Coursera Capstone : Bakery in Paris region

Business problem

- Find an optimal location for a new bakery in Paris Region
- Paris suburb is full of existing bakeries.
- Other venues can sell bread and pastries, like supermarkets,…

- Who is interested ?
  - Investors, to invest on a right place
  - Future manager of the bakery, to have good sales.
  - Public managers (mayor), to convince investor to open a new bakery in their town

Methodology

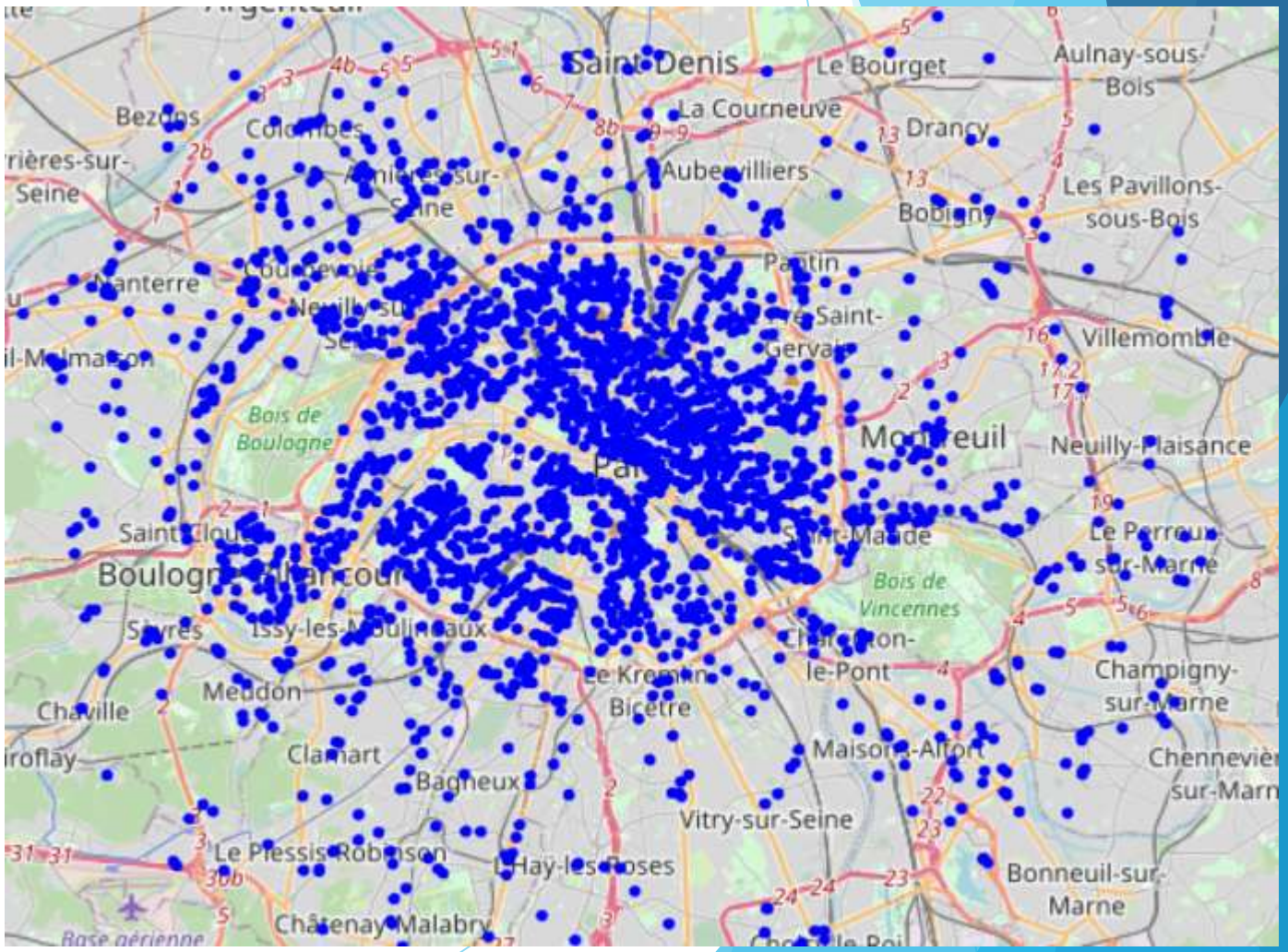
- Find all existing bakeries in Paris Region
  - Using Foursquare API
  - Filtering by « bakery » venue
  - Foursquare answer is 50 venues maximum, so create a grid of location and interrogate foursquare on each location.
- Analysing by visualizing all bakeries
  - Using folium and heatmap
  - Define a promising region
- Filter on every location of the promising region
  - Define a point every 300 m
  - Calculate distance if the nearest bakery for each point.
  - Filter the points with nearest bakery > 500m
- Cluster all resulting points to find candidates locations.
  - Using Scikit Learn and k-means clustering

Data – Foursquare interrogation

- Creating a grid of points within 15km form paris city centre, separated by 1000 m
- Interrogating Foursquare about bakeries within 1000m of each point.
- Adding all results in a dictionnary to avoid duplicates.

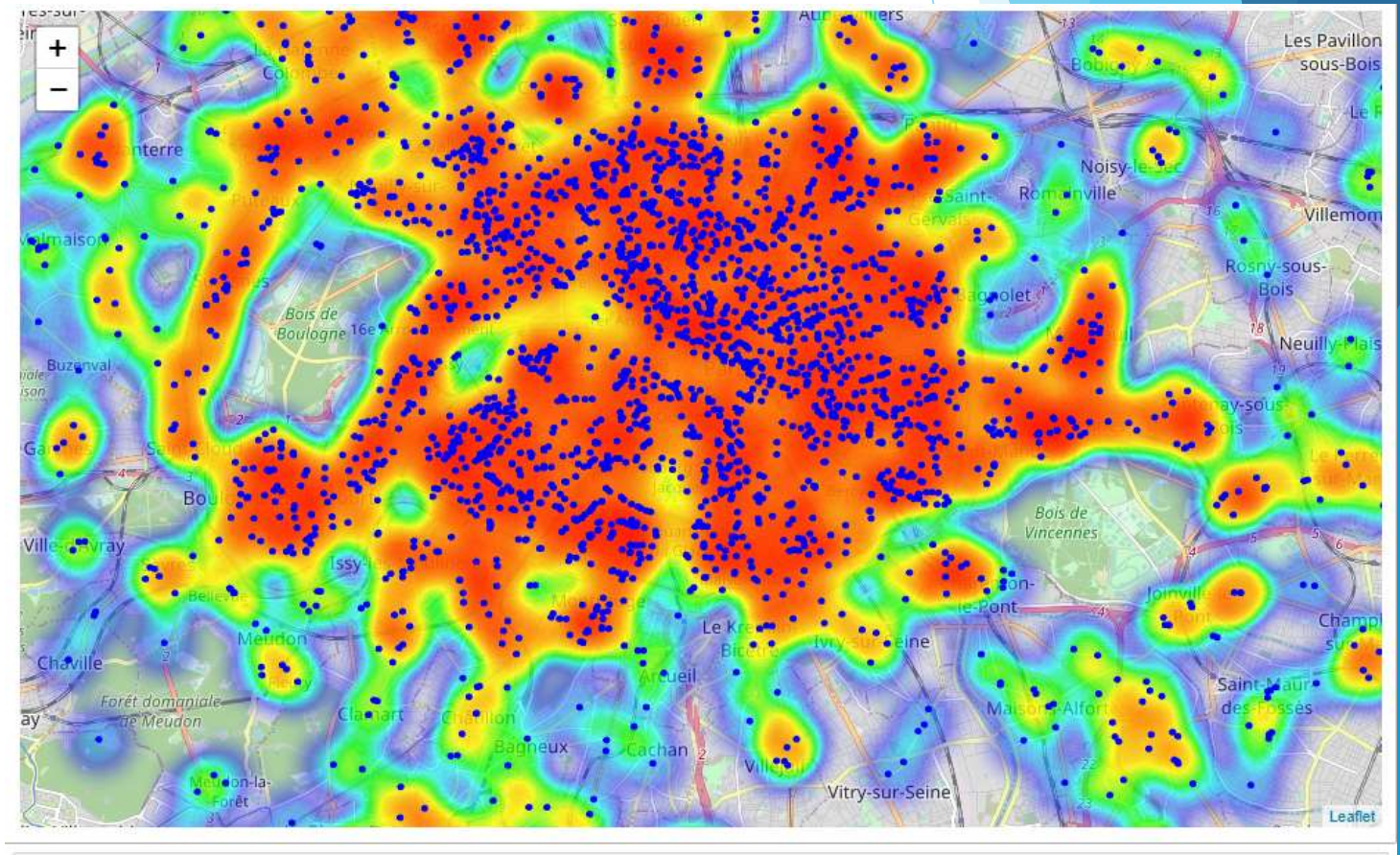
Data – Foursquare bakeries

- Venues that are bakeries or bakeries-like.
- 2 821 venues in the area



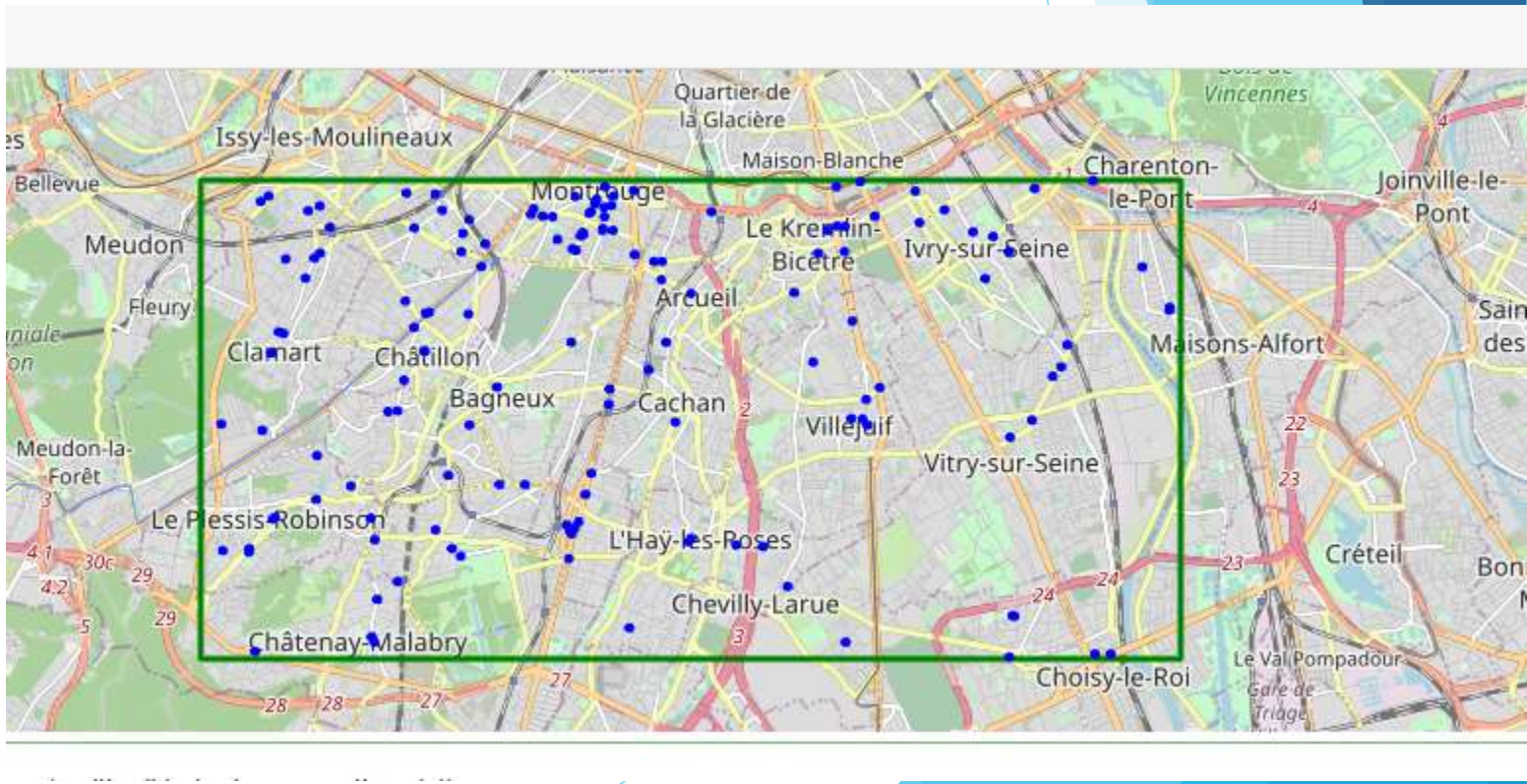
Analysing – bakeries locations

- Heatmap shows high density and low density bakeries neighborhoods in Paris and suburb
- Most of bakeries are in Paris City
- High density is in the west suburb, east suburb and south-west (Boulogne)
- South suburb of Paris seems to be a low density zone.



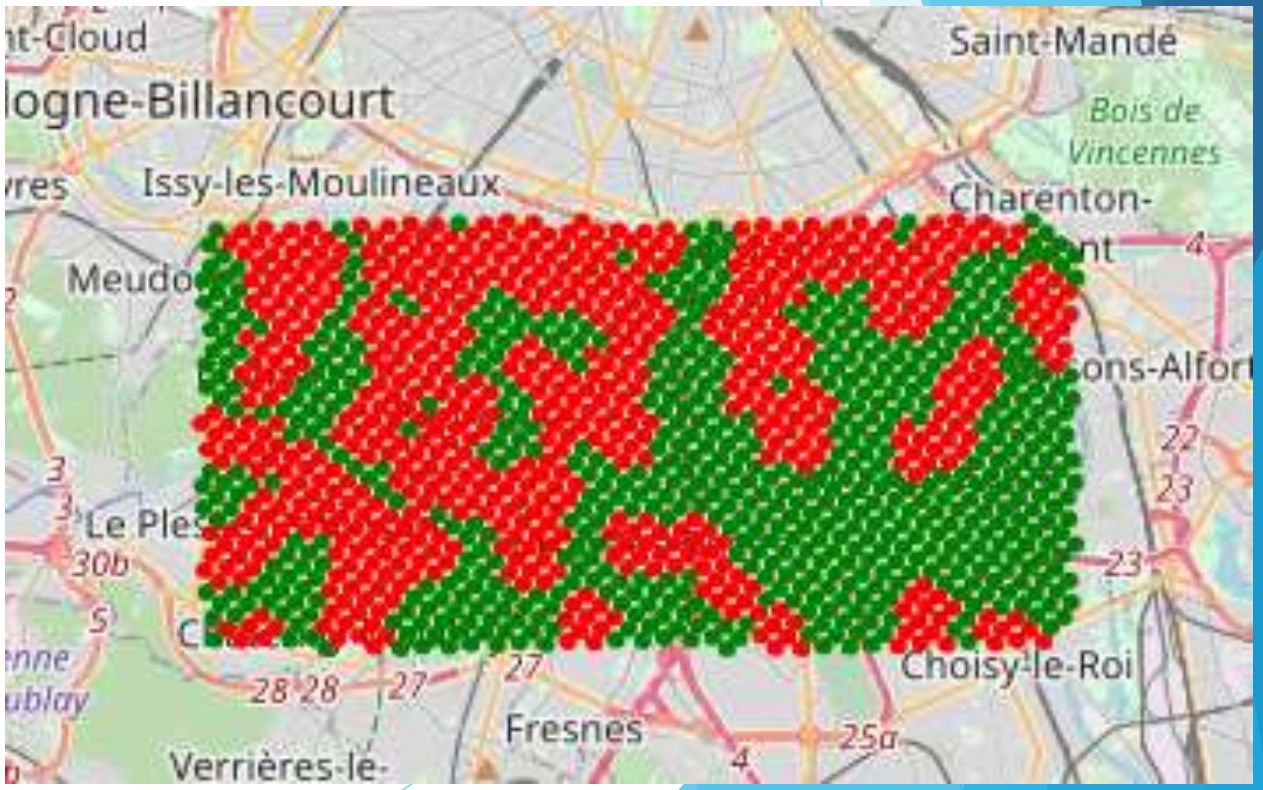
Analysing – restricting to the south suburb

- Restricting to the south suburb.
- Low density of bakeries.
- High density of population



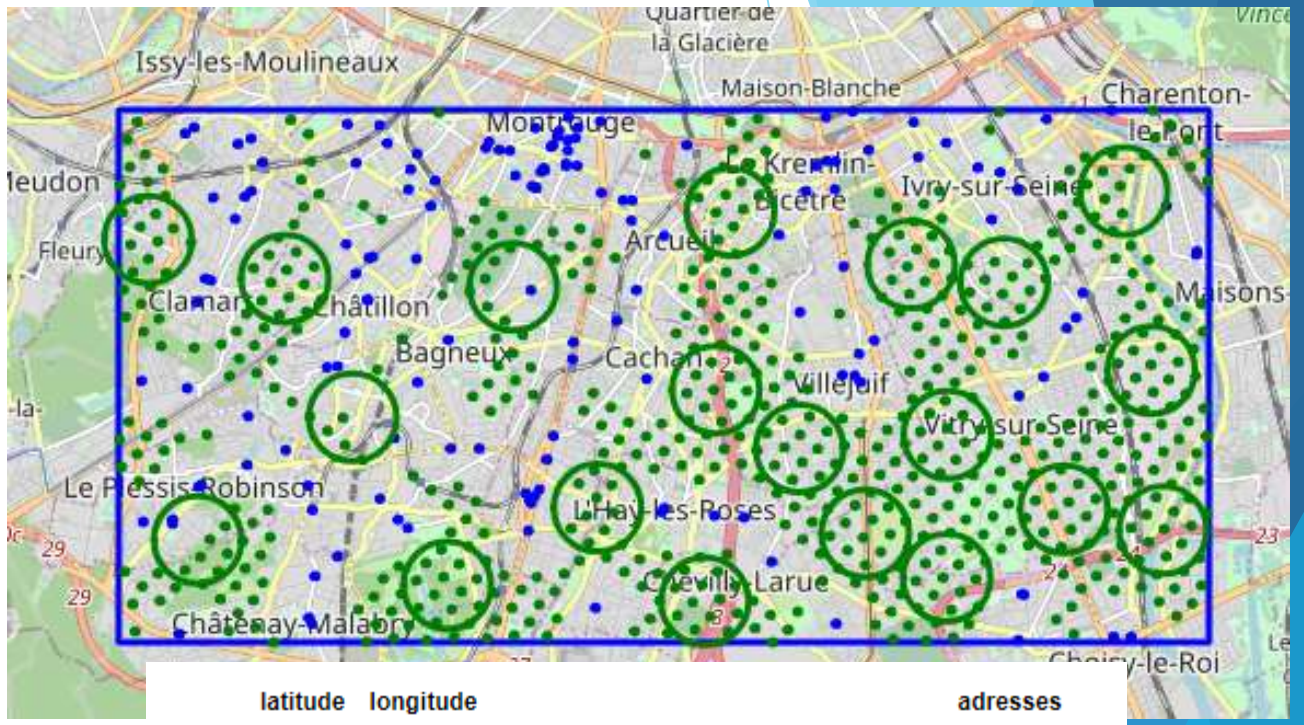
Analysing – interesting points.

- Within the region
- Create points each 300 m
- Calculate nearest bakery for each point
- Filter the points with no bakery within the 500m
  - Green points : nearest bakery >= 500m
  - Red points : nearest bakery < 500m



Analysing – clustering points.

- Points with no bakery within 500m
- K-means clustering (scikit learn)
- 20 clusters defined.
- Adress calculated for each cluster
  - Google geocode API



	latitude	longitude	adresses
0	48.791115	2.343770	148 Rue Gabriel Péri, 94230 L'Haÿ-les-Roses, F...
1	48.802196	2.388969	16 Rue Jean-Baptiste Renoult, 94200 Ivry-sur-S...
2	48.806606	2.254747	43 Avenue Adolphe Schneider, 92140 Clamart, Fr...
3	48.776600	2.413875	Seine Bridge, A86, 94600 Choisy-le-Roi, France
4	48.788246	2.286838	14 Rue André Neyts, 92260 Fontenay-aux-Roses, ...
5	48.771688	2.379975	14 Rue Jean Mermoz, 94320 Thiais, France
6	48.809479	2.346087	30 Rue Labourse, 94250 Gentilly, France
7	48.801727	2.311990	29 Rue de Verdun, 92220 Bagneux, France
8	48.775592	2.262536	47 Rue du Moulin Fidel, 92350 Le Plessis-Robin...

Conclusion

- 20 adresses obtained, with fulfilled criteria
  - In south suburb
  - With no bakery within 500m
- Theses addresses have to be further analysed with other criteria:
  - Attractiveness, proximity of roads or railway station;
  - Population