FINAL REPORT ON THE COURSERA CAPSTONE PROJECT

This notebook is about the final report on the capstone project. It will contain all the steps of the required exercise.

I- INTRODUCTION / BUSINESS PROBLEM

The owner of a business like a restaurant needs to fix all the aspects in order to ensure good results. One of the major aspects is the location of the restaurant as if the restaurant is well located, it will have a positive impact in terms of affluence. On the other hand, if the restaurant is badly located, it may negatively impact the restaurant's attendance, then the business recipes. The question is thus "Where is the best place in the city to install my new restaurant?".

We then fall on a descriptive model. With the foursqaure API, many things can be explore to locate that place depending of the city. We can explore the venues there, the ratings of these venues. Depending of a venue in the city, we can look at how many restaurants do already exist there? what are their types? what about the trending venues there? What informations do we get from tips?

The goal at the end is to cross all these informations and with some maps, to give a reliable answer to our client who needs to know where to install his restaurant. The audience of this project is all the people interested in starting a business such as opening a restaurant and wanting to know the best place in the city where to install it. The city we choose to treat is Clermont-Ferrand and any type of the restaurant.

II- DATA

This section presents the data we will use to solve the problem. As said in the introduction, we assume the city is the clermont-ferrand town and we am looking for the best place for a new restaurant.

Factors that will influence our decission are :

- number of existing restaurants in the neighborhood (any type of restaurant)
- distance of neighborhood from city center

We first choose to get geographical coordinates of Clermont-Ferrand city, second generate some candidates locations nearby the city center, and third choose to explore some data with the Foursquare API.

1. Data acquisition

The dataset that we will be using is comin from the Foursquare API where a lot of data could be found about the city of Clermont-Ferrand. We first get the coordinates of the city, then we we download and treat some data from Foursquare.

a. City (Clermont-Ferrand) search for latitude and longitude coordinates

City (Clermont-Ferrand) search for latitude and longitude coordinates :

```
45.7774282 3.0813291
```

After having extact the coordinates of the studied city (clermont_Ferrand, We will explore many locations in that city. As an example, We will show you the explorations of the venues.

b. Definding an url

For the url, we fixed the radius to 600 and defined the url as follows :
radius = 600
url='https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&v={ }&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, clermont_latitude, clermont_longitude, VERSION, radius, LIMIT)

2. Data cleaning

After applying the step of data acquisition with a GET request of the last url, we proceed to the data cleaning step. That consists on geutting relevant part of JSON and transforming it into a *pandas* dataframe. Then, we assigned relevant part of JSON to venues.

	categories	hasPerk	id	location.address	location.cc	location.city	location.country	location.crossStreet	loc
0	[{'id': '4bf58dd8d48988d164941735', 'name': 'P...	False	4b1fedf8f964a520f32a24e3	Place de Jaude	FR	Clermont-Ferrand	France	NaN	
1	[{'id': '4bf58dd8d48988d10c941735', 'name': 'F...	False	4c7bf69776ce9c741bddbb0c	1 Rue Blatin	FR	Clermont-Ferrand	France	14 Place de Jaude	
2	[{'id': '52e81612bcbc57f1066b79f1', 'name': 'B...	False	514db4a5e4b0e7ec7079ad5a	Rue Des Minimes	FR	Clermont-Ferrand	France	NaN	
3	[{'id': '4bf58dd8d48988d10c941735', 'name': 'F...	False	4ec40a8802d5ad633a8c047b	NaN	FR	Clermont-Ferrand	France	NaN	
4	[{'id': '4bf58dd8d48988d16e941735', 'name': 'F...	False	4ba948d6f964a520521b3ae3	51-53 Avenue des Etats-Unis	FR	Clermont-Ferrand	France	NaN	
5	[{'id': '4bf58dd8d48988d107951735', 'name': 'S...	False	4f12ed7de4b0804e8bdb507f	NaN	FR	NaN	France	NaN	
6	[{'id': '4bf58dd8d48988d132941735', 'name': 'C...	False	51502a29e4b02a09caced3ff	NaN	FR	NaN	France	NaN	
7	[{'id': '52f2ab2ebcbc57f1066b8b4f', 'name': 'B...	False	506d77ffe0e2b36a5b745a87	18 rue Blatin	FR	Clermont-Ferrand	France	NaN	

3. Features selection

After the step of data cleaning, we defined information of interest, filter dataframe shown before., and keep only columns that include venue name, and anything that is associated with location.

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng
0	Place de Jaude	Plaza	Place de Jaude	FR	Clermont-Ferrand	France	NaN	90	[Place de Jaude, 63000 Clermont-Ferrand, France]	[{'label': 'display', 'lat': 45.77681306296083...	45.776813	3.082093
1	Le Faisan Doré	French Restaurant	1 Rue Blatin	FR	Clermont-Ferrand	France	14 Place de Jaude	54	[1 Rue Blatin (14 Place de Jaude), 63000 Clerm...	[{'label': 'display', 'lat': 45.77698454161444...	45.776985	3.081626
2	Le Bistrot d'à Côté	Bistro	Rue Des Minimes	FR	Clermont-Ferrand	France	NaN	258	[Rue Des Minimes, 63000 Clermont-Ferrand, France]	[{'label': 'display', 'lat': 45.77874923383751...	45.778749	3.084074
3	Le Bouchon de Jaude	French Restaurant	NaN	FR	Clermont-Ferrand	France	NaN	31	[Clermont-Ferrand, France]	[{'label': 'display', 'lat': 45.77767578339747...	45.777676	3.081120
4	McDonald's	Fast Food Restaurant	51-53 Avenue des Etats-Unis	FR	Clermont-Ferrand	France	NaN	69	[51-53 Avenue des Etats-Unis, 63000 Clermont-F...	[{'label': 'display', 'lat': 45.77798357612858...	45.777984	3.081738
5	Minelli	Shoe Store	NaN	FR	NaN	France	NaN	85	[France]	[{'label': 'display', 'lat': 45.77739079520978...	45.777391	3.082423
6	Église Saint-Pierre-des-Minimes	Church	NaN	FR	NaN	France	NaN	77	[France]	[{'label': 'display', 'lat': 45.77732989315853...	45.777330	3.080340
7	Arrêt Jaude ®© [9,10]	Bus Stop	18 rue Blatin	FR	Clermont-Ferrand	France	NaN	47	[18 rue Blatin, 63000 Clermont-Ferrand, France]	[{'label': 'display', 'lat': 45.77704, 'lng': ...	45.777040	3.081064
8	Brasserie Le Lion	Bar	NaN	FR	NaN	France	NaN	71	[France]	[{'label': 'display', 'lat': 45.77684526562149...	45.776845	3.081701

We got some examples data on the city of Clermont-Ferrand that we turn into a dataframe. For the data, we fixed the raduis to 500. With this data, we can explore many locations to know more about the places there. We can see in our dataframe that there are many categories (hotel, art gallery, residential Building, etc). We will investiguate (for more data of clermont-ferrand city near the city center) some of that categories in relationship to the restaurant category to select our best place for the restaurant.

III- METHODOLOGY

1. Description

The methodology to know where is the best place to install the new restaurant is to investigate locations where we still don't have a great number of restaurant. The second criterion, is to look at venues around the city center. We will limit our analysis to area ~6km around city center.

We will first generate a grid of candidates of areas around the city center.

Second, We will look at the distribution of restaurants around the city center to see where are located areas with minimum restaurants.

Third, We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

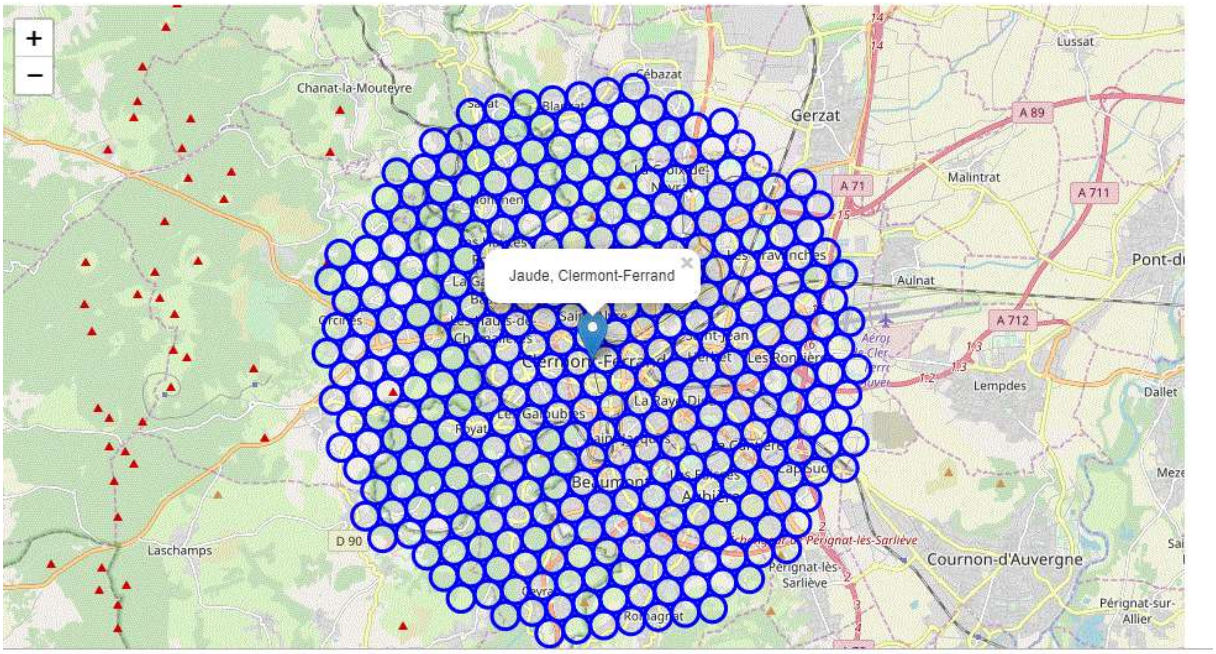
2. Analysis

Let's perform some basic explanatory data analysis and derive some additional info from our raw data.

First let's create an Hexagonal grid of candidates of locations. We offset every other row, and adjust vertical row spacing so that every cell center is equally distant from all it's neighbors.

	Latitude	Longitude	X	Y	Distance from center
0	45.724696	3.069564	-428167.930747	5.133149e+06	5992.495307
1	45.725495	3.077108	-427567.930747	5.133149e+06	5840.376700
2	45.726294	3.084652	-426967.930747	5.133149e+06	5747.173218
3	45.727092	3.092196	-426367.930747	5.133149e+06	5715.767665
4	45.727890	3.099741	-425767.930747	5.133149e+06	5747.173218
5	45.728687	3.107286	-425167.930747	5.133149e+06	5840.376700
6	45.729483	3.114832	-424567.930747	5.133149e+06	5992.495307
7	45.728072	3.057259	-429067.930747	5.133669e+06	5855.766389
8	45.728872	3.064803	-428467.930747	5.133669e+06	5604.462508
9	45.729671	3.072348	-427867.930747	5.133669e+06	5408.326913

OK, we now have the coordinates of centers of neighborhoods/areas to be evaluated, equally spaced (distance from every point to it's neighbors is exactly the same) and within ~6km from Clermont-ferrand center, Jaude.Let's visualize the data we have so far: city center location and candidate neighborhood centers :



Now that we have our location candidates, let's use Foursquare API to get info on restaurants in each neighborhood.

We're interested in venues in 'food' category, but only those that are proper restaurants - coffe shops, pizza places, bakeries etc. are not direct competitors so we don't care about those. So we will include in out list only venues that have 'restaurant' in category name, and we will be careful to check locations with at most 2 restaurants nearby.

We first go over our neighborhood locations and get nearby restaurants;

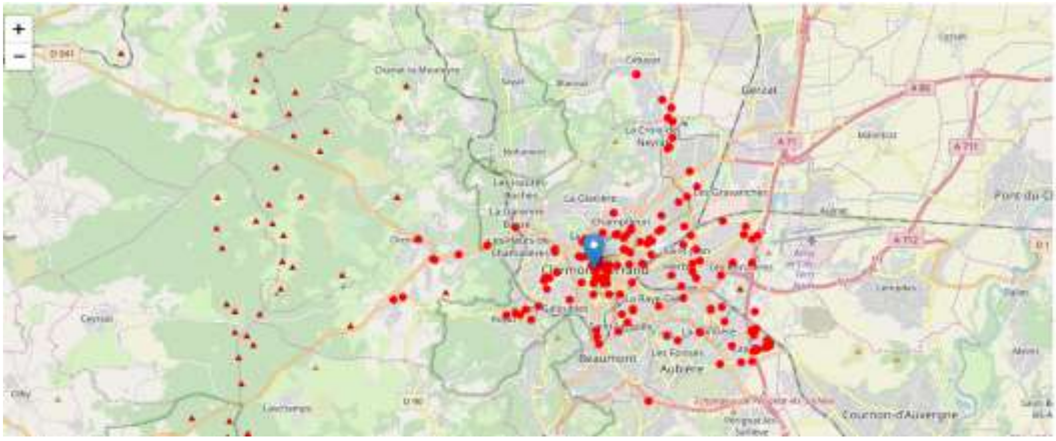
```
import numpy as np

print('Total number of restaurants:', len(restaurants))
print('Total number of african restaurants:', len(african_restaurants))
print('Percentage of african restaurants: {:.2f}%'.format(len(african_restaurants) / len(restaurants) * 100))
print('Average number of restaurants in neighborhood:', np.array([len(r) for r in location_restaurants]).mean())
```

```
Total number of restaurants: 152
Total number of african restaurants: 0
Percentage of african restaurants: 0.00%
Average number of restaurants in neighborhood: 0.3543956043956044
```

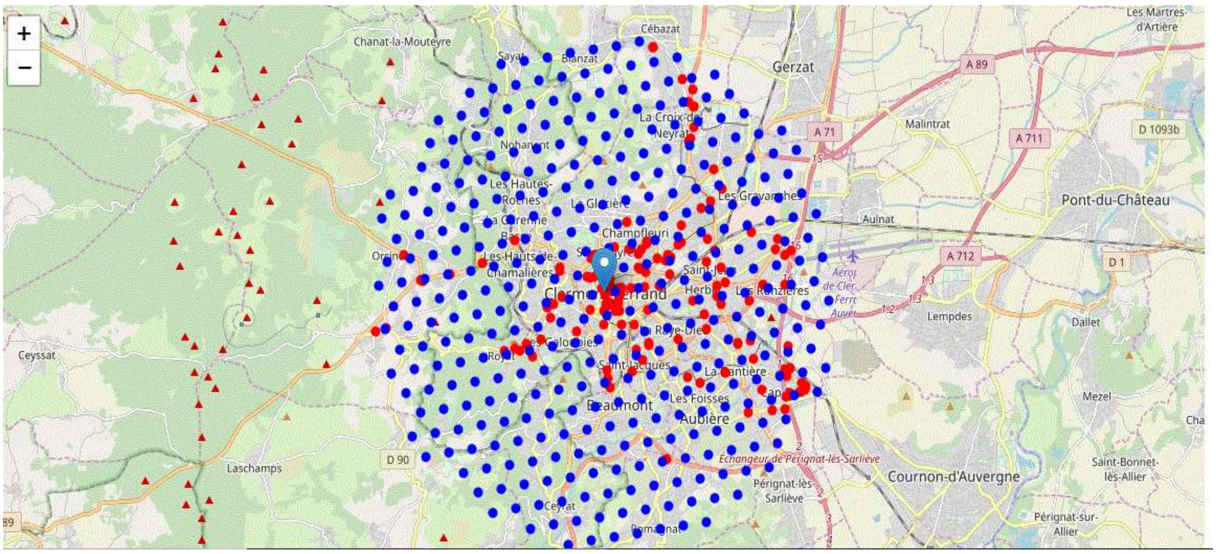
We also maintain a dictionary of all found restaurants .

```
List of all restaurants
-----------------------
('584003a16f20d11b223c07ad', 'CBH Cuisines Professionnelles', 45.74258002149345, 3.1018352508544917, '13 bis rue Cugnot, 63540 ROMAGNAT, France', 117,
False, -425359.0246018602, 5134755.757146825)
('5270fa6111d2c97eeaa622cc', 'Rouge Tendance', 45.75312997180263, 3.140796381479397, 'France', 241, False, -422154.3514284204, 5135470.703306506)
('4f390034e4b0948a83a52600', 'Cafeteria Auchan', 45.752402506847694, 3.128564517897922, 'France', 218, False, -423117.1267175402, 5135533.023429573)
('5494146c498ec92862e2d1ea', '5ème Saison', 45.75284957885742, 3.136699914932251, '34 Avenue De COURNON, 63170 Aubiere, France', 257, False, -422477.3
976580292, 5135487.476067202)
('4e674d62ae609d64bcd17c1c', 'La Distillerie', 45.75727138796016, 3.1330977411735095, '39 rue des Sauzes, 63170 Aubière, France', 258, False, -422683.
44720677566, 5136020.504812379)
('4ca8c8668882199c6fc744d4', 'Oïshi', 45.756805078836095, 3.1455516915886217, '49 avenue Lavoisier, 63170 Aubière, France', 161, False, -421723.433013
1897, 5135823.117757798)
('4dfc7b84aeb7594e86202206', 'Hippopotamus', 45.75650578573906, 3.1422940195290896, 'Aubière, France', 137, False, -421981.587862151, 5135827.96678592
45)
('4dc839ce7d8b9580aa3e1320', 'La Boucherie Clermont Ferrand Aubière', 45.75598094266169, 3.1423767942512923, '26.0 avenue Lavoisier, 63170 Aubiere, Fr
ance', 116, False, -421983.91983094194, 5135768.732919965)
('4d9219219d0f721ede642573', 'La Rotisserie', 45.756336194934114, 3.1424527502166617, 'France', 118, False, -421972.08471901366, 5135807.283916024)
('4dbee9ca93a08f9274f6e378', 'La Boucherie Sud', 45.75623365994492, 3.1427834803765595, '63000 Clermont-Ferrand, France', 90, False, -421948.09534682
6, 5135792.034837911)
...
Total: 152
```

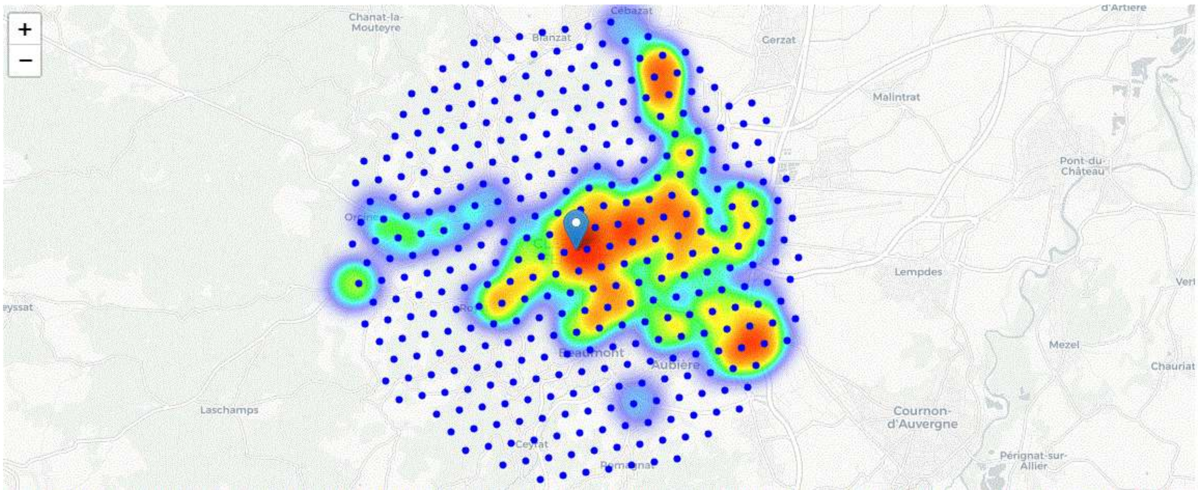
We then Visualize restaurants in the vicinity of the city center.



Now that we have the distribution of the restaurants in the Clermont-city in a radiums of 6km from the city center jaude, we can visualize them with the locations



We can also show on a map the distribution of resturants using a HeatMap.



We continue the analysis by couting the number of restaurants per location.

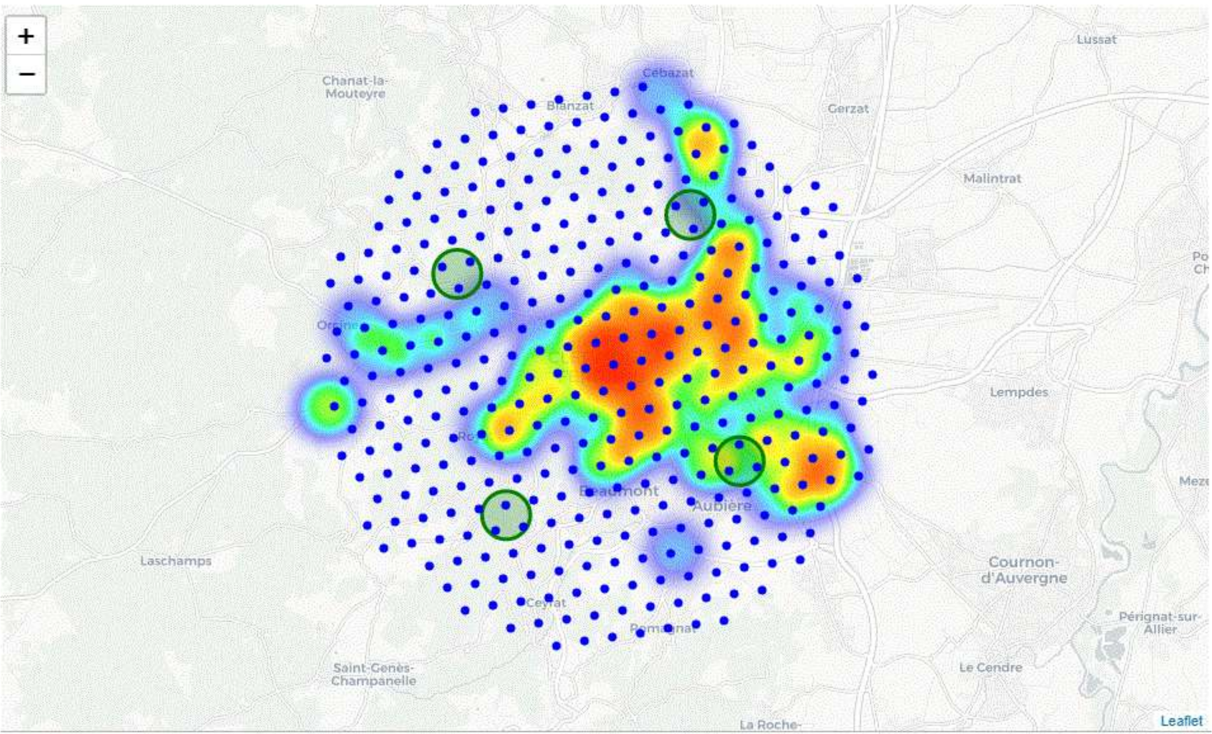
	Latitude	Longitude	X	Y	Distance from center	Restaurants in area
0	45.724696	3.069564	-428167.930747	5.133149e+06	5992.495307	0
1	45.725495	3.077108	-427567.930747	5.133149e+06	5840.376700	0
2	45.726294	3.084652	-426967.930747	5.133149e+06	5747.173218	0
3	45.727092	3.092196	-426367.930747	5.133149e+06	5715.767665	0
4	45.727890	3.099741	-425767.930747	5.133149e+06	5747.173218	0
5	45.728687	3.107286	-425167.930747	5.133149e+06	5840.376700	0
6	45.729483	3.114832	-424567.930747	5.133149e+06	5992.495307	0
7	45.728072	3.057259	-429067.930747	5.133669e+06	5855.766389	0
8	45.728872	3.064803	-428467.930747	5.133669e+06	5604.462508	0
9	45.729671	3.072348	-427867.930747	5.133669e+06	5408.326913	0
10	45.730470	3.079892	-427267.930747	5.133669e+06	5273.518749	0
11	45.731269	3.087437	-426667.930747	5.133669e+06	5204.805472	0
12	45.732067	3.094982	-426067.930747	5.133669e+06	5204.805472	0
13	45.732865	3.102528	-425467.930747	5.133669e+06	5273.518749	0
14	45.733662	3.110074	-424867.930747	5.133669e+06	5408.326913	0
15	45.734458	3.117620	-424267.930747	5.133669e+06	5604.462508	0

We check the locations with more than 2 restaurants in their vicinity

	Latitude	Longitude	X	Y	Distance from center	Restaurants in area
75	45.755948	3.143876	-421867.930747	5.135748e+06	5474.486277	6
93	45.760127	3.139118	-422167.930747	5.136267e+06	4938.623290	4
104	45.758725	3.081510	-426667.930747	5.136787e+06	2100.000000	3
139	45.765477	3.056884	-428467.930747	5.137826e+06	2343.074903	3
162	45.772852	3.082318	-426367.930747	5.138346e+06	519.615242	6
167	45.776842	3.120075	-423367.930747	5.138346e+06	3044.667470	3
179	45.775428	3.062452	-427867.930747	5.138865e+06	1500.000000	3
182	45.777828	3.085105	-426067.930747	5.138865e+06	300.000000	12
201	45.782004	3.080340	-426367.930747	5.139385e+06	519.615242	3
203	45.783602	3.095444	-425167.930747	5.139385e+06	1307.669683	3
204	45.784400	3.102996	-424567.930747	5.139385e+06	1873.499400	3
209	45.788384	3.140764	-421567.930747	5.139385e+06	4828.043082	3
222	45.787779	3.090679	-425467.930747	5.139904e+06	1374.772708	4
331	45.817631	3.107412	-423667.930747	5.143022e+06	4956.813493	3

Then We check the locations with less than 2 restaurants in their vicinity

We then cluster with the k-means algortihm the areas with areas less or equals to 2 restaurants. We choose to have 4 candidates clusters and plot their centers. K-means algorithm is chose because of its simplicity of implementation.



The cluster centers are :

```
[(3.042817670315043, 45.794928060853266),
 (3.105863098291043, 45.80606883790682),
 (3.119188603231289, 45.759568680689846),
 (3.0560261375214237, 45.74935328906884)]
```

The names of the clusters centers are (obtained by the Google Place API) :

```
(3.042817670315043, 45.794928060853266) = 'Route de Ternant, 63830 Durtol, France'
(3.105863098291043, 45.80606883790682) = 'Fédération PCF 63, 34 Rue des Clos, 63100 Clermont-Ferrand, France'
(3.119188603231289, 45.759568680689846) = '18 Allée du Capitaine Diéderich, 63170 Aubière, France'
(3.0560261375214237, 45.74935328906884) = '39 Avenue du Chorigier, 63122 Ceyrat, France'
```

IV- RESULTS

We conduct our analysis for the town of Clermont-Ferrand. We fixed to search locations around near the city centercalled Jaude. In terms of affluence, the city center will have more population and transports.

The analysis of the data shows that if we limit the research of the area of 6km around the city center, we have many candidates of locations.

When directing our attention on the distribution of restaurants in these locations. The map reveal that the density of restaurants is higher in the west part of the vicinity of the city center, than other part of the ciy center.

We then decided to look at locations with less or equals number of restaurants in the vicinity to 2. After having the candidates, we clustered them into 4 using the K_means algorithm. The results are 4 clusters centers where the centers are good candidates to be the best place for the new restaurant. The corresponding address of these places are :

- Route de Ternant, 63830 Durtol, France
- Fédération PCF 63, 34 Rue des Clos, 63100 Clermont-Ferrand, France
- 18 Allée du Capitaine Diéderich, 63170 Aubière, France
- 39 Avenue du Chorigier, 63122 Ceyrat, France

V- DISCUSSION

Our results list 4 possible locations to where the new restaurant could be installed (Route de Ternant, 63830 Durtol, France Fédération PCF 63, 34 Rue des Clos, 63100 Clermont-Ferrand, France,18 Allée du Capitaine Diéderich, 63170 Aubière, France, 39 Avenue du Chorigier, 63122

Ceyrat, France). An other analysis can focus for example on candidates locations having at most 3 or 4 restaurants nearby and those locations may be clustered into 10 for example.

Another thing that can be investigated is the type of the restaurants (asian, exotic, italian, african french, ...) depending of the type of the new restaurant to be installed.

We should also consider recommended zones as a starting point for more detailed analysis. It can be reasons of why the density of restaurants is low for some locations candidates.

VI- CONCLUSION

The aim of this project was to help stakeholders finding optimal location for a new restaurant in the city of Clermont-Ferrand in France.

We firts generated extensive collection of locations which satisfy some basic requirements like being of the vicitiny of the city center (6km). We investigated this collections of locations and look at caracteristics like the restaurant density distribution around the city center which may be more affluent (citizens, tourists, Transports,..). Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations with actually low restaurant density), and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

The Final decission on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.