Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Pizza Shop in Chennai, India

By: Karthick Arulraj

May 2020

INTRODUCTION

For many Food lovers, visiting Pizza shops is a great way to relax and enjoy the taste of pizza. Pizza is one of the famous Food all over the world especially among kids. The reason for pizza spread is that you can add anything to it and eat it anytime of the day. It was originally dough with topping of any ingredients as meat or vegetables that baked in the oven. People can have different variety of pizza in whatever price range they need. For Entrepreneur , the central location and the large crowd place provides a great opportunity to Run a pizza shop. As a result, there are many Pizza shops in the city of Chennai and many more are being built. Opening Pizza shop allows Entrepreneur to earn consistent money. Of course, as with any business decision, opening a new Pizza shop requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the Pizza shop is one of the most important decisions that will determine whether the Pizza shop will be a success or a failure.

Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Chennai , 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 Chennai, India, if a Entrepreneur is looking to open a new Pizza shop, where would you recommend that they open it?

Target Audience

The entrepreneur who wants to find the location to open a New Pizza shop.

Data

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

• List of Neighbourhoods in Chennai, India. This defines the scope of this project which is confined to the city of Chennai, India.

• 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/List_of_neighbourhoods_of_Chennai) contains a list of Neighbourhoods in Chennai. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and Beautiful-soup 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 shop 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.