COURSERA CAPSTONE PROJECT

IBM APPLIED DATA SCIENCE PROJECT

Opening a New Pizza Shop in Mumbai, India

By : Shantanu Sharma

Feb 2020

Introduction

The pizza market in India stands at Rs 3,500 crore according to a report by Technopak, offering a big room for investors. There has been a spurt in quick service restaurants in India. Initially, pizza was a difficult sector to enter as entry barriers were high and scaling up in this category was difficult. But now, there are several new brands that are at par with the already established ones and not all can sustain for a span of 10 to 20 years. So, more the number of franchisors, the better for the industry! Competition in the pizza segment is getting intense as franchisors are significantly ramping up their presence across tier-II and tier-III cities of India. There exists a tremendous growth opportunity for franchisors in these areas as educated middle class population is becoming more sophisticated with respect to standardised food consumption.

Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Mumbai, India to open a new Pizza shop. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Mumbai, India , if a property developer is looking to open a new Pizza Shop, where would you recommend that they open it?

Target Audience

 India has the largest population of youth (under 25 years) in the world and this sector will have more spending power in 2017 and years ahead. So, the growth prospects lie manifold for the franchisors. This project also aims to target big pizza chains such as Pizza Hut and Domino's, for opening a new joint in certain parts of Mumbai to increase their sales.

Data

To solve the problem, we will need the following data:

• List of neighbourhoods in Mumbai. This defines the scope of this project which is confined to the city of Mumbai, the financial capital city of the country of India in Asia.

• Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.

• Venue data, particularly data related to Pizza shops. We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai) contains a list of neighbourhoods in Mumbai, with a total of 39 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Pizza Place category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.

Methodology

Firstly, we need to get the list of neighbourhoods in the city of Mumbai, India. Fortunately, the list is available in the Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai). We will do web scraping using Python requests and beautifulsoup packages to extract the list of neighbourhoods data. However, this is just a list of names. We need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the wonderful Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the data into a pandas DataFrame and then visualize the neighbourhoods in a map using Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the city of Mumbai . Next, we will use Foursquare API to get the top 100 venues that are within a radius of 2000 meters. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each neighbourhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analysing the "Pizza place" data, we will filter the "Pizza place" as venue category for the neighbourhoods. 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. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighbourhoods into 3 clusters based on their frequency of occurrence for "Pizza place". The results will allow us to identify which neighbourhoods have higher concentration of Pizza places while which neighbourhoods have fewer number of Pizza places. Based on the occurrence of Pizza places in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new Pizza places.

Results

The results from the k-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for "Pizza place":

• Cluster 0: Neighbourhoods with moderate number of Pizza place

• Cluster 1: Neighbourhoods with low number to no existence of Pizza place

 • Cluster 2: Neighbourhoods with high concentration of Pizza places

 The results of the clustering are visualized in the map below with cluster 0 in red colour, cluster 1 in purple colour, and cluster 2 in mint green colour
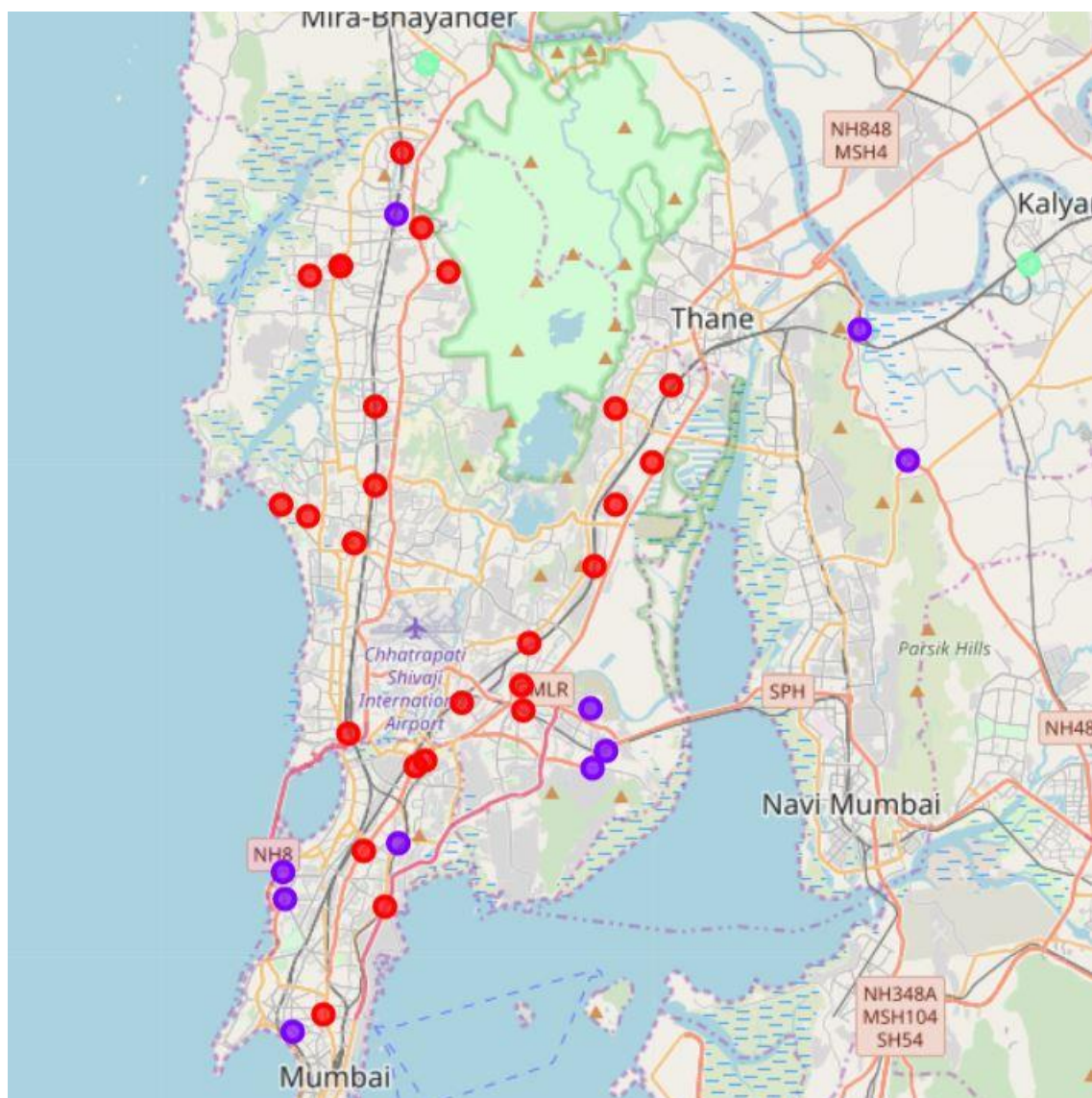
Discussion

As observations noted from the map in the Results section, most of the Pizza places are concentrated in the central area of Mumbai city, with the highest number in cluster 2 and moderate number in cluster 0. On the other hand, cluster 1 has very low number to no Pizza place in the neighbourhoods. This represents a great opportunity and high potential areas to open new Pizza places as there is very little to no competition from existing shops. Meanwhile, Pizza places in cluster 2 are likely suffering from intense competition due to oversupply and high concentration of Pizza places. From another perspective, the results also show that the oversupply of Pizza places mostly happened in the central area of the city, with the suburb area still have very few Pizza places. Therefore, this project recommends property developers to capitalize on these findings to open new Pizza places in neighbourhoods in cluster 1 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new Pizza places in neighbourhoods in cluster 0 with moderate competition. Lastly, property developers are advised to avoid neighbourhoods in cluster 2 which already have high concentration of Pizza places and suffering from intense competition.

Limitations and Suggestions for Future Research

In this project, we only consider one factor i.e. frequency of occurrence of Pizza Places, there are other factors such as population and income of residents that could influence the location decision of a new pizza shops . However, to the best knowledge of this researcher such data are not available to the neighbourhood level required by this project. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new pizza shops. In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Future research could make use of paid account to bypass these limitations and obtain more 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. property developers and investors regarding the best locations to open a new Pizza place. To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighbourhoods in cluster 1 are the most preferred locations to open a new Pizza shop. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new Pizza shop.

References

Category: Suburbs in Mumbai, India . Wikipedia. Retrieved from

https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai

Foursquare Developers Documentation. Foursquare. Retrieved from
https://developer.foursquare.com/docs

Indian Pizza market offers a mouth watering opportunity for investors(2017-02-15)

https://www.franchiseindia.com/content/Indian-Pizza-market-offers-a-mouth-watering-

opportunity-for-investors.9128