IBM CAPSTONE PROJECT – The Battle of Neighborhoods:
Cluster Analysis of Paris

Introduction: Business Problem

In this project I would like to find an optimal location for a dessert shop in Paris.

The idea is to find a place in the city center, near some clothes shop or landmark in order to have more tourists. It's important that this shop is far from other similar shops.

I'll use web scraping in order to get information about possible neighborhoods and Foresquare in orger to get information about shops.

The target audience of this report is any one that is interested in opening a dessert shop but have no idea in which arrondissement.

Data

The most important criteria is the number of existing dessert shops in the neighborhood. I'll divide Paris according to its arrondissements to define neighborhoods, centered around city center; for each arrondissement I will find the dessert shops in order to calculate the best location for the new shop.

Each arrondissement will be approximated using Google Maps API geocoding, every activity its type and location using Foursquare API. The center of Paris will be approximated using Google Maps API too.

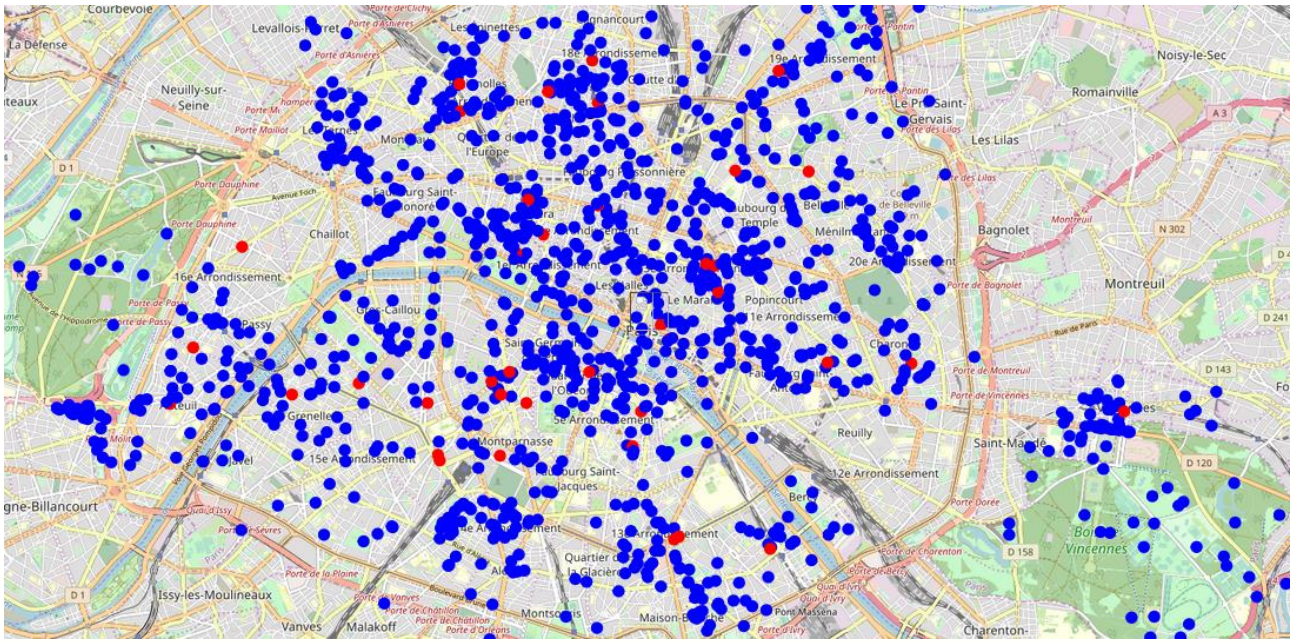
Data Exploration

After extracting and reading the data, we will translate the above data into a Pandas data frame for processing which would look like this. These are the data elements that are needed when we call Foursquare web service call in order to get the venues available in that arrondissement.

	Arrondissement	Name	Arrondissement_number	Latitude	Longitude
0	1st (Ier) R	Louvre	1st	48.862158	2.337036
1	2nd (IIe) R	Bourse	2nd	48.867684	2.343126
2	3rd (IIIe) R	Temple	3rd	48.862683	2.358685
3	4th (IVe) R	Hôtel-de-Ville	4th	48.854156	2.356789
4	5th (Ve) L	Panthéon	5th	48.845419	2.352582
5	6th (VIe) L	Luxembourg	6th	48.849392	2.332260
6	7th (VIIe) L	Palais-Bourbon	7th	48.854914	2.312860
7	8th (VIIIe) R	Élysée	8th	48.877480	2.317650
8	9th (IXe) R	Opéra	9th	48.877070	2.337921
9	10th (Xe) R	Entrepôt	10th	48.876008	2.360445
10	11th (XIe) R	Popincourt	11th	48.860071	2.378143
11	12th (XIIe) R	Reuilly	12th	48.835200	2.445136
12	13th (XIIIe) L	Gobelins	13th	48.829853	2.363026
13	14th (XIVe) L	Observatoire	14th	48.829567	2.323962
14	15th (XVe) L	Vaugirard	15th	48.841430	2.296165
15	16th (XVIe) R	Passy	16th	48.857151	2.256731
16	17th (XVIIe) R	Batignolles-Monceau	17th	48.887439	2.306523
17	18th (XVIIIe) R	Butte-Montmartre	18th	48.892126	2.348178
18	19th (XIXe) R	Buttes-Chaumont	19th	48.887219	2.383103
19	20th (XXe) R	Ménilmontant	20th	48.861827	2.401073

Foursquare

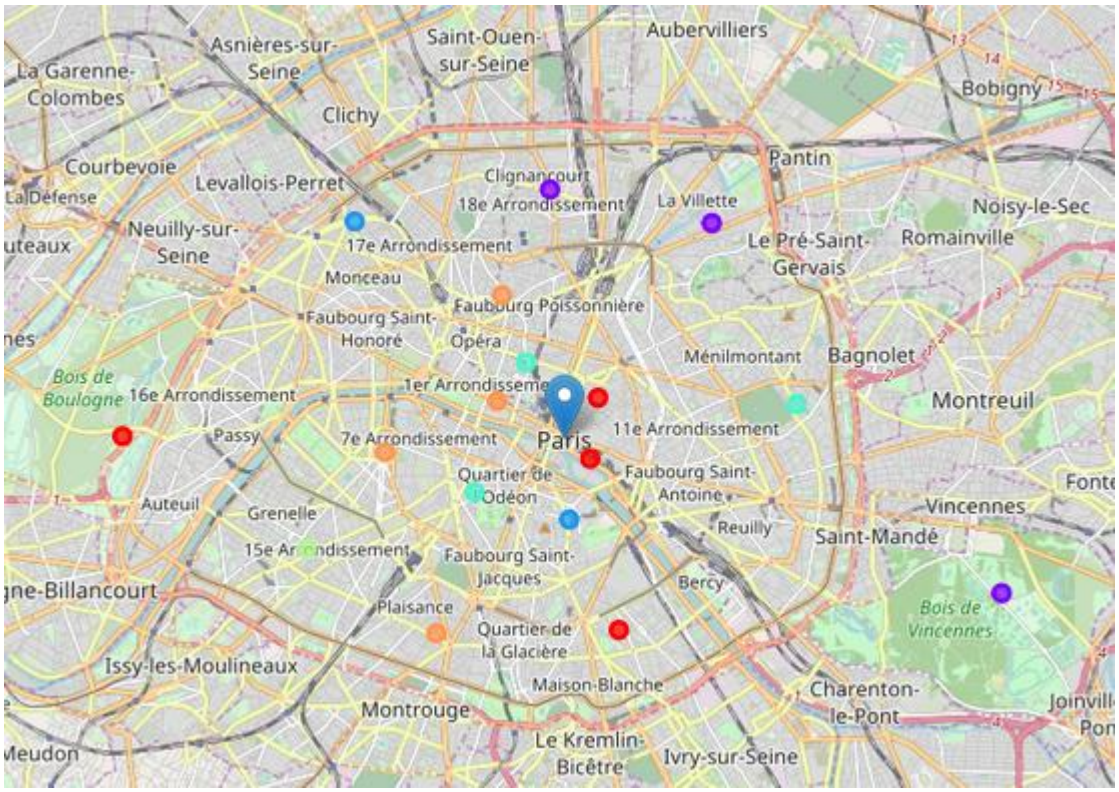
I use Foursquare API to get info on each activity in each neighborhood.
I'm interested in activity that sell dessert, cupcake, chocolate...
In the map all this kind of activity are in red



I prefer to concentrate only in dessert shops so I filter the data and I try to identify cluster for each arrondissement.

Arrondissements 8,10,11 have not any dessert shop so are perfect place.

However I visualize cluster related to the other arrondissemen



As we can see most of them are concentrated in the city center buti f the client dosen't matter about the distance from the center in the rest of the city there a lot of good place near other activity as we have seen above.