Coursera Capstone

IBM Applied Data Science Capstone

Finding a Neighborhood to move-in in a new City

Bangalore, India

By: Prashanth Ravi Shankar

June 2020

Introduction

A migrant worker is a person who either migrates within their home country or outside it to pursue work. Migrant workers usually do not have the intention to stay permanently in the country or region in which they work. [Wikipedia] Millions of skilled and un-skilled workers migrated to another city every year. When a skilled migrant worker moves to another city in a country. He/She finds it difficult to find a safe and a neighborhood with all basic and common amenities in them. This is a hypothetical clustering analysis of such migrant worker who is planning to move to a new city and need to find a neighborhood to move-in and eventually settle down there.

Business Problem

The Objective of this capstone project is to select the optimal location close by to office and at the same time finding a neighborhood that has all the common amenities like parks, hotels, etc. With the help of data science methodology and machine leaning technique like clustering analysis, this projects aims to solve a problem that is faced by every migrant worker faces when He/She moves to a new city. The major question is, which is the best neighborhood to stay?

Data

The data for this project has been gathered from multiple sources and processed for clustering.

The Bangalore neighborhood data is not readily available, the data is extracted from Wikipedia page using BeautifulSoup library for Python and scraped the required data from the webpage. This data does not have latitude and longitude coordinates with it, so Google's Geocoding API is used to gather latitude and longitude information.

After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest database of 105MM places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data and we use the available information for each neighborhood for this clustering.