

# Introduction:

Starting a new business venture of any kind can be immensely challenging owing to the huge number of factors which has to be taken under consideration. The decisions made at the initial stages proves to be crucial in determining the success of the venture. That is why the foundations must be laid out very carefully to ensure success in the long run. A problem of similar manner will be explored in this project.

We have a client who is the owner of a popular Indian restaurant chain and has branches over various parts of India. The client wishes to open another branch in Mumbai, arguably among the busiest and liveliest cities of India. Selection of the venue will go a long way in helping the business to prosper in the future. So, in this project, we explore the location data for the city to find out the best possible location for establishing the restaurant.

The target audience (so to speak) of this project are those people who wish to determine the optimum location for establishing a business venture in any region, provided that some data regarding the surrounding areas are available for analysis and determination of suitable location. Anybody looking forward to start a store-based business (where the venue is of utmost importance) may refer this project and borrow ideas as to what factors should be taken into account in determination of optimum location and how those factors influence the process of decision making.

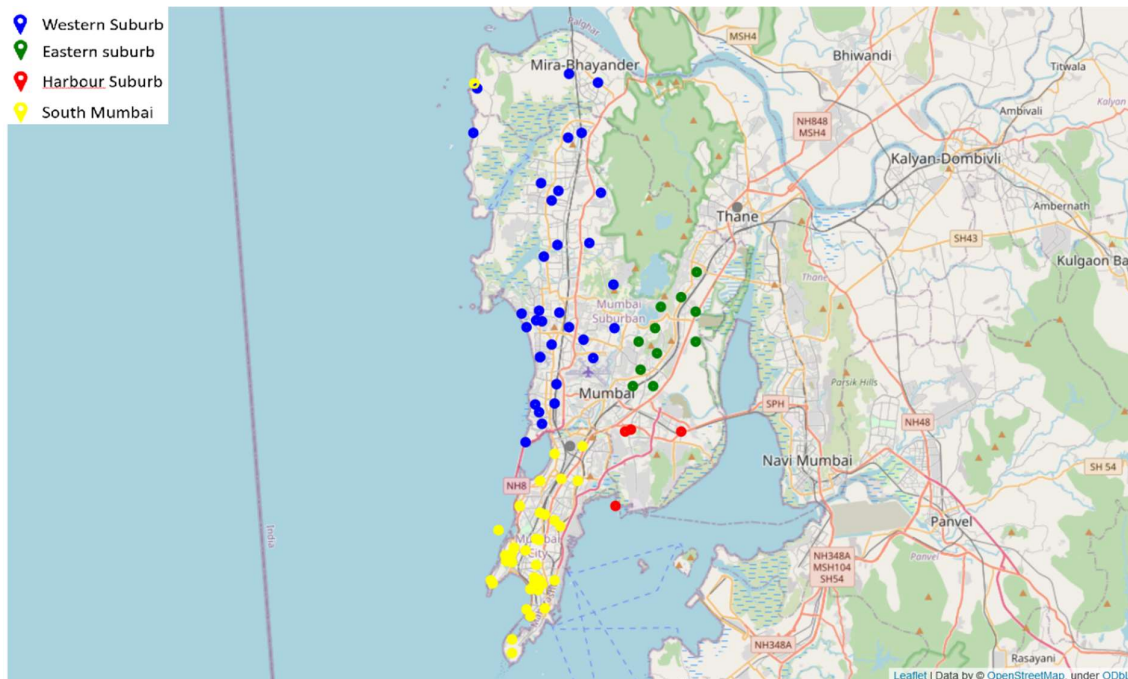
# Data:

Data has been collected for the Mumbai City district and Mumbai Suburban district, which form two separate regions of the state of Maharashtra. Together, they come under the administration of BMC. The total area can be divided into four primary divisions viz. **Western Suburbs**, **Eastern Suburbs**, **Harbour Suburbs** and **South Mumbai**. The data consists of areas under each of these four divisions along with their latitude and longitude values. A glimpse of the collected data is given below--

Area	Location	Suburbs	Latitude	Longitude
Amboli	Andheri	Western Suburbs	19.1293	72.8434
Asalfa	Ghatkopar	Eastern Suburbs	19.091	72.901
Deonar	Govandi	Harbour Suburbs	19.05	72.89
Navy Nagar	Colaba	South Mumbai	18.9012	72.8101

(source: [https://en.wikipedia.org/wiki/List\\_of\\_neighbourhoods\\_in\\_Mumbai](https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai))

The plot of the areas collected for this project are illustrated below--



We make queries using **Foursquare API** to find out surrounding venues within a 500-meter range for the above locations. The venue names, venue categories and location coordinates are stored in a data frame for the purpose of further analysis. The data consists of the location variables from the previous table along with venues in and around the areas along with their respective geo-locations. An example of the venue data is given below

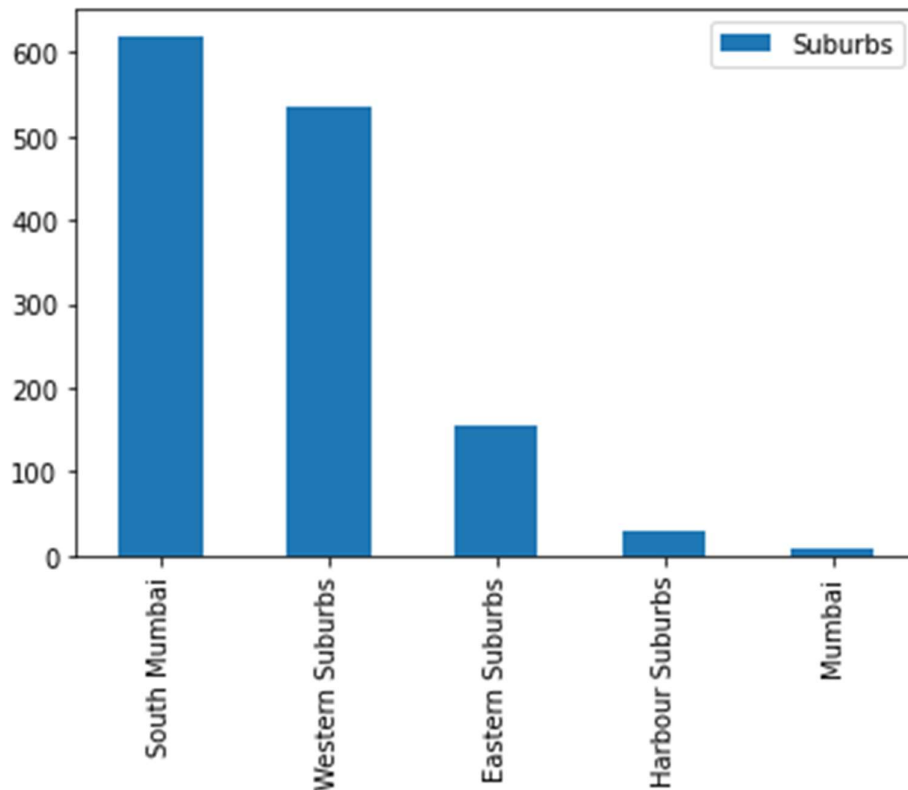
Area	Location	Suburbs	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Amboli	Andheri	Western Suburb	19.1293	72.8434	Cafe Coffee Day	19.12775	72.84466	Coffee Shop
Hiranandani Gardens	Powai	Eastern Suburb	19.118986	72.911767	Meluha, The Fern	19.11773	72.91111	Hotel
Mankhurd	Harbour Suburb	Harbour Suburb	19.05	72.93	Cafe X0	19.04797	72.93131	Coffee Shop
Carmichael Road	South Mumbai	South Mumbai	18.9722	72.8113	Maji Sagar	18.97268	72.81468	Fast Food Restaurant

We considered Movie Theatre, Playground, Park, Garden, Coffee shop, General Entertainment, Stadium, Petrol pump, Aquarium, Hotel, Beach, Recreation Centre, etc. venues. These venues are places that many people usually visit them for entertainment and hence we will have good demand for restaurants around them.

We can perform clustering on this data based on areas and obtain groups of areas which are similar in some respects. We can also study the density of different categories of venues across the locations to look for regions with less competitors. This study can also reveal areas with more frequently distributed venues which implies that demand for a restaurant will be more in those regions.

# Methodology:

Our objective is to find an optimum location for setting up a new restaurant. We begin by visualising the data to get an idea of the distribution of the venues over the city of Mumbai.



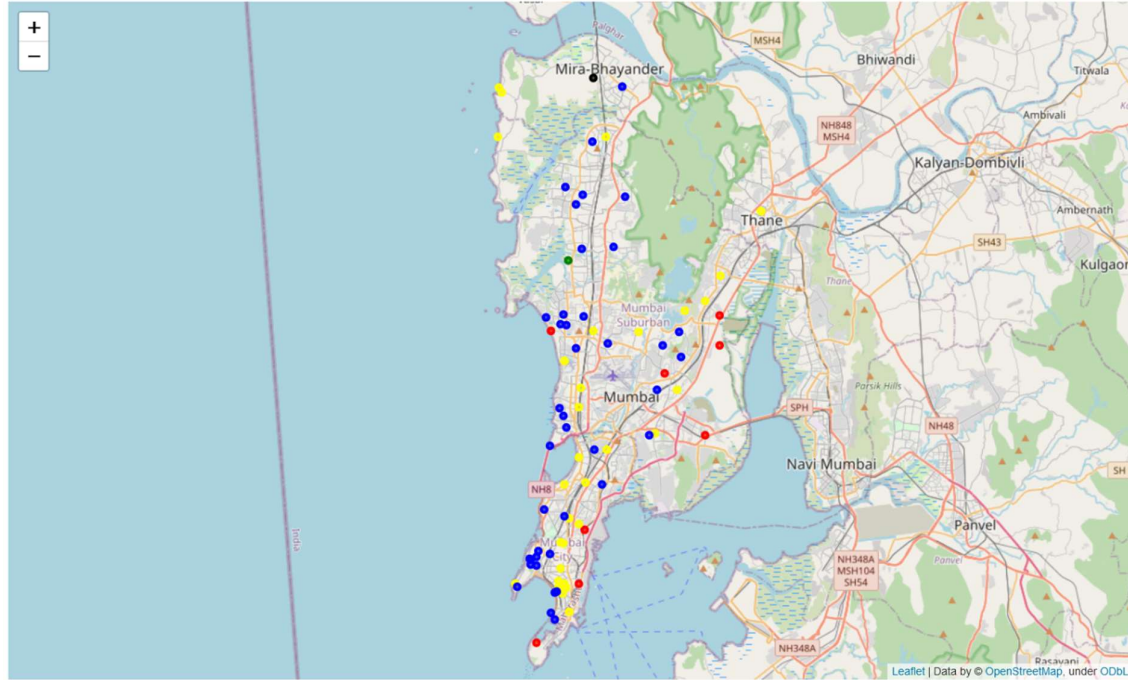
From the above plot we see that the majority of the venues are located in South Mumbai, followed by the other suburbs.

Now, we would like to categorise the areas based on the availability of venues around them so as to group the similar areas into homogenous groups. For this purpose, we convert the venue dataframe into dummy variables depending on the presence or absence of a certain venue category. With this data, we perform the k-means clustering process and divide the areas into 5 clusters.

The cluster with the least number of restaurants but the greatest number of other venues will be ideal as demand for a restaurant will be more in and around the other venue categories. Also, we will prefer areas from the most crowded suburb for the very similar reason.

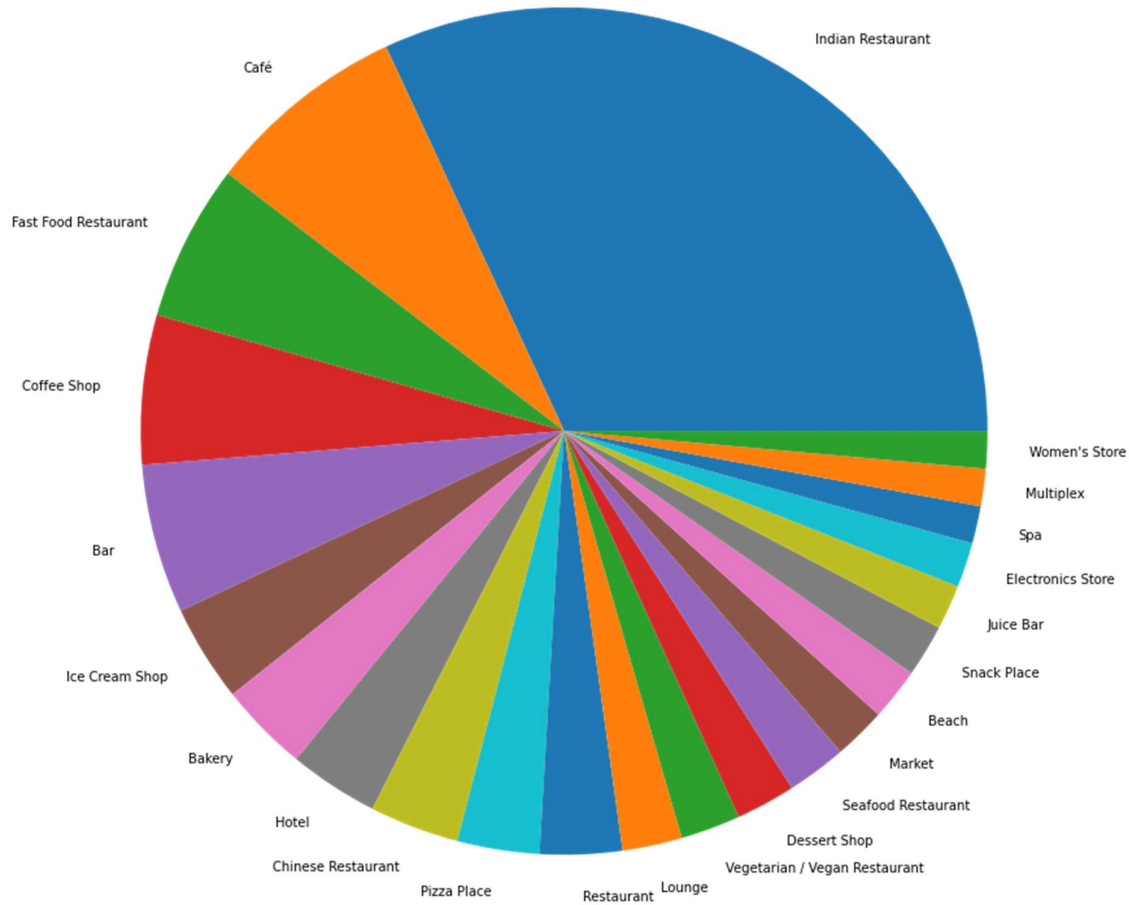
# Results & Discussion:

We have clustered the areas into five groups based on the venues which is illustrated in the plot below. Each different colour represents a different cluster.



Now we will take a detailed look into each of the clusters to determine the optimum location. We observe the distribution of venue categories in each cluster along with the densities.

## Cluster 1

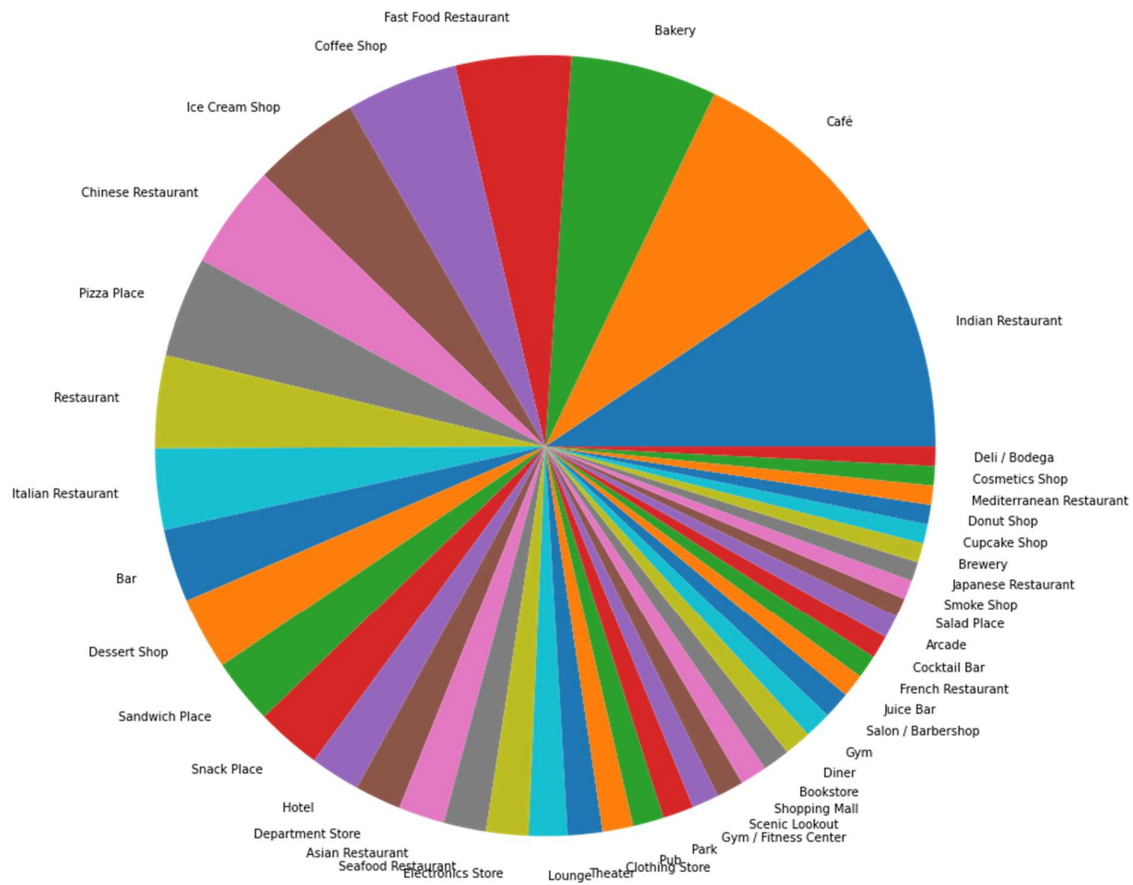


Observe that majority of this cluster is covered by restaurants of different types such as Indian Restaurant, Café, Fast Food Restaurant, Coffee Shop, Bar, Ice Cream Shop, Bakery, Chinese Restaurant, Pizza Place, Restaurant, Vegetarian / Vegan Restaurant, Dessert Shop, Seafood Restaurant, Snack Place, Juice Bar etc. Clearly, this cluster is already saturated with restaurants and hence is not suitable for a new restaurant from the business perspective.

## Cluster 3

This cluster consists of a single area Banqur Nagar, Goregaon of the Western Suburbs, the only venue here being "Katta aka Narayan", a **Smoke Shop**. Clearly this area is not ideal for opening a new restaurant.

## Cluster 2

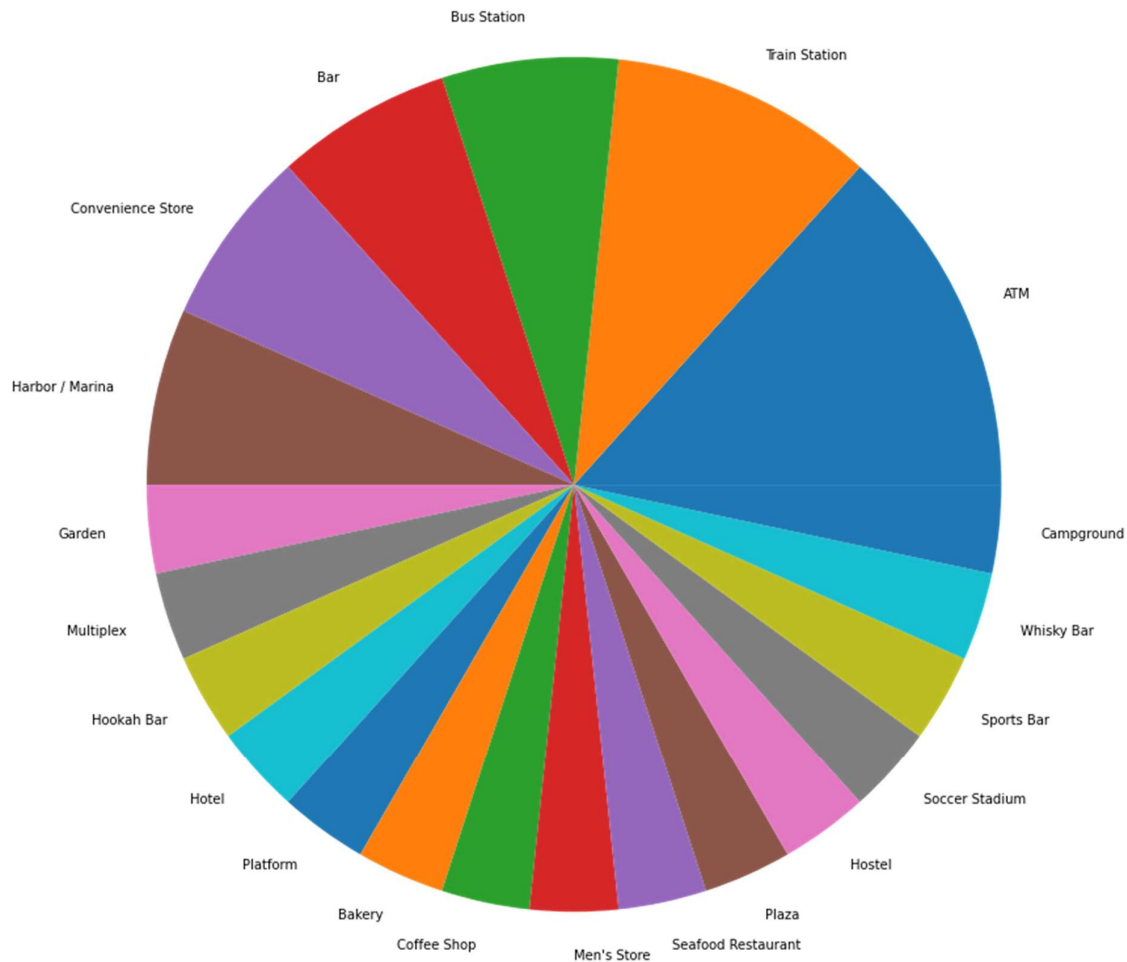


Although this cluster is a little behind cluster 1 with respect to the densities of restaurants, it is clear that this cluster too is heavily saturated with various types of restaurants. As such, restaurants take up the majority of venue categories. Thus, following a similar logic, this cluster is not suitable for opening a new restaurant.

## Cluster 5

This cluster consists of a single area Bhayandar, Mira-Bhayandar of the Western Suburbs, the only venue here being “D Mart”, a **Shipping Store**. Clearly this area is not ideal for opening a new restaurant.

## Cluster 4



In this cluster, we observe that there are various types of venues present in these areas. But, a very minor sector the venue categories are covered by restaurants. This makes the cluster ideal for starting a new restaurant business, since there are sufficient number of other venues implying that the areas are heavily crowded, but there aren't enough restaurants which means that the demand for a new restaurant will be quite high in these areas.

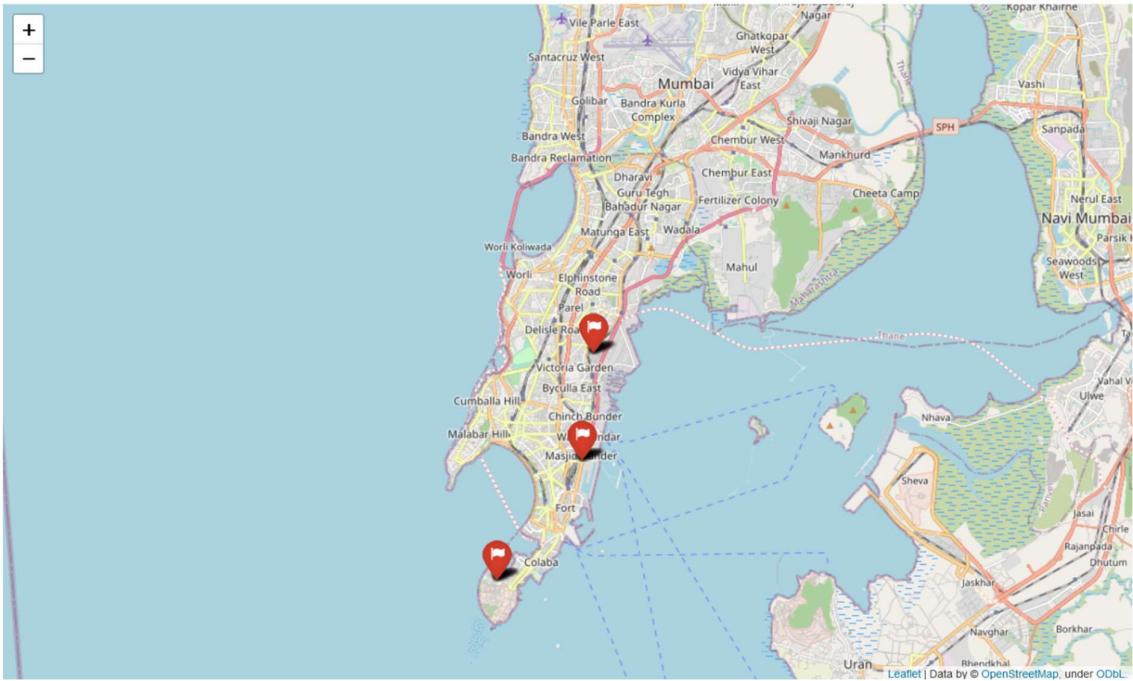
We have already seen that South Mumbai is the most densely distributed suburb. So, we proceed to find those areas in South Mumbai which belong to cluster 4.



These resulting areas are given below--

Area	Location	Latitude	Longitude
Ballard Estate	Fort	18.950000	72.840000
Cotton Green	South Mumbai	18.986209	72.844076
Cuffe Parade	South Mumbai	18.910000	72.810000

The areas are shown on a map in the following diagram--



# Conclusion:

So, we began our search for the ideal location for opening a new restaurant by collecting locational data for the city of Mumbai. By clustering the areas and observing the distribution of venues, we have narrowed our search and arrived at the conclusion that for opening a new restaurant in Mumbai, the best possible areas are Ballard Estate, Cotton Green and Cuffe Parade in South Mumbai. We considered the presence of competition in the areas, possible demand for a restaurant and some similar factors to arrive at this conclusion.

## Further Scope

This project, however has some limitations. Firstly, we have limited the number of search results for venues (the limit being 100). Inclusion of more venues may lead to a more accurate conclusion. Secondly, we were unable to consider the varying expenditure of opening restaurants over the different areas, which would definitely influence the decision. Thirdly, all the areas were given equal weightage for being considered as a possible venue. But it would be better suited if we could give more weightage to the tourist spots etc.