

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.