IBM Applied Data Science Capstone Project

Analysis of Restaurants near Hospitals in Bangalore, India

Maria Franklin Judia

August 2020

Introduction

Hospitals are one of the most essential services that is required for the proper functioning of any community. There are people travelling even from different countries in order to access specialized medical services provided by the respective Hospitals. In such a scenario, extended periods of stay are often envisioned and good restaurants are sought out near the Hospitals for the food requirements of family members and visitors. While nutritious, appealing food in hospitals may not have yet evolved to the point that all stakeholders would like, advances are being made. Concerns persist with respect to many issues including insufficient budgets and human resources; local and sustainable food procurement challenges; ensuring food safety and sustainability; balancing nutrition and taste; plate waste; and barriers to patient eating.

Business Problem

The objective of this capstone project is to analyse Hospitals in the city of Bangalore, India to get a statistics of the Restaurants around it. 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 Bangalore, if a patient's family or bystander is looking for food venues, do they have enough options and if not, for investing in a new restaurant, where can we recommend it be opened?

Target audience for this project

The target audience of this project is twofold: First, the family and friends of patients visiting, who would get a clear idea of the restaurants around them and second, to property developers and investors looking to open or invest in new restaurants in Bangalore. We can analyse the top venues around the Hospitals and check the significance of each place relative to the Hospital. Comfortable and spacious dining areas with a choice of healthy foods would be of good demand near Multi-Speciality Hospitals. These places will give ample options for people visiting, if they prefer not to eat Hospital food for the various reasons discussed above.

Data

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

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

• Latitude and longitude coordinates of those Hospitals. This is required in order to plot the map and also to get the venue data.

• Venue data, particularly data related to Restaurants. We will use this data to perform clustering on the Hospitals.

Sources of data and methods to extract them

The following Wikipedia page: (https://en.wikipedia.org/wiki/Category:Hospitals_in_Bangalore) contains a list of Hospitals in Bangalore, with a total of 27 Hospitals. 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 Hospitals using Python Geocoder package which will give us the latitude and longitude coordinates of the Hospitals. After that, we will use Foursquare API to get the venue data for those Hospitals. 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 Restaurant 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.