Coursera Capstone
IBM applied Data science Capstone

Battle of Neighborhoods in Mumbai city, India for opening New Shopping mall.

By Saurabh Mahadane
June 2021

Introduction

For many shoppers, 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 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 city of Mumbai 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 Problem

The objective of this capstone project is to analyse and select the best locations in the city of Mumbai to open a new shopping mall. Using data science methodology and machine learning techniques -k means clustering, 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 shopping mall, 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 shopping malls in the capital city of Maharashtra, MUMBAI, India.

Data Collection

The data required for this project is as follows has been collected from multiple sources. The following data is required for the project:
1) Neighborhood data of Mumbai
2) Geographical coordinates of Mumbai and all neighborhoods in Mumbai
3) Venue data for neighborhoods in Mumbai

Neighborhoods Data

The data of the neighborhoods in Mumbai was scraped from https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai. The data is read into a pandas data frame using the read_html () method. The main reason for doing so is that the Wikipedia page provides a comprehensive and detailed table of the data which can easily be scraped using the read_html () method of pandas.

Geographical Coordinates

The geographical coordinates for Mumbai data has been obtained from the GeoPy library in python. This data is relevant for plotting the map of Mumbai using the Folium library in python. The geocoder library in python has been used to obtain Latitude and longitude data for various neighborhoods in Mumbai.

Venue Data

The venue data has been extracted using the Foursquare API. This data contains Venue recommendations for all neighborhoods in Mumbai and is used to study the popular venues of different neighbourhood.