Location for a new Indian restaurant in Mumbai

1. Introduction / Business Problem

I live in the city of Mumbai in India, I have therefore chosen a project that is based in Mumbai.

Mumbai, also known as **Bombay** (the official name until 1995), is the capital city of the Indian state of Maharashtra. As of 2011 (last census) it is the most populous city in India with an estimated city proper population of 12.4 million. The larger Mumbai Metropolitan Region is the second-most-populous metropolitan area in India, with a population of 21.3 million as of 2016. Mumbai lies on the Konkan coast on the west coast of India and has a deep natural harbour. It is also the wealthiest city in India, and has the highest number of millionaires and billionaires among all cities in India.

In Mumbai there are 24 wards. For the convenience of city administration, wards have been decentralized. Each ward has its own ward office with the Ward Officer who is responsible for the municipal services under his area.

1.1. Problem Definition

Taking the municipal wards as the unit, the aim is to find the most optimum ward to open a new Indian restaurant. We would like to choose a ward which is popular for its eateries, has Indian restaurants that are reasonably popular, but as yet not the top in popularity. This will provide us with a good location where the new Indian restaurant can shine.

1.2. Target Audience

A person / group that is planning to open a new Indian restaurant in Mumbai city

2. Data

We need ward wise geometry so that we can get the latitude and longitude for the various wards from it. For this the geoson file was taken from the following location which has the geometry for various municipalities in India
https://github.com/datameet/Municipal_Spatial_Data/blob/master/Mumbai/BMC_Wards.geojson

The geometry of each ward is a polygon with multiple vertices. To obtain the center of the various wards (to use as input to four square api) the mean of all the vertices was used.

Then a dataframe was created with three columns – Ward, Latitude, Longitude

	Ward	Latitude	Longitude
0	A	18.920981	72.827472
1	B	18.956941	72.839720
2	C	18.951097	72.827200
3	D	18.955231	72.808287
4	E	18.973439	72.843214
5	F South	18.998980	72.853965
6	G South	19.004169	72.820140
7	F North	19.031895	72.870740
8	G North	19.030693	72.844038
9	N	19.086679	72.920185
10	R Central	19.233691	72.829735
11	S	19.133590	72.921913
12	T	19.167511	72.937554
13	K West	19.133594	72.781897

To get more meaningful names for the wards the location area name for the wards was taken from http://www.demographia.com/db-mumbaidistr91.htm

The first two columns was taken from this, and made into a csv file through Excel and then read into a dataframe

	Ward	Area
0	A	Colaba
1	B	Sanhurst Road
2	C	Marine Lines
3	D	Grant Road
4	E	Byculla
5	F South	Parel
6	F North	Matunga
7	G South	Elphinstone
8	G North	Dadar/Plaza

The two dataframe was then merged to create the following database

	Ward	Area	Latitude	Longitude
0	A	Colaba	18.920981	72.827472
1	B	Sanhurst Road	18.956941	72.839720
2	C	Marine Lines	18.951097	72.827200
3	D	Grant Road	18.955231	72.808287
4	E	Byculla	18.973439	72.843214
5	F South	Parel	18.998980	72.853965
6	F North	Matunga	19.031895	72.870740
7	G South	Elphinstone	19.004169	72.820140
8	G North	Dadar/Plaza	19.030693	72.844038
9	H East	Khar/Santacruz	19.077906	72.859472

The ward identifier and ward area name was then merged into a single column to be used as label in maps

	WardArea	Latitude	Longitude
0	Ward A - Colaba	18.920981	72.827472
1	Ward B - Sanhurst Road	18.956941	72.839720
2	Ward C - Marine Lines	18.951097	72.827200
3	Ward D - Grant Road	18.955231	72.808287
4	Ward E - Byculla	18.973439	72.843214
5	Ward F South - Parel	18.998980	72.853965
6	Ward F North - Matunga	19.031895	72.870740
7	Ward G South - Elphinstone	19.004169	72.820140
8	Ward G North - Dadar/Plaza	19.030693	72.844038
9	Ward H East - Khar/Santacruz	19.077906	72.859472

This Latitude and longitude will be used with a radius of 500m to get the venues from the Four Square API.

	Ward	Ward Latitude	Ward Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ward A - Colaba	18.920981	72.827472	Theobroma	18.919298	72.829185	Dessert Shop
1	Ward A - Colaba	18.920981	72.827472	New Martin	18.918624	72.829512	Indian Restaurant
2	Ward A - Colaba	18.920981	72.827472	Mad Over Donuts	18.919209	72.829427	Donut Shop
3	Ward A - Colaba	18.920981	72.827472	Café Basilico	18.918609	72.830484	Diner
4	Ward A - Colaba	18.920981	72.827472	Piccadilly	18.921425	72.830936	Falafel Restaurant

3. Methodology

3.1.Exploratory Analysis of Data obtained from Four Square API

The venues obtained from the Four Square APO was then categorized by type like café, Indian restaurant, Chinese restaurant, beach, etc and then sorted based on number of occurrences to determine the most popular location in each ward. A one hot dataframe will be created which can then be analyzed for most popular location in each ward based on frequency.

	Ward	American Restaurant	Arcade	Art Gallery	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bank	Bar	Beach	Bed & Breakfast	Beer Garden	Bistr
0	Ward A - Colaba	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00000
1	Ward B - Sanhurst Road	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00000
2	Ward C - Marine Lines	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00000
3	Ward D - Grant Road	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00000
4	Ward E - Byculla	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00000
5	Ward F North - Matunga	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00000
6	Ward F South - Parel	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00000
7	Ward G North - Dadar/Plaza	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.047619	0.047619	0.000000	0.000000	0.00000
8	Ward G South - Elphinstone	0.000000	0.000000	0.25	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00000
9	Ward H East - Khar/Santacruz	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.333333	0.000000	0.000000	0.000000	0.00000

```
----Ward A - Colaba----
               venue  freq
0  Indian Restaurant  0.15
1               Café  0.12
2              Diner  0.09
3        Coffee Shop  0.06
4  German Restaurant  0.03


----Ward B - Sanhurst Road----
                    venue  freq
0       Indian Restaurant  0.29
1       Convenience Store  0.14
2         Harbor / Marina  0.14
3  Furniture / Home Store  0.14
4          Sandwich Place  0.14
```

Based on the frequency we can create a dataframe of the top 10 popular locations which will be used for further analysis.

	Ward	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	Ward A - Colaba	Indian Restaurant	Café	Diner	Coffee Shop	German Restaurant	Chaat Place	Cosmetics Shop	Middle Eastern Restaurant	Mediterranean Restaurant	Juice Bar
1	Ward B - Sanhurst Road	Indian Restaurant	Furniture / Home Store	Harbor / Marina	Sandwich Place	Smoke Shop	Convenience Store	Diner	Cosmetics Shop	Cupcake Shop	Dance Studio
2	Ward C - Marine Lines	Indian Restaurant	Women's Store	Multiplex	Food	Fast Food Restaurant	Bus Station	Jewelry Store	Diner	Cupcake Shop	Dance Studio
3	Ward D - Grant Road	Indian Restaurant	Coffee Shop	Park	Restaurant	Gastropub	Ice Cream Shop	Clothing Store	Food & Drink Shop	Theater	Bakery
4	Ward E - Byculla	Chinese Restaurant	Dessert Shop	Ice Cream Shop	Indian Restaurant	Women's Store	Donut Shop	Cosmetics Shop	Cupcake Shop	Dance Studio	Deli / Bodega

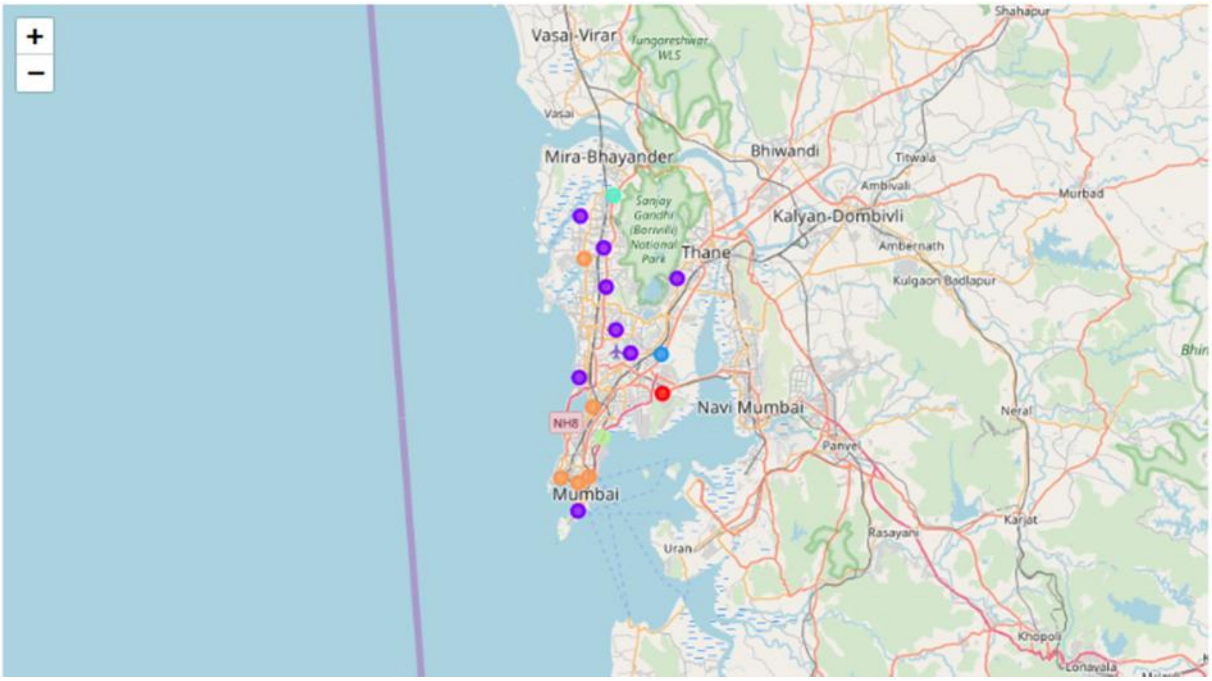
3.2. Clustering

It was decided to use KMeans to group similar wards together. After looking at the optimal number of clusters required a cluster size of 5 was chosen. On creating the 5 clusters and analyzing the same it was found that majority of the wards fell in a single cluster, namely cluster number 0. The other 4 clusters were found to be non-happening areas which was not in our interest

It was therefore decided to extract cluster number 0 as a separate dataframe and to do a clustering exercise again on the subset

4. Results

The second cycle of clustering on the subset from the 1st cycle gave the following clusters



Based on above cluster number 1 best suited for our criteria of having popular Indian restaurants but have a ward that was one where Indian restaurant was popular but not the top venue.

Cluster 1 - Restaurants and eateries

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

	WardArea	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	Ward A - Colaba	Indian Restaurant	Café	Diner	Coffee Shop	German Restaurant	Chaat Place	Cosmetics Shop	Middle Eastern Restaurant	Mediterranean Restaurant	Juice Bar
10	Ward H West - Bandra	Indian Restaurant	Café	Chinese Restaurant	Dessert Shop	Gym / Fitness Center	Arcade	Asian Restaurant	Bakery	Bar	Bookstore
11	Ward K East - Andheri (East)	Hotel	Indian Restaurant	Multiplex	Vegetarian / Vegan Restaurant	Pizza Place	Department Store	Hotel Bar	Burger Joint	Diner	Cosmetics Shop
13	Ward L - Kurla	Café	Indian Restaurant	Coffee Shop	Clothing Store	Fast Food Restaurant	Chinese Restaurant	Restaurant	Dessert Shop	Beer Garden	Shopping Mall
17	Ward P South - Goregaon	Indian Restaurant	Smoke Shop	Chinese Restaurant	Dance Studio	Fast Food Restaurant	Breakfast Spot	Donut Shop	Cosmetics Shop	Cupcake Shop	Deli / Bodega
19	Ward R South - Kandivalli	Park	Moving Target	Café	Ice Cream Shop	Snack Place	Diner	Convenience Store	Cosmetics Shop	Cupcake Shop	Dance Studio
20	Ward R Central - Borivali West	Café	Restaurant	Indian Restaurant	Arcade	Chinese Restaurant	Fast Food Restaurant	Pub	Bus Station	Multiplex	Women's Store
23	Ward T - Mulund	Indian Restaurant	Pizza Place	Fast Food Restaurant	Clothing Store	Cosmetics Shop	Coffee Shop	Food Court	Mobile Phone Shop	Café	Shopping Mall

5. Discussion

Cluster number 1 with the following wards met our criteria the best

- Ward K East – Andheri East – where Indian restaurant was the 2nd most popular
- Ward L – Kurla – where Indian restaurant was the 2nd most popular
- Ward R Central – Borivali West – where Indian restaurant was the 3rd most popular

Cluster 5 was rejected as of the 5 wards in that cluster 4 already had Indian restaurant has the most popular (so could be already saturated) and the 5th did not have any Indian restaurants at all in the top 10 popular venues of that ward

Based on this it is recommended to open a new Indian restaurant in **Borivali West**

As a secondary alternative **<u>Andheri East / Kurla</u>** in this cluster could be considered

6. Conclusion

Ward wise venues were obtained for each ward using Four Square API.

After clustering and analyzing for popular location that had eateries, it is recommended that the best location to open a new Indian restaurant would be **Borivali West**

As a secondary location **<u>Andheri East / Kurla</u>** could also be considered