Coursera Capstone week5 Project

Location identification for opening a new mall in Mumbai city

**Introduction:**

The shopping is an important activity of consumers’ lives and it is continually changing, making the investigation and understanding of this field important in order to create a pleasant shopping experience and achieve customers’ satisfaction. Today’s customers tend to purchase products while spending time in a shopping mall. Shopping malls are characterized as venues that enable a comfortable shopping experience and have turned into social centers and recreational facilities for various activities . At the beginning of their life cycle, shopping malls were primarily economic entities that provide shopping mall customers with a wide selection of stores and merchandise at a single location. Shopping malls have adapted to new designs and tenant varieties to meet the changes in consumers’ needs, desires, values, and lifestyles.

Shopping mall customers visit shopping malls not only for searching for products, but they also view these visits as an entertainment activity that provides fun and pleasure from the shopping experience Shopping mall customers can visit shopping malls to watch movies at movie theatres and watch soccer matches at different restaurants

This gives further benefits shopping mall managers and retailers by providing recommendations on how to best satisfy customers’ needs and achieve customers’ satisfaction and loyalty at their respective shopping malls. Shopping mall is an emerging and lucrative business in India. However, the business demands a right strategic planning with financial and marketing planning.

Shopping malls are getting tremendous popularity these days. And definitely, there are several reasons behind it. Generally, shopping malls offer a comfortable and luxury shopping experience. 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/Target audience**

This project best describes on how to identify location for opening a shopping mall in Mumbai city,India . Clustering technique can be used to solve this problem Shopping malls are getting tremendous popularity these days. And definitely, there are several reasons behind it. According to the retail industry experts, the demand for the mall will more increase in coming 5 years. And we can expect a significant growth in the shopping mall business throughout the country. And we can expect a significant growth in the shopping mall business throughout the country This project is particularly useful to property developers and investors looking to open or invest in new shopping malls in Mumbai. A business plan is the most crucial part of starting this business. And each and every decision makes a huge difference in the success of the business. The world of retail is changing dramatically, but the mall still can have a central role in urban and suburban societies

**Data:**

The data that will be used to solve the problem is mentioned below <https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai>

We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests . Foursquare is a social location service that allows users to explore the world around them. Foursquare location data would be used to solve this problem . Geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods. 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).

**Methodology:**

Data is being used from below Wikipedia link <https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai>

We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests

Geographical coordinates like latitude and longitude are required a in order to be able to use Foursquare API. Foursquare is a social location service that allows users to explore the world around them . Geocoder package will be used that will allow to convert address into geographical coordinates. After gathering the data, we will populate the data into a pandas data frame and then visualize the neighborhoods in a map using Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the city of Mumbai. Next, we will use Foursquare API to get the top 100 venues that are within a radius of 2000 meters. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighborhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each neighborhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyze each neighborhood by grouping the rows by neighborhood and taking the mean of the frequency of occurrence of each venue category. We will filter the “Shopping Mall” as venue category for the neighborhoods. Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighborhoods into 3 clusters based on their frequency of occurrence for “Shopping Mall”.

**Results**

The results from the k-means clustering show that we can categorize the neighborhoods into 3 clusters based on the frequency of occurrence for “Shopping Mall”:

Cluster 0: Neighborhoods with moderate number of shopping malls

Cluster 1: Neighborhoods with low number to no existence of shopping malls

Cluster 2: Neighborhoods with high concentration of shopping malls

The results will allow us to identify which neighborhoods have higher concentration of shopping malls while which neighborhoods have fewer number of shopping malls. Based on the occurrence of shopping malls in different neighborhoods, it will help us to answer the question as to which neighborhoods are most suitable to open new shopping malls

**Discussion**

As observations noted from the map in the Results section, most of the shopping malls are concentrated in the central area of Mumbai, with the highest number in cluster 2 and moderate number in cluster 0. On the other hand, cluster 1 has very low number to no shopping mall in the neighborhoods. This represents a great opportunity and high potential areas to open new shopping malls as there is very little to no competition from existing malls. Meanwhile, shopping malls in cluster 2 are likely suffering from intense competition due to oversupply and high concentration of shopping malls. From another perspective, the results also show that the oversupply of shopping malls mostly happened in the central area of the city, with the suburb area still have very few shopping malls. Therefore, this project recommends property developers to capitalize on these findings to open new shopping malls in neighborhoods in cluster 1 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new shopping malls in neighborhoods in cluster 0 with moderate competition. Lastly, property developers are advised to avoid neighborhoods in cluster 2 which already have high concentration of shopping malls and suffering from intense competition.

**Conclusion :**

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. property developers and investors regarding the best locations to open a new shopping mall. To conclude :The neighborhoods in cluster 1 are the most preferred locations to open a new shopping mall.