

Coursera Capstone

IBM Applied Data Science Capstone

Opening a new Bakery in Mumbai, India

By: Ameya Thosar

July 2020



Introduction

Bakeries are an important part of our life as on occasion of Birthdays, Anniversary, a normal day etc we often go to bakery for buying Cakes, cookies, desserts. Having a Bakery in nearby location is good. In Mumbai there are many new bakeries but when you want to start your own you need to look for a location where in neighborhood there are minimum number of bakeries. Each bakery typically offers breads (bagels, buns, rolls, biscuits and loaf breads), cookies, desserts (cakes, cheesecakes and pies), muffins, pizza, snack cakes, sweet goods (doughnuts, Danish, sweet rolls, cinnamon rolls and coffee cake) and tortillas. So opening a new bakery is a good decision as people love to visit them.

Business Problem

The objective of this capstone project is to analyse and select the best location in Mumbai, India to open a new Bakery. Using Data Science Methodology and Machine Learning techniques like clustering, this project will provide a solution to the problem. In Mumbai, India if anyone wants to open a new Bakery we can suggest him where he can open it. Because when you are new in business its better that you have less competitors so that your business can grow at faster rate.

Target Audience

In this project we clearly aim at the person who want to open a new Bakery. Using this we will help him to choose the best location.

Data

To solve this problem we need following data:

- List of Neighborhood's in Mumbai.
- Latitude and Longitudes of those neighborhood's. This will be required to get plot the map and get venue data.
- Venue data particularly related to Bakery. We will use this data for clustering.

Source of data and method to extract it:

The Wikipedia page that contains a list of neighborhoods in Mumbai, (https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai) with a total of 40 neighborhood's. We will use web scraping techniques to extract the data from the Wikipedia page. By this we will only get the names of the neighborhood's but with tat we also need their coordinates. For that we will use Python page, with the help of Python requests and beautiful soup Geocoder package which will give us the latitude and longitude coordinates of the neighborhood.

After that, we will use Foursquare API to get the venue data for those neighborhood's. 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 Bakery category in order to help us to solve the business problem put forward. This capstone project 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).