**Coursera Capstone**

**IBM Applied Data Science Capstone**

# *Opening a New Restaurant In Mumbai,India*

By: Sumit

September 2019



# Introduction

As far back as I can remember, I have always liked going out to eat .

For most of us, visiting restaurants is a great way to relax and enjoy themselves during weekends and holiday. For retailers, the central location. As a result, there are many restaurants in the city of Mumbai and many more are being built. Opening restaurants allows property developers to earn consistent rental income. Now, as with any business decision, opening a new restaurant requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the restaurant is one of the most important decisions that will determine whether the mall will be a success or a failure.

## Business Problem

Goal of this capstone project is to analyse the best locations in the city of Mumbai,India to open a new Restaurant. Using machine learning techniques like clustering,and also the concepts of data science this project aims to provide solutions to answer the business question: In the city of Mumbai,India, if a property developer is looking to open a new restaurant, where would you recommend that they open it?

## Target Audience of this project

This project is particularly useful to property developers and investors looking to open or invest in new restaurants in the Mumbai,India.

# Data

**To solve the problem, we will need the following data:**

* List of neighbourhoods in Mumbai,India. Scope of the project is confined to the city of Mumbai India
* Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.
* Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighbourhoods.

## Sources of data and methods to extract them

This Wikipedia page =https://en.wikipedia.org/wiki/Category:Suburbs\_of\_Mumbai[)](https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur)

Web scrapping using BeautifulSoup.

Foursquare API

Folium

Methodology

Firstly, we need to get the list of neighbourhoods in the city of Mumbai,India. The List is available at Wikipedia page (<https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai>).

We will use Beautifulsoup to extract the list of neighbourhoods.This will provide us a list of names.Then we will use foursquare api to get coordinates of these locations. To do so, we will use the wonderful Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude.Then we will create a pandas dataframe to store the data.Then we will visualize the data of neighbourhoods in a map.Then using foursquare api we will get top venues of the city within a specifies radius.after retrieving data from foursquare we will check how manay unique categories are present.analyse the data Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category.we are also preparing the data for use in clustering. we will filter the “Restaurant” as venue category for the neighbourhoods. Lastly, we will perform clustering on the data by using k-means clustering. We will cluster the neighbourhoods into 5 clusters based on their frequency of occurrence for “Restaurants”. The results will allow us to identify which neighbourhoods have higher concentration of shopping malls while which neighbourhoods have fewer number of shopping malls. Based on the occurrence of shopping malls in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new shopping malls.

Results

K-means clustering shows 5 clusters based on the frequency of occurrence for “Restaurant”:  
Cluster 0 and cluster 4 have the most number of restaurants while clusters 1,2 and 3 have moderate no of clusters.Restaurants in cluster 0 and 4 have the most competitionwhile restaurants in clusters 1,2 and 3 have moderate competition.So this represents a great opportunity for a person to open a restaurant in these areas.While Clusters 0 and 4 have a oversupply of restaurants they may have adequate amount of customers due to that area being well developed.But the other areas are also well developed and have a undersupply of restaurants.So it is a great opportunuty to open a restaurant there.

Discussion

Group 0 and group 4 have the most number of restaurant while bunches 1,2 and 3 have moderate no of clusters.Restaurants in bunch 0 and 4 have the most challenge while restaurant s in clusters 1,2 and 3 have moderate competition.So this speaks to an extraordinary open door for an individual to open an restaurant in these areas.While Clusters 0 and 4 have an oversupply of restaurant they may have sufficient measure of clients because of that region being admirably developed.But different regions are additionally very much created and have an undersupply of restaurants.So it is an incredible opportunuty to open a restaurant there.

Limitations and Suggestions for Future Research

In this project, we only consider one factor i.e. frequency of occurrence of shopping malls, there are other factors such as population and income of residents that could influence the location decision of a new shopping mall. However, to the best knowledge of this researcher such data are not available to the neighbourhood level required by this project. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new shopping mall. In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Future research could make use of paid account to bypass these limitations and obtain more results.

Conclusion

In this undertaking, we have experienced the way toward distinguishing the business issue, determining the information required, removing and setting up the information, performing AI by bunching the information into 5 groups dependent on their similitudes, and in conclusion giving proposals to the pertinent partners for example property designers and speculators in regards to the best areas to open another restaurant. The discoveries of this task will assist the pertinent partners with capitalizing on the open doors on high potential areas while staying away from stuffed regions in their choices to open another eatery