Opening a Traditional French Bakery in Paris

Opening a new traditional Bakery in Paris

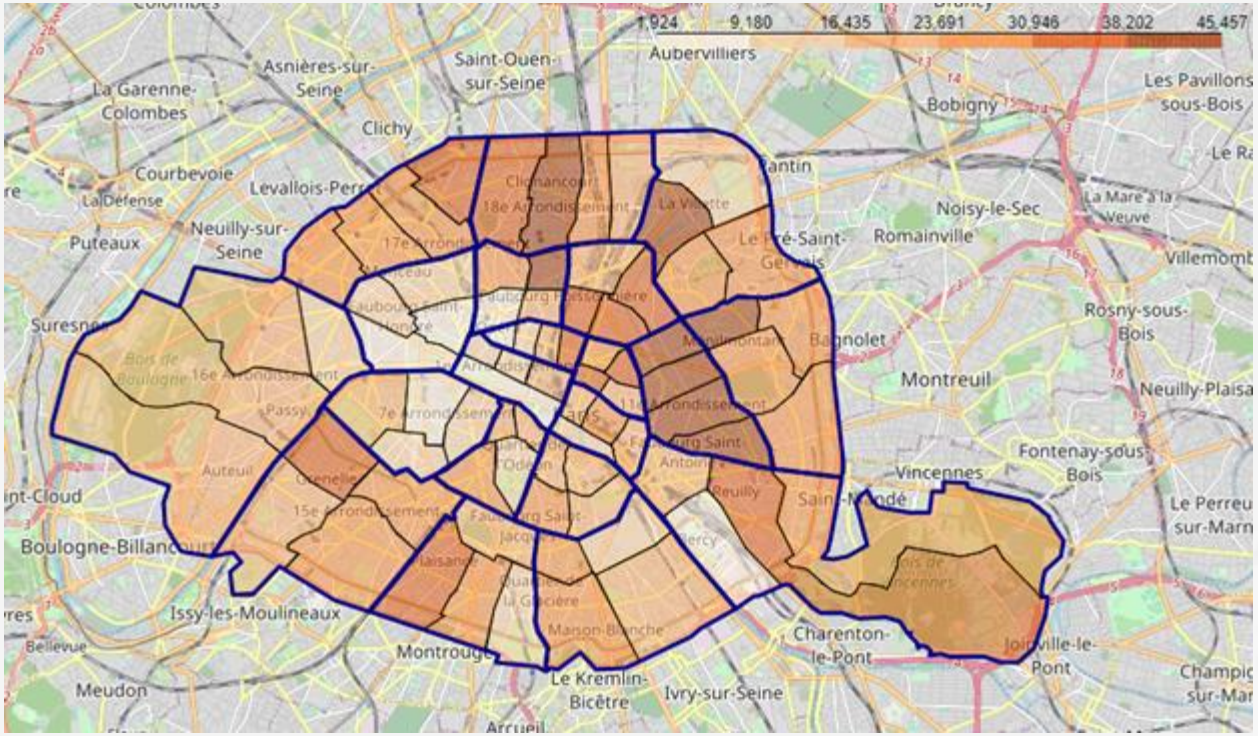
- French people love bread and croissants, but Paris is already full of bakeries of all kinds
- **How to find the most suitable area to open a new high-standing, traditional bakery?**
- Among all possible criteria we choose 3:
  - Density of population: people go to the bakery by foot, so we look for a populous area
  - Number of bakeries per inhabitant: with less competitors, the market share is higher
  - Distance to a “top” bakery: is there a top-tiered competitor in the area?

Data Sources

- Free data easily available on the internet:
  - **Population density** ⇨ Wikipedia (Paris administrative districts)
  - **Paris districts shape** ⇨ opendata.paris.fr
  - **Existing bakeries** (names, geo-localization, rating) ⇨ Foursquare API

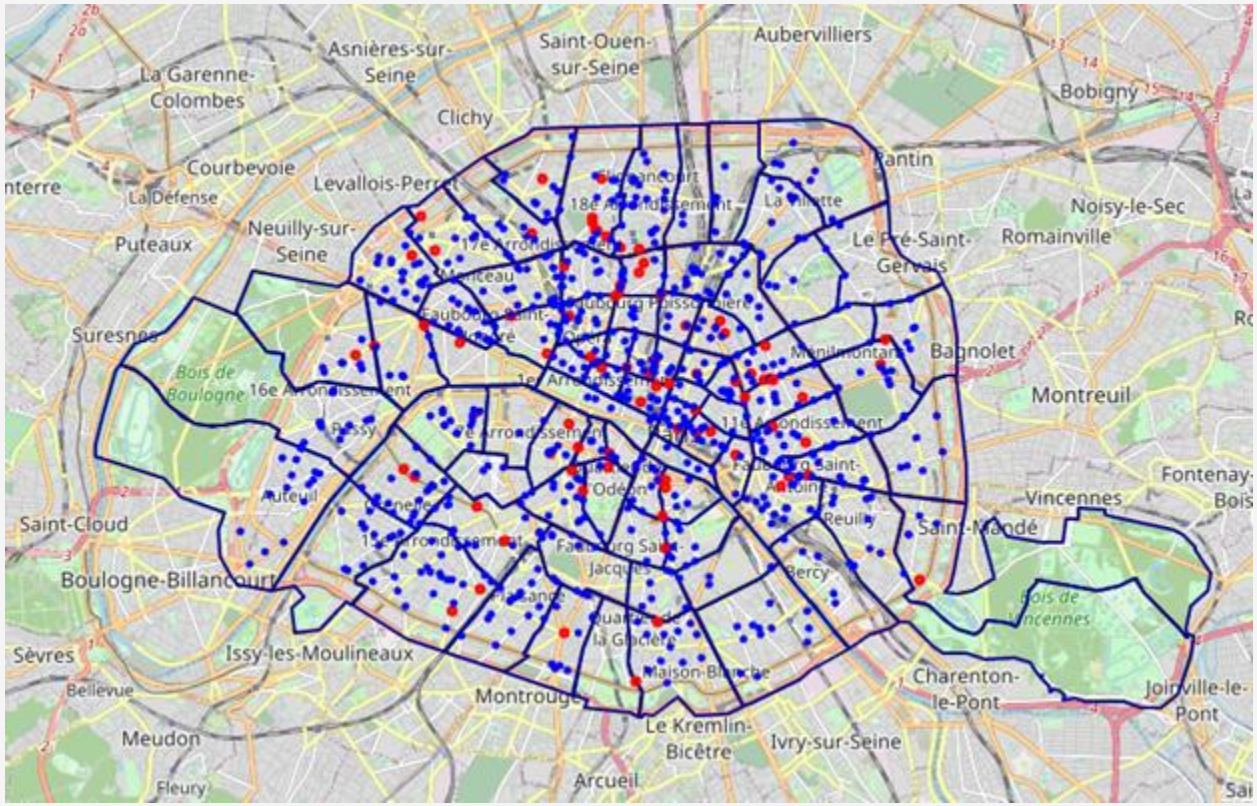
Data Acquisition 1/2

- Paris districts shapes and population density:
  - Paris is divided into 80 administrative districts

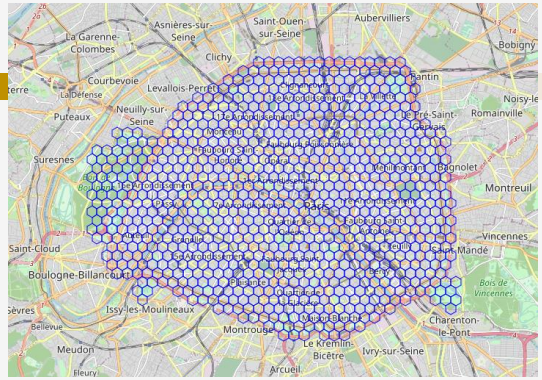


Data Acquisition 2/2

- Existing bakeries (in red top bakeries with rating ≥ 8.0)
  - 804 bakeries, o/w 82 “top”

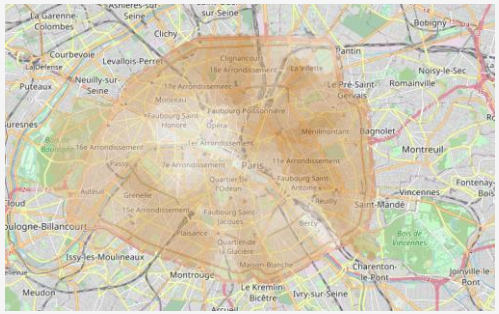


Methodology

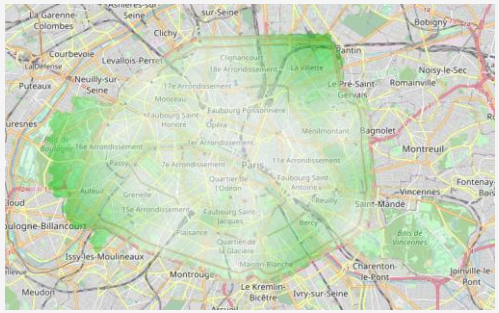


- Split Paris into small hexagonal shapes ⇨ 898 shapes defined

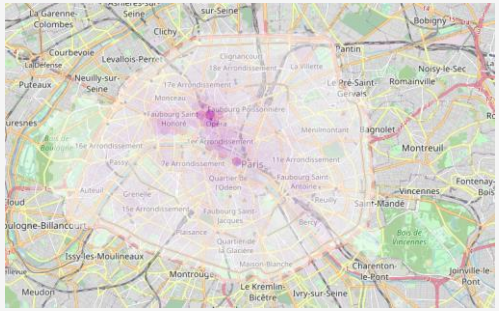
- Estimate the 3 metrics for each shape



density



distance to “top” bakery



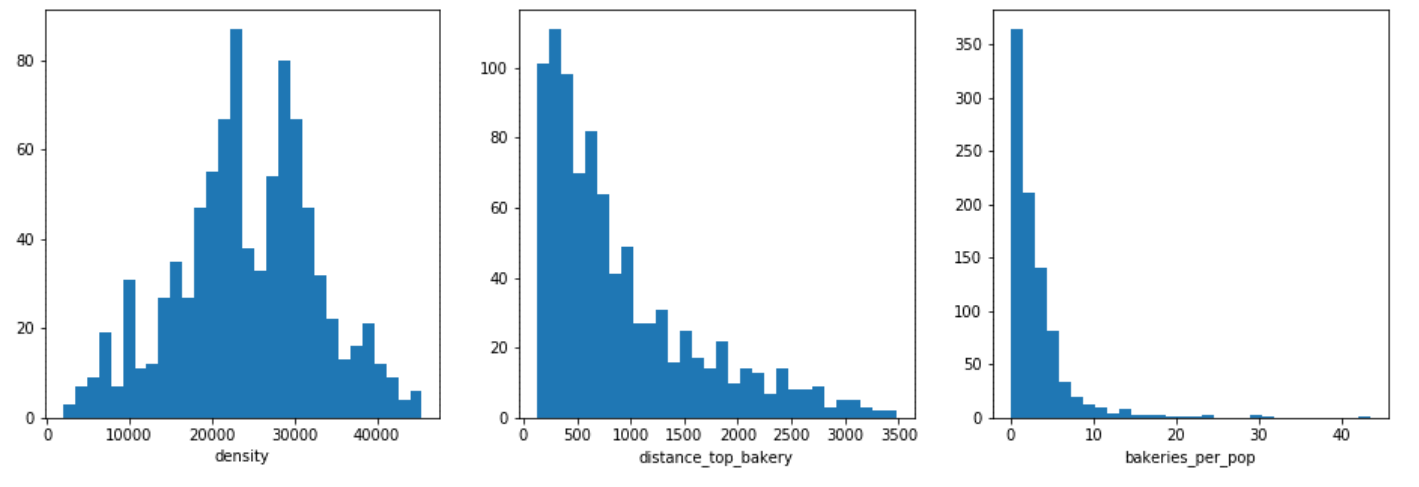
bakeries per inhabitant

- Apply Clustering and Scoring

Analysis – data distribution

- Distribution of the 3 metrics over the hexagonal shapes

	density	distance_top_bakery	bakeries_per_pop
count	898.000000	898.000000	898.000000
mean	24261.788885	910.327130	2.948793
std	8521.976380	720.490596	3.781869
min	1924.000000	125.727683	0.000000
25%	18816.000000	363.176230	0.747819
50%	23983.285714	661.965405	1.961644
75%	29984.464286	1248.814720	3.868790
max	45457.000000	3474.529448	43.496302



Clustering 1/2

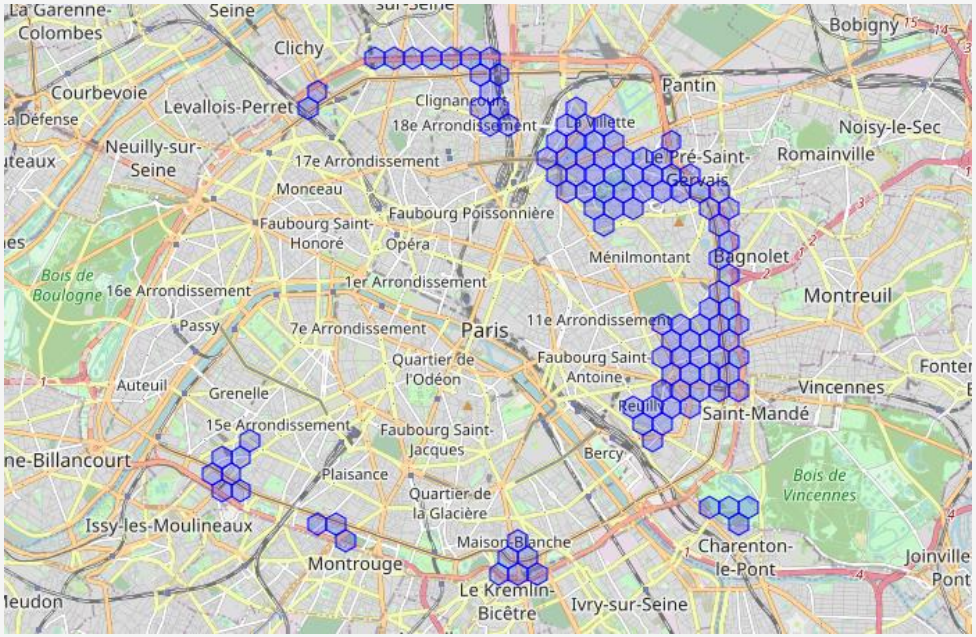
- Apply K-Means algorithm with 10 clusters:

clusters	x	y	district	bakeries_inside	bakeries_around	density	distance_top_bakery	bakeries_per_pop
0	27.442105	15.652632	59.905263	0.357895	2.315789	21317.438471	1830.262240	1.029695
1	33.658537	20.926829	19.463415	1.902439	13.951220	11432.466899	308.167401	11.853174
2	55.178571	19.678571	67.410714	0.339286	2.723214	31454.471747	1277.111190	0.815633
3	38.808989	17.393258	31.303371	2.235955	14.000000	21797.162119	347.242952	6.132117
4	32.341270	14.134921	50.809524	0.452381	3.507937	20533.763190	844.895242	1.642722
5	36.000000	21.000000	21.250000	2.500000	17.375000	5829.910714	332.808815	28.040943
6	26.901639	21.262295	66.491803	0.147541	1.131148	17694.702966	2724.606822	0.763300
7	49.411765	22.847059	52.082353	1.588235	10.764706	38361.710924	390.393212	2.718413
8	36.200000	15.373333	36.960000	0.386667	4.186667	11451.390476	655.932549	3.432419
9	37.985437	16.169903	56.878641	1.004854	6.762136	29113.982108	482.156598	2.236587

- Cluster #2 scores well on the 3 criteria

Clustering 2/2

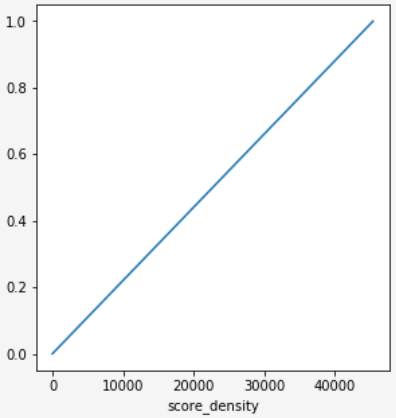
- Localization of the areas belonging to Cluster #2:

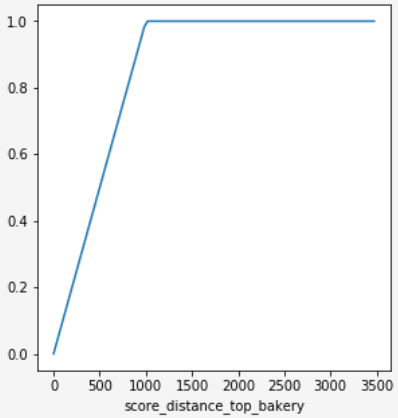


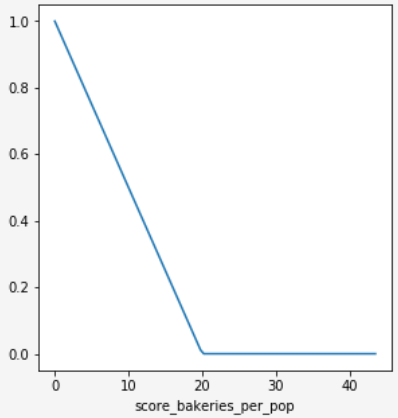
- We need scoring to narrow down list of optimal places

Scoring

- Score hexagons using scoring functions per feature





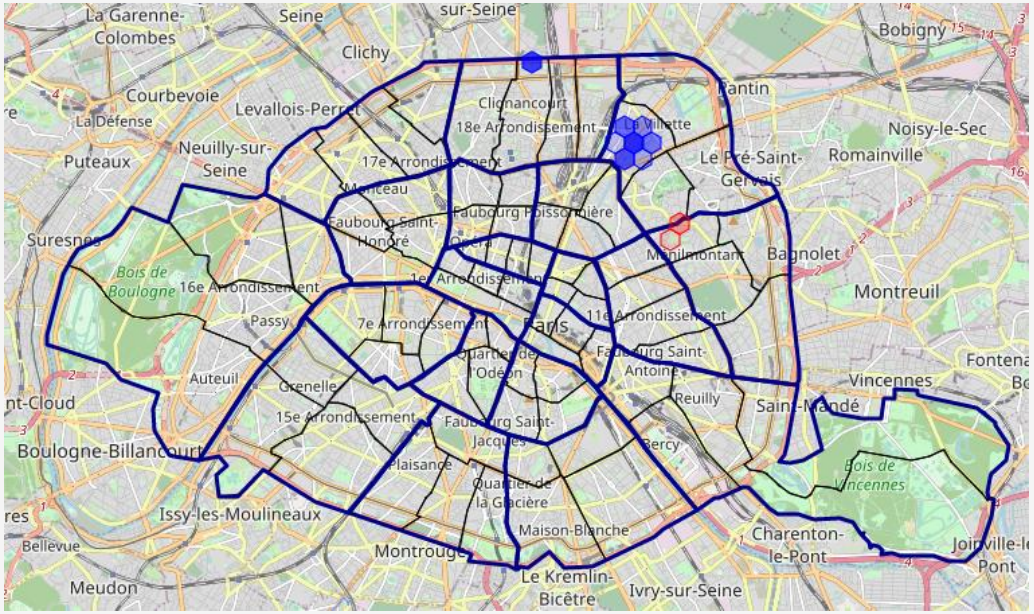


- Results: top-10

x	y	density	distance_top_bakery	bakeries_per_pop	clusters	score_density	score_distance_top_bakery	score_bakeries_per_pop	total_score
54	28	41718.000000	1824.576153	0.922624	2	0.917746	1.000000	0.953869	2.871615
43	33	39243.000000	1202.971214	0.245203	2	0.863299	1.000000	0.987740	2.851039
53	27	41718.000000	1479.405036	1.383936	2	0.917746	1.000000	0.930803	2.848550
53	29	41718.000000	2017.519234	1.383936	2	0.917746	1.000000	0.930803	2.848550
56	28	41718.000000	2009.207142	1.383936	2	0.917746	1.000000	0.930803	2.848550
55	29	41718.000000	2170.098599	1.614592	2	0.917746	1.000000	0.919270	2.837017
59	23	40264.857143	950.043531	0.477960	7	0.885779	0.950044	0.976102	2.811925
55	27	38400.571429	1668.152062	0.751747	2	0.844767	1.000000	0.962413	2.807180
52	28	38305.714286	1679.872141	0.753609	2	0.842680	1.000000	0.962320	2.805000
58	22	44328.000000	849.940465	0.651225	7	0.975163	0.849940	0.967439	2.792543

Results

- Top-10 hexagons on the map:



- Best area = **19th borough, district of La Villette**

Discussion - Conclusion

- Study performed using only 3 features and free data

- In real life, many more criteria should be considered:
  - ***Specificities of the neighborhood***: pedestrian area, marketplace, landmarks, subway etc.
  - ***Type of local population***: more traditional French or immigrants? How likely they will but bakery products?
  - ***Economics***: price to buy / rent a shop? Income level of local population?

- However, the methodology could be applied:
  - In this scenario, using more metrics and more accurate data
  - In in any other situation where a competitive business wants to settle down in a big city