**Determining best location for a fast food establishment**

Vishaal Rao

16th April 2019

## Introduction

### **Background**

Let us consider a business owner planning to open a fast food restaurant in Mumbai. She has approached us to recommend an ideal location to open this restaurant. In this use case, an ideal location would be one that can maximise footfall in the restaurant in order to maximize her profits.

Additionally, once provided an ideal location, she would also like to acquire:

1. Average price range of the restaurants nearby: This will allow her to design her price her offerings competitively.
2. Average rating of the restaurants nearby: This will allow her to gauge her competition. For e.g. Highly rated restaurants means going up against well established players.
3. Combined total number of ratings of restaurants nearby: This will not give the exact footfall but could be used to compare 2 locations. Location with greater number of ratings in most cases implies greater footfall.

### **Our Approach:**

Before we get started, I would like to incorporate a few of my experiences/ideas to optimise our approach to finding an ideal location:

1. Mumbai has one of busiest commuter rail systems with over 7.5 million commuters daily. Locating a restaurant near a rail station can result in a high footfall. I personally have visited some of the popular restaurants near rail stations and have always found them jam packed!
2. We should also optimise for locations with high densities of fast food restaurants. This idea is based on the concept of game theory and Nesh equilibrium theorem, which imply that when shops of the same kind open close to one another, these shops tend to enjoy better profits.

### **Interest**

The following methodology could be used for any establishment in order maximize the number of customers a business could have. By having an outlet at a place filled with passers-by, there is more of a chance to gain a customer. Hence this project deals with the methodology to use the railway stations as a way achieve higher number of consumers.

### **Methodology**

Incorporating the ideas, our solution will consist of the following steps:

1. Acquire the locations of fast food restaurants in Mumbai, along with their price-range, ratings. We shall utilise Foursqaure API for this.
2. Acquire the locations of rail stations in Mumbai using the foursquare API.
3. Study the price-range, ratings and number of likes given to these establishment in order to gt a better understanding of the competition
4. Clustering and understanding the locations with the high density of fast food restaurants using KMeans.
5. Removing all the outlier’s railway station which aren’t anywhere close to the clusters and marking the ones which are within 2 km of the cluster centres.
6. Since there is no dataset about the footfall in the stations, we shall data scrape that information off a blog post and draw up a conclusion.

## Data Cleansing

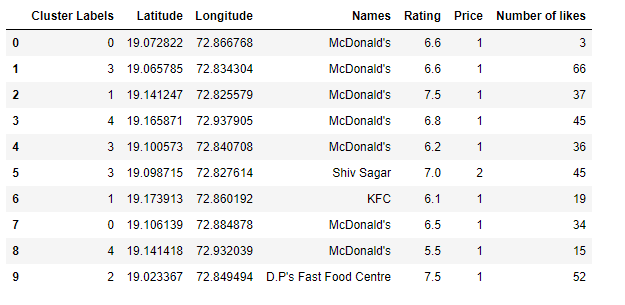
Data has been collected mainly from Foursquare API, and it does seem that Mumbai is not as well documented as western cities. So, in a few cases where the ratings were missing, were replaced with the average of the rest of the values.

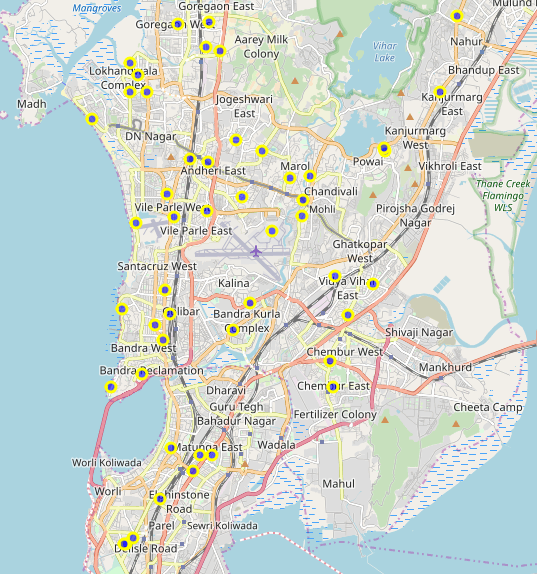
One of the stations, Goregaon station, had multiple entries of fast food restaurants near it but wasn’t present in the Foursquare database. Thus, I manually input and append the entry to the table.

Once I had looked through the locations of restaurants and railway station, I decided to remove some of the outlier railway stations to only consider the ones within Mumbai suburban.

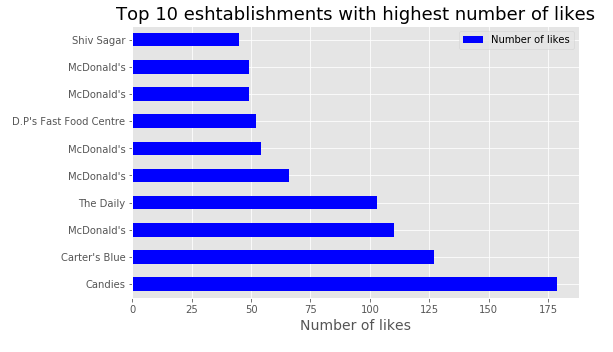
## ****Data exploration****

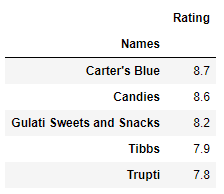
**After importing data from foursquare API, the 1st 10 results are displayed as following:**

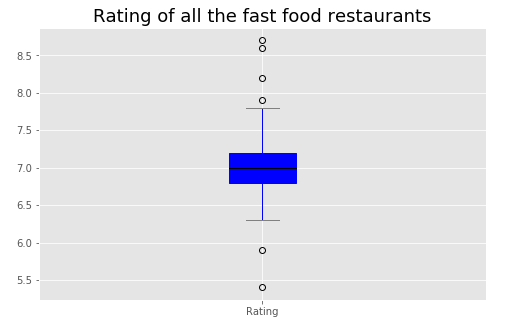
**Upon mapping these restaurants, the locations were spread out around Mumbai as follows:**



**‘Number of likes’ given by users is taken as an estimate for the footfall an establishment receives. Upon observation it was noticed that individual restaurants performed better than chain of restaurants such as McDonalds.**



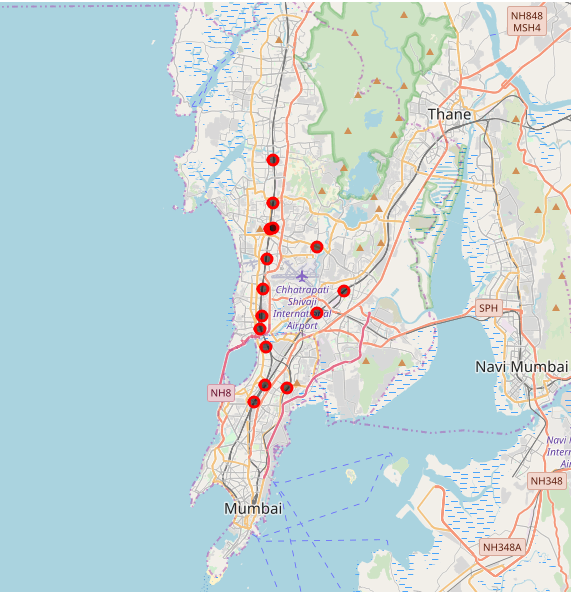
When ratings of the establishments were being observed it was noticed that most of the restaurants are above the median rating of around 7. With restaurants such as Candies and Carter blue having a higher rating.



We have also calculated the average rating and the average pricing in the whole of the area to get a better idea about the competitors the business woman might be facing.



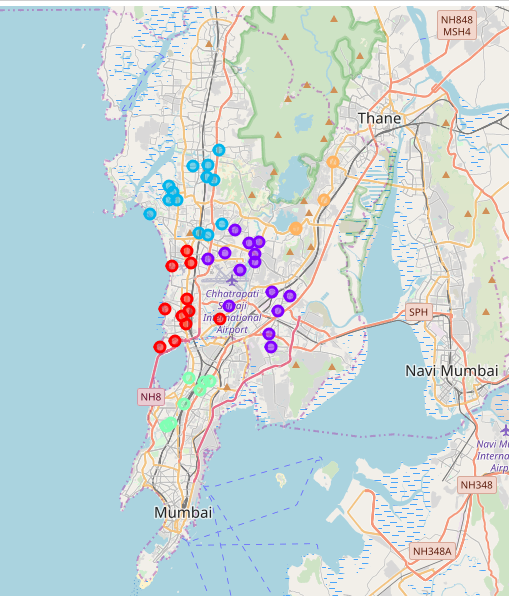
We have mapped out the railway stations which are in proximity to the fast food restaurants. We removed any outliers beyond certain stations.



## Clustering

By observing the fast food restaurant map it was noticed that the restaurant locations could be clustered into 5 parts. Hence a value of 5 was considered for the K means clustering.

And this was the result obtained.

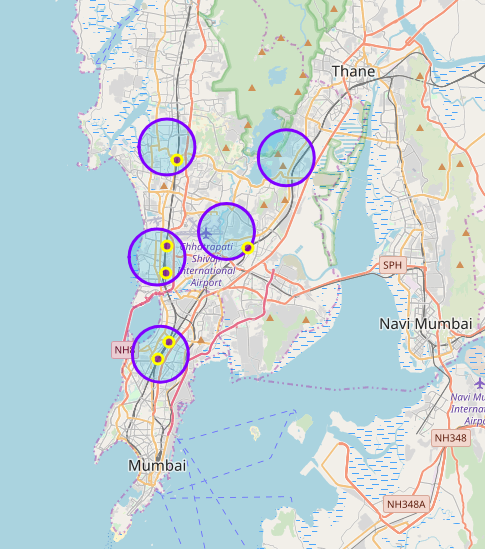


Based on the clusters formed we will be looking for stations which are within 2 km of these cluster centre.

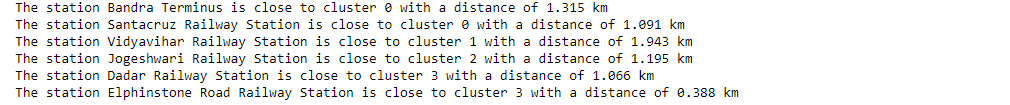
## Selecting suitable stations

By following the above step, we will be completing both the objectives of listing out all the stations which are relatively popular when compared to other stations. And know the locations with a high density of fast food restaurants which is follows the nesh theory.

Upon mapping the results are as follows:

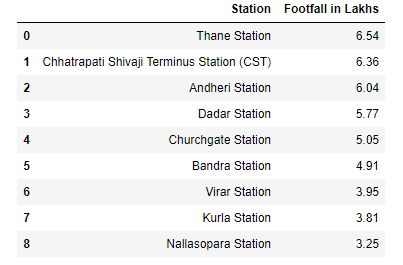


The following stations were noticed to be in the 2km radius of a cluster centre.



## To find the optimal station based on footfall.

Earlier we had data scraped the top 10 busiest stations in Mumbai.



Comparing the railway stations and the footfall accounts it is noticed that **Bandra** and **Dadar** stations would be the ideal places to open a fast food restaurant.

## Conclusion

* In this study, I collected the data for the location, rating, likes, price range and the locations of railway stations around Mumbai. I identified that individual restaurants have more of a preference than chain of restaurants such as McDonalds. The individual restaurants generally have a higher rating and number of likes in comparison to the chain, while having a cheap price range.
* I performed K Means clustering in order to know the density of fast food restaurants based on which an ideal location could be decided.
* Dadar and Bandra were found to be the best locations in comparison to the other stations.
* These models can be very useful in order to determine the position of any establishment based on other services such as schools, offices, tourist destination etc.
* A similar method with different deciding variables can be used for other establishments such as hospitals, Fire station, police station etc.

## Further Enhancements

This methodology only considers only the proximity of railway stations. I wish to add more deciding factors such as places of tourism, average cost per square feet in the area, types of fast food which are doing well.

Using all the above factors we can perform regression with holding rating as the dependent variable.

Or if we could get the footfall data for every restaurant in the area, then that might be a better substitute for dependent variable.