Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Cafe in Mumbai, India

By: Nishchay Nagpal

July 2020

Introduction

When I travel, wherever I travel, more than museums, monuments, temples and, even restaurants, I seek out a café.

The reasons are many. Nowhere quite represents a place like its cafés. What would Vienna, Rome and Paris be without theirs? It's not just the history or the opulence. It's the window on the world. You see ordinary people – locals as well as visitors – coming and going, taking time out between work and home, meeting friends or colleagues, joining families. If they're not right beside you in the café, you watch them on the street or from your shaded terrace.

I, myself am a huge caffeine fanatic which led me to a question of where can one open a Café shop in Mumbai, the iconic city of India.

Business Problem

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

Target Audience

The target audience for this project would be twofold. Firstly, caffeine enthusiasts visiting Mumbai to get to know which neighborhoods have higher number of Cafes. Secondly, a company or a person looking to open a café shop in Mumbai can use the information provided here.

Data

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

- List of neighborhoods in Mumbai: This defines the scope of this project which is confined to the city of Mumbai, the iconic city of India
- Latitude and longitude coordinates of those neighborhoods: This is required in order to plot the map and to gather the venue data from foursquare for the neighborhoods.
- Venue data, particularly data related to café shops. We will use this data to perform clustering on the neighborhoods

Sources of data and methods to extract them:

The Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai) contains a list of neighborhoods in Mumbai, with a total of 41 neighborhoods. We will be using web scraping techniques utilizing the python packages - requests and beautifulsoup, to extract the data from the Wikipedia page. Then we will be getting the geographical coordinates of the

neighborhoods using the Python Geocoder package. After that, we will be using the Foursquare API to get the venue data for all neighborhoods. Foursquare has one of the largest databases of 100+ million places and is used by over a large number of developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Cafe 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.