

APPLIED DATA SCIENCE
CAPSTONE PROJECT

Opening New Gyms in Delhi, India

By: Prashant Thakur

(Aug-2020)



Introduction

As everyone is aware that Indian food is prepared using large amount of spices and oil. According to a recent survey by National Restaurant Association of India (NRAI) shows that people in Delhi on an average dine out six times in a month, which is the most in India compared to other cities. Across India, 41percent people love North Indian food, followed by 27 per cent Chinese, 23 per cent South Indian, 22 per cent Mughlai and 16 per cent Italian. It was also found that almost two-thirds (64%) of Indians say that they don't exercise. They further revealed that Indians tend to choose very basic forms of exercise - as many as two-thirds (67%) of Indians who exercise typically do brisk walking. This is followed by 26% of Indians who do yoga/pilates/CrossFit, 11% who do cardio and team sports (e.g. football, cricket), respectively, and 10% who do bodyweight exercises (e.g. push-ups, pull-ups, squats). They reason for choosing very basic forms of exercise is the fact that exercises involving gyms and fitness classes tend to be more expensive.

Business Problem

The objective of this project is to filter the locations in Delhi, India to open low budget gyms. Using data science methodology and machine learning techniques like clustering to answer the business problem: what are the recommended places in Delhi, India to open a low budget gym?

Data

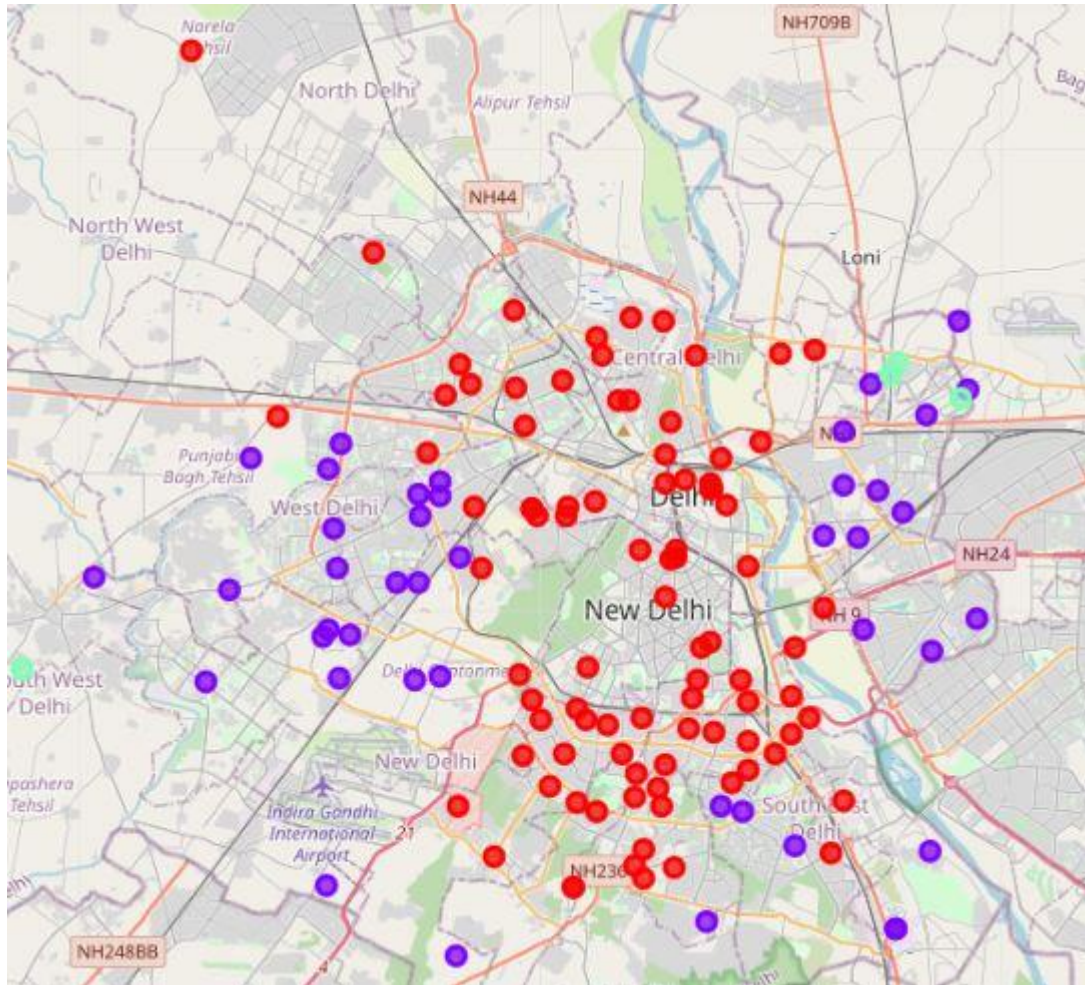
Information about the neighbourhoods of Delhi was scrapped from Wikipedia.com and Foursquare API is used to get the information about the location of Gyms in different parts of Delhi. Also, information from a survey of NRAI and newspaper report of Times of India about a survey by market intelligence agency, Mintel is used to conclude the business problem. Features like 'Latitude', 'Longitude' of Neighbourhoods and 'VenueName', 'VenueLatitude', 'VenueLongitude', 'VenueCategory' will be extracted using Foursquare API. After filtering the 'gyms' using 'VenueCategory', 'K-means' clustering algorithm will be used to segment neighbourhoods in 3 clusters (High, Moderate and Low) density areas on the basis of 'gyms'. This project can also be used by retailers of sports and fitness equipments, by advertising in high density areas of 'gyms' they will be able to advertise to large target audience.

Example: In high density area of gyms it will hard to acquire customers whereas in Low density areas it will be easier to acquire customers due to less or no competitors.

Methodology

Using the Library 'BeautifulSoup' details of neighbourhoods in delhi is scrapped from Wikipedia and stored in dataframe. Then 'Latitude' and 'Longitude' of each neighbourhood is extracted and visualized on a map using 'Folium' Library. Then Foursquare API is used explore the venues in all the neighbourhood and extract the relevant data. Gyms in each neighbourhood are filtered out from the venues using 'VenueCategory' feature. Finally, 'K-means' clustering is used to partition the neighbourhoods into high, moderate and low density of gyms.

Results



In the above figure, neighbourhoods with 'low' density of gyms are depicted by red spots. In those areas setting up gyms will be recommended. Whereas neighbourhoods with 'moderate' and 'high' density of gyms are depicted by purple & green spots. In these areas, retailers of sports and fitness equipments will be recommended to advertise to grab attention of larger target audience.

Discussion

On the basis of our model's result and surveys by National Restaurant Association of India (NRAI) and Mintel, it can be observed that large amount of Neighbourhoods in Delhi have 'low' density of

gyms. Since 67% of people of India are not exercising in which many consider 'price' as hindrance. Also, poor eating habits of people of Delhi as compared to other Indian cities makes it the best place to open 'budget gyms' to cater masses.

Conclusion

cluster 0

```
In [47]: dl_merged.loc[dl_merged['Cluster Labels'] == 0]
```

```
Out[47]:
```

	Neighborhood	Gym	Cluster Labels	Latitude	Longitude
68	Mehrauli	0.0	0	28.521130	77.180170
95	Patel Nagar	0.0	0	28.647830	77.164490
93	Pandav Nagar	0.0	0	28.614580	77.275740
91	Palika Bazaar	0.0	0	28.631560	77.219590
89	Old Delhi	0.0	0	28.654340	77.232580
88	Okhla	0.0	0	28.532470	77.278390
87	Nizamuddin West	0.0	0	28.590820	77.244340
86	Nizamuddin East	0.0	0	28.601240	77.264521
85	Nigambodh Ghat	0.0	0	28.664710	77.236330
84	New Moti Bagh	0.0	0	28.581010	77.181830
83	New Friends Colony	0.0	0	28.578120	77.269990
82	New Delhi	0.0	0	28.630950	77.217210

In this project, I have identified the 91 neighbourhoods in Delhi, in which setting up new budget gyms will be recommended to cater highly price sensitive Indian market as no gyms are available in 10 kilometres vicinity.

References

1. <https://timesofindia.indiatimes.com/life-style/health-fitness/health-news/64-per-cent-indians-dont-exercise-study/articleshow/70038656.cms>

2. <https://nrai.org/delhiites-dine-out-six-times-a-month-but-bangaloreans-spend-the-most/>
3. https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Delhi