

# Paris and its near suburb

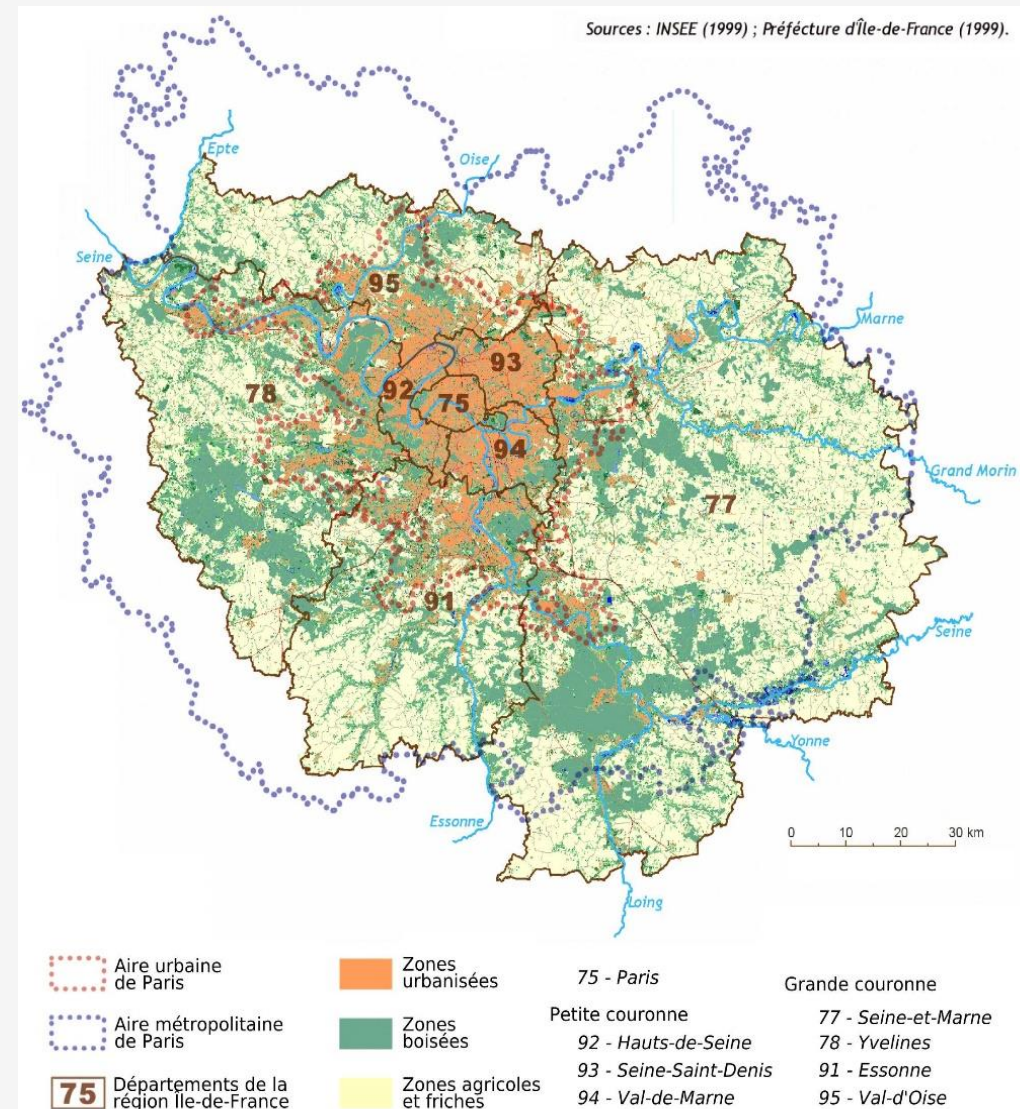
An analyze of population, level of life, and foursquare  
recommendation

# Paris and its regions

Paris is the capital of France, it is in the region name Ile de France, the region represent:

- Smallest region in France
- About 18% population of France
- About 28.8% GDP of France
- 8 departments in the regions, 4 with small surface (inner ring), and 4 with large surface(outer ring)

Name	Department Code	Surface(km2)	Demography 2017
Paris	75	105.4	2 187 526
Hauts-de-Seine	92	176	1 609 306
Seine-Saint-Denis	93	236	1 623 111
Val-de-Marne	94	245	1 387 926
Seine-et-Marne	77	5915	1 403 997
Yvelines	78	2281.4	1 438 266
Essonne	91	1804	1 228 618
Val-d'Oise	95	1246	1 165 397



# Data Source

---

- Insee: <https://www.insee.fr/fr/statistiques/2521169>

The above data source provide by national statistics institute, contain information about each city, department code, region, its population, and it life level (which is a indicator related to the household median income).

- Open data of Ile de France:  
<https://data.iledefrance.fr/explore/dataset/base-comparateur-de-territoires/>

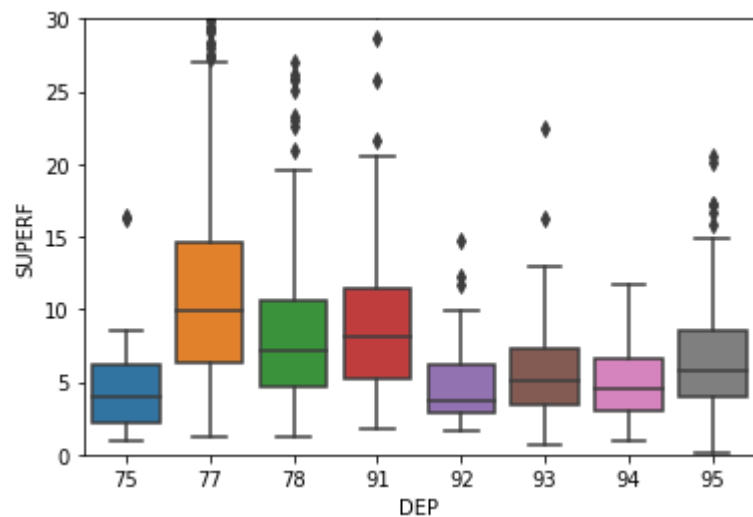
The above data is provided by reginal government, the data contains the insee data for Ile France part and geographic data (e.g GPS data)

- Foursquare Venues API explore to get the recommendation data

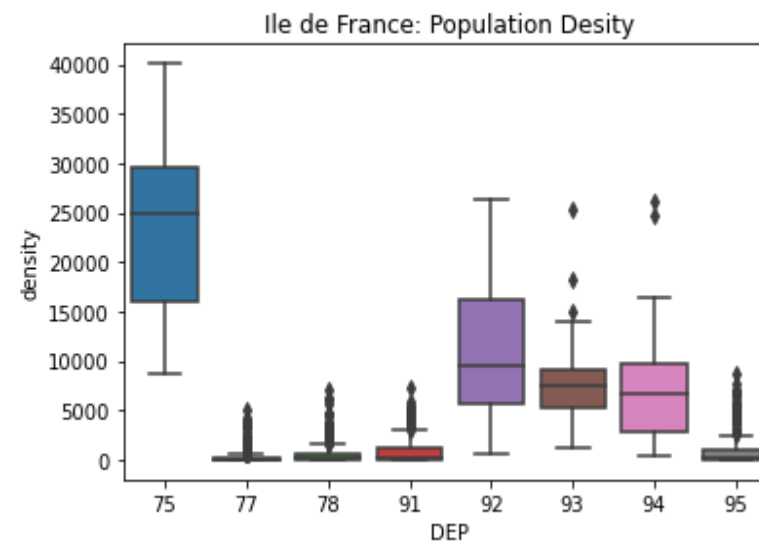
# Paris Inner Ring (petite couronne)

Paris is surrounded by 3 small department, together, they are call Inner ring (petite couronne in French)

Small cities in inner ring area (department 75, 92-94)



High population in the inner ring aera (department 75, 92-94)

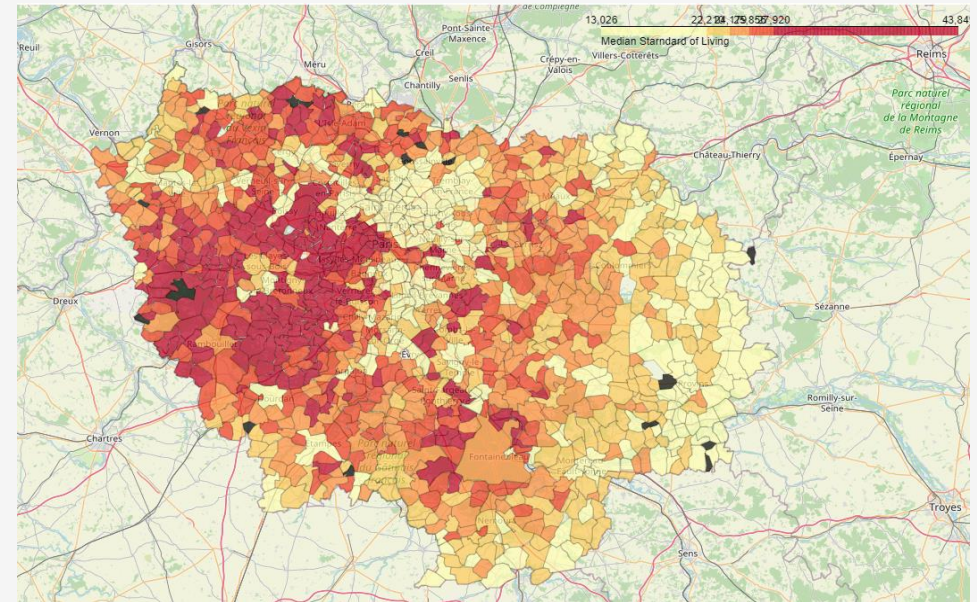
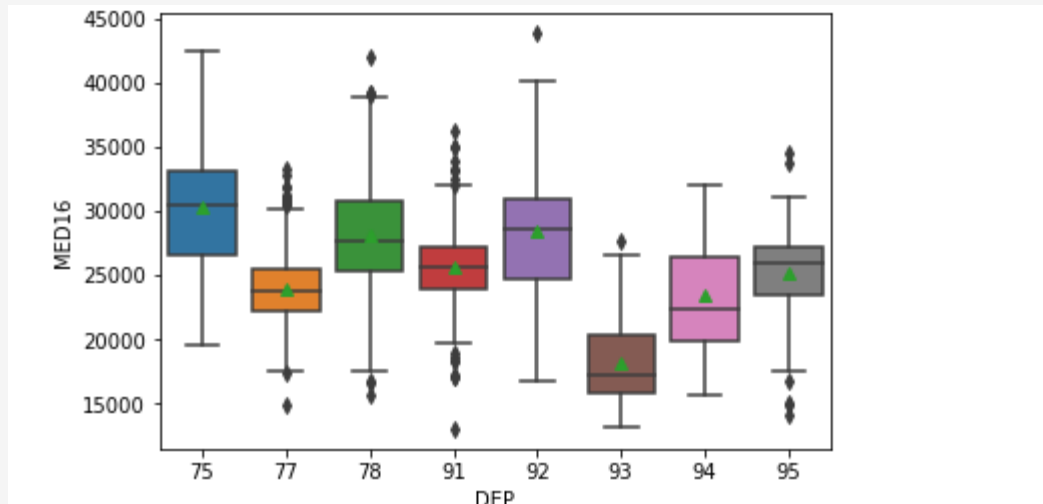


# Median stand of living(revenue indicator)

Median standard of living, it equal to median household fiscal revenue divided by household fiscal part: (generally, 1 adult adds 1 part, 1 child adds 0.5 part (age $\geq$ 14yrs) or is 0.3 part(age<14yrs)

Paris and department 92 is far away before the department 93

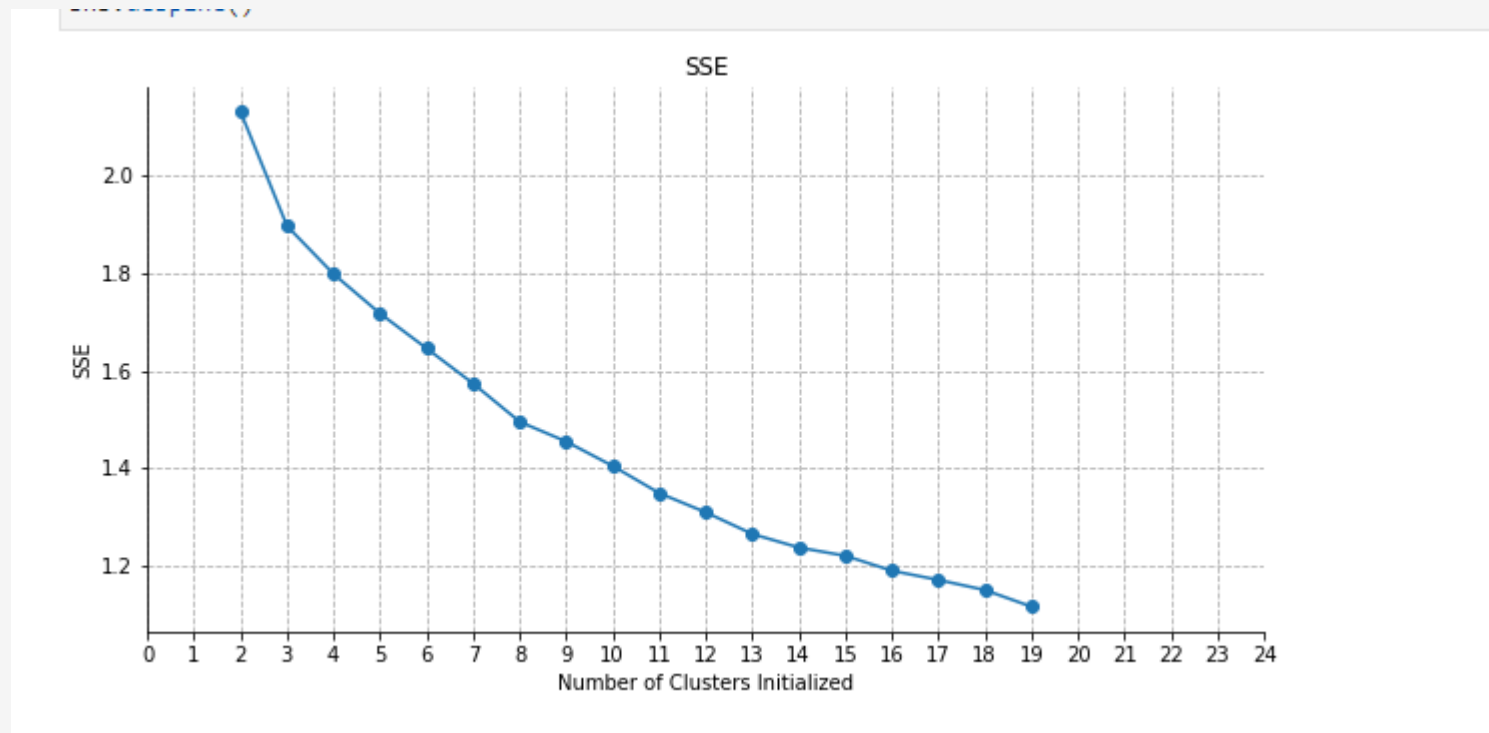
Geographically, west part the region has higher stand of living than east part



# Foursquare explore API data clustering

---

We use the elbow method to choose the number of cluster to be used in the k-means clustering, 3 looks like an appropriate value



# Clustering result

---

Cluster 0:

Top 10 venues in the cluster 0

Venue Category	
French Restaurant	0.163934
Hotel	0.047690
Italian Restaurant	0.045604
Bakery	0.039344
Japanese Restaurant	0.036066
Bar	0.025633
Plaza	0.024143
Bistro	0.022951
Park	0.021461
Supermarket	0.020566

Cluster 1:

Top 10 venues in the cluster 1

Venue Category	
Supermarket	0.118201
Fast Food Restaurant	0.089948
Train Station	0.057851
Hotel	0.055545
Shopping Mall	0.033634
French Restaurant	0.029022
Park	0.027869
Clothing Store	0.027484
Furniture / Home Store	0.023448
Bakery	0.021334

Cluster 2:

Top 10 venues in the cluster 2

Venue Category	
Supermarket	0.076766
Hotel	0.068388
French Restaurant	0.064991
Park	0.041440
Japanese Restaurant	0.035100
Bakery	0.034420
Italian Restaurant	0.032835
Sandwich Place	0.022645
Train Station	0.019928
Pizza Place	0.017663

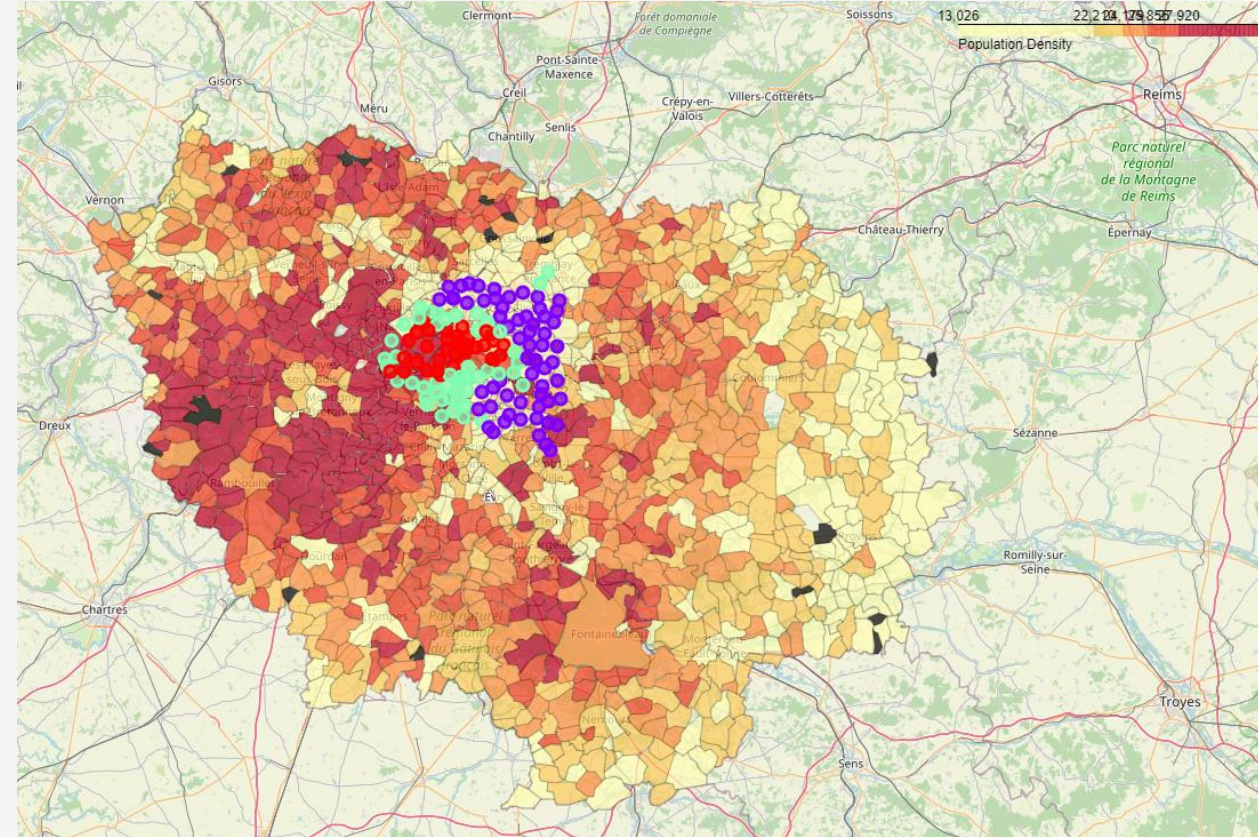


# Geolocation of the cluster

Cluster 0 (red) is mostly Paris and its west suburbs

Cluster 1 (purple) is mostly east side suburbs (but not the near suburbs)

Cluster 2 (green) is most nearest suburbs of Paris, and surround the cluster 0, there are outliers in the maps, they are cities where Paris CDG airport is situated





# Median stand of living of cluster

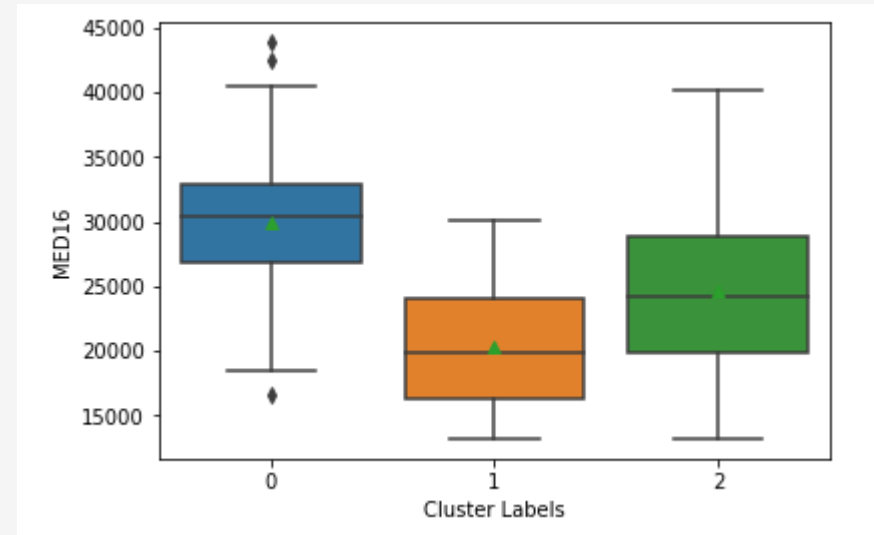
---

The box plot show that

Cluster 0 has the highest median stand of live, with some outlier

Cluster 1 has the lowest media stand of live

cluster 2 has the last range of upper and down limit (low limit is almost same as cluster 0 and up limit is almost same as cluster 0



# The outliers

For cluster 0, the outliers are:

CODGEO	LIBGEO	DEP	SUPERF	MED16
93006	Bagnolet	93	2.57	16583.000000

93048	Montreuil	93	8.92	18428.000000
-------	-----------	----	------	--------------

CODGEO	LIBGEO	DEP	SUPERF	MED16
92051	Neuilly-sur-Seine	92	3.73	43848.666667

75107	Paris 7e Arrondissement	75	4.09	42465.555556
-------	-------------------------	----	------	--------------

For cluster 2, the cities in the top 25 are :

CODGEO	LIBGEO	DEP	SUPERF	MED16
92047	Marnes-la-Coquette	92	3.48	40190.400000

92076	Vaucresson	92	3.08	38694.444444
-------	------------	----	------	--------------

92033	Garches	92	2.69	33548.571429
-------	---------	----	------	--------------

92071	Sceaux	92	3.60	33503.000000
-------	--------	----	------	--------------

92014	Bourg-la-Reine	92	1.86	31038.666667
-------	----------------	----	------	--------------

92044	Levallois-Perret	92	2.41	30930.666667
-------	------------------	----	------	--------------

94052	Nogent-sur-Marne	94	2.80	30241.666667
-------	------------------	----	------	--------------

92063	Rueil-Malmaison	92	14.70	30197.142857
-------	-----------------	----	-------	--------------

# Conclusion and future directions

---

- Cluster 0 is high median stand of living cities
- Cluster 1 is low median stand of living cities
- Cluster 2 has the most variable median stand of living
- Inner ring area is a highly centralized area, there is a disparities between its west part and its east part

We can use the distance to Paris as a feature for further model