Capstone Project | Explore Segmenting and cluster the neighborhood of | Delhi | India

IBM Applied Data Science Capstone Project

By : Gautam Sharma

July 2021

INTRODUCTION:

- For Many Shoppers, Visitings Shopping Malls is a great way to relex and enjoy themselves during wekkend and holidays. They can do grocery Shopping , dine at reaturents , ahop at the various fashion outlets , wathch movie and perform many more activities .
- Shopping Mallls are like one - stop destination for all types of shoppers. for retails , the central location and the large crowd at the shopping malss provides a great distribution channels to marlet their products and service.
- Proverty developers are also taking advatages of this trend to build more shopping malss to cater to the demand. Opening Shopping mall serious Consideration and is a lot more complicates than it seems. Particularly, the location of the shopping mall is none of the most important decisions that will determine the mall will be success. As the many shopping malls present in Delhi , India. so developer think whuch is best place for built the Shopping Mall.

BUSINESS PROBLE:

The objective of this projecct is to analyse and select the best locations in the city of Delhi , India to opening a new Shopping Mall. Using data science and Machine Learning techniques likes Clustering , this project aims to provoide solution to answer the business Questions . Which is the best location in delhi to built the shopping Mall?

DATA:

To solve the problem, we will need the following data

- List of neighborhood in Delhi. This defines the scope of this project whch confined to the the city of Delhi.
- Latitude and Longitude coordinates of those neighborhoods . This 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 tis data toperform clustering on the neighborhood

Source:

https://en.wikipedia.org/wiki/Category:Delhi_geography_stubs

Methodlology:

Firstly, we need to get the list of neighborhoods in the city of Delhi , India. Fortunately , the list is available in the wikipedia page https://en.wikipedia.org/wiki/Category:Delhi_geography_stubs

we will do web scraping using pythn requests and beautifulsoup packages to extract the list of neighborhoods data . however , this just a list of names.

we will need to get the geographical coordinates in the form of latitude nad longitude . after we will populate the data into a pandas dataframe and the visulalize the neighborhoods in a map using folium package . This allows us to perform a sanity check to make the geographical coordinates return by geocoder are correctlt plotted in the city of Delhi, India.

Next, we will use foursquare api to get the top 200 that are within a radius of 5000 meters .we register with the foursquare client_id and client_sceret as we registered in the foursquare website.

then we will analyse each of occurance of each vanue category. By doing sp, we are also preparing the data for use in clusstering .Since weare analysing the shopping Mall data, we will filter the shopping mall as venue category as the neighborhood.

Lastly , we will perform clustering on the data by using k-means clustering algorithm identifies k number of centroid, and the allocats every data points to the nearest cluster, while keeping the centroids , as small as possible.

Results:

The results from the k-means clustering show that we can categorized into 4 clusters on the freuency of occurance of Shopping Mall

- Cluster 0: Neighborhood with moderate number of number of shopping mall

- cluster 1: Neighborhood with Low number of number of shopping mall
- cluster 2: Neighborhood with Large number of number of shopping mall
- cluster 3: Neighborhood with Large number of number of shopping mall