**Location for a new Indian restaurant in Mumbai**

# 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.

## 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.

## Target Audience

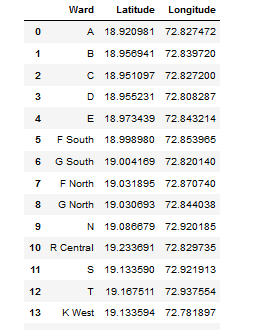
A person / group that is planning to open a new Indian restaurant in Mumbai city

# 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>

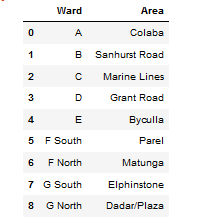
The geometry of each ward is a polygon with multiple vertices. To obtain the center of the various wards (to use as input to four square api) the mean of all the vertices was used.

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

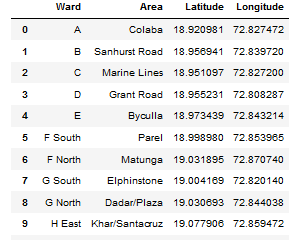


To get more meaningful names for the wards the location area name for the wards was taken from <http://www.demographia.com/db-mumbaidistr91.htm>

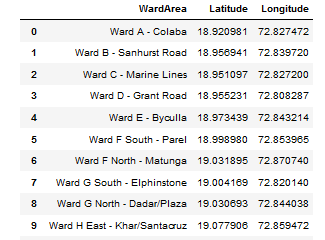
The first two columns was taken from this, and made into a csv file through Excel and then read into a dataframe



The two dataframe was then merged to create the following database



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



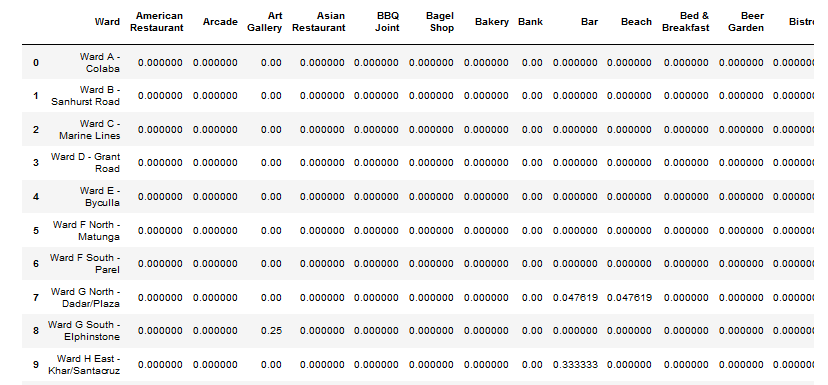
This Latitude and longitude will be used with a radius of 500m to get the venues from the Four Square API.

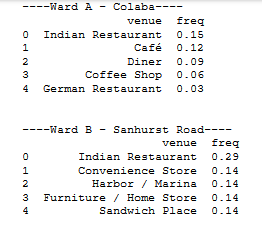


# Methodology

## 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.





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



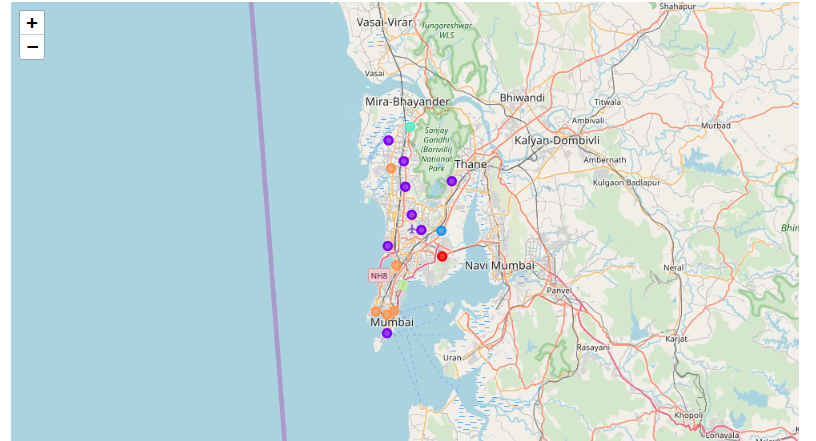
## Clustering

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

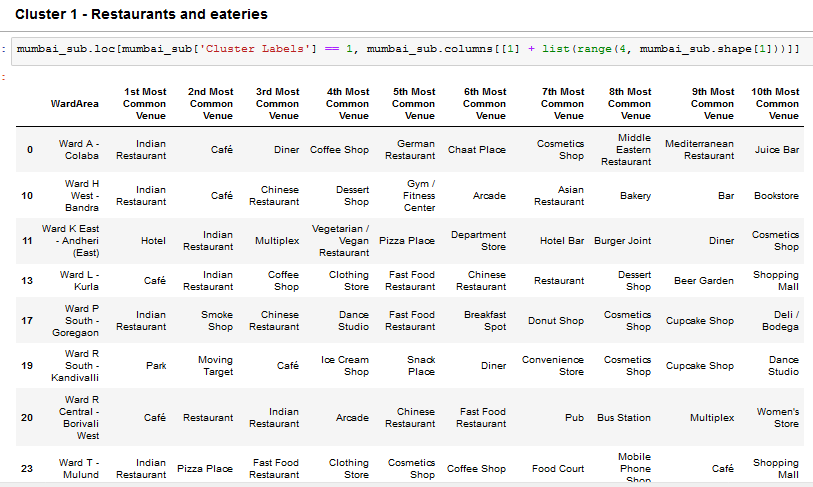
It was therefore decided to extract cluster number 0 as a separate dataframe and to do a clustering exercise again on the subset

# 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.



# 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 **Andheri East / Kurla** in this cluster could be considered

# 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 **Andheri East / Kurla** could also be considered