Coursera Capstone – Final Project
Finding a suitable place to open a shopping mall in Delhi.

By Nimesha Dilini

Introduction/ Business Problem

The shopping mall is an emerging and lucrative business in India. However, the business demands the right strategic planning with financial and marketing planning. Shopping malls are getting tremendous popularity these days. And definitely, there are several reasons behind it. According to the retail industry experts, the demand for the mall will more increase in the coming years.

A business article presented that(How to Start a Shopping Mall Business In India - NextWhatBusiness) when you are building a mall or shopping center, entire process starts with market research and ends with marketing and maintenance. First of all, you must do some market surveys or market research and then Finding the location and acquiring the land. Therefore, you should have a proper way to select a land for opening a shopping mall. You have to identify the competition and business opportunities.

Objective

The main goal of this research is to analyze and select the best locations in Delhi to open a new shopping mall. Data science and machine learning techniques like clustering can be used in this case. foursquare location API used to get venue details. Therefore, the project gives the solution to answer the business question: When a businessman looking for a land to a shopping mall in new Delhi, where would you recommend opening it?

Data

For finding an answer to the above problem, we have to use following datasets,

- List of boroughs and neighborhoods dataset with longitudes and latitudes and
- Venue data of shopping malls in the Delhi area (use the foursquare API to find all types of restaurants within a 1000-meter radius for every neighborhood.)

Sources of Data:

The List of boroughs and neighborhoods dataset was taken form Kaggle. (Delhi Neighborhood Data | Kaggle)

	Borough	Neighborhood	latitude	longitude
0	North West Delhi	Adarsh Nagar	28.61419	77.07154
1	North West Delhi	Ashok Vihar	28.69945	77.18483
2	North West Delhi	Azadpur	28.70766	77.17555
3	North West Delhi	Bawana	28.79966	77.03288
4	North West Delhi	Begum Pur		
5	North West Delhi	Dhaka	39.03171	-90.2612
6	North West Delhi	Jahangirpuri	28.72597	77.16266
7	North West Delhi	Karala	28.73514	77.03251
8	North West Delhi	Keshav Puram	28.68893	77.16168
9	North West Delhi	Kingsway Camp	28.61426	77.20155
10	North West Delhi	Kohat Enclave	28.69804	77.14054
11	North West Delhi	Model Town	39.03171	-90.2612
12	North West Delhi	Narela	28.84261	77.09184

This has 4 columns namely, Borough, Neighborhood, latitude and longitude.

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.'(Wikipedia)

Here we are using the explore API call and filtering the search only to find venues that are identified as restaurants.

Methodology

1. Loading Dataset

First, I had to find a neighborhood data set for Delhi. Fortunately, I founded a dataset from Kaggle. I have read that csv file using pandas. There were some NaN values. Therefore, I have dropped rows with null values.

```
[2]  #loading the neighborhood data set. I get the data set from kaggle
     delhi = pd.read_csv('/content/sample_data/delhi_dataSet.csv',skipinitialspace=True, usecols=['Borough','Neighborhood','latitude','longitude'])

[3]  #Drop rows containing NaN values
     delhi = delhi.dropna()

[4]  delhi.head()
```

	Borough	Neighborhood	latitude	longitude
0	North West Delhi	Adarsh Nagar	28.614192	77.071541
1	North West Delhi	Ashok Vihar	28.699453	77.184826
2	North West Delhi	Azadpur	28.707657	77.175547
3	North West Delhi	Bawana	28.799660	77.032885
5	North West Delhi	Dhaka	39.031714	-90.261223

These neighborhood data visualized in a map using Folium library.
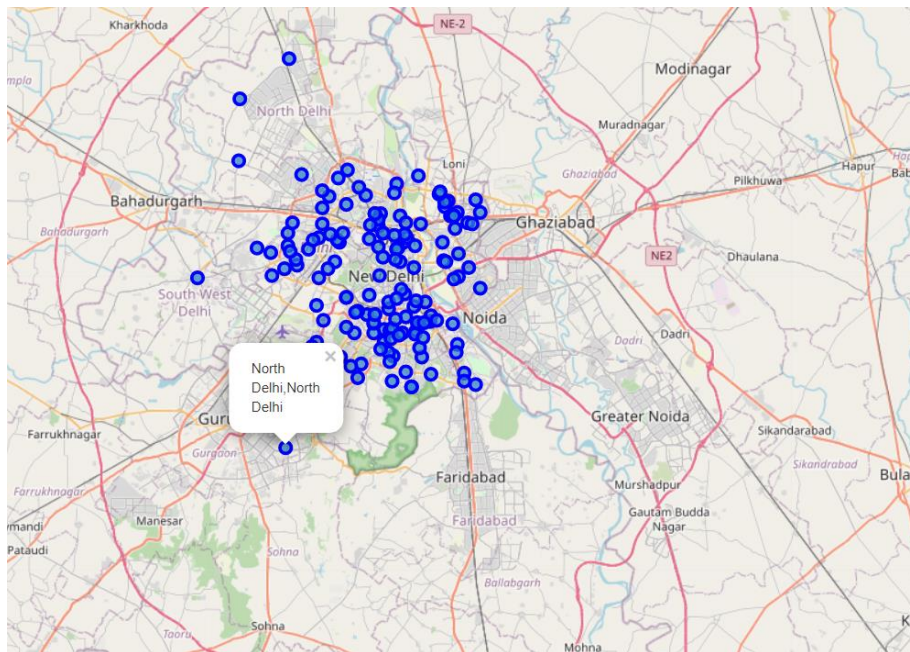
2. Explore Venues by using Foursquare API

The Foursquare API allows application developers to interact with the Foursquare platform. The API itself is a RESTful set of addresses to which you can send requests. Top 100 venues that are within radius of 2000 meters are retrieved using Foursquare API. There you need to have a Foursquare developer account and obtain the Foursquare ID and Foursquare secret key for making API requests. Then we can make API calls by passing the list of neighborhoods by using a python loop.

Foursquare return venue data in JSON format and we have to extract the venue name, venue category, venue latitude and venue longitude. Then we can check how many venues are returned for each neighborhood and examine how many unique categories can be found.

```
[14]  venues_df['VenueCategory'].unique()[:50]

      array(['Indian Restaurant', 'Café', 'Pizza Place', 'Coffee Shop',
             'Sandwich Place', 'Gym', 'Pool', 'Bakery', 'Fast Food Restaurant',
             'Gym / Fitness Center', 'Train Station', 'Asian Restaurant',
             'Snack Place', 'Chinese Restaurant', 'Department Store',
             'Donut Shop', 'South Indian Restaurant', 'Food Truck',
             'Convenience Store', 'Bus Station', 'Lake', 'Dessert Shop',
             'Market', 'Restaurant', 'Clothing Store',
             'North Indian Restaurant', 'Multiplex', 'Light Rail Station',
             'Shopping Mall', 'Playground', 'Business Service', 'Dive Bar',
             'Hotel', 'ATM', 'Athletics & Sports', 'Diner',
             'Arts & Crafts Store', 'Smoke Shop', 'Music Venue',
             'History Museum', 'Spa', 'Garden', 'Karnataka Restaurant',
             'BBQ Joint', 'Northeast Indian Restaurant', 'Hotel Bar', 'Lounge',
             'Sculpture Garden', 'Planetarium', 'Moroccan Restaurant'],
            dtype=object)
```

3. Analyze Venues.

One hot encoding is applied to venues of neighborhood. Each neighborhood is analyzed by taking the mean frequency of occurrences of each venue category. These processing are done to prepare data for K-Means Clustering. Since the problem interested in 'Shopping Mall', filtered the neighborhood with 'Shopping Mall' venues.

4. K – Means Clustering on Data

we need to cluster all the neighbourhoods into different clusters. The results will allow us to identify which neighbourhoods have a higher concentration of shopping malls while which neighbourhoods have a fewer number of shopping malls. Based on the occurrence of shopping malls in different neighbourhoods, it will help us answer the question as to which neighbourhoods are most suitable to open new shopping malls.

We set the number of clusters to 3 and run the algorithm. After applying the K-Means clustering algorithm, all the neighbourhoods get segregated and form different clusters.

	Neighborhood	Shopping Mall	Cluster Labels
0	Adarsh Nagar	0.000000	1
1	Alaknanda	0.000000	1
2	Anand Vihar	0.055556	0
3	Ashok Nagar	0.000000	1
4	Ashok Vihar	0.000000	1

Here the Shopping Mall column represents the number of shopping malls in that particular area and Cluster Labels represents the cluster number (either 0 or 1 or 2)

Results

There 125 places in Cluster 1 and that is the cluster with highest number of places. Cluster 0 have 32 places and Cluster 2 having only 6 places. Cluster 0 represents which do not have shopping malls. That means there we can see *32* places which do *not have shopping malls*, *125* places *having one shopping mall* and *6* places with *2 or more shopping malls*.

The results from the K-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for "Shopping Mall":
• Cluster 0: Neighbourhoods with very a smaller number of shopping malls
• Cluster 1: Neighbourhoods with a moderate concentration of shopping malls
• Cluster 2: Neighbourhoods with a high concentration of shopping malls

In the following graph Cluster 0 is in red color, Cluster 1 in purple color and Cluster 2 in mint green color.
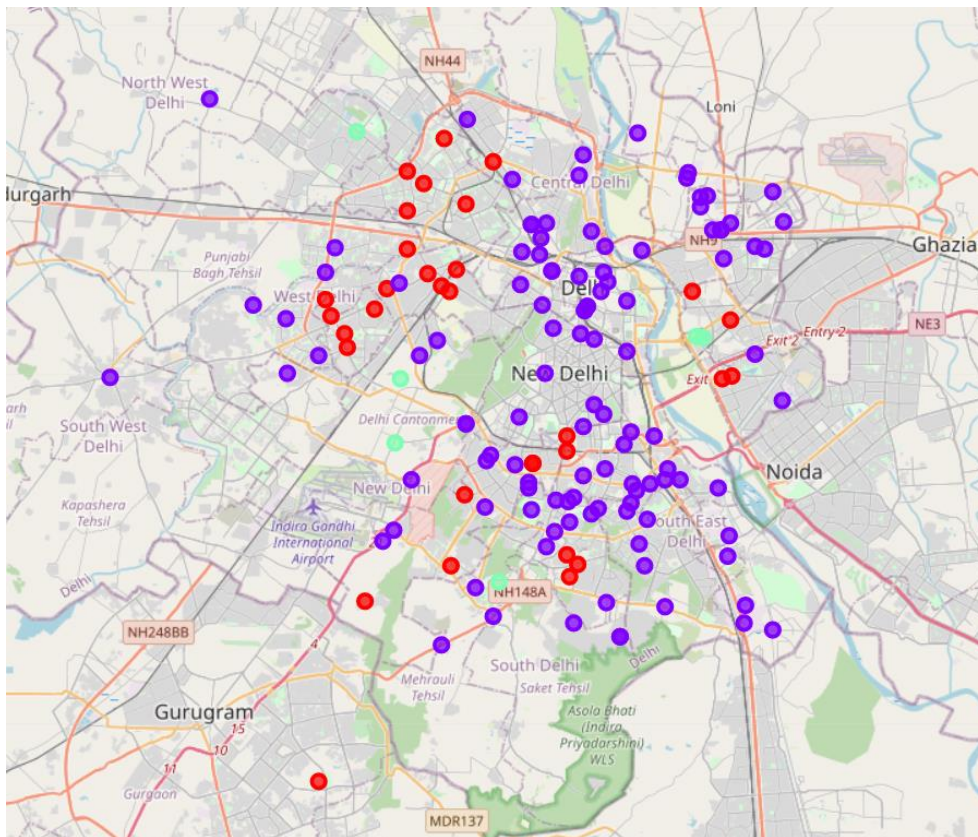
Discussion

Most of the shopping malls are concentrated to the center of Delhi. You can see there have at least one shopping mall in most of the neighborhoods. Only 32 neighborhoods not having a shopping mall. There is a lack of opportunities to start a shopping mall. But we can select a

neighborhood from cluster 0. For that we have to consider more attributes like population of that areas to get a clear decision. But with the available data final conclusion the opportunities to start a shopping mall in Delhi is very less. But this conclusion may vary when considering some more factors.

…