REPORT

ON

PROJECT

BEST PLACE TO OPEN A GYM

Author:
Prince

INTRODUCTION/BUSINESS PROBLEM

The objective of this capstone project is to find the best place to open a gym in Delhi. Delhi , capital of India have high population of people and opening a new business here can be a great success if any businessman /investor know the suitable place for business which give them an advantage on competitors.

This project is particularly useful to :-

- Businessman ,
- property dealers
- fitness freaks
- People want to open a gym

as it helps them to know about number of Gyms in each localities of Delhi and find the localities(Neighborhood) with least number of Gyms.

DATA SECTION

For this project, I used two data sets to execute my idea :-

1. Delhi's places data
2. Foursquare Location Data

Delhi's places data - This data is collected from the Kaggle site. It contain all Neighborhoods' name , Borough of each Neighborhood and its coordinates.

I cleaned and processed the data , and display the map of Delhi containing all Neighborhoods in data.

Foursquare Location Data - For getting the venues information nearby each Neighborhood I used Foursquare API , to get Venue's name , Venue's category and other things.
This data helps me to identify the number of gyms near each Neighborhood.

METHODS USED TO SOLVE PROBLEM

For, the Project's Problem to be solved I have used many techniques.

- First, to get the list of Neighborhoods data of Delhi I convert the .csv file to Pandas Dataframe.

- Then, I use *Geocoder Library* to get the Latitudes and Longitudes of each Neighborhood and dropping the previously present Coordinates to use only correct data for further analysis.

- After, the cleaning of data our Dataframe look like this : -

```
[42]: delhi_new.head()
```

[42]:

	Borough	Neighborhood	latitude	longitude
0	North West Delhi	Adarsh Nagar	28.614193	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
4	North West Delhi	Begum Pur	28.725503	77.058371

```
[44]: delhi_new.shape
```

```
[44]: (107, 4)
```

- Then , we will visualize the map of Delhi with its Neighborhood using *Folium Library* as this library is very effective and easy to use.

- To get the all nearby venues of each Neighborhood I call the *Foursquare API* .

I , get the venues of each Neighborhood which present under 1000 meter of range and the maximum venue can be get is limit to 100.

- Then , I analyse each neighborhood by grouping the rows by neighborhood and taking the mean of the frequency of each occurrence of venue having gym/fitness Center category.

- Then I grouped the columns of gym and fitness center to one as they direct to one category.

Now we will merge these two columns as both directs to a gym or place which have a gym and other things.

```
[115]: venues_gym['GYM'] = venues_gym['Gym'] + venues_gym['Gym / Fitness Center']

[121]: venues_gym = venues_gym.drop(["Gym" ,"Gym / Fitness Center"] , axis = 1)

[122]: venues_gym.head()
```

[122]:

	Neighborhood	GYM
0	Adarsh Nagar	0.25
1	Anand Vihar	0.00
2	Ashok Nagar	0.00
3	Ashok Vihar	0.00
4	Azadpur	0.00

- 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. We will cluster the neighborhoods into 3 clusters based on their frequency for “Gym”. The results will allow us to identify which neighborhoods have higher concentration of gyms while which neighborhoods have fewer number of gyms.

The code is like :-

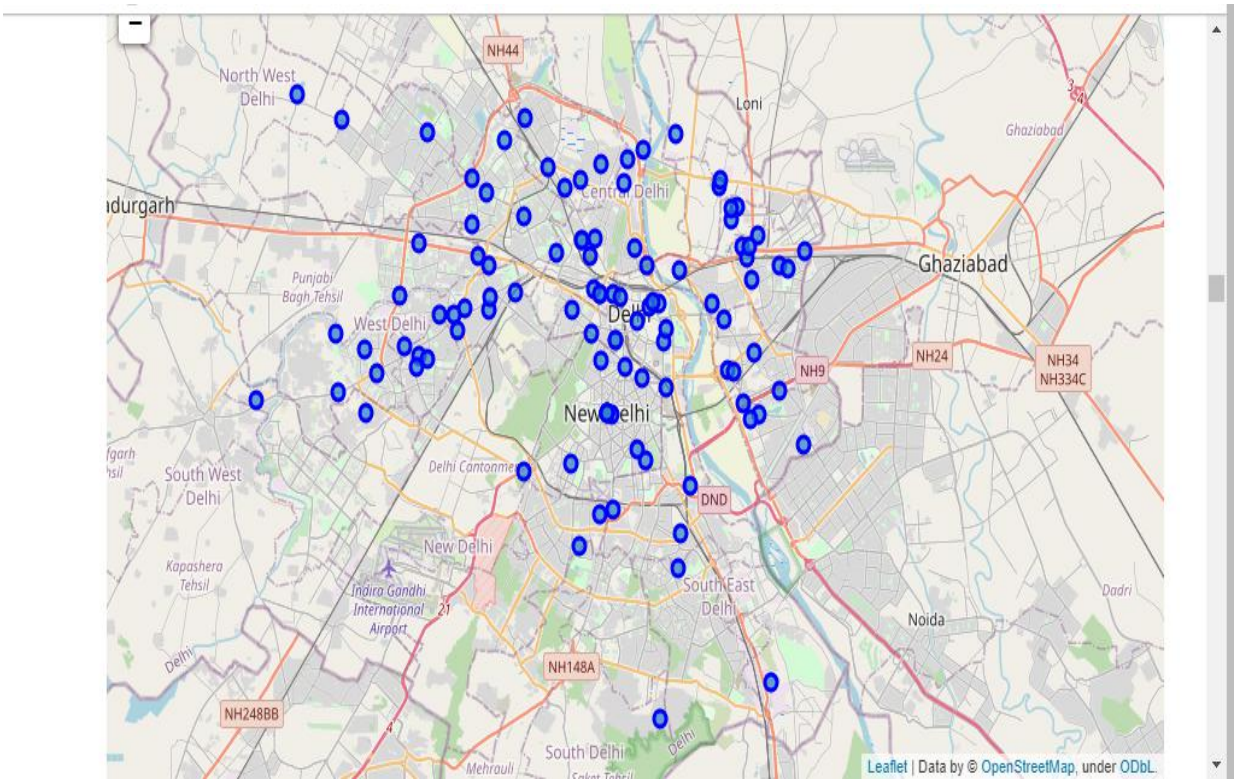
```
n_clusters = 3

delhi_venues_clustering = venues_gym.drop('Neighborhood' , axis =1)

kmeans = KMeans(n_clusters = n_clusters ,random_state =10).fit(delhi_venues_clustering)

kmeans.labels_
```

- After this , I made a map to show the neighborhoods with their labels(its cluster label) to easily know which Neighborhood is best for opening a new gym.

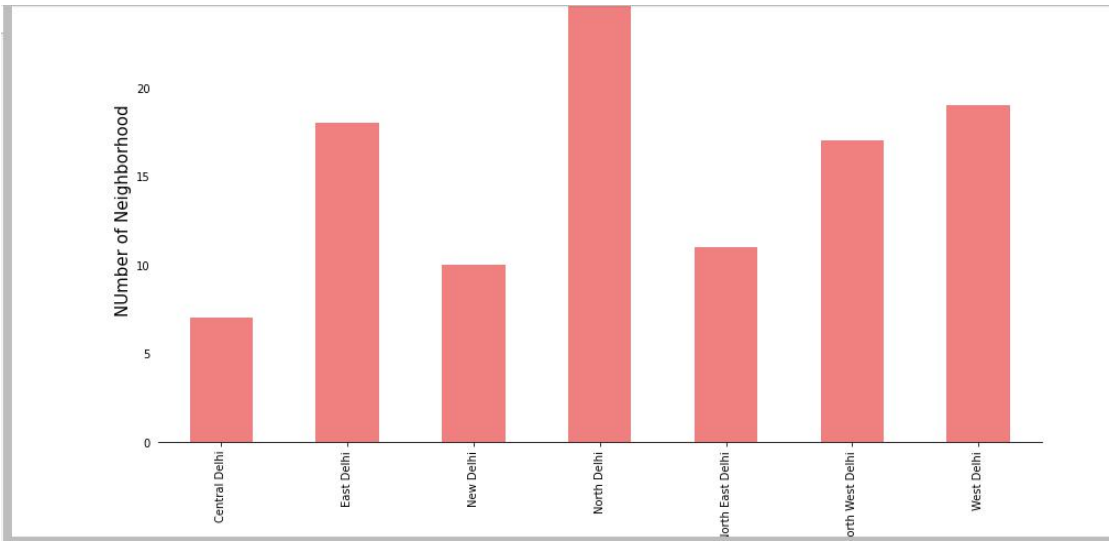
ANALYSIS SECTION

In this project , we analyzed many things.

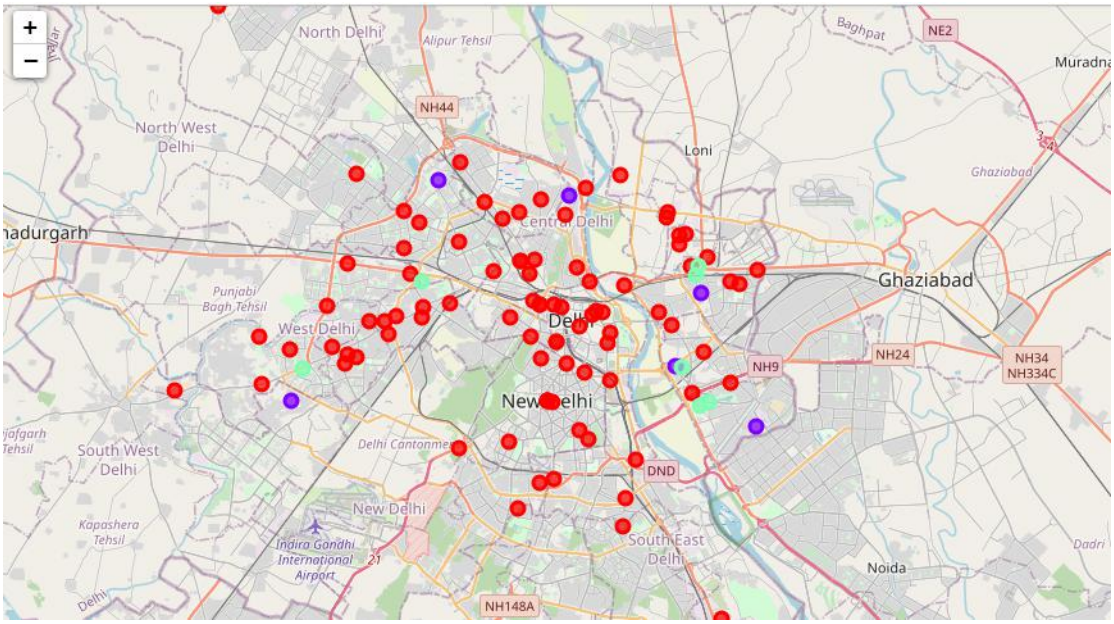
1. Visualizing all Neighborhood in Delhi using Folium map.



2. Number of Neighborhood in each Borough.



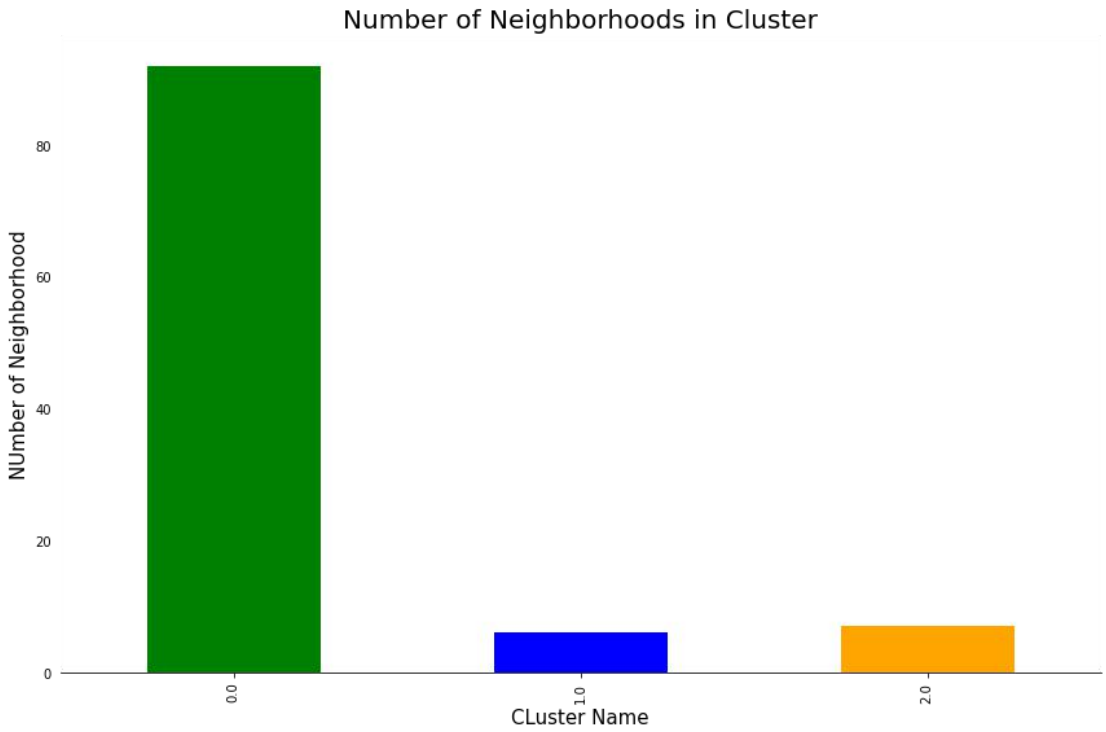
3. After analyzing and making cluster, making map of the Neighborhoods with its cluster label.



Here,

Red color marker denotes ----> cluster 0
Blue color marker denotes ----> cluster 1
Green color marker denotes ----> cluster 2

4. Number of Neighborhood in each cluster type.



RESULT SECTION

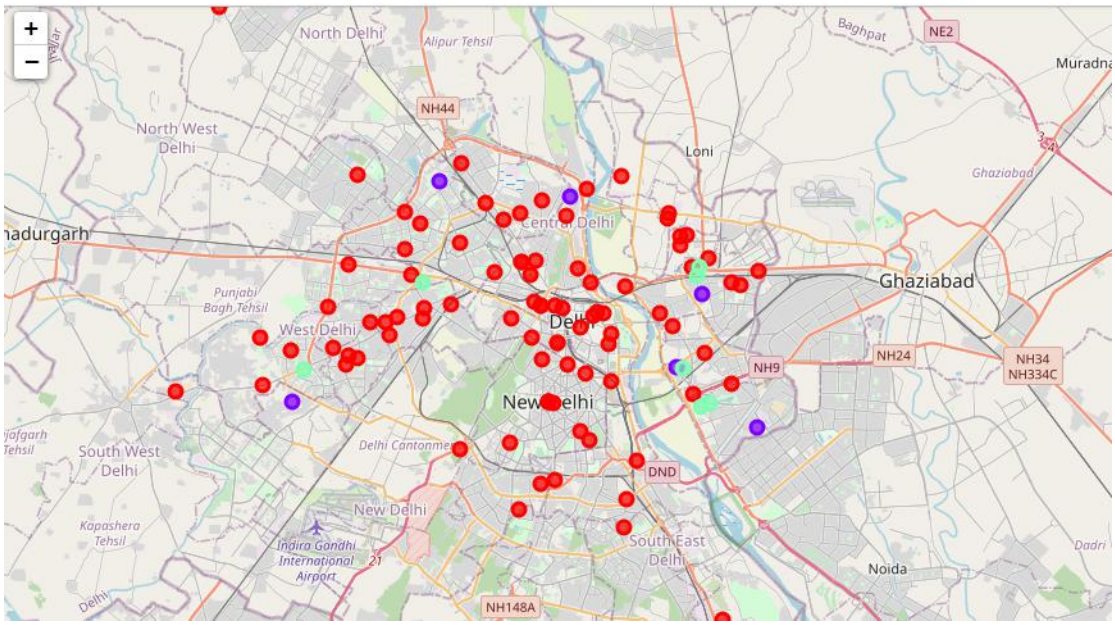
The results from the k-means clustering shows that we can categorize the neighborhoods into 3 clusters based on the frequency of occurrence of “Gym”:

· Cluster 0: Areas with no Gyms.

· Cluster 1: Areas with low number of Gyms

· Cluster 2: Areas with moderate number of Gyms.

The results of the clustering are visualized in the map below. The Red color is Cluster 0 , the blue color is Cluster 1 and the green color is cluster 2.



So, the best place to open a new gym will be in Cluster 0.

DISCUSSION SECTION

In , this project we only consider one factor i.e frequency of occurrence of gyms in a neighborhood , there are also other factors like population of people in neighborhood , rating of gyms in each locality , rate of land or rent for

opening a gym that could influence the decision of opening a new gym. However , this type of data is couldn't obtained easily.So, it is out of scope of this project.

Also, we don't include the Neighborhoods of Southern Delhi region as it will make the visualization of maps a little bit confusing.
Future research may can bypass these limitations by getting paid-information and obtain more and accurate 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 3 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. businessman and investors regarding the best locations to open a new gym. To answer the business question that was raised in the introduction session , the answer proposed by this project is: --> **The Areas in cluster 0 are the most preferred locations to open a new gym.**