**Coursera Capstone Project Of Neighbourhoods**

**IBM Applied Data Science Capstone**

****

**Introduction:**

A **shopping mall** is a specially built covered area containing **shops**

and restaurants which people can walk between, and where cars are

not allowed. In any city Shooping malls are the best way for people

to relax and for shopping in Weekends or Holidays. People can chill

and relax in Shopping malls .Malls are different from bazaars where

the shops are not tiny booths. In Malls each shop has its own space.

Shopping malls consists of shops i.e fashion stores, Theatres and

What not. In Shopping malls people can enjoy relax and they can do

Shopping. So, the shopping mall should be located in the centre of

city or the place with good neighborhoods. Opening shopping malls

should make developers to earn their consistent rental income.

Literally, the location of Shopping mall plays a crucial role in deter-

Ming whether it is success or not. So, after the great scrutiny of

Neighborhoods the location of Shopping mall should be decided.

**Business Problem:**

The Ultimate aim of this **Capstone** is to select the best Locations in

The city of Banglore, India to open a new Shopping mall. Using the

Data science methodology and Machine learning techniques like

**Clustering K-means .** The main aim of this capstone is to answer the

Business question is:

If the some property Developer wants to build a shopping mall in the

City of **Banglore, India,** where would you suggest to open it?

**Target audience of this Project:**

This project is mainly useful for the property developers and the

Investors who are looking to open new shopping malls in Banglore.

**Data Required:**

->List of Neighborhoods in Banglore,India. This defines the scope of

Project which is confined to the city of Banglore, this city is the third

Most populated city in India.

->Latitude and Longitude from geopy which is used to get the venues

and plot in the Map.

->Venues or venue data which is used in to perform clustering of the

neighborhoods.

**Source of Data and Methods for Extracting Them**

This Wikipedia(“<https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Bangalore>”)

page consists of nearly 140 neighborhoods in Banglore. By using web

scraping and Beautiful soup and also Python requests we will extract

the list of Neighborhood. Then we will get the geographical co-ordina

tes of the neighborhoods such as Latitude and Longitudes with the

help of python geocoder package.

After that, we will use Foursquare API to get the venue data for

those neighbourhoods. Foursquare has one of the largest database

of 105+ million places and is used by over 125,000 developers. Four

square API will provide many categories of the venue data, we are

particularly intere sted 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.

**Methodology:**

First we need to get the neighborhoods of Banglore from the page:

(<https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Bangalore>) of Wikipedia.Then we have to retrieve the list of neighborhoods

using requests and Beautiful soup methods. These are just the names

of neighborhoods.

we have to retrieve the **geographical coordinates** of the locations

using the geopy package which is useful to convert address into

latitude and longitude. Using the geographical coordinates and the

folium package we have to visualize them in folium maps. Using the

four square UPI client id and client server we had to get the top 100

venues of radius 500 metres.

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. By doing so, we are also preparing the data

for use in clustering. Since we are analyzing the “Shopping Mall”

data, we will filter the “Shopping Mall” as venue category for the

neighborhoods. Then using the K-Means Clustering identifies k

number of centroids, and then allocates every data point to the

nearest cluster, while keeping the centroids as small as possible.

We will cluster the neighborhoods into 3 clusters based on their

frequency of occurence of the shopping mall in Neighborhoods.

Then we had to examin the clusters as cluster 0 and cluster1 and

Cluster2 to get the frequency of clusters.

**Results:**

* 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
* Red-Cluster0, Blue-Cluster1, Green-Cluster2

**Discussions:**

* As we seen or observed in the examine of clusters that most of the shopping malls are situated in the places of cluster 2 and moderate number in cluster1. But there were no shopping malls in the places of cluster0.
* So, it is advisible for the developers or Investors that malls should be located either in cluster0 places with no competition or cluster 1 places with moderate competition.
* It is not suggestible to open the shopping mall in cluster0 areas or neighborhoods.

**Recommendations:**

* Here we considered only the frequency of occurance of Shopping malls but there were other factors other than this such as **Population etc;**

**Conclusion:**

* In this capstone we have gone long through i.e extracting a data from webpage using requests and beautiful soup.
* Foursquare api requesting for geographical coordinates and mapping them.
* Analyzing the neighbourhood and examining the clusters for the recommendation of best places to Investors to open new Shopping mall.
* The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new shopping mall.