COURSERA CAPSTONE

IBM APPLIED DATA SCIENCE CAPSTONE

Project- Opening a new Italian or Indian restaurant in Mumbai, India.

by- Rishika Shrivastava

BUSINESS PROBLEM-

- Our customer wants to enter the culinary business in the city of Mumbai, India. Being new to the city he wants to establish a name for himself but he doesn't know which area to start.
- There are many factors which determine the success of a restaurant. These include cost, ambience, quality, quantity among many. One of the major aspects would be location and the competition around.
- Such a business cannot be started in impulse and all characteristics must be analysed carefully.

DATA COLLECTION-

For this analysis we needed the following data-

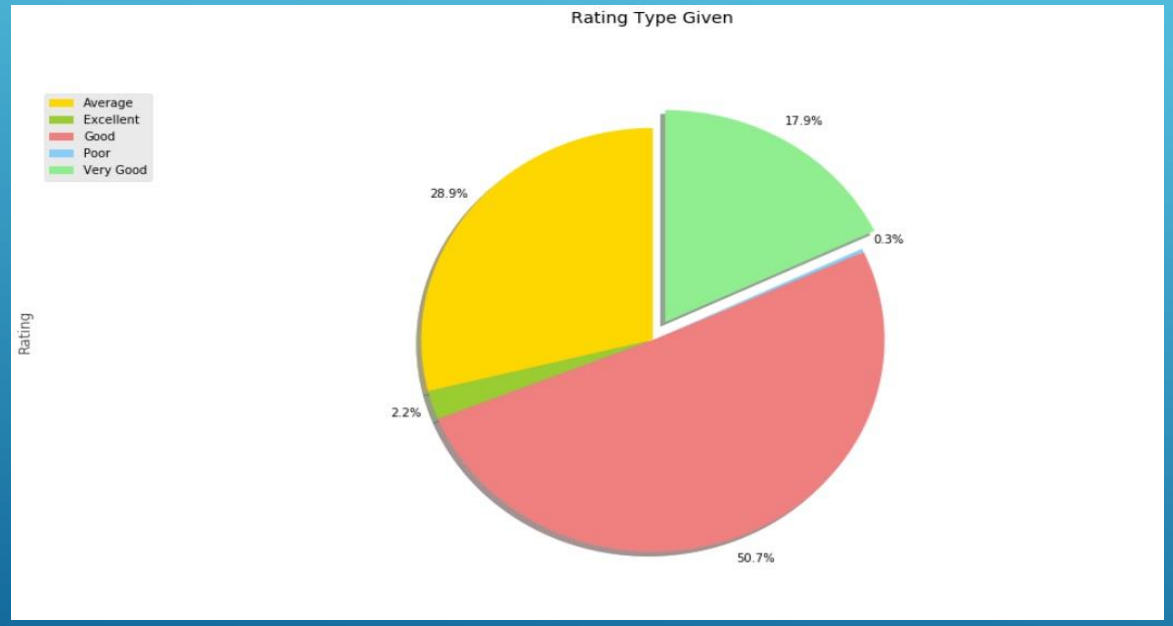
- List of restaurants with their cost, location, ratings, votes among many parameters.
- The latitude and longitude of the neighbourhoods in Mumbai.
- The venue and it's surrounding data, primarily that of restaurants.

We collected the data from multiple sources-

- For the initial analysis we used the Zomato, Mumbai database, it consisted of the specifications of the restaurants.
- We used Geocoder to retrieve the coordinates ie. The latitudes and the longitudes
- We used the Foursquare API to get the venue features.

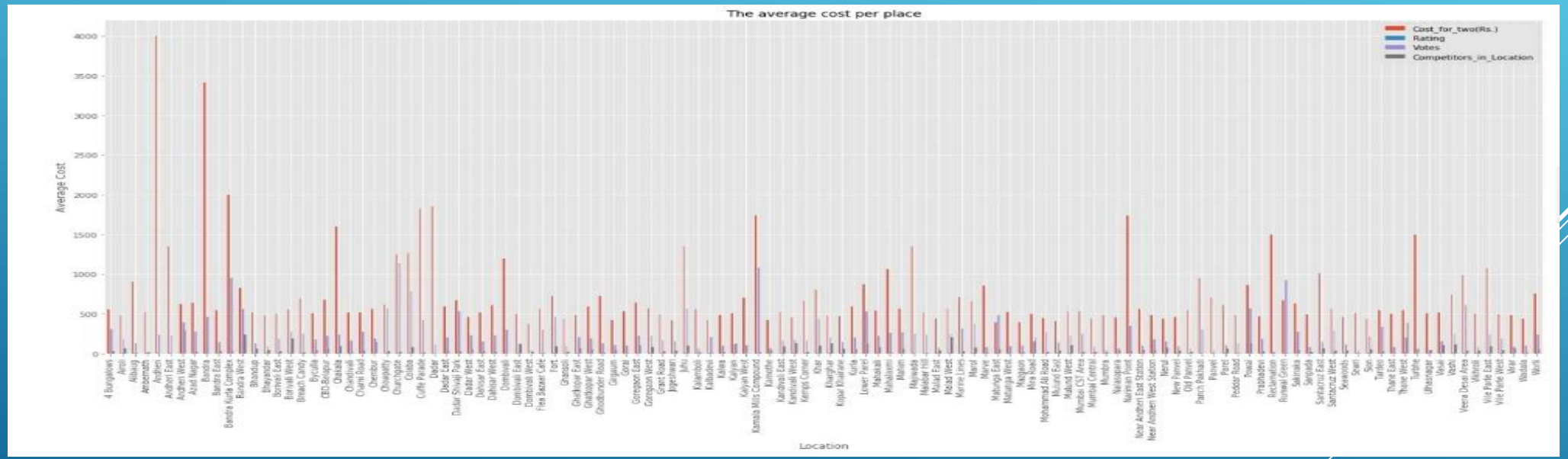
METHODOLOGY-

- We initially begin with the Zomato dataset and drop off the unnecessary data.
- We initially calculate what the standard of most restaurants is by taking all of the rating categories and creating a pie chart for visualization.



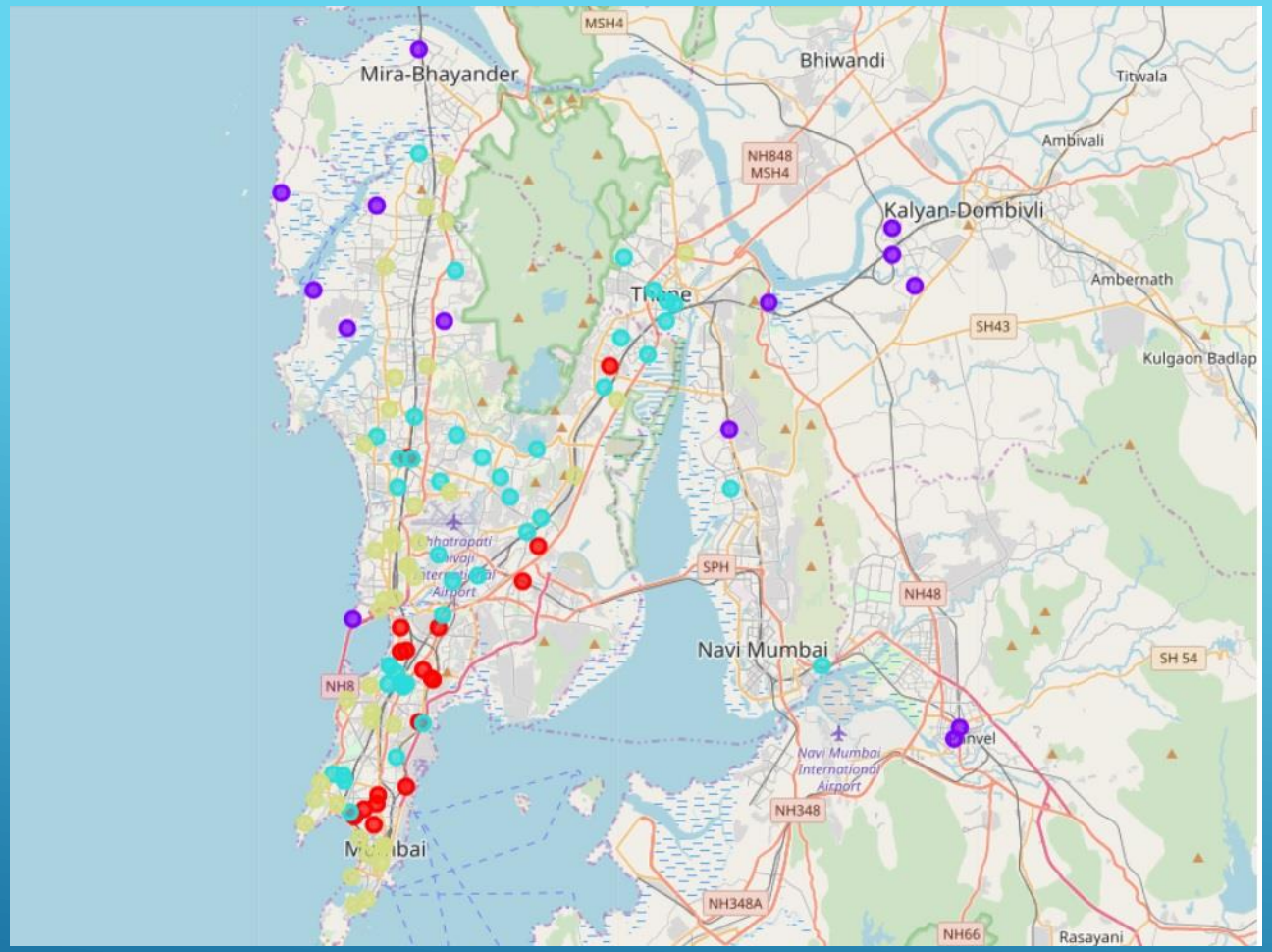
METHODOLOGY CONTD.

- We then calculate the average of the necessary parameters in order to compare the various locations based on boundaries like cost, rating, competitiors and rating. We then construct a bar graph to weigh the criterions accordingly.



METHODOLOGY CONTD.

- We the use Geocoder to get the latitude and the longitude of the neighbourhoods.
- We use the Foursquare API to retrieve the venue data.
- We group the data based on neighbourhoods.
- We filter the data for Indian and Italian restaurants, because that's what we have assumed the customer requirement to be.
- We carry out clustering using the K-means clustering procedure.
- Folium is used to create a map for visualizing our data.



RESULTS-

- We distribute the data into 4 different clusters.
- Each cluster has varying number of Italian and Indian restaurants.
- Cluster 1 displays the least number of both Italian and Indian restaurants whereas Cluster 2 displays the maximum number of both of them.

DISCUSSION-

- We notice that most of the restaurants are located on the southern side.
- Maximum number of the restaurants are in Cluster 2 and the minimum number of restaurants are in Cluster 1. Cluster 3 and 4 display moderate number od restaurants.
- South Bombay seems to flourish with high end restaurants.

RECOMMENDATIONS

- Cluster 1 would be the safest bet to open a new restaurant owing to the less density of existing eateries.
- Cluster 2 would be the least recommended due to intense competition and well established restaurants.
- If the customer has a well known restaurant chain and wants to expand or has a unique proposal, he can try out in Cluster 3 and 4 where the completion is moderate.
- The customer can approach Cluster 2 if he has a well known chain in other areas of Mumbai, to increase his credibility.

CONCLUSION-

- According to our deduction, Cluster 1 would be the best option for our customer.
- These conclusions will assist our customer to make his final decision on where to open a new restaurant.

THANK YOU FOR YOUR TIME