Capstone Project – IBM Coursera

**Clustering Neighborhoods of Bangalore, IN based on Venue Information**

Jayadev K A – 2019

1. **Introduction**

India is a land of flavors, since independence the country has undergone a drastic transformation both in terms of Economic development and the very fabric of Indian lifestyle. With tremendous influence from the west, arising due to the rapid influx of technological jobs and the deep rooted family values the evolution has been remarkable. As the younger generation gets exposed to new trends, concepts such as Pubs, Café, Fast food, Mall’s etc. once believed to be only accessible on other countries are quickly becoming a reality and constitute a majority of the business in several metro cities. Bangalore, being the IT capital of India has numerous multinational companies employing thousands. This ultimately has led in the rise of Pubs, Café, Fast food, Mall’s etc. As these venue’s become more and more common there is an issue of selecting the right locality to maximize the sales. Most often than not Entrepreneurs are in a dilemma about the location to setup their new ventures either a Restaurant, Fast food or shop etc.

This study tries to collect information on various venues located in the city of Bangalore and assist New Entrepreneurs in deciding on the location for their business ideas. Once, all the information has been collected based on the Foursquare location data for the various geo locations in Bangalore, the study classifies different locations as major Food, sports, shopping hub etc. All the venue categories are further classified into major buckets like Café, Fast Food, Entertainment\Recreation\Sports, Restaurant, Sweets/Bakery, Shopping, Pharmacy, Transport, and Pub’s for better clarity. The final result will help guide new Business owners in selecting the neighborhood respectively based on the Type of business i.e. Restaurants, Fast food etc.

1. **Data Description**

The study derives information from secondary data that was scrapped from sites mentioned below:

1. **Data Source:**
   1. Bangalore Location Data : <https://finkode.com/ka/bangalore.html>
   2. Bangalore Venue Data: <https://foursquare.com/>
2. **Data Pre-Processing:**

Locations data was derived based on the postal information from site “a”. Once the location information was scrapped using the “BeautifulSoup” library in Python, the “geopy” library was used to retrieve the latitude and longitudes of every location. As postal data comprises of location having the same geo-locations but different PIN codes, the data was cleaned to remove any duplicate geo-locations. It was also noticed that couple of stray geo-locations were populated, therefore using the quartile outlier method using the latitude metric these locations were successfully eliminated from the final list of neighborhoods. The final list contained 90 neighborhood data for which venue information would be collected.

Using the Foursquare location data, with radius set to 500 meters and limit of venues as 100 Venue details for various locations were collected. The resulting dataset spanning 900 plus rows as shown in the sample table (Table 2.1).



Table 2.1 – Sample Data Set

The Foursquare resulted in 170 plus different venue categories, since this study tries to cluster the neighborhoods into shopping or Restaurant hubs, Top categories called as “Major Categories” were creates as follows (Café, Fast Food, Entertainment\Recreation\Sports, Restaurant, Sweets/Bakery, Shopping, Pharmacy, Transport, and Pub’s) and mapped with their respective venue information from Foursquare.