

Opening a Restaurant in Paris

Salah eddine KOUTA

27 April 2020

I. Introduction:

a. Problem Background:

Paris , the city of light, is one the biggest capitals in Europe. it attracts so many people willing to open a business or commerce . Its multicultural aspect can be noticed through the diversity of the restaurants' cuisines. But most importantly , "Les parisiens" adores going out to eat in restaurant , especially the weekend. Therefore , this raise an important question: how a "parisien" choose his/her restaurant. That's is the question which has to be addressed before opening one. Geographically The city is split into three parts , which are defined by the seine. Each region concentrate a number of population , have its own markets and hold its own events. So in which neighborhood , it is better to open a restaurant ? which attributes can used for the model ? how the city is segmented giving the attributes ?

In this project , we will try to answer these questions and have a better understanding of the city of Paris.

b. Problem Description:

Opening a restaurant has never been easy. Opening a restaurant in Paris is far more challenging as the city is famous for its restaurant and cuisine . As the catering market is very competitive , one would like beforehand to understand the market dynamic in the city. And one of the vital aspect to begin with is the opening location. To this end , through this project , we will try to spot the best promising neighborhoods using a machine learning algorithm. We will perform this analysis based on :

- Paris Population
- Farmer markets density in Paris
- Events Density during the year in Paris
- Distribution of different restaurants in Paris (competition)

The clustering will convey a better understanding of catering distribution on Paris' neighborhoods.

c. Target :

The aim of this study is to help whoever want to take on the adventure of opening a restaurant in Paris. This is part is mainly concentrated in the location and further studies are needed to complete the analysis.

II. Data description

In this analysis, the data used is open source on the city of Paris <https://opendata.paris.fr/pages/home/>. We will go through the data set used in details.

a. Data Set 1 : Borough and neighborhoods

The first data set that we need in this analysis is the borough split and the neighborhood segmentation. This first data set groups the 20 arrondissements (borough) and the neighborhoods associated (80), we added afterward the coordinates which will be used to plot the map.

Source :

- <https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e>
- <https://opendata.paris.fr/explore/dataset/marches-decouverts/download/?format=json&timezone=Europe/Berlin'>

b. Data Set 2 : Population

We will include the population for every Borough. As stated in the description , this attribute will give us a better understanding of the population distribution in the city

borough	City	Region	population	postal_code	geo_point_2d
0	1	PARIS	ILE-DE-FRANCE	17.6	[48.8626304851685, 2.336293446550539]
1	2	PARIS	ILE-DE-FRANCE	22.4	[48.86790337886785, 2.344107166658533]
2	3	PARIS	ILE-DE-FRANCE	35.7	[48.86305413181178, 2.359361058970589]
3	4	PARIS	ILE-DE-FRANCE	28.2	[48.854228281954754, 2.357361938142205]
4	5	PARIS	ILE-DE-FRANCE	61.5	[48.844508659617546, 2.349859385560182]
5	6	PARIS	ILE-DE-FRANCE	43.1	[48.84896809191946, 2.332670898588416]
6	7	PARIS	ILE-DE-FRANCE	57.4	[48.85608259819694, 2.312438687733857]
7	8	PARIS	ILE-DE-FRANCE	40.3	[48.87252726662346, 2.312582560420059]
8	9	PARIS	ILE-DE-FRANCE	60.3	[48.87689616237872, 2.337460241388529]
9	10	PARIS	ILE-DE-FRANCE	95.9	[48.87602855694339, 2.361112904561707]
10	11	PARIS	ILE-DE-FRANCE	152.7	[48.85941549762748, 2.378741060237548]
11	12	PARIS	ILE-DE-FRANCE	142.9	[48.83515623066034, 2.419807034965275]
12	13	PARIS	ILE-DE-FRANCE	182.0	[48.82871768452136, 2.362468228516128]
13	14	PARIS	ILE-DE-FRANCE	137.2	[48.82899321160942, 2.327100883257538]
14	15	PARIS	ILE-DE-FRANCE	236.5	[48.84015541860987, 2.293559372435076]
15	16	PARIS	ILE-DE-FRANCE	169.4	[48.86039876035177, 2.262099559395783]
16	17	PARIS	ILE-DE-FRANCE	168.5	[48.88733716648682, 2.307485559493426]
17	18	PARIS	ILE-DE-FRANCE	200.6	[48.892735074561706, 2.348711933867703]
18	19	PARIS	ILE-DE-FRANCE	184.8	[48.88686862295828, 2.384694327870042]
19	20	PARIS	ILE-DE-FRANCE	197.1	[48.86318677744551, 2.400819826729021]

The table presented show the population of the 20 boroughs (it is presented in 1000 unit)

Source:

- <https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e>

c. Data Set 3 :Events

We will include also the number of events/activity. The data set again is an open source and can be found in the “opendata.paris” websites. This set gives us the number of events held through the year on every borough. We have even the address of the event. This is presented as such :

	access_link	access_type	address_city	address_name	address_street	address_zipcode	blind	category
0	https://exploreparis.com/fr/138-la-gare-du-nor...	reservation	Paris	Gare du Nord	112 rue de Maubeuge	75010	0.0	Animations -> Visite guidée
1	http://resas.scribeparis.org	conseillée	Paris	Espace Protestant Marc Boegner	27, rue de l'Annonciation	75016	0.0	Spectacles -> Théâtre
2	https://www.lebaisersale.com/agenda?event=2020...	conseillée	Paris	Le Baiser Salé	58 rue des Lombards	75001	0.0	Concerts -> Jazz
3	http://www.promothéâtre.com/marais_(théâtre_du...	conseillée	Paris	Théâtre du Marais	37 rue volta	75003	0.0	Spectacles -> Humour
4	https://theatredebelleville.mapado.com	conseillée	Paris	Théâtre de Belleville	16 Passage Piver	75011	0.0	Spectacles -> Théâtre

The data has to be cleaned and restructured to something consumable. I have did a count of number per borough.

Source :

- https://opendata.paris.fr/explore/dataset/que-faire-a-paris-/export/?disjunctive.category&disjunctive.tags&disjunctive.address_zipcode&disjunctive.address_city&disjunctive.access_type&disjunctive.price_type&basemap=jawg.transports&location=7.47.73195.2.41324

d. Data Set 4 : Farmer markets

This data set gives the list of the farmer markets in Paris. Again we will use just a counting variable linked to that data set.

Borough	dimanche	geo_point_2d	geo_shape	gestionnaire	h_deb_sem_1	h_fin_sem_1	id_marche	jeudi	jours_tenue	lineaire	localisation
0	9	0.0	[{"type": "Polygon", "coordinates": [[[2.344507...}}]	Dadoun	15:00	20:30	12.0	0.0	vendredi apm	176.0	place d'Anvers le long du square d'Anvers en v...
1	12	1.0	[{"type": "Polygon", "coordinates": [[[2.387303...}}]	Bensidoun	10:00	20:30	20.0	0.0	mercredi , dimanche	142.0	entre le n° 14 place Lachambeaudie et le n°11 ...
2	16	0.0	[{"type": "Polygon", "coordinates": [[[2.27489208238...}}]	Dadoun	07:00	14:30	48.0	0.0	mardi, vendredi	352.0	du n° 31 rue Gros au n° 1 rue la Fontaine
3	12	0.0	[{"type": "MultiPolygon", "coordinates": [[[2.39241681845...}}]	Bensidoun	07:00	14:30	71.0	0.0	mardi, vendredi	1182.0	sur les trottoirs pair et impair du boulevard ...
4	18	0.0	[{"type": "Polygon", "coordinates": [[[2.3533114763...}}]	Dadoun	07:00	14:30	54.0	0.0	mercredi, samedi	637.0	terre-plein du boulevard de la Chapelle entre ...

III. Methodology :

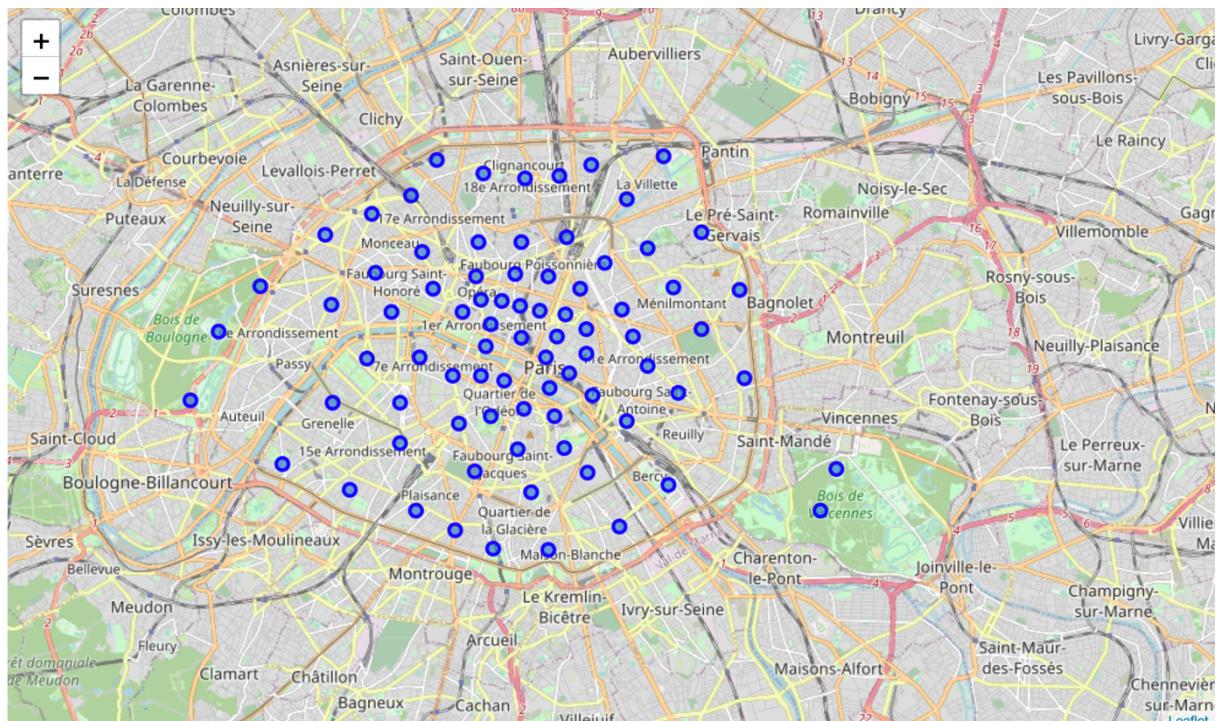
a. Business understanding:

The objective is to spot the promising neighborhood in Paris to open a restaurant.

b. Analytic approach:

1. Neighborhoods :

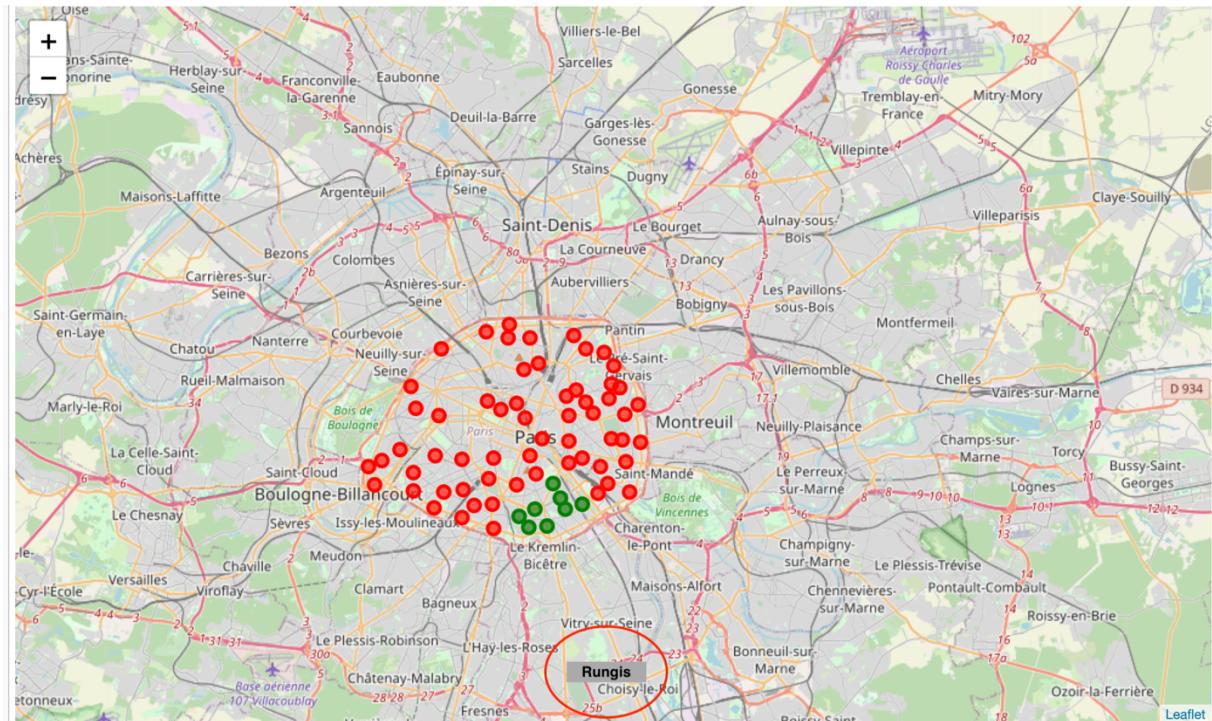
As an analysis , we have retrieved the set of Paris' arrondissements and linked that the list of neighborhoods. The following map give us the set of data which will be used for clustering.



As someone living in Paris , there is the center of Paris , the inner circle , and the outer circle or what we call (the first crown). Actually Paris is expanding more on the second crown which is out of our scope for now.

2. Farmer markets :

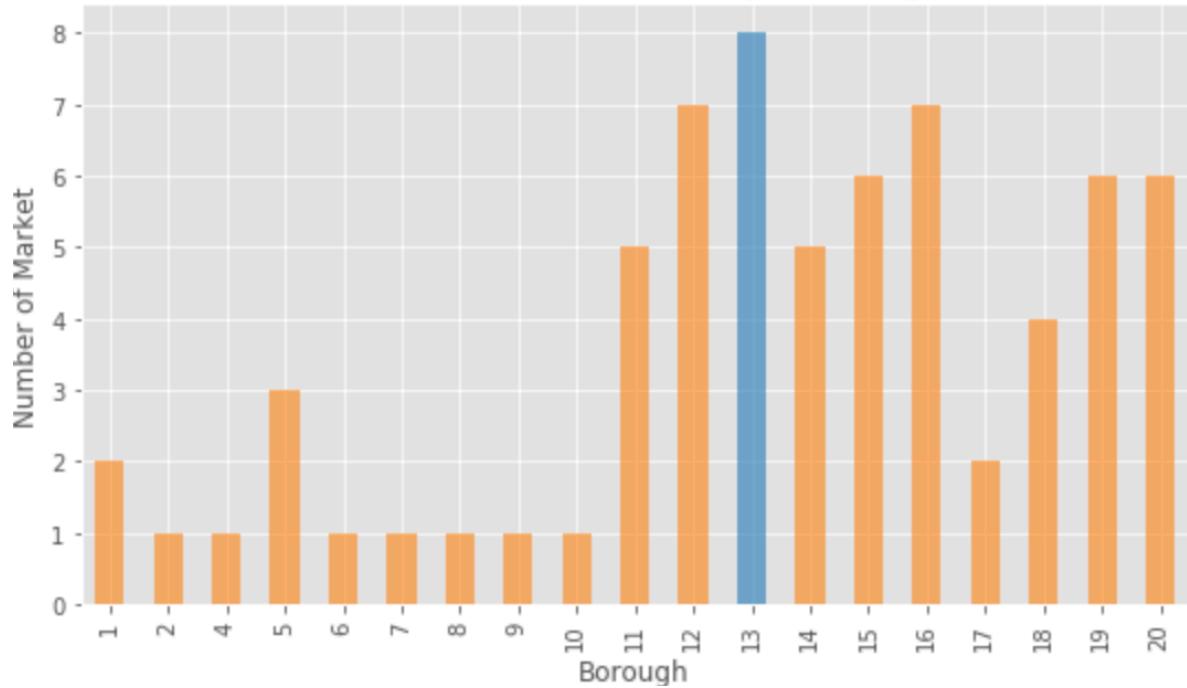
In this section , we wanted to analyze the area where farmer market are more dominant. An area where the market are more present is better for a restaurant as it is a source of fresh ingredients , but most importantly it is easier to be shipped as there exist already many solid supply chains.



We can see that there is a high concentration on the 13th arrondissement. Actually it is the neared to Rungis market which might explain this observation. Rungis is one of the biggest market in Europe.

From the graph we see clearly that the higher number of market in Paris is in the 13th arrondissement .

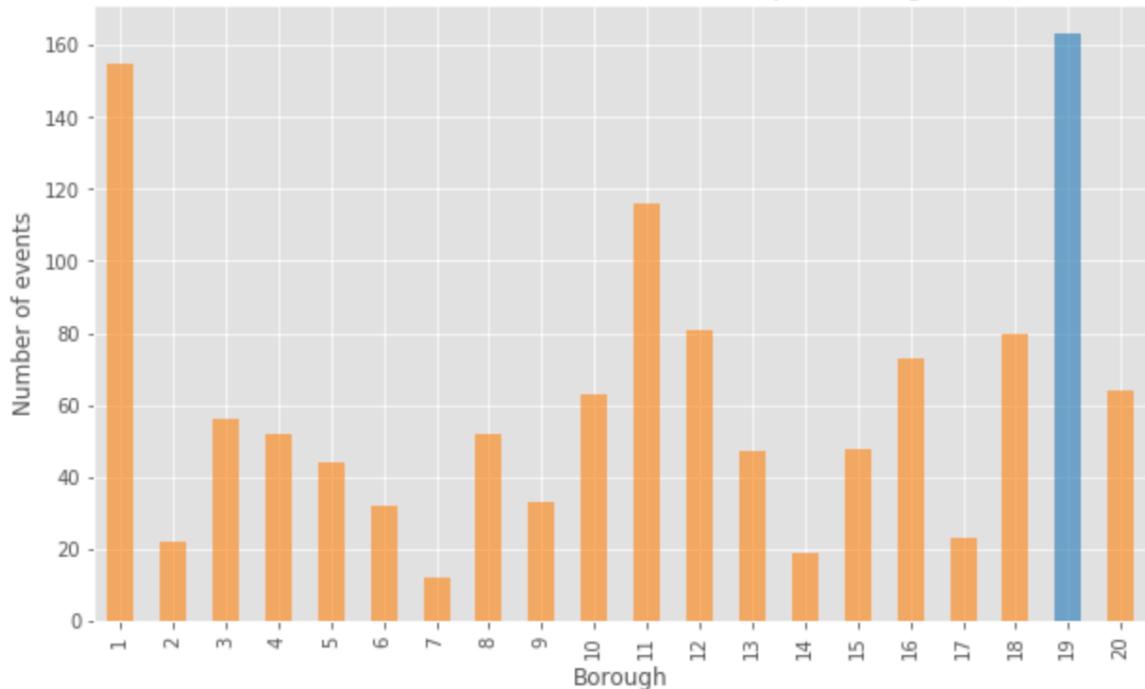
Count number of markets per Borough



3. Events/activities :

In this section we had analyzed the density of events held in Paris. We counted the number of events/activities per arrondissement.

Count number of events/activities per Borough



We can see that much more events are organized in the 19th arrondissement.

As for the first conclusion, the density is on the outer circle , but in this case more on the north of Paris. We start sensing a certain trend. From the two analysis, two of the borough that stand out are from the outer circle around Paris. Unfortunately , we can't draw any conclusion based

on the these two tests. And that's why we're going to use a machine learning clustering algorithm to cluster based on these variables .

c. Machine learning model:

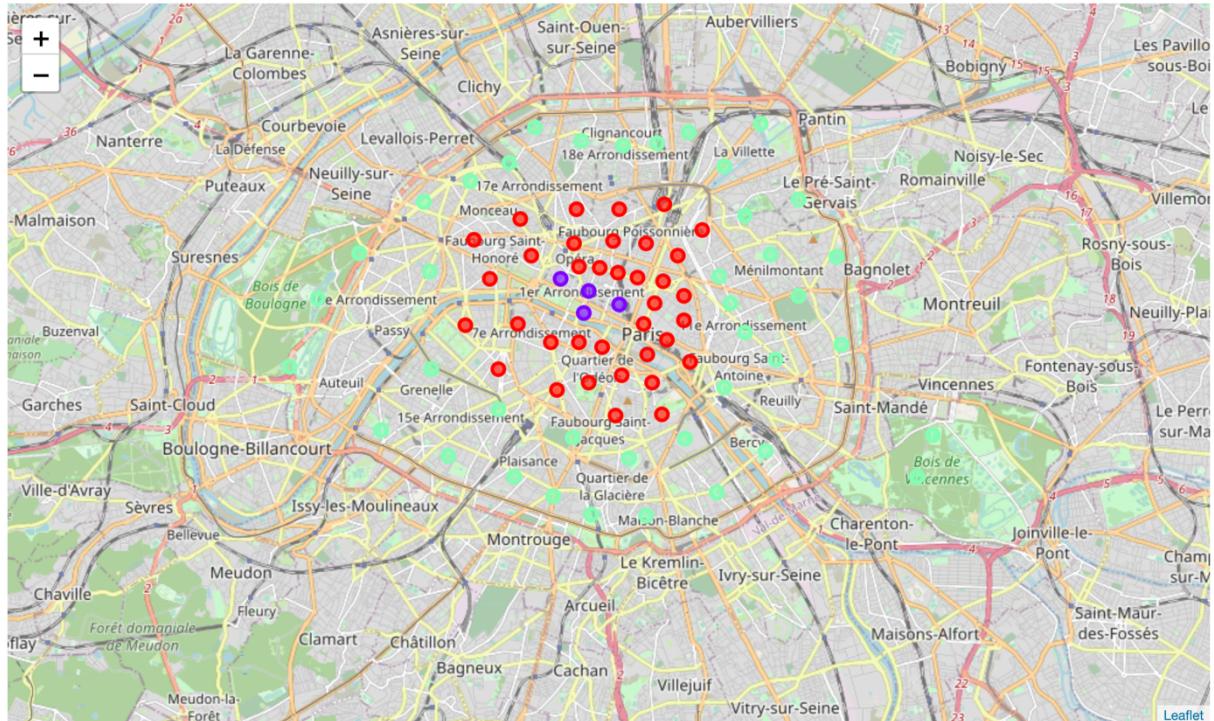
In order to draw this segmentation frame. We are going to use all the data stated before , not forgetting about the results above. as we don't have any idea about how to cluster using these variables we will use an unsupervised algorithm to divide data into similar clusters.

IV. Result :

For a “parisien” the results are quiet expected. The algorithm was done using 3 clusters (the number of clusters was set based on the efficiency of the algorithm given the silhouette Coefficient.

```
For n_clusters=2, The Silhouette Coefficient is 0.5781036415849897
For n_clusters=3, The Silhouette Coefficient is 0.5977237415941316
For n_clusters=4, The Silhouette Coefficient is 0.5625446287118456
For n_clusters=5, The Silhouette Coefficient is 0.5169129007639898
For n_clusters=6, The Silhouette Coefficient is 0.5390982155134438
For n_clusters=7, The Silhouette Coefficient is 0.4297378539851772
For n_clusters=8, The Silhouette Coefficient is 0.5263126539588534
For n_clusters=9, The Silhouette Coefficient is 0.465705046016261
For n_clusters=10, The Silhouette Coefficient is 0.47984229064121975
For n_clusters=11, The Silhouette Coefficient is 0.5140911302563455
For n_clusters=12, The Silhouette Coefficient is 0.5025074787260118
For n_clusters=13, The Silhouette Coefficient is 0.5299456007664407
For n_clusters=14, The Silhouette Coefficient is 0.5266277511095112
For n_clusters=15, The Silhouette Coefficient is 0.5374362541976391
For n_clusters=16, The Silhouette Coefficient is 0.5458264138667417
For n_clusters=17, The Silhouette Coefficient is 0.5476567923127329
For n_clusters=18, The Silhouette Coefficient is 0.5591625582366856
For n_clusters=19, The Silhouette Coefficient is 0.5552599490337641
```

The results are as follow:



- ***Cluster 0 is marked in red:*** This cluster unites the neighborhoods with the highest number of restaurants, but with a low number of markets and not a very significant number of events held during the year. Probably if we used the density of tourism, we would understand the high number of restaurants.
- ***Cluster 1 is marked in purple :*** This cluster reflects low population, but very high event density. Most people go around these areas to attend events during the year, which might drive up the customer flow and hence increase the number of dishes served.
- ***Cluster 2 is marked in purple :*** This cluster reflects a compromise between the three features. It is situated in an area dense with population, adding the population movements coming from events/activities and the advantages of having many markets. It seems it is the best area to open a restaurant.

V. Conclusion :

We can conclude that based only on this model set (restricted model because of lack of data), we can say that the idea of investing in a restaurant is more interesting around Paris than in the center of Paris. The main reason is that population boosted by the movement generated by events and other activities is situated around Paris. Also the number of restaurants is not too saturated as it is the case in the Center of Paris, in other words the competition is too harsh in the center of Paris.