Coursera Capstone Project: Applied Data Science

Abhishek Kumar
abhiavisekkr@gmail.com

July, 2020

Opening a new Shopping Mall in Delhi, India

1. Introduction

The importance of shopping malls as retailing formats has become increasingly remarkable, and today malls plays a significant role in consumers' lifestyle. But nowadays shopping malls have become not just a place to shop it has become a place where social factors get deeper. Shopping malls are like one-stop destination for not only various types of shoppers but also for dine at restaurants, watch movies, celebrate etc. This gives retailers a central location and large crowd at the shopping mall which in turn provides a great distribution channel to market their products and services. Real Estate, builders are also taking interest to build shopping malls to cater the demand. This also becomes a consistent rental income for the owners. As a result, there are many shopping malls in the Delhi and more and more malls are being built. But to open a shopping and such that the shopping mall succeeds lot of factors come into play. Among them one of the major factors is the location of the shopping mall.

2. Business Problem

Since, shopping malls are main interest to various groups therefore they are built and real estate investors invest in these projects. But for shopping malls to attract large crowd there are few major factors, one of those is location. The objective of this capstone project is to analyse and select the best locations in Delhi, India to open a new shopping mall with high chances of success using data science methodology and machine learning techniques. We will make this decision in this project. This will specially benefit the real estate builders since Indian retail sector has metamorphosed significantly over last few decades. Rapid urbanization and digitization, rising disposable incomes and lifestyle changes of particularly the middle-class has led to a major revolution in the retail sector, projected to grow from US $672 billion in 2017 to US $1.3 trillion in 2020. Evolving rapidly from usual 'kirana shops' to large multi-format stores offering global experience to the e-commerce model that is highly technology-driven, the Indian retail sector has evolved.

3. Data

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

- List of neighbourhoods in Delhi. This defines the scope of this project which is confined to Delhi, the capital of India.
- Latitude and Longitude coordinates of the neighbourhoods. This required to plot the map and get the venue data
- Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighbourhoods.

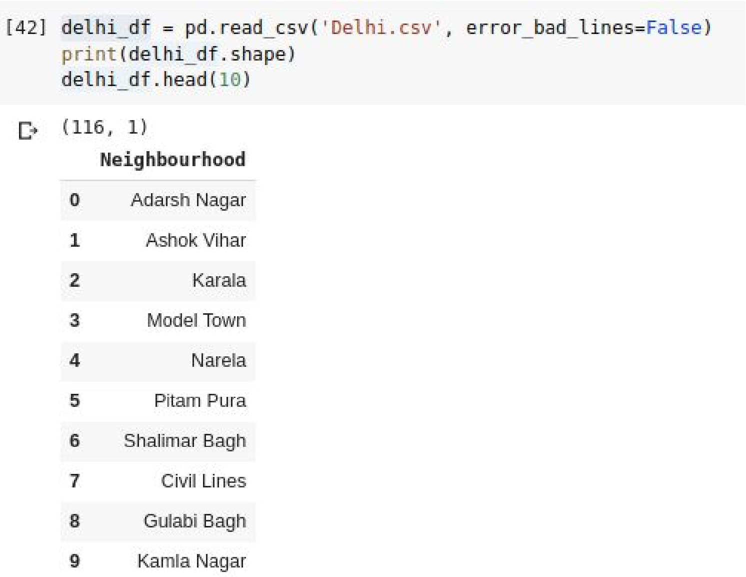
Data Sources:

- The data of the neighbourhoods in Delhi can be extracted from Wikipedia page. (https://en.wikipedia.org/wiki/Neighbourhoods_of_Delhi)
- Then the latitude and longitude data can be retrieved from Python geocoder package.
- Then using latitude and longitude data venues can be fetched from Foursquare API.

4 Methodology

Method to extract data:

- We do web scraping using BeautifulSoup a library of python to get the neighbourhoods from the Wikipedia page. We will send get request to get the html page and then extract the list and store it in a csv file. Shape of our data frame is (116 × 1)



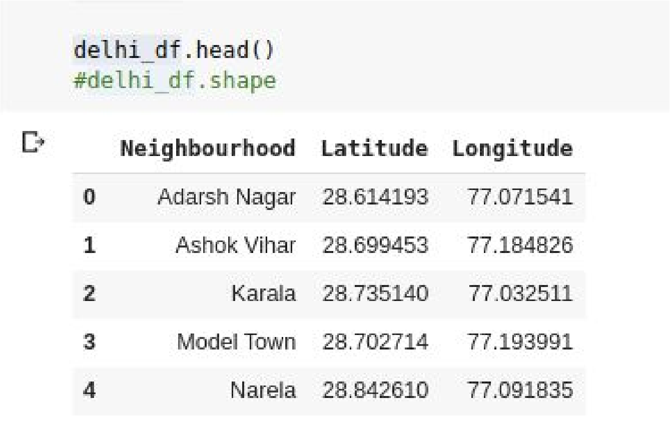
```
[42] delhi_df = pd.read_csv('Delhi.csv', error_bad_lines=False)
     print(delhi_df.shape)
     delhi_df.head(10)
```

```
(116, 1)
```

	Neighbourhood
0	Adarsh Nagar
1	Ashok Vihar
2	Karala
3	Model Town
4	Narela
5	Pitam Pura
6	Shalimar Bagh
7	Civil Lines
8	Gulabi Bagh
9	Kamla Nagar

Method get Longitude and Latitude and Visualize :

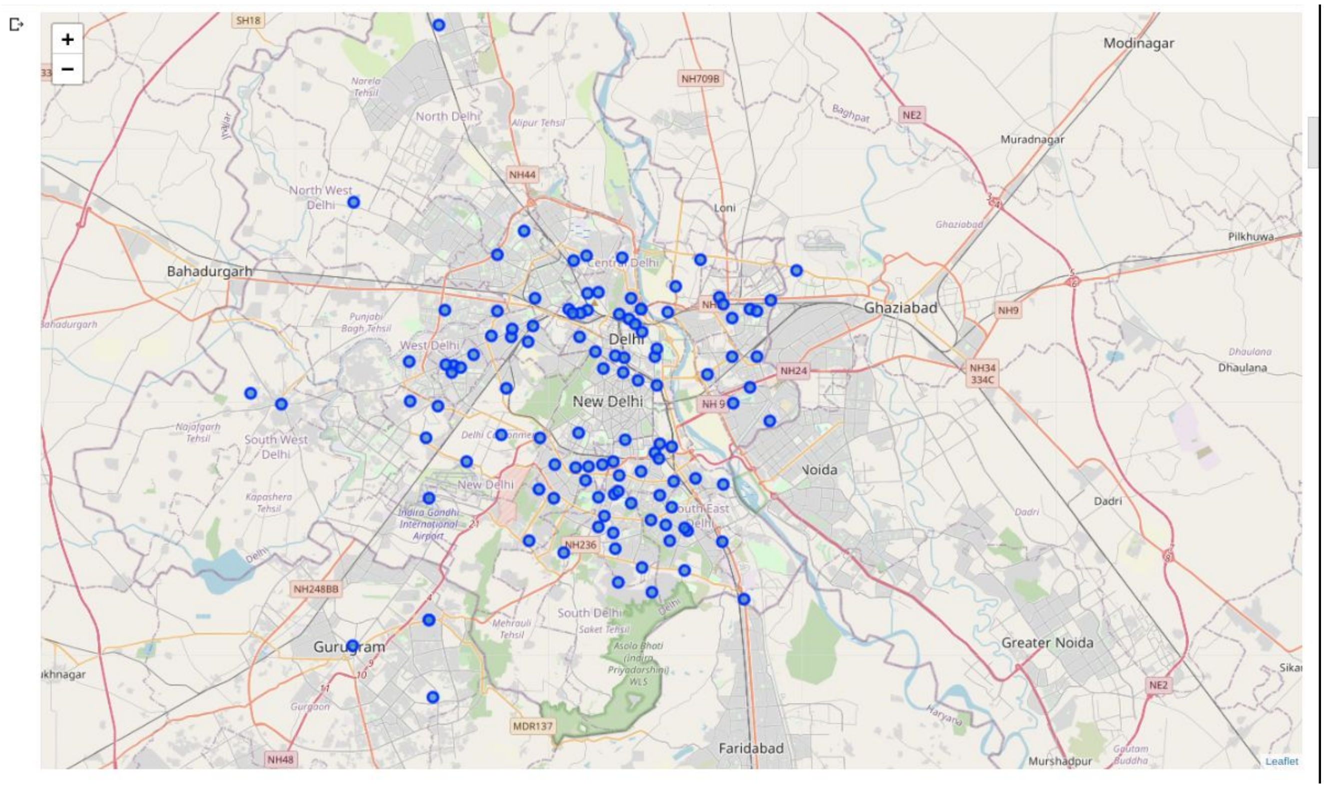
- The list will be just names of the places but we will need their geographical coordinates so we will pass the data from the csv file to Geocoder to get the latitude and longitude. Then we will append the latitude and longitude of individual location to the dataframe.



```
delhi_df.head()
#delhi_df.shape
```

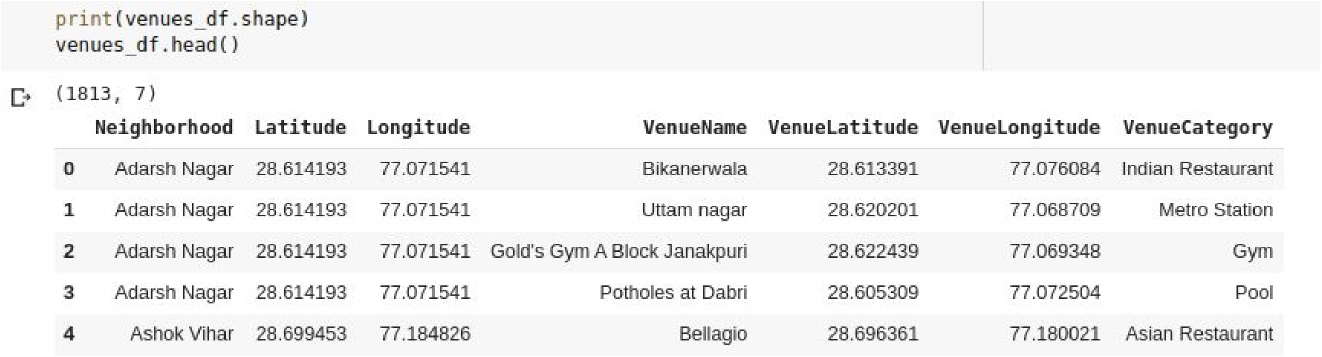
	Neighbourhood	Latitude	Longitude
0	Adarsh Nagar	28.614193	77.071541
1	Ashok Vihar	28.699453	77.184826
2	Karala	28.735140	77.032511
3	Model Town	28.702714	77.193991
4	Narela	28.842610	77.091835

- Now we can visualize these locations on map using Folium. This allows us to perform a sanity check to make sure that the geographical coordinates returned by Geocoder are correctly plotted.



Method to get venue data from FourSquare API:

- Next, we will use Foursquare API to get the top 30 venues that are within a radius of 1000 meters. For this a Foursquare developer account is needed.
- We make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python. We use the venues/explore API endpoint to request the data. Foursquare returns the venue data in JSON format which is then decoded to extract venue names, venue category, venue longitude and venue latitude.



```
print(venues_df.shape)
venues_df.head()
```

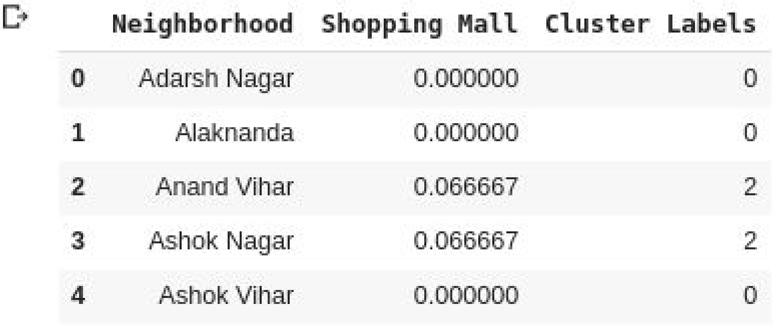
```
(1813, 7)
```

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Adarsh Nagar	28.614193	77.071541	Bikanerwala	28.613391	77.076084	Indian Restaurant
1	Adarsh Nagar	28.614193	77.071541	Uttam nagar	28.620201	77.068709	Metro Station
2	Adarsh Nagar	28.614193	77.071541	Gold's Gym A Block Janakpuri	28.622439	77.069348	Gym
3	Adarsh Nagar	28.614193	77.071541	Potholes at Dabri	28.605309	77.072504	Pool
4	Ashok Vihar	28.699453	77.184826	Bellagio	28.696361	77.180021	Asian Restaurant

Grouping data and Performing Clustering:

- 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 algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping centroid as small as possible. In this project we have clustered the neighbourhoods into 3 clusters based on their frequency of occurrence for “Shopping Mall”. Now, lets go towards the results and discuss them.



	Neighborhood	Shopping Mall	Cluster Labels
0	Adarsh Nagar	0.000000	0
1	Alaknanda	0.000000	0
2	Anand Vihar	0.066667	2
3	Ashok Nagar	0.066667	2
4	Ashok Vihar	0.000000	0

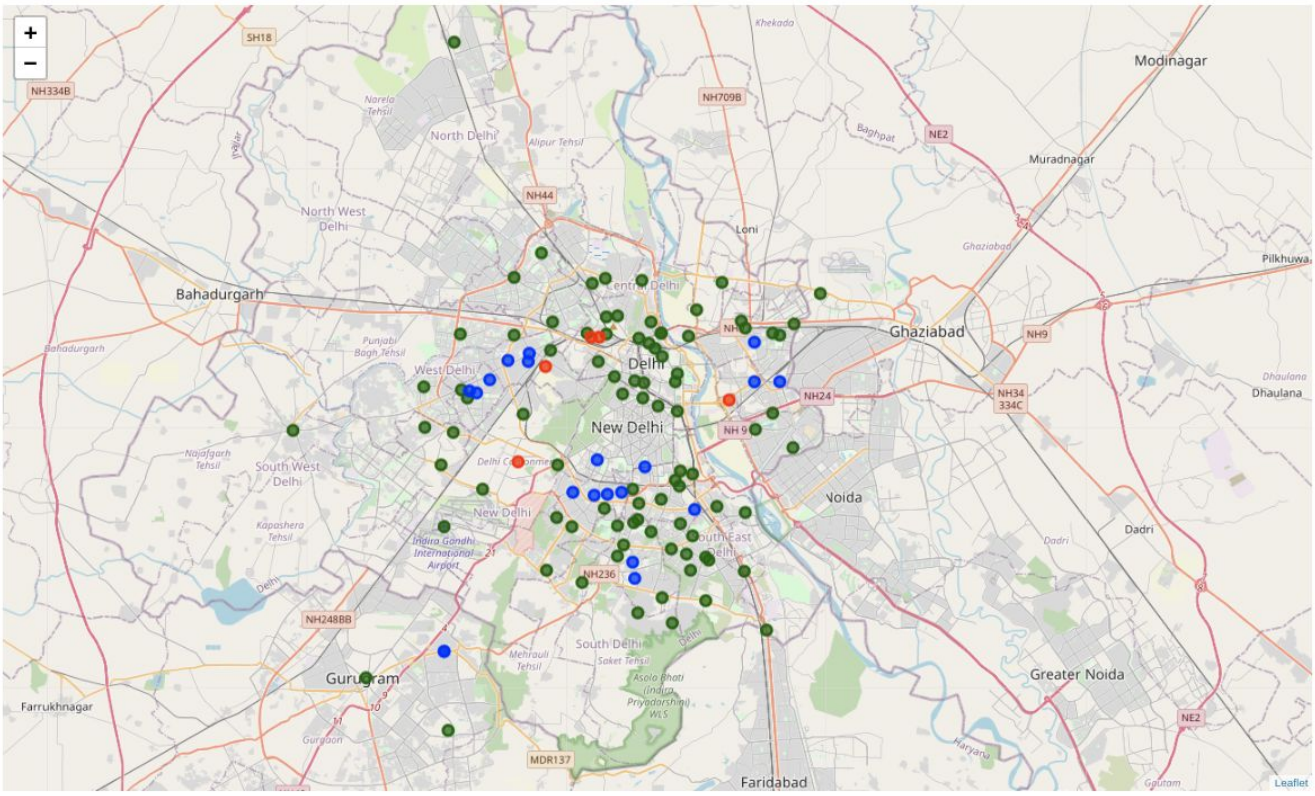
5 Result:

Results from K-means Clustering show that:

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

Colour Code

- Cluster 0 – Dark Green
- Cluster 1 – Red
- Cluster 2 – Blue



6 Discussion

Most of the shopping malls are concentrated in the central east, southern and few on the northern part of Delhi, with

- The highest number in cluster 1
- Moderate number in cluster 2.
- On the other hand, cluster 0 has a very low number to total 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 1 are likely suffering from intense competition due to oversupply and high concentration of shopping malls. From another perspective, this also shows that the oversupply of shopping malls mostly happened in the central eastern and southern part of the city, with the suburb area still having a 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 2 with moderate competition.

Lastly, property developers are advised to avoid neighbourhoods in cluster 1 which already have high concentration of shopping malls and suffering from intense competition.

7 Conclusion

In this project, we have gone through all the data science methodology. We first identified a business problem, then collected the required data, preparing the data and then performing machine learning by clustering the data into 3 clusters based on their similarities.

Finally, we have also provided recommendations to the relevant stakeholders regarding the best locations to open a new shopping mall.

To answer the problem which was raised at the Introduction: The neighbourhoods in cluster 0 are most preferred location to open a new shopping mall.

Appendix

Cluster 0

- Adarsh Nagar
- Patel Nagar
- Paschim Vihar
- Palam
- Paharganj
- Old Delhi Railway Station
- Okhla
- Nizamuddin West
- New Usmanpur
- New Friends Colony
- Pitam Pura
- New Delhi Railway Station
- Neeti Bagh
- Naveen Shahdara
- Narela
- Naraina
- Najafgarh
- Munirka
- Wazirabad
- Mori Gate
- Model Town
- Nehru Place
- Pragati Maidan
- Punjabi Bagh
- Sadar Bazaar
- Vivek Vihar
- Vikaspuri
- Vasundhara Enclave
- Vasant Vihar
- Vasant Kunj
- Tughlaqabad
- Tis Hazari
- Timarpur
- Tilak Nagar
- South Extension
- Siri Fort
- Shastri Park
- Shastri Nagar
- Shalimar Bagh
- Shakti Nagar
- Shahdara
- Sarvodaya Enclave
- Sarita Vihar
- Sangam Vihar
- Sainik Farm
- Safdarjung Enclave
- Mehrauli
- Mayur Vihar
- Yamuna Vihar
- Ghaziabad
- Hauz Khas
- Gurugram
- Gulmohar Park
- Gulabi Bagh
- Civil Lines
- Green Park
- Greater Kailash
- Chittaranjan Park
- Govindpuri
- Connaught Place
- Dabri
- Daryaganj
- Defence Colony
- Fateh Nagar
- East of Kailash
- Dhaula Kuan
- Gole Market
- East Vinod Nagar
- Hazrat Nizamuddin Railway Station
- Indira Gandhi International Airport
- Alaknanda
- Ashok Vihar
- Lajpat Nagar
- Kotwali
- Badarpur
- Khanpur
- Kashmiri Gate
- INA Colony
- Barakhamba Road
- Karol Bagh
- Kamla Nagar
- Kalkaji
- Jor Bagh
- Jhilmil Colony
- Jhandewalan
- Chandni Chowk
- Jangpura
- Kashmiri Gate
- Dilshad Garden

Cluster 1

- Delhi Cantonment
- Moti Bagh
- Laxmi Nagar
- Pandav Nagar
- Sarai Rohilla

Cluster 2

- Sarojini Nagar
- Lodi Colony
- Vishwas Nagar
- Anand Vihar
- Ashok Nagar
- Malviya Nagar
- Kirti Nagar
- Bali Nagar
- Moti Nagar
- Netaji Nagar
- Tihar Village
- Sriniwaspuri
- Noida
- Preet Vihar
- Rajouri Garden
- Rama Krishna Puram
- Faridabad
- Saket
- Chanakyapuri
- Laxmibai Nagar