

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.