IBM Applied Data Science Capstone Course by Coursera

Finding Suitable Location for Shopping mall In Chennai city India

1. Introduction

Shopping malls can provide the best shopping experiences such as social gatherings, entertainment, performances, product launches, promotions and festivals. Malls can be very helpful because all the needs are in one building. Shopping malls tend to be a major tourist attraction. Shopping malls are becoming popular destinations because these are multi-storeyed air-conditioned buildings with shops on different floors. People have a large variety of goods to choose from whether branded or non- branded products.

Benefits of Shopping Malls.

1. They have their own parking facility.
2. There is a wide variety of products available.
3. There are products from competing producers available under one roof. ...
4. They have facilities such as restrooms.
5. They have gaming zones.
6. There are food courts with a wide variety of cuisine.
7. There are movie theatres in shopping centres.

Opening a shopping mall allows property owner to earn consistent source of Income. Main criteria to open a mall is location of the mall. Success or failure of that mall is depend on the location of the mall.

Business Problem

The objective of the capstone project to find out location, to open a new shopping mall in the city of chennai.

Why Chennai City?

I have selected Chennai as a city to open a new mall for various reasons. Few are discussed below.

Chennai, on the Bay of Bengal in eastern India, is the capital of the state of Tamil Nadu. Chennai is ranked as a beta-level city in the Global Cities Index, and was ranked the best city in India by India in the 2014 annual Indian city survey. Chennai, the capital city of Tamil Nadu, attracts many visitors either as tourists or as part of its large workforce. Chennai is popular for IIT and for being an IT and industrial hub. The city is also renowned for its passion for music.

Expected / Interested Audience

70% local and 30% foreign peoples visit Chennai once in a year as a tourist or on a business tour. Some people stay for couple of days or more. They are in search of multi cuisine restaurants and shopping hubs. Chennai Suburbs is saturated by Multinational IT industries and various multinational manufacturing industries. But number of shopping complex are less as compared to the population of suburbs Chennai.

The challenge is to find a suitable location for opening a new Shopping mall that will attract to local and foreign people.

2 Data Section

To solve the problem we need following data.

- List of neighbourhood in the city of Chennai which is located in Tamilnadu state of India.
- latitude and longitude of the neighbourhood to map the data and to select the venue.
- To perform clustering on neighbourhood we need venue data related to shopping mall.

2.a What data is used?

wikipedia page https://en.wikipedia.org/wiki/Category:Suburbs_of_Chennai

It contains list of neighbourhoods in the chennai city which contain 68 neighbourhoods. we will use webscraping technique to extract data from wikipedia page with the help of python request and beautifulsoup package. We will use python

geocoder to get geographical data i.e. latitude and longitude coordinates of neighbourhoods.

We will use *Foursquare API* to get venue data of the neighbourhood. The Foursquare Places API provides location based experiences with diverse information about venues, users, photos, and check-ins. The API supports real time access to places, Snap-to-Place that assigns users to specific locations, and Geo-tag. Additionally, Foursquare allows developers to build audience segments for analysis and measurement. JSON is the preferred response format. These features will help us to solve our problem to find good location to open shopping mall.

2.b Importing Liabraries

We will import basic python liabraries along with folium , geocoder,beautifulsoup,json etc.

2.c Credentials and Core location, 2.d Search for Mall, 2.e Location of Mall

To fulfil above requirement we will use data science techniques from web scraping (wikipedia), working with API (Foursquare), Data cleaning, data wrangling, machine learning (K means Algorithm) and map visualisation (Folium).

Some example how to download some liabraries!

```
import numpy as np
import pandas as pd
pd.set_option("display.max_columns", None)
pd.set_option("display.max_rows", None)

from geopy.geocoders import Nominatim
try:
    import geocoder
except:
    !pip install geocoder
    import geocoder

import requests
from bs4 import BeautifulSoup

try:
    import folium
except:
```

```
    !pip install folium
    import folium
    from sklearn.cluster import KMeans

from sklearn import preprocessing

from sklearn.model_selection import train_test_split

from sklearn.neighbors import KNeighborsClassifier

from sklearn import metrics

from sklearn.cluster import KMeans

from pandas.io.json import json_normalize # tranform JSON file into a pandas dataf
rame

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors
print ('Liabraries Downloaded')
```

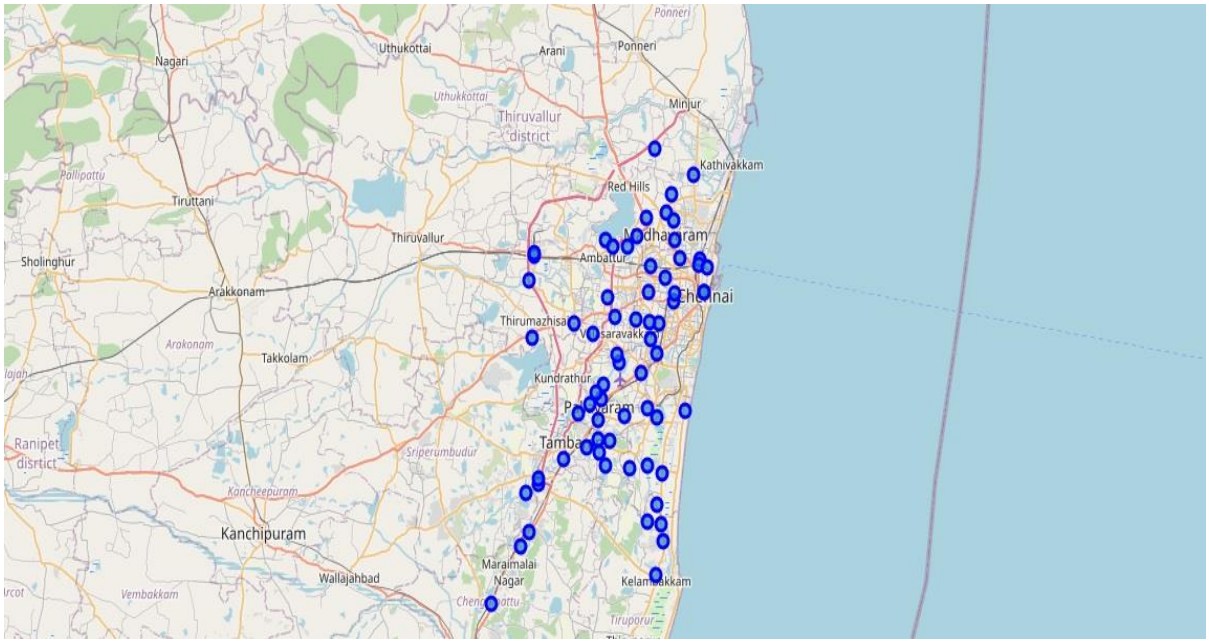
3. Methodology

Firstly, we need to get the list of neighbourhoods in the city of Chennai. Fortunately, the list is available in the Wikipedia
Page https://en.wikipedia.org/wiki/Category:Suburbs_of_Chennai
We will do web scraping using Python requests and beautifulsoup packages to extract the list of neighbourhoods data.
We need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the data into a pandas Data Frame and then visualize the neighbourhoods 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 Chennai
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 neighbourhoods 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 neighbourhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyse each neighbourhood by grouping the rows by neighbourhood 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 analysing the Shopping Mall" data, we will filter the "Shopping Mall" as venue category for the neighbourhoods.

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. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighbourhoods into 4 clusters based on their frequency of occurrence for "Shopping Mall". The results will allow us to identify which neighbourhoods have higher concentration of shopping malls while which neighbourhoods have fewer number of shopping malls. Based on the occurrence of shopping malls in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new shopping malls.

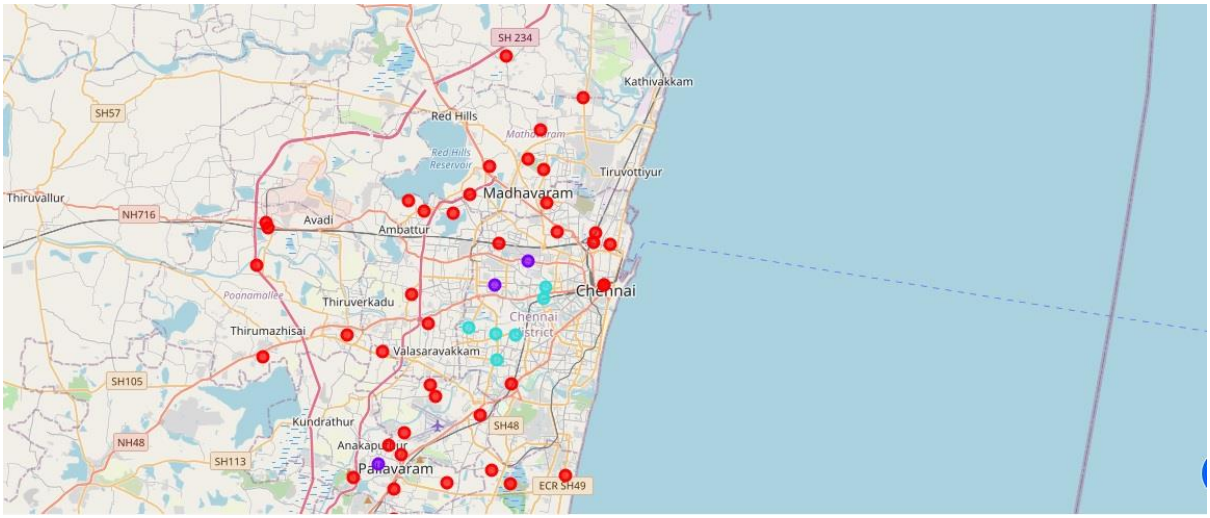
4. Results



The above map shows the Chennai suburbs neighbourhood. There are total 68 locations that we are going to analyse for finding correct location.

The results from the k-means clustering show that we can categorize the neighbourhoods into 4 clusters based on the frequency of occurrence for “Shopping Mall”:

- Cluster 0: Neighbourhoods with low number to no existence number of shopping malls
- Cluster 1: Neighbourhoods with moderate number of shopping malls
- Cluster 2: Neighbourhoods with high concentration of shopping malls
- cluster 3: Neighbourhoods with high concentration of shopping malls



The results of the clustering are visualized in the map below with cluster 0 in red colour, cluster 1 in purple colour, cluster 2 in mint green colour and cluster 3 in yellow.

```
kl_merged.head()
```

Out[74]:

	Neighborhood	Shopping Mall	Cluster Labels
0	Alandur	0.000000	0
1	Anna Nagar	0.030000	1
2	Ashok Nagar, Chennai	0.014085	2
3	Assisi Nagar	0.000000	0
4	Ayanavaram	0.032258	1

Above image shows some neighbourhood with frequency of availability of malls with cluster number.

5. Discussion

As observations noted from the map in the Results section, most of the shopping malls are concentrated in the central area of Chennai city, with the highest number in

cluster 2 and 3 , moderate number in cluster 1. On the other hand, cluster 0 has very low number to no shopping mall in the neighbourhoods. 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 neighbourhoods in cluster 0 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new shopping malls in neighbourhoods in cluster 0 with moderate competition. Lastly, property developers are advised to avoid neighbourhoods in cluster 2,3 which already have high concentration of shopping malls and suffering from intense competition.

6. Limitations and Suggestions for Future Research

In this project, we only consider one factor i.e. frequency of occurrence of shopping malls, there are other factors such as population and income of residents ,density of houses,that could influence the location decision of a new shopping mall. However, to the best knowledge of this researcher such data are not available to the neighbourhood level required by this project. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new shopping mall. In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Also having some limitation in getting sufficient data from Chennai location. Future research could make use of paid account to bypass these limitations and obtain more results.

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 4 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 answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighbourhoods in cluster 0 are the most preferred locations to open a 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.

References

1. Category: Suburbs in Chennai. Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Category:Suburbs_of_Chennai
2. Foursquare Developers Documentation. Foursquare. Retrieved from https://developer.foursquare.com/docs
3. Chennai malls information from https://en.wikipedia.org/wiki/Category:Shopping_malls_in_Chennai