Capstone Project - The Battle of Neighborhoods

IBM Applied Data Science Capstone

Opening of a new Pizza Place near Mumbai's Tourist Places

Author: Kunal Tale
Date: 23rd April 2020

Introduction:

Mumbai city is the financial capital of India and is one of the most visited places in India by both domestic and international tourists for business and recreational activities. Mumbai city has various popular tourist places which are surrounded by popular restaurants, cafes, food court, pubs, etc. Tourist usually visit one of the nearby recreational places near the tourist place for refreshment and hangout at one of such destination to meet there near and dear ones. This Project is designed to analysis the potential Tourist places in Mumbai city where a Pizza Place can be established by a Pizza Place firm. The Analysis will be helpful for a firm which pan to establish its presence in Mumbai for its pizzeria chain and help the firm to gain higher footfall as these localities are near popular tourist attractions in the city and these places have some popular pizzerias around. This analysis will focus on identifying popular tourist destinations in Mumbai city where there already exists a presence of pizza culture in the neighbourhood place.

Business Plan

The objective of this capstone project is to analyse and identify tourist places in Mumbai which have an existing popular pizza cafe near their neighbourhood as one of the top 10 most popular attraction. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city Mumbai, India, if a Pizza place firm is looking to open a new Pizza Place near a popular tourist place, where would you recommend that they should open it?

Target Audience

This project is useful to Pizza Place firms, Restaurant & café owners who are looking for entering in Mumbai city with their Pizza Place unit or any existing Pizza Place firm in Mumbai city who are interested in expanding their presence and grow their business within the city.

Data

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

• List of tourist places in Mumbai and its geographical coordinates

• List of popular neighbourhood venues near the tourist places within 3 KM range

• Venue data, particularly data related to tourist places and any existing pizza places near it

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/List_of_tourist_attractions_in_Mumbai) contains a list of tourist places in Mumbai, India. 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 Places category around the tourist destination in Mumbai 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

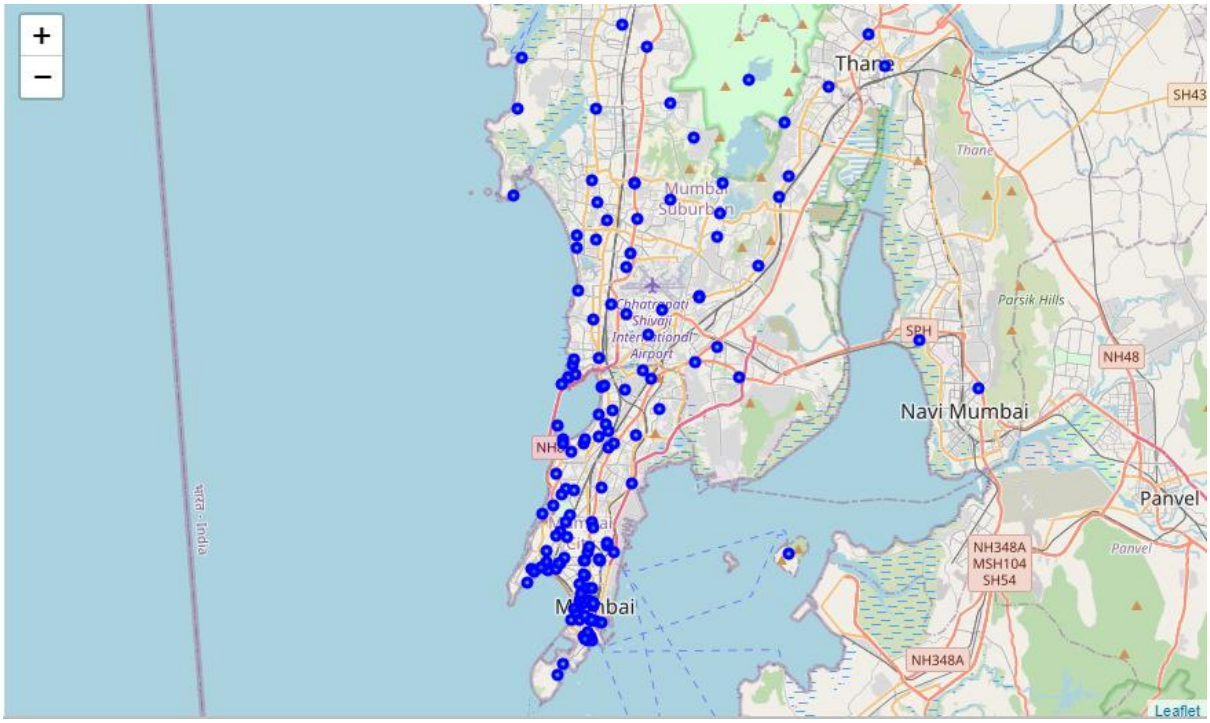
Firstly, we need to get the list of all the tourist places in the city of Mumbai. Fortunately, the list is available in the Wikipedia page List of tourist attractions in Mumbai. We will do web scraping using Python requests and BeautifulSoup package to extract the list of tourist places data. However, this is just a list of names of 162 tourist places in Mumbai. 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 tourist places 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, India.

Next, we will use Foursquare API to get the top 100 venues that are within a radius of 3000 meters of all the tourist places in Mumbai. 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 of the tourist places 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 Tourist places into 9 clusters based on their frequency of occurrence for various venues as part of top 10 most common places near the tourist destination. The results will allow us to identify which neighbourhoods of the tourist places 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 of Tourist destination, it will help us to answer the question as to which neighbourhoods are most suitable to open new Pizza Place.

Results

Map of Tourist places of Mumbai



The results from the k-means clustering show that we can categorize the neighbourhoods into 9 clusters as follows:

- **Cluster 1:** This cluster consist of 4 Tourist places *(Metro Junction Mall ; Growel's 101 ; Dadaji Kondadev Stadium)* having Indian Restaurant as the 1st most common venue destination and all of the tourist places in cluster 1 have Pizza Places a one of the top 10 most common venue near the tourist places. Hence this is a viable option to open a pizza place.
- **Cluster 2:** This cluster consis of 63 Tourist places *(Wadia Movietone ; Vasai Fort ; Tulshi Dam ; St. Thomas Cathedral ; St. Joseph's Church, Juhu ; Sion Hillock Fort ; Siddhivinayak Temple ; Shreebalajimandir ; Sardar Vallabhbhai Patel Indoor Stadium ; Sacred Heart Church, Santacruz ; Riwa Fort ; Ranjit Studios ; R-Mall ; Princess Street ; Portuguese Church ; Phoenix Marketcity ; Oberoi Trident ; New Empire Cinema ; Nehru Science Centre ; Nariman House ; Mount Mary Church, Bandra ; Modak Sagar ; Metro Big Cinemas ; Mazagon Fort ; Mani Bhavan ; Mahindra Hockey Stadium ; Mahatma Jyotiba Phule Mandai ; Mahalaxmi Racecourse ; Mahalakshmi Temple ; Lohar Chawl ; Liberty Cinema ; Knesset Eliyahoo ; Kamalistan Studios ; Kadeshwari Devi Temple ; Joseph Baptista Gardens ; Jama Masjid ; ISKCON temple ; INS Vikrant (R11) ; Horniman Circle Gardens ; Holy Cross Church, Kurla ; Haji Ali Dargah ; Gowalia Tank ; Gokul ; Ghodbunder Fort ; Gate of Mercy Synagogue ; Fort George, Bombay ; Fashion Street ; Dr. Bhau Daji Lad Museum ; Dava Bazaar ; Cross Maidan ; Cowasji Jehangir Hall ; Coronation Cinema ; Cooperage Ground ; Colaba Woods ; Church of Our Lady of Mount Carmel, Bandra ; Chhatrapati Shivaji Maharaj Vastu Sangrahalaya ; Bombay Castle ; Belapur Fort ; Bassein Fort ; Babulnath ; B.P.T. Ground ; National Gallery of Modern Art)*. All have Indian Restraunt followed by Café/Ice cream shop/ pub as the 1st and 2nd most common venues. All these places have mean of 0.02 for Pizza Places near the Tourist place and only one tourist place is there having pizza place as one of the top 10 most common venues.

As the number of Tourist places in this cluster with pizza places as one of the top 10 common venues is very less, we do not recommend this cluster to open a new pizza place.

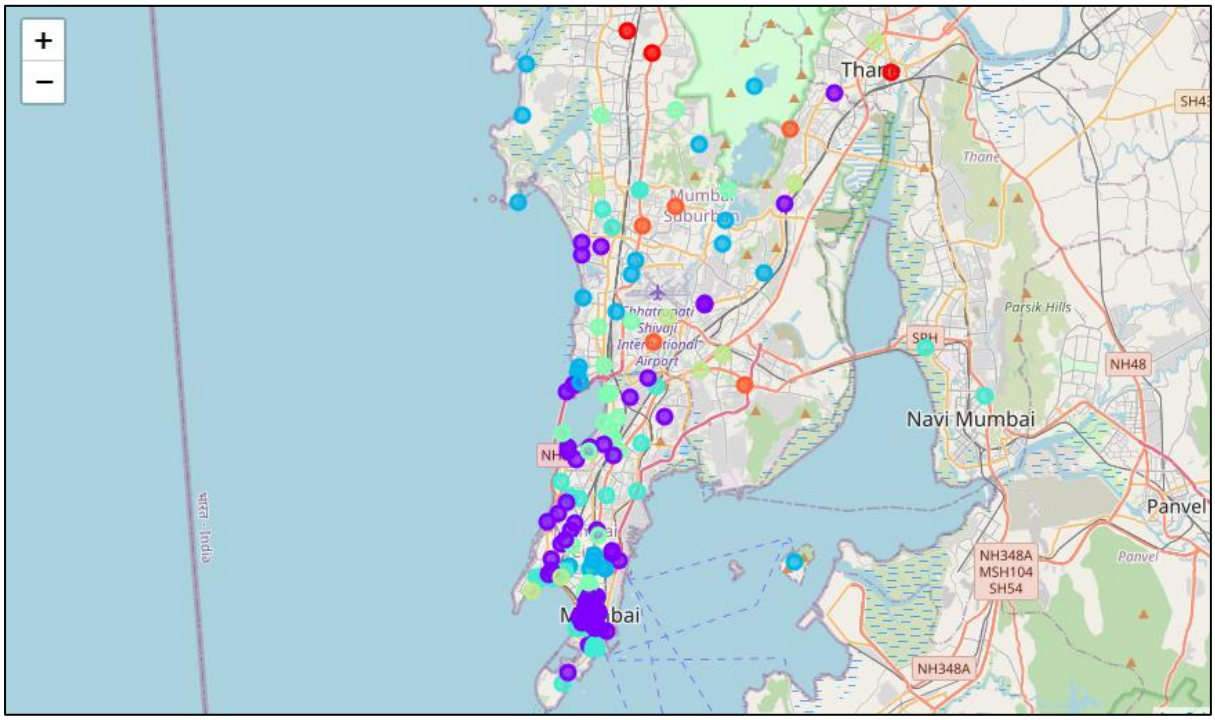
- **Cluster 3:** This cluster consist of only Watson's Hotel as the tourist place and it is having Pizza Place as on of the top 10 most common venues near it. Hence we would recommend this as one of the viable cluster to open a pizza place.
- **Cluster 4:** This cluster consist of 28 tourist places *(Tulsi Lake ; St. John the Baptist Church ; St. Andrew's Church ; Raudat Tahera ; Rajkamal Kalamandir ; R City Mall ; Powai Lake ; Oval Maidan ; Mehboob Studio ; MarvÃ© Beach ; Mandapeshwar Caves ; Makhdoom Ali Mahimi ; Magen David Synagogue (Byculla) ; Madh Fort ; Kalamb Beach ; Juhu Beach ; Jogger's Park ; Irani cafÃ© ; Elephanta Caves ; Dongri Fort ; Dadasaheb Phalke Chitranagari ; Castella de Aguada ; Bombay Talkies ; Annapurna Studios ; Aksa Beach ; Adlabs Imagica ; Capitol Cinema)* and none of them are having pizaa place near them (mean pizza place= 0). Hene we would not recommend cluster 4 likely destination to open a pizza place.
- **Cluster 5:** This cluster consist of 28 Tourist places *(The Table ; Sewri Fort ; Regal Cinema ; Raghuleela Mall, Vashi ; Punjabi Chandu Halwai Karachiwala ; Leopold Cafe ; Lamington Road ; Lalbaugcha Raja ; Kamala Nehru Park ; Jogeshwari Caves ; Jogeshwari Caves ; High Street Phoenix ; Hanging Gardens of Mumbai ; Girgaum Chowpatty ; Gilbert Hill ; Four Seasons Hotel Mumbai ; Filmistan ; Eros Cinema ; DY Patil Stadium ; Colaba Causeway ; Church of Our Lady of Good Counsel & Shrine of St. Anthony, Sion ; Church of Our Lady of Dolours, Wadala ; Cathedral of the Holy Name ; Cafe Mondegar ; Brabourne Stadium ; Andheri Sports Complex ; Afghan Church)* with oaa of them having 0.02+ mean for a pizza place near the tourist destinationa and 12 of the 28 tourist places have pizza place as one of the top 10 most common venue near them. Hence we would recommend this as one of the viable cluster to open a pizza place.
- **Cluster 6:** This cluster consist of 21 tourist places (*Wagheshwari Temple ; Vihar Lake ; Sterling Cineplex ; St. Michael's Church ; Shri Swaminarayan Mandir ; Shivaji Park ; Plaza cinema ; Mumba Devi Temple ; Maratha Mandir ; Mahim Fort ; Linking Road ; Kohinoor Square ; Jivdani Mata ; Inorbit Mall ; Grand Hyatt Mumbai ; Gloria Church ; Church of Our Lady of Health, Cavel ; Chor Bazaar ; Bandra Talao ; Antarang â€“ Sex Health Information Art Gallery*) all having 0.01 around mean for pizza place and none of the tourist place amont this cluster has pizza place as one of the topp 10 most common venue near them. Hene we would not recommend cluster 6 likely destination to open a pizza place.
- **Cluster 7:** This cluster consist of 11 tiurist places (*Viviana Mall ; Shaneshwar Sansthan ; Royal Opera House ; Ramdev Film Ciity ; Our Lady of Egypt Church ; Neptune Magnet Mall ; Middle Income Group Club Ground ; Kanheri Caves ; EsselWorld ; Crossroads Mall*) with 9 of them having pizza place as one of the top 10 most common venues near them. Hence we would recommend this as one of the viable cluster to open a pizza place.
- **Cluster 8:** This cluster consist of only 'Our Lady of Immaculate Conception Church' as a tourist place and it has pizza placeas one of the top 10 most common venue Hence we would recommend this as one of the viable cluster to open a pizza place.
- **Cluster 9:** This cluster consist of 5 tourist places *(R. K. Studio ; Mahakali Caves ; Jijamata Udyaan ; Bandra Kurla Complex Ground*) with all of them having pizza place as one of the top 10 most common venue. Hence we would recommend this as one of the viable cluster to open a pizza place.

Cluster Consolidated Data:

	Cluster	Total Tourist Places	Tourist Places	Presence of Pizza Place	%Presence of Pizza Place	Viable Option
0	Cluster 1	4	Metro Junction Mall ; Growel's 101 ; Dadaji Ko...	4	100.000000	Yes
1	Cluster 2	63	Wadia Movietone ; Vasai Fort ; Tulshi Dam ; St...	1	1.587302	No
2	Cluster 3	1		1	100.000000	Yes
3	Cluster 4	28	Tulsi Lake ; St. John the Baptist Church ; St....	0	0.000000	No
4	Cluster 5	28	The Table ; Sewri Fort ; Regal Cinema ; Raghul...	12	42.857143	Yes
5	Cluster 6	21	Wagheshwari Temple ; Vihar Lake ; Sterling Cin...	0	0.000000	No
6	Cluster 7	11	Viviana Mall ; Shaneshwar Sansthan ; Royal Ope...	9	81.818182	Yes
7	Cluster 8	1		1	100.000000	Yes
8	Cluster 9	5	R. K. Studio ; Mahakali Caves ; Jijamata Udyaa...	5	100.000000	Yes

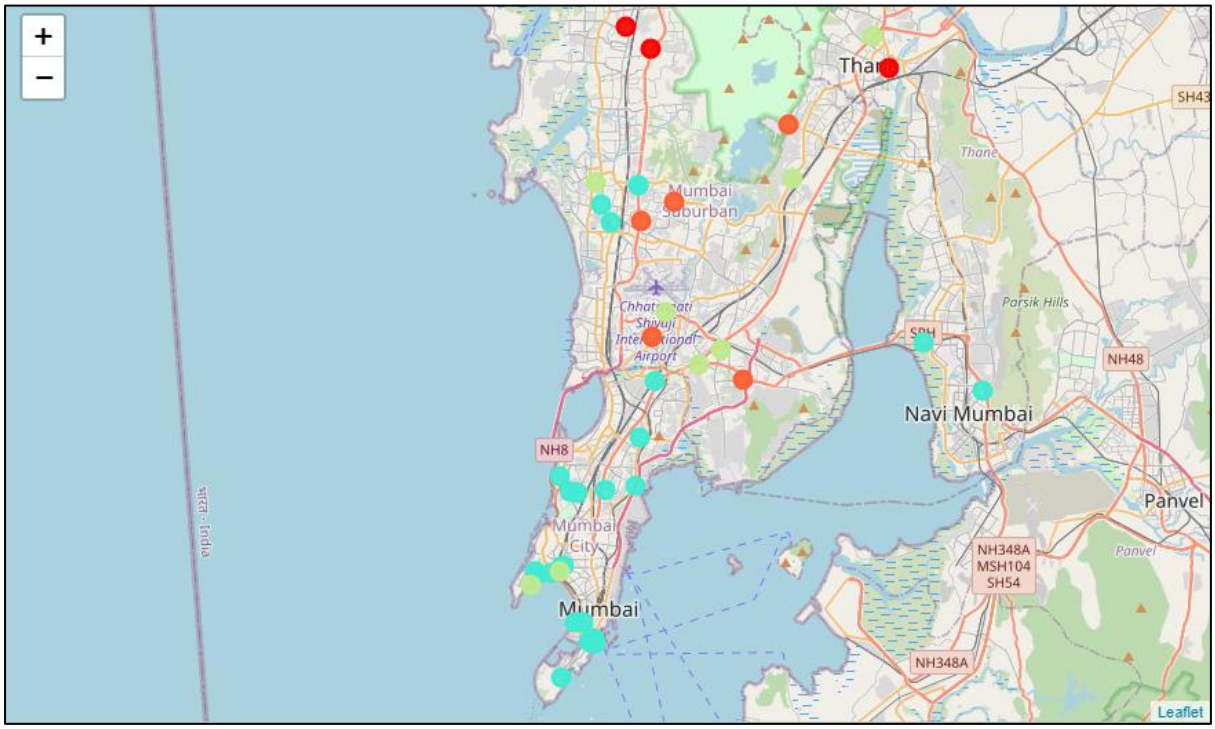
This Table give us a consolidate vie of all the clusters and recommends the viable cluster to open a new pizza place with scores for business analyst. Hence this data can be used by business to analyse their model and plans for setting-up a new pizza place in the cluster.

Clustered Map of all tourist places:



This map shows the location of all Tourist places in Mumbai with color coding mapping to its respective cluster

Recommended Cluster Map for new Pizza Place



The map above shows the tourist place of recommended clusters which are viable to open a new Pizza place for a business.

Discussion

We got a glimpse of the Pizza Places in Mumbai around its Tourist Destinations and were able to find out some interesting insights which might be useful to a Pizza Place firm looking for expansion in Mumbai City as well as people with business interests. Let's summarize our findings:

- The Tourist destination in Mumbai are classified under 9 Clusters
- 6 out of the 9 Clusters have Pizza Places as significant contribution as one of its top 10 most common venues near their Tourist Places
- Cluster 4 (light blue) has the highest number of neighbourhood popularity in Pizza Places followed by Cluster 8 (orange)
- Cluster 2 (purple) is located with long distant Tourist places and all of them have Pizza Places as one of the most popular venue of attraction near the tourist place
- Cluster 9 (red) and Cluster 7 (light green) had low number of Tourist places but they still have a great popularity for Pizza Places near its tourist places

Lastly, Pizza Places firms are advised to focus on neighbourhoods in Cluster 4 and Cluster 8 which already have high concentration of Tourist places hence more footfall can be anticipated thought it will have an intense competition due to its popularity for Pizza. The interested Pizza Place firm is advised to maintain high quality standards and wider menu options to attract more customers and grow its business.

Limitations and Suggestions for Future Research

In this project, we only consider one factor i.e. frequency of occurrence of Pizza place, there are other factors such as population and income of residents and tourist visiting the locality to have food and their taste preferences that could influence the location decision of a new pizza place. 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 place. 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 9 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders regarding the best locations to open a new Pizza Place near Mumbai's tourist destination. To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighbourhoods near tourist destination of Cluster 1, Cluster 3, Cluster 5, Cluster 7, Cluster 8 and Cluster 9 are the most preferred locations to open a new pizza place. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding analysis their plan to expand their pizza place business in the city of Mumbai.

Data Reference

- Category: List of Tourist places in Mumbai
  https://en.wikipedia.org/wiki/List_of_tourist_attractions_in_Mumbai
- Neighbourhood venue data source:
  https://developer.foursquare.com