Coursera Capstone Project

Café Recommendation

Jaynik Gaglani

12 - 02 - 2020

1. Introduction:

1.1 Background:

Opening a new business has to take many factors into consideration. Some of the factors include market, number of competitors, property rent, etc. A place with less competition may result into high profit since the business would be a monopoly. It is very important to get the data about such places for the business to flourish. Hence, it is important to build a system through which we can find the perfect place to open our business.

1.2 Problem Definition:

Creating a system for finding how many businesses of same category are present in an area and doing analysis upon that to acquire the perfect location for starting a business. The implementation is done keeping only cafes in mind but it can be extended to other types of business too.

1.3 Scope:

Current scope of the project finds the density of cafes in the city of Mumbai. It shows the neighborhood with the density of cafes present. The project can be implemented for any type of business by just changing one parameter. Café parameter was selected since the data availability was high.

2. Data Acquisition and Cleaning:

2.1 Data Source:

The data was collected from Wikipedia- https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai.
It consists of list of neighborhoods in Mumbai. The data was scraped using BeautifulSoup library of python. Geocoder library of python was used to obtain the coordinates of each neighborhood. The dataframe consisted of Neighborhood, Latitude and Longitude.

2.2 Data Cleaning:

Before creating the dataframe, the data was cleaned. Since the scraped data contained numbers along with the names, all the numbers were removed from each row. Extra rows like references, others, etc. were also removed. For removing the numbers, a regular expression was used. Hence, re python library was also used. The index was reset after deleting the rows. After cleaning, the coordinates were added to the dataframe with individual columns of Latitude and Longitude.