CAPSTONE PROJECT FOR FINDING OPTIMIZED VENUE AT FRANCE

Rajavelu Angalan

2019

Problem Description

In this project, the problem is to find the optimal location or finding the city of 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 onon their preferred venues(eg.Tourists).

In order to access Foursquare API data, Actual data from https://simplemaps.com/data/fr-cities is transformed into below format for processing.

	A	B	C	D
1	city	lat	lng	country
2	Paris	48.866667	2.333333	France
3	Lyon	45.748457	4.846711	France
4	Marseille	43.285413	5.37606	France
5	Lille	50.632971	3.058585	France
6	Nice	43.713644	7.25952	France
7	Toulouse	43.599516	1.433188	France
8	Bordeaux	44.840439	-0.5805	France
9	Rouen	49.433333	1.083333	France
10	Strasbourg	48.600381	7.787355	France
11	Nantes	47.216509	-1.552379	France
12	Metz	49.115461	6.175875	France
13	Grenoble	45.171546	5.722387	France
14	Toulon	43.117705	5.941712	France
15	Montpellier	43.61092	3.87723	France

Once the data pre-processed can be visualized as below in map.

The cities were plotted into map of Franc based on the co ordinates.

Then the venues of near by cities at France are extracted and frequency of venue will be calculated.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Paris	48.866667	2.333333	Pierre Hermé	48.868222	2.333397	Pastry Shop
1	Paris	48.866667	2.333333	Le Roch Hotel & Spa Paris	48.866200	2.332995	Hotel
2	Paris	48.866667	2.333333	Cantine California	48.867401	2.332017	Food Truck
3	Paris	48.866667	2.333333	Boulangerie Aki	48.866211	2.335458	Bakery
4	Paris	48.866667	2.333333	Brasserie Réjane	48.865486	2.334824	Restaurant

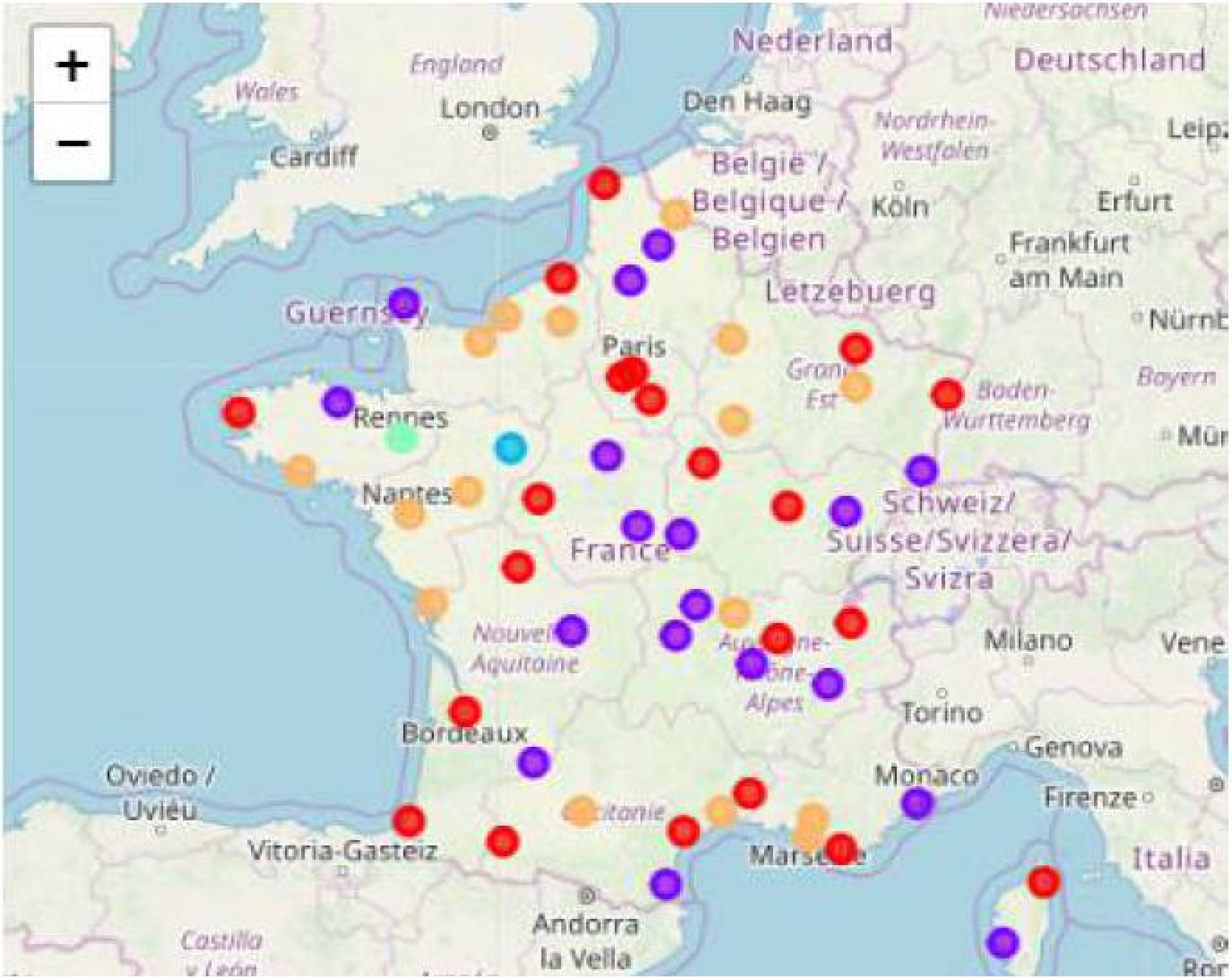
Once the venues of each cities obtained then calculate the frequency of occurrence of each venue.

```
----Agen----
               venue  freq
0        Supermarket  0.33
1       Dance Studio  0.33
2               Park  0.33
3  Accessories Store  0.00
4        Optical Shop  0.00
```

Cluster of cities based on Foursquare API venues will be measured using Kmeans algorithm and displayed as below

	CITY	LATITUDE	LONGITUDE	COUNTRY	POPULATION	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Paris	48.866667	2.333333	France	9904000	0	Japanese Restaurant	Hotel	French Restaurant
1	Lyon	45.748457	4.846711	France	1423000	0	Restaurant	Diner	Bistro
2	Marseille	43.285413	5.376060	France	1400000	4	Plaza	Bus Stop	Lounge
3	Lille	50.632971	3.058585	France	1044000	4	French Restaurant	Bar	Japanese Restaurant
4	Nice	43.713644	7.259520	France	338620	1	French Restaurant	Plaza	Seafood Restaurant

To cntd..


+
−
Nederland
Deutschland
England
Wales
London
Den Haag
Nordrhein-Westfalen
Cardiff
België / Belgique / Belgien
Köln
Erfurt
Frankfurt am Main
Letzebuerg
Guernsey
Paris
Bayern
Rennes
Baden-Württemberg
Nantes
Schweiz/ Suisse/Svizzera/ Svizra
France
Nouvelle-Aquitaine
Milano
Torino
Bordeaux
Genova
Oviedo / Uviéu
Monaco
Firenze
Vitoria-Gasteiz
Italia
Andorra la Vella
Castilla y León

Finally, User input has applied on the data and optimized clusters displayed.

Conclusion :

As far as we can see with this data, some of the clusters are not populated because of user filter.
It is highly possible that user_cluster 1 & 5 has more cities which has the user preferences of hotel cluster. If the user input data should perform with more data and logic also framed in proper way then we can provide more accurate Output .

References

https://developer.foursquare.com/docs/api/venues/
https://simplemaps.com/data/fr-cities
https://www.coursera.org/