The Battle of the Neighbourhoods

Determining locations for gym's in Delhi NCR, India

Author :- Gaurav

Introduction

Obesity brings with it a host of health risks like heart disease, stroke and high blood pressure. The survey of over 1,000 respondents in Delhi-NCR, both men and women in age group of 18-45, also found that poor self-image and lifestyle disorders was a common worry among the obese people. Obesity in India has reached epidemic proportions in the 21st century, with morbid obesity affecting 5% of the country's population. India is following a trend of other developing countries that are steadily becoming more obese.

In India, the prevalence of overweight and obesity has increased rapidly in recent decades. Given the association between overweight and obesity with many non-communicable diseases, forecasts of the future prevalence of overweight and obesity can help inform policy in a country where around one sixth of the world's population resides.

Business Problem

In Delhi, NCR, India where 45.5% of the male population and 49.8% of the female population are suffering from obesity, which is capital of India and place where major working population resides and yet they have few gyms around them to keep them healthy. A company *XYZ* wanted to start a gym chain in Delhi so we might suggest them some place for their gyms. Where they can at least work on their health and keep them fit because according to Centre of Disease Control(CDC), 45-60 minutes are enough for one them to keep them healthy. As these gyms will be available to them in their neighbourhood so they can easily reach out their.

The target market for these gyms would be the entire community in which it resides, as the gym would provide ample variety in activities that all age groups from adolescent to senior citizen would be able to partake and benefit from the physical activity, promoting healthier lifestyles and aiding in the prevention of obesity in Delhi.

Methodology

Jupyter Notebook, will be used to utilize Python 3 programming code, which will be the primary means of analysing, visualizing, and reviewing the data. The Pandas Library, which offers data structures and operations for manipulating data, will be used to import the datasets from CSV downloaded from the internet with data containing to Delhi neighbourhood.

GeoPy, will be used with Nomatim API package, which will be used to determine the coordinates of Delhi, India. These coordinates will then be used to generate a map with help of the Folium package in Python. Folium is a library that allows for a visualization of interactive spatial data. Once this map is created, markers will be added to the map to indicate the exact location of all the zip codes within Delhi neighbourhood.

Next the Foursquare API will be accessed to search the venues within each neighbourhood. This list will be converted into a data frame, and one-hot encoding will be used for determining the number of occurrences of each venue category within each neighbourhood. The mean frequency of each category will then be taken and the updated data frame will group the rows of the data frame by neighbourhood and the resulting mean frequency of each venue category.

To manipulate, visualize, and analyse the above mentioned datasets, the above mentioned resources will be used.

Analysis

1. Data Preprocessing

|	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
4	North West Delhi	Begum Pur	NaN	NaN

2. Data Visualisation
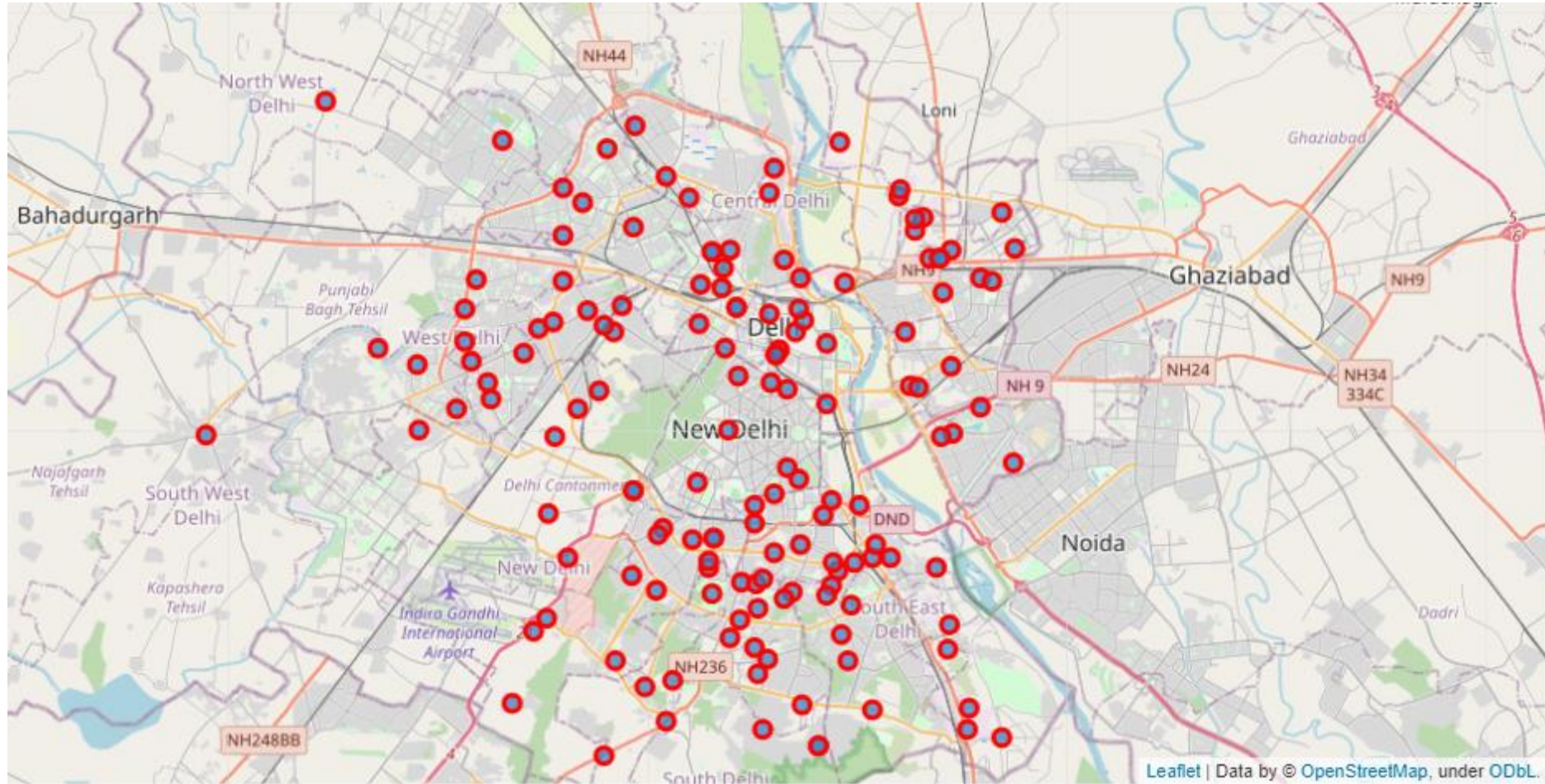
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adarsh Nagar	28.614192	77.071541	Bikanerwala	28.613391	77.076084	Indian Restaurant
1	Ashok Vihar	28.699453	77.184826	Nat Khat Caterers	28.699630	77.187832	Indian Restaurant
2	Ashok Vihar	28.699453	77.184826	ELIXIR FERTILITY CENTRE	28.700034	77.188093	Health & Beauty Service
3	Ashok Vihar	28.699453	77.184826	Bakers Stop	28.700495	77.188716	Bakery
4	Ashok Vihar	28.699453	77.184826	Invitation Banquet	28.696018	77.185953	Diner

	Neighborhood	Restaurant	Cluster Labels	Borough	latitude	longitude
0	Adarsh Nagar	0.0	0	North West Delhi	28.614192	77.071541
1	Alaknanda	0.0	0	South Delhi	28.529336	77.251632
2	Anand Vihar	0.0	0	East Delhi	28.625691	77.101941
3	Ashok Vihar	0.0	0	North West Delhi	28.699453	77.184826
4	Azadpur	0.0	0	North West Delhi	28.707657	77.175547

4. Output Cluster
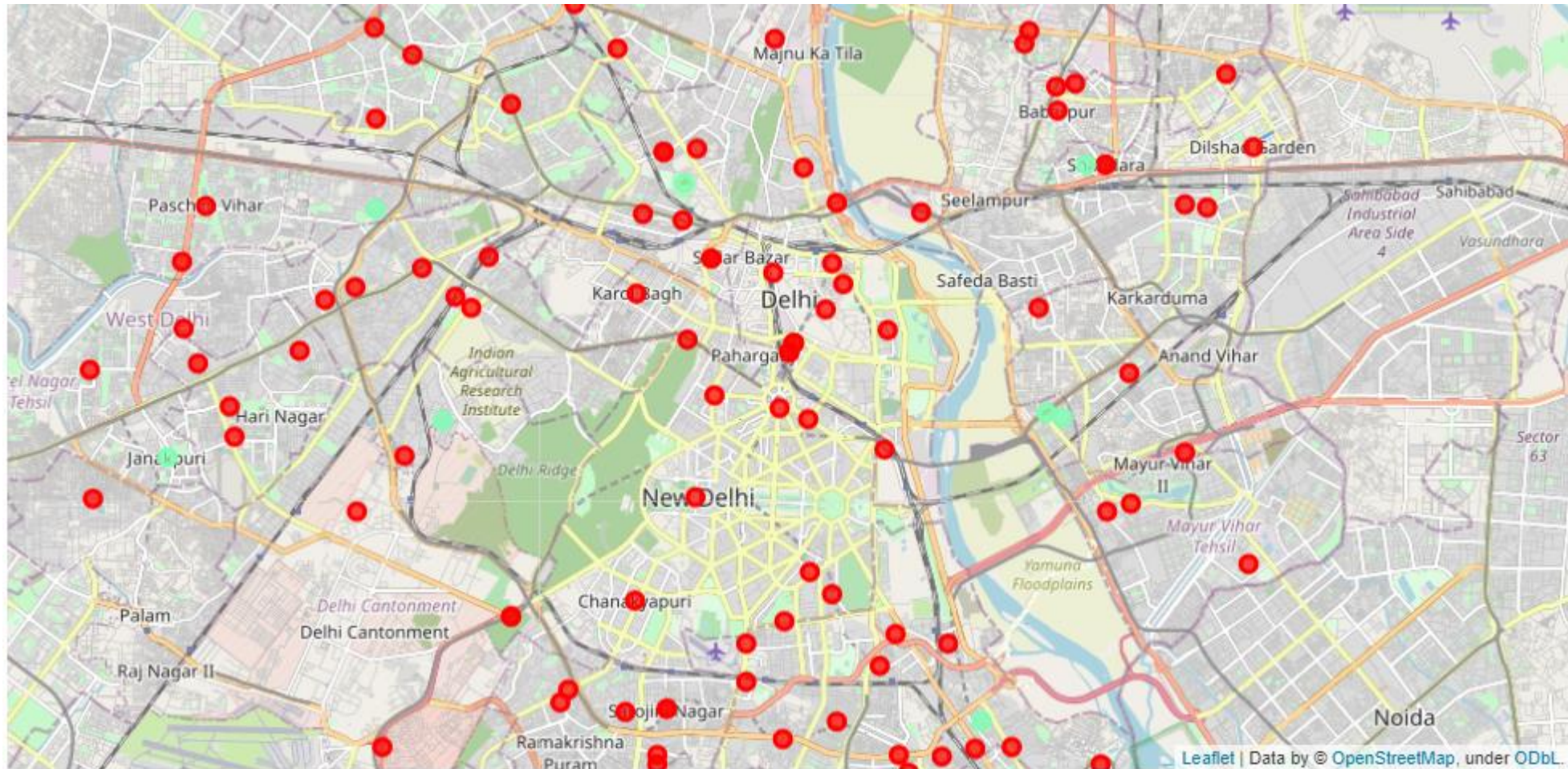
Result And Discussion

Of the 3 clusters that were generated utilizing k-means clustering with the Scikit learn library, the some neighbourhood were clustered more heavily in one of the three cluster which Cluster 0. The remaining clusters had much lower neighbourhood attributed to them. This was due to the very low level of mean frequency of occurrence of gyms in the neighbourhood contained within Cluster 0, and a relatively higher mean frequency of occurrence in Clusters 1 and 2.

Given this, Cluster 0 would be the optimal cluster to isolate as potential location for the new gyms. With the low overall average of the mean frequency of occurrence of a gym occurring in Cluster 0, this would be the target cluster proposing a new location for the gyms in Delhi. Based solely on the information presented, several neighbourhood that would potentially provide optimum return on investment by generating higher membership rates could be considered, providing a greater chance of exposure to the general public.

Additional research could be performed to incorporate average household income, demographic breakdown, and commercial real estate rental and purchase rates within each zip code to make a better assessment of the ideal location to open a gym by further identifying the target market.

Conclusion

This project used Python in Jupyter Notebook to access and manipulate datasets pertaining to Delhi adult obesity rates, neighbourhood and associated census and geographic data. This data was manipulated and analyzed using various Python libraries, such as Pandas, Matplotlib, Scikit Learn, and GeoPy.

Additionally, the Foursquare API was utilized to identify the frequency of occurrence of gyms within each neighbourhood in Delhi, India. Using k-means clustering, 3 clusters were created from the 163 neighbourhood within Delhi. Using this information, it was identified that based on frequency of gym occurrence alone, the ideal location for a gym would be in one of the attributed to Cluster 0. From there, it was suggested that one of the three highest populated neighbourhood within Cluster 0 be further explored as the potential site of a new gym.