

Coursera Capstone

IBM Applied Data Science Capstone

Opening a Shopping Mall in correct place in India

By: Sanyam Srivastava
May 2020

Introduction

For many people, visiting Shopping Mall and eating or shopping or exploring is a great way to relax and enjoy themselves during weekends and holidays. They can dine at restaurants, pack it for home, do shopping, etc. Of course, as with any business decision, opening a Shopping Mall requires serious consideration and is a lot more complicated than it seems. Particularly, the location of it, the competition in that area are one of the most important decisions that will determine whether the Shopping Mall will be a success or a failure. This becomes way more important when you are to open one of the biggest shopping mall.

Business Problem

The objective of this capstone project is to analyse and select the best locations in India to open a one of the biggest shopping mall. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In which city and where in that city, 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 mall in the capital city of India. This project is timely as the city is currently suffering from oversupply of such places.

Data to be used

To solve the problem, we will need the following data:

- The wikipedia page(https://en.wikipedia.org/wiki/List_of_shopping_malls_in_India) containing info about all the big malls of india.
- Latitude and longitude coordinates of capital city and then the existing malls of it. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to restaurants. We will use this data to perform clustering on the area.

METHODOLOGY

This Wikipedia page (https://en.wikipedia.org/wiki/List_of_shopping_malls_in_India) contains a list of all malls of India. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and **beautifulsoup packages**. After importing this data into a dataframe, we will conclude that delhi does not have a big shopping mall. Delhi being the national capital and a right amount of population to have one of the nation's biggest mall. Hence, we will now work on Delhi.

We will get the geographical coordinates of the Delhi and it's neighbour Gurugram and Gaziabad. Using **Python Geocoder** package which will give us the latitude and longitude coordinates of the neighbourhoods.

After that, we will use **Foursquare** API to get the venue data of Malls for those areas. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested 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).

Working:

We have after doing scraping and then combining the mall data of the states. The output is:

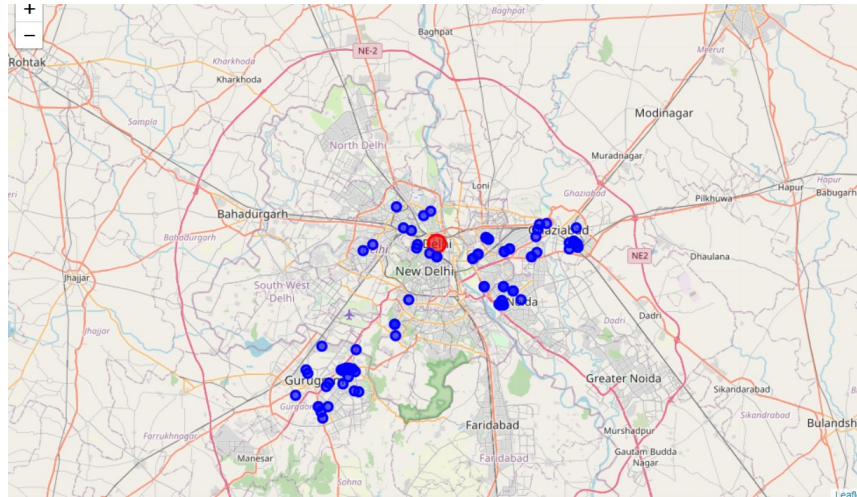
```
In [144]: dataframe_filtered
```

```
Out[144]:
```

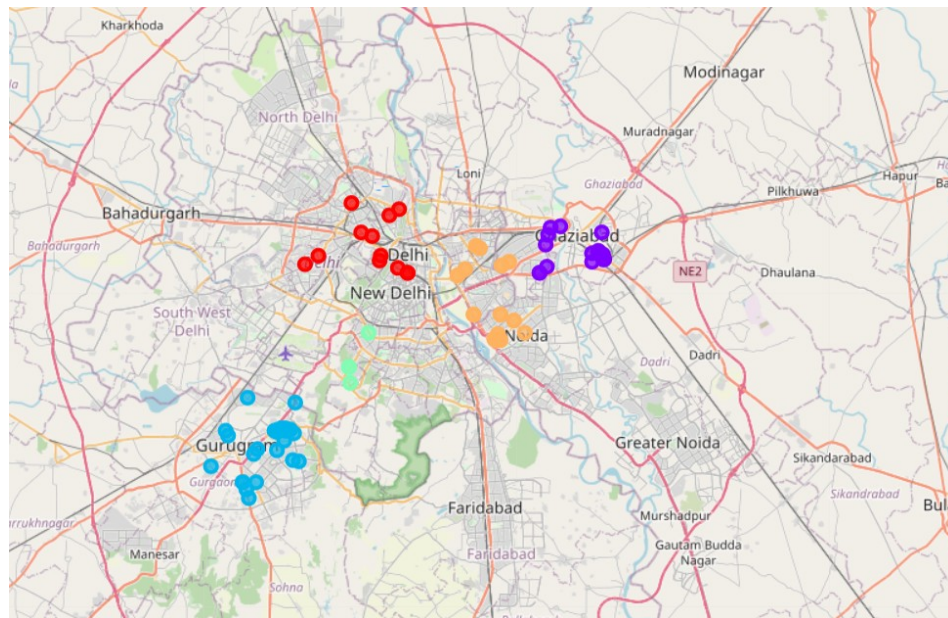
| | name | categories | lat | lng | distance |
|-----|--------------------------------------|----------------|-----------|-----------|----------|
| 0 | Wave Cinemas, Gaur Central Mall, RDC | Multiplex | 28.646943 | 77.440607 | 3879 |
| 1 | Shoppers Stop Shipra Mall | Clothing Store | 28.634457 | 77.369911 | 5798 |
| 2 | Mahagun Metro Mall | Shopping Mall | 28.645045 | 77.335597 | 8012 |
| 3 | Mall Road, Shimla, Himachal Pradesh | Scenic Lookout | 28.655501 | 77.435200 | 2855 |
| 4 | Gaur Central Mall | Shopping Mall | 28.673960 | 77.440178 | 2766 |
| ... | ... | ... | ... | ... | ... |
| 85 | Manyavar @ MGF Metropolitan Mall | Men's Store | 28.481079 | 77.080292 | 5258 |
| 86 | Lift @ MGF Mall | None | 28.477965 | 77.071177 | 4302 |
| 87 | Cross Point Mall | Shopping Mall | 28.468426 | 77.083126 | 5224 |
| 88 | DLF City Center Mall | Shopping Mall | 28.479007 | 77.076236 | 4807 |
| 89 | Pantaloons @ Sahara Mall | Clothing Store | 28.478193 | 77.085285 | 5624 |

90 rows × 5 columns

Then, we using folium plotted these malls on a map. The output was :



We then performed K means on this. We selected 5 clusters. We then appended this column to our main data frame and then plotted it as well. The result was:

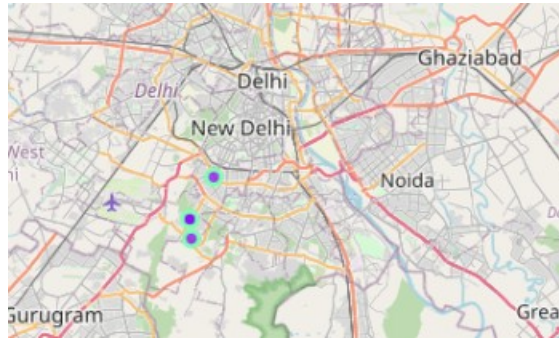


Now, we count the number entries of each cluster. It is clear that the one with least number will be our target and that cluster is '2'.

```
: dataframe_filtered['clusterLabels'].value_counts()
: 3    29
  1    28
  0    17
  4    12
  2     4
   Name: clusterLabels, dtype: int64
```

Result:

We have concluded that the cluster 2 will be our place. For better understanding, here is a visualization of it.



So, judging by the map, the area between New Delhi and Gurugram seems like the perfect place to open a mall.