IBM – Coursera

Data Science Specialization

Capstone Project – Project Report

CAPSTONE PROJECT FOR FINDING OPTIMIZED VENUE AT FRANCE

Rajavelu Angalan

2019

Table of Content:

Problem Description :

This report is for the final course of the Data Science Specialization. A 9-courses series created by IBM, hosted on Coursera platform. The problem and the analysis approach are left for the learner to decide, with a requirement of leveraging the Foursquare location data to explore or compare neighborhoods or cities of your choice or to come up with a problem that you can use the Foursquare location data to solve.

In this project, the problem is to find the optimal location or finding the cityof cluster which has user preferred venue eg. BAR,PLAZA and GYM in France. To achieve this task, an analytical approach will be used, based on advance machine learning techniques and data analysis,concretely clustering and perhaps some data visualization techniques.

So can the city surrounding has user preferred venues ?
If so, what types of venues cluster has the most affect, both positively and negatively?

The Target Audience for this project is for who prefer to stay in hotel based on on their preferred venues(eg.Tourists).

Data Preparation:

France cities were chosen as the observation target due to the following reasons:

With more than 10 million tourists a year, the French Riviera (French: Côte d'Azur), in southeastern France, is the second leading tourist destination in the country, after the Parisian region.

The availability of geo data which can be used to visualize the dataset onto a map.

France City Coordinates have been taken from below website as a csv file
https://simplemaps.com/data/fr-cities

- FourSquare API which provides the surrounding venues of a given coordinates.

The process of collecting data as follows:

• Load the city co ordinates into data frame and clean the data by removing unnecessary filed's.
• Pass the obtained co rdinates to the foursquare API.The "explore" endpoint will return a list of surrounding venues in a pre-defined radius.
• Count the occurrence of each venue type in a neighborhood. Then apply one hot encoding to turn each venue type into a column with their occurrence as the value.

	A	B	C	D	E	F	G
1	city	lat	lng	country	iso2	capital	population
2	Paris	48.866667	2.333333	France	FR	primary	9904000
3	Lyon	45.748457	4.846711	France	FR	admin	1423000
4	Marseille	43.285413	5.37606	France	FR	admin	1400000
5	Lille	50.632971	3.058585	France	FR	admin	1044000
6	Nice	43.713644	7.25952	France	FR	927000	338620
7	Toulouse	43.599516	1.433188	France	FR	admin	847000
8	Bordeaux	44.840439	-0.5805	France	FR	admin	803000
9	Rouen	49.433333	1.083333	France	FR	admin	532559
10	Strasbourg	48.600381	7.787355	France	FR	admin	439972
11	Nantes	47.216509	-1.552379	France	FR	admin	438537
12	Metz	49.115461	6.175875	France	FR	minor	409186
13	Grenoble	45.171546	5.722387	France	FR	388574	158552
14	Toulon	43.117705	5.941712	France	FR	357693	168701
15	Montpellier	43.61092	3.87723	France	FR	minor	327254