

**Coursera Capstone**  
**IBM Data Science Capstone Project**

**Opening of a new Shopping Mall in  
Mumbai, India.**

**By:** Ruchit Bagade.  
May 2020



## **Introduction**

For many people, visiting shopping malls is a great way to relax and enjoy themselves during weekends and holidays. They can do grocery shopping, dine at restaurants, shop at the various fashion outlets, watch movies, and perform many more activities. Shopping malls are like a one-stop destination for all types of shoppers. For retailers, the central location and the large crowd at the shopping malls provides a great distribution channel to market their products and services. Property developers, builders, real-estate businessmen are also taking advantage of this trend to build more shopping malls to cater to the demand. As a result, there are many shopping malls in the Mumbai City and many more are being built. Opening shopping malls allows property developers to earn consistent rental income. Of course, as with any business decision, opening a new shopping mall requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the shopping mall is one of the most important decisions that will determine whether the mall will be a success or a failure.

## **Business Understanding**

The objective of this capstone project is to analyze and select the best locations in Mumbai, India to open a new shopping mall. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In Mumbai, if a property developer, builders or real-estate businessmen is looking to open a new shopping mall then which location will be the best location for the mall to open ? What would you recommend?

## **Target Audience of this project**

This project is particularly useful to property developers and investors looking to open or invest in new shopping malls in Mumbai. This project is particularly based on the density of the existing malls in the respected areas. There could be other factors which could influence the opening of shopping mall in the particular suburbs such as student areas, residential societies, etc. but for the property developers reducing the competition in the Mall business will be the priority and hence this factor proves promising to take decisions on opening a mall in the area with less existing malls.

## Data

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

- List of neighborhoods in Mumbai, India. This defines the scope of this project which is confined to the Mumbai City, the financial capital of India.
- Latitude and longitude coordinates of those neighborhoods. 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 neighborhoods.

## 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_of_Mumbai)") contains a list of neighborhoods in Mumbai City, with a total of 42 neighborhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and BeautifulSoup packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods. After that, we will use Foursquare API to get the venue data for those neighborhoods.

Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Shopping Mall category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).

In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.