Predicting the …….

Zayneb AMINE

July,1,2020

I : Introduction

I.1 : Background

Mumbai, also known as Bombay, is the capital city of the Indian state of Maharashtra. According to United Nations, as of 2018, Mumbai was the second-most populous city in India after Delhi and the seventh-most populous city in the world with a population of roughly 20 million. As per Indian government population census of 2011, Mumbai was the most populous city in India with an estimated city proper population of 12.5 million living under Municipal Corporation of Greater Mumbai. Mumbai is the centre of the Mumbai Metropolitan Region, the sixth most populous metropolitan area in the world with a population of over 23 million. A such populated city has so many shops and restaurants.

I.2 : Problem

How to search for a specific type of venues, to explore a particular venue, to explore a foursquare user, to explore a geographical location, and to get a trending venues around location. Also how to use the visualization library, folium, and visualize the result. How a person lost in the city can found shops near his position.

I.3 : Data interest

There is so many people in Mumbai, it could be a real interest for the user to find a shop. Also the city administration can learn how the population live and consumes.

II : Data aquisition and cleaning

II.1 : Data sources

Most mumbai suburb stats are available in the Mumbai Suburb dataset here https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai it is an old dataset dated from 2009.

II.2 : Data cleaning

Data downloaded from source will be combined into one table. Then I will parse data from the html into an object and create a list to store neighborhood data and finally create a dataframe from the list. The I will create another dataframe from the coordinates data. And finally merge it in the first dataframe.

II.3 : Features selection

After cleaning the data there are 199 uniques categories of shop (restaurant). So there is 2738 samples and 199 features.