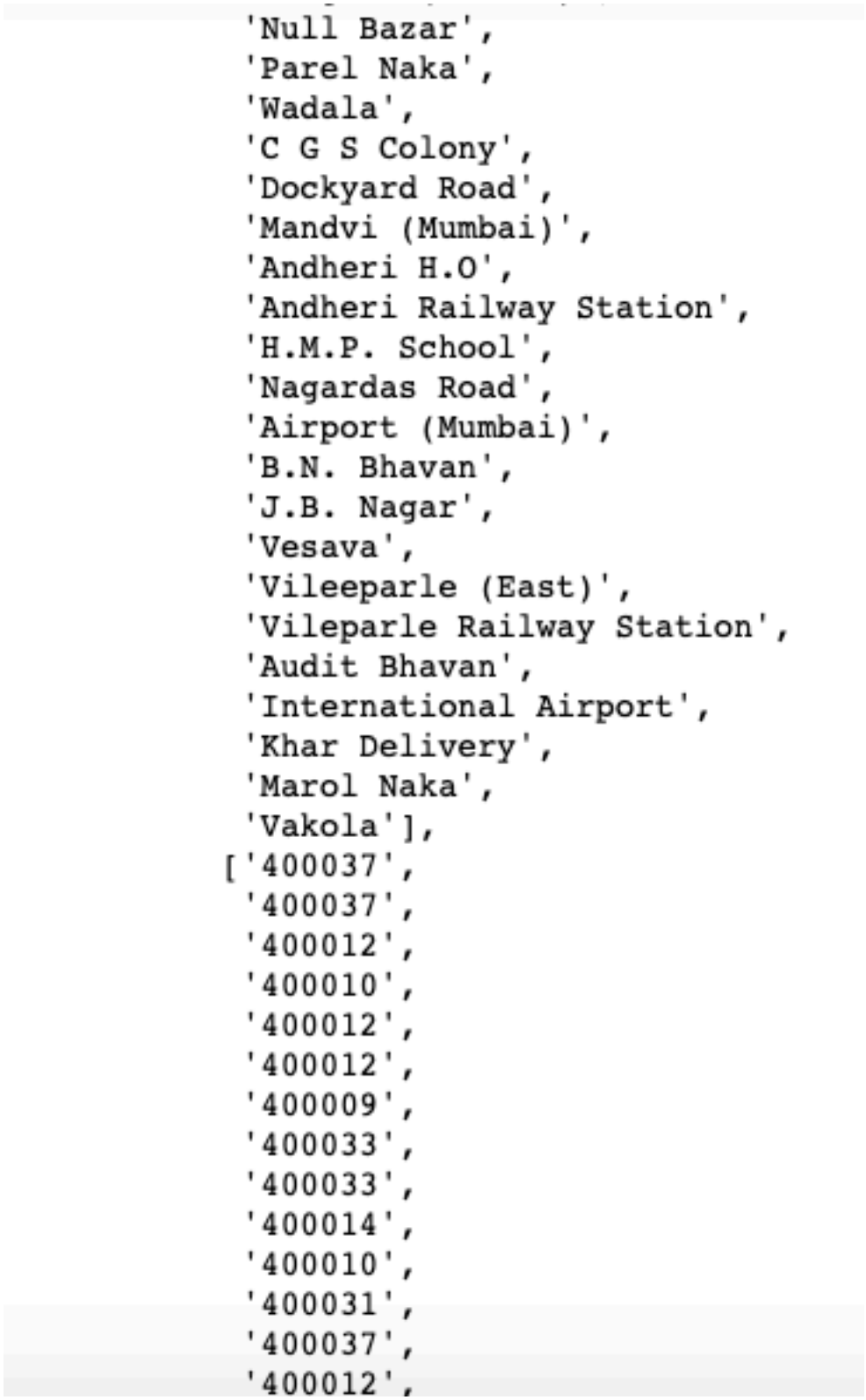

Fig 3:Separate lists for Neighborhoods and Postal Codes

The only thing left is to arrange them in a tabular form which would then be used as a basis for further analysis.

Out[5]:

	Postal Code	Neighborhood
0	400037	Antop Hill
1	400037	B P T Colony
2	400012	Haffkin Institute
3	400010	Mazgaon Dock
4	400012	Lal Baug
5	400012	Parel Rly Work Shop
6	400009	Princess Dock
7	400033	Reay Road
8	400033	Cotton Exchange
9	400014	Dadar Colony
10	400010	V K Bhavan
11	400031	Wadala Rs
12	400037	Wadala Truck Terminal
13	400012	BEST STaff Quarters
14	400014	Dadar H.O
15	400033	Kalachowki
16	400009	Noor Baug
17	400015	Sewri
18	400027	V J B Udyan
19	400003	B.P.Lane
20	400009	Chinchbunder H.O

Fig 4: Table containing Postal Code and Area

1.3 Adding Features:

After the data has been cleansed, we need to find longitude and latitude for each of our neighbors in order to carry out our analysis which can easily be obtained by using various geocoding APIs which are out there but for this experiment I have decided to use OpenCage geocoding API. So, by using the appropriate code and passing down the keys I was able to obtain the list which contains both longitude and latitude.

```
Out[7]: ([19.0230741,
          19.0230741,
          18.996311,
          18.9716102,
          19.0011989,
          18.996311,
          18.9579025,
          18.9818519,
          18.9818519,
          19.010619,
          18.9685234,
          19.022172,
          19.0230741,
          19.0011989,
          19.010619,
          18.9818519,
          18.9579025,
          18.9997454,
          18.9769335,
          18.9571871,
          18.9579025,
          18.9558649,
          18.9716102,
          19.0011989,
          18.9818519,
          18.9817798,
          18.9716102,
          19.0011989,
          19.022172,
          19.0212398,
          18.9558649,
          19.0011989,
          19.022172,
          19.0230741,
          18.9716102,
```

Fig 5:List of Longitude and Latitude

After this just append these into the original table so that we can move towards our analysis part.

Out[8]:

	Postal Code	Neighborhood	Latitude	Longitude
0	400037	Antop Hill	19.023074	72.867622
1	400037	B P T Colony	19.023074	72.867622
2	400012	Haffkin Institute	18.996311	72.842493
3	400010	Mazgaon Dock	18.971610	72.835622
4	400012	Lal Baug	19.001199	72.840356
5	400012	Parel Rly Work Shop	18.996311	72.842493
6	400009	Princess Dock	18.957902	72.835418
7	400033	Reay Road	18.981852	72.840574
8	400033	Cotton Exchange	18.981852	72.840574
9	400014	Dadar Colony	19.010619	72.846936
10	400010	V K Bhavan	18.968523	72.837844
11	400031	Wadala Rs	19.022172	72.857293
12	400037	Wadala Truck Terminal	19.023074	72.867622
13	400012	BEST STaff Quarters	19.001199	72.840356
14	400014	Dadar H.O	19.010619	72.846936
15	400033	Kalachowki	18.981852	72.840574
16	400009	Noor Baug	18.957902	72.835418
17	400015	Sewri	18.999745	72.851490
18	400027	V J B Udyan	18.976934	72.835154
19	400003	B.P.Lane	18.957187	72.832541

Fig 6:Final Base Table

3.Data Analysis:

3.1 Nearby Values:

In order to find the suitable location for a restaurant we need to first define our ideal location. An ideal location for our restaurant startup would be where we can expect customers that have come primarily to eat in the neighborhood. Since our restaurant is new, we can't expect our customers to go out of their way and try it out but what we can expect is that if customers are in the neighborhood to eat at their usual places and they get to see our restaurant then they may be tempted to try it out. So our primary objective should be to find neighborhoods that has food related venues as their most common venues.

For this, we need to first find all the nearby venues for each neighborhood. This was done using Foursquare API by defining a function. By passing all the required credentials of the API in the function I was able to find all the nearby venues of each neighborhood with limit set to 100 venues along with their longitude and latitude followed by the category under which they fall.

Out[13]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Antop Hill	19.023074	72.867622	naaz supermarket	19.022164	72.863284	Grocery Store
1	B P T Colony	19.023074	72.867622	naaz supermarket	19.022164	72.863284	Grocery Store
2	Haffkin Institute	18.996311	72.842493	ITC Grand Central	18.998469	72.838433	Hotel
3	Haffkin Institute	18.996311	72.842493	Terrace Garden	18.998119	72.838529	Roof Deck
4	Haffkin Institute	18.996311	72.842493	Hornby's Pavilion	18.998141	72.838419	Restaurant

Fig 7:Table consisting of all nearby venues

After this, we can group each neighborhood and count the total number of venue category under each of these neighborhoods followed by finding out the frequency of each this category with which they occur. With this, the top 10 most common venues was found out for each neighborhood.

Out[19]:

	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	Airport (Mumbai)	Ice Cream Shop	Indian Restaurant	Chinese Restaurant	Multiplex	Vegetarian / Vegan Restaurant	Asian Restaurant	Cocktail Bar	Shopping Mall	Food Truck	Camera Store
1	Andheri H.O	Indian Restaurant	Bakery	Dessert Shop	Donut Shop	Pizza Place	Café	Falafel Restaurant	Fast Food Restaurant	Cupcake Shop	Snack Place
2	Andheri Railway Station	Coffee Shop	Sandwich Place	Vegetarian / Vegan Restaurant	Ice Cream Shop	Falafel Restaurant	Cupcake Shop	Department Store	Dessert Shop	Diner	Donut Shop
3	Antop Hill	Grocery Store	Zoo	Flea Market	Cupcake Shop	Department Store	Dessert Shop	Diner	Donut Shop	Falafel Restaurant	Farmers Market
4	Audit Bhavan	Fast Food Restaurant	Pizza Place	Spa	Café	Indian Restaurant	Zoo	Cupcake Shop	Department Store	Dessert Shop	Diner

Fig 8:Table containing the most common venues.

3.2 Clustering:

After finding out the top 10 venues for each neighborhood, we can cluster them using K-means and through each cluster point we can visualize these venues and try to find out the most optimum location based on features that each location offers and thereby narrowing down our search. For clustering, we first need to set a cluster label for each of this neighborhoods that can be seen as follows:

Out[21]:

	Postal Code	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
0	400037	Antop Hill	19.023074	72.867622	2	Grocery Store	Asian Restaurant	Zoo	Fast Food Restaurant	Cricket Ground	Cupcake Shop	Department Store	Dessert Shop	Diner
1	400037	B P T Colony	19.023074	72.867625	2	Grocery Store	Asian Restaurant	Zoo	Fast Food Restaurant	Cricket Ground	Cupcake Shop	Department Store	Dessert Shop	Diner
2	400012	Haffkin Institute	18.996311	72.842493	4	Coffee Shop	Lounge	Chinese Restaurant	Hotel	Hotel Pool	Restaurant	Roof Deck	Cricket Ground	Cupcake Shop
3	400010	Mazgaon Dock	18.968523	72.837844	1	Garden	Snack Place	Café	Ice Cream Shop	Indian Restaurant	Zoo	Falafel Restaurant	Cricket Ground	Cupcake Shop
4	400012	Lal Baug	18.996311	72.842493	4	Coffee Shop	Lounge	Chinese Restaurant	Hotel	Hotel Pool	Restaurant	Roof Deck	Cricket Ground	Cupcake Shop

Fig 9:Setting Cluster Labels for each neighborhood

After this we can plot a map using Folium and visualize our cluster points to make it easier in narrowing down our search for selection of location.

Fig 9:Map containing all Cluster Points

4.Result:

Now since we have a clear map of all cluster points, it has now come down to observation and having a basic geographic knowledge about locations thereby aiming for factors that we would need in order to make our restaurant a success.

First we need to narrow down our search to locations that have most food related venues as their most common venues for reasons we discussed earlier and second we need our cluster point to be close to railway stations because when we talk about Mumbai, it is also famous for it's 'Local trains' which are cheap and easy means of travel around the city for almost everybody and lakhs of people use it almost every single day. Now, there are 3 lines through which these local trains travel. These are Western Line, Central Line & Harbour Line. Now, We would want to go with Western Line cause it has a good mix of Urban and Suburban Neighborhoods.

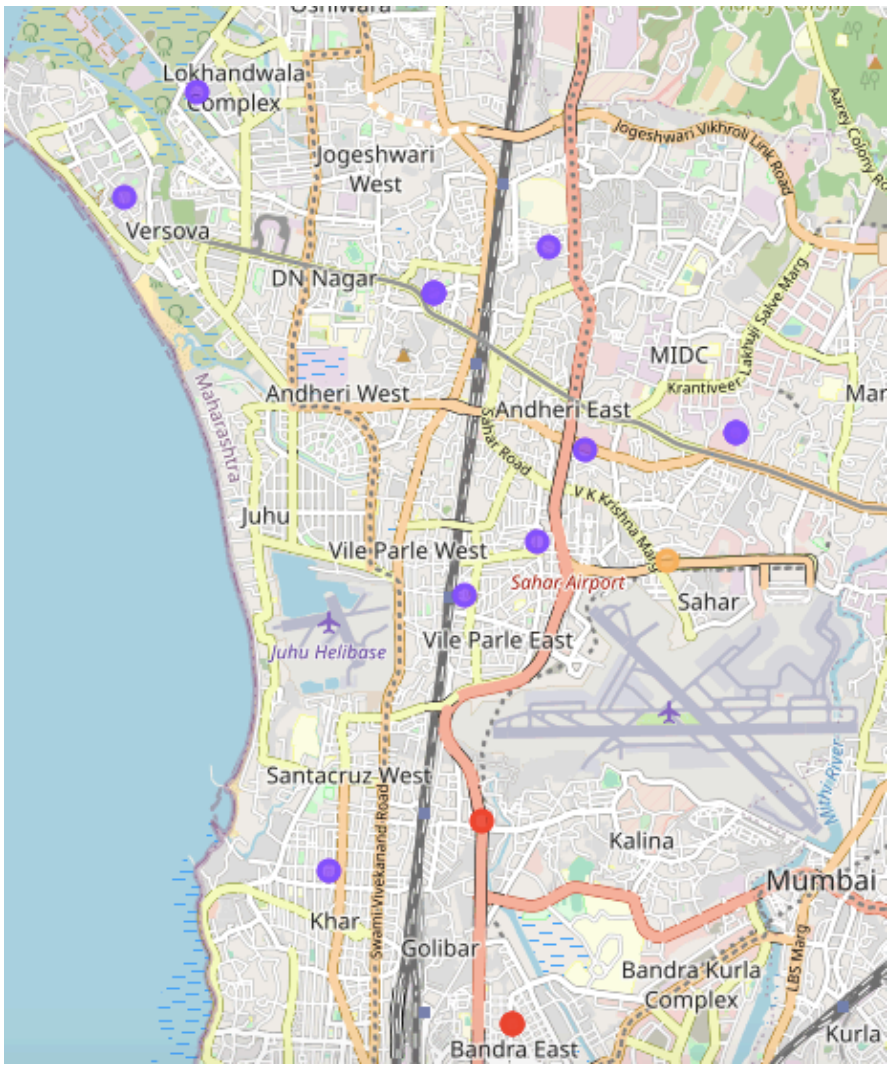

Fig 10:Western Line Cluster Points

As we can see from the map, we do have some potential locations that are relatively close to the railway line meaning they do have good connectivity. But we would want to focus on the area named ‘Vile Parle’ because when we look closely there are two cluster points that are quite close to the actual ‘Vile Parle Station’.

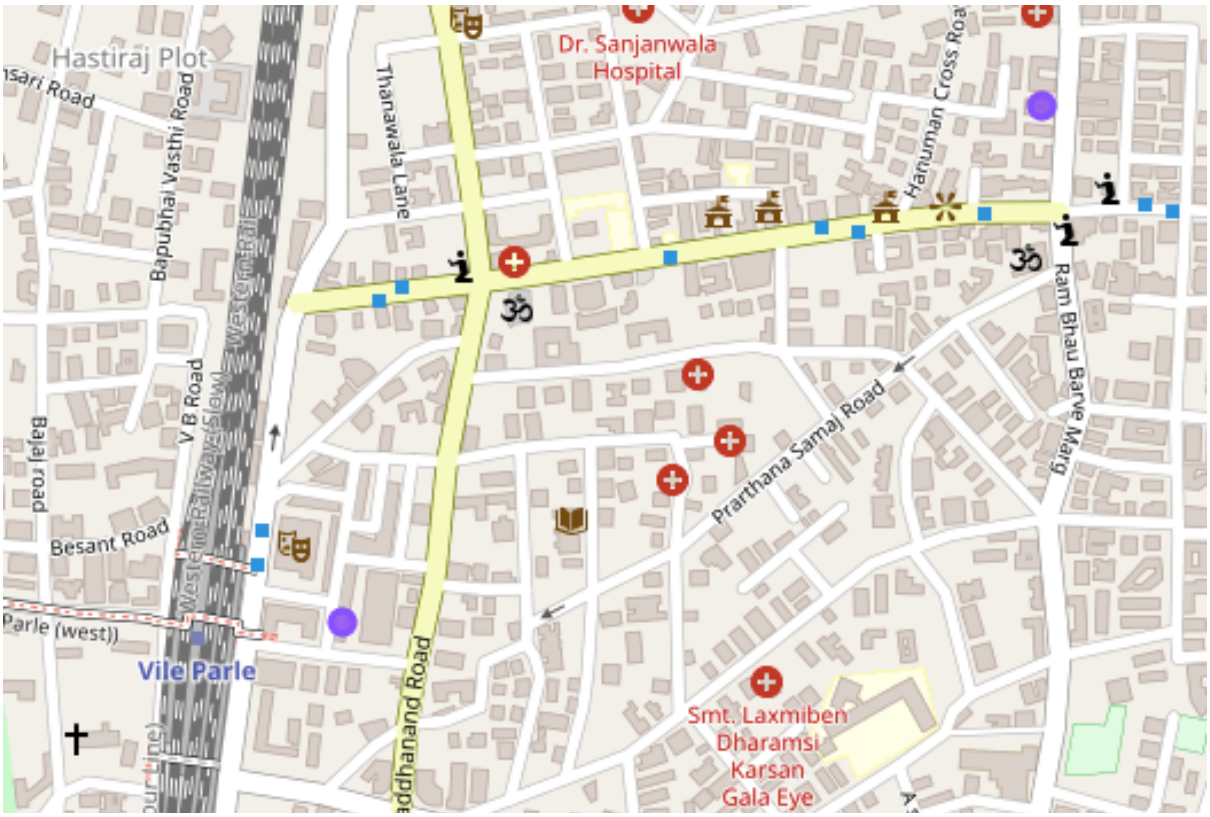

Fig 11:Vile Parle Cluster points

If we look at the most common venues that these points contain then we can observe that both of them have for most food related venues meaning it can be the food locality that we were looking for. On further searching through the area we can also find out that the west side has some well-known colleges which are recognised throughout the city and country as well and there are Hostels and Rented apartments in this area.Apart from this, there is also a famous tradtional drama theatre for which many people attend. So it can be inferred that this can an ideal location for setting up our restaurant.

400057	Vileeparle (East)	19.100066	72.845356	1	Indian Restaurant	Fast Food Restaurant	Café	Sandwich Place	Hotel	Market	Maharashtrian Restaurant	Lounge	Diner	Gift Shop
400057	Vileparle Railway Station	19.104612	72.851914	1	Indian Restaurant	Tea Room	Snack Place	Seafood Restaurant	Juice Bar	Chinese Restaurant	Gym	Gym / Fitness Center	Martial Arts Dojo	Smoke Shop

Fig 12.Most Common Venues of Vile Parle

5.Conclusion:

Thus, based on analysis and then further narrowing the results down based on several factors. I am able to infer that ‘Vile Parle’ might be a suitable location if one wishes to start a restaurant since it does provides an opportunity to grow and does looks promising.