Opening a Local Food Delivery Company in Paris, France

1. Introduction

The city of Paris is an international city which has a lot to offer to its residents and visitors alike, its rich cultural background and diverse community make it one of best places to live in. It is the most populous city of France, with a population of 2,148,271 residents. Paris is the center and seat of government of the Île-de-France, or Paris Region, which has an estimated official 2020 population of 12,278,210, or about 18 percent of the population of France. The economy of the City of Paris is based largely on services and commerce; of the 390,480 enterprises in the city, **80.6** percent are engaged in commerce, transportation, and diverse services, 6.5 percent in construction, and just 3.8 percent in industry.

The city of Paris is divided into twenty administrative districts, more simply referred to as **"arrondissements".** The arrondissements are arranged in the form of a clockwise, starting from the middle of the city, with the first on the Right Bank (north bank) of the Seine. The number of the arrondissement is indicated by the last two digits in Parisian postal codes (75001 up to 75020). The figure below shows the distribution of the Paris's arrondissements.

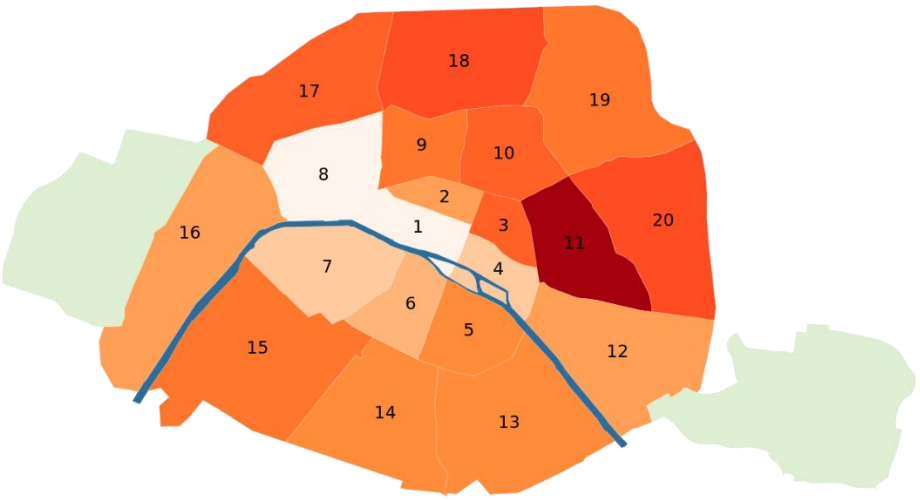


Figure 1: The distribution of Paris's administrative districts[1]

[1] By Paris 16 - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=38412578

- **Business Problem**

A local food delivery company is about to open in the city, it is a freshly created company with a limited startup budget. In its early years of creation, the company's motivation is to partner up with the **biggest number of restaurants and supermarkets** while offering the top-notch customer service to their clients. Because of its limited resources, this company cannot cover all the twenty districts of Paris, this will require more capital and investment that the company is willing allocate in the long term. Hence, their business strategy for the next five years is to focus on only **two districts**- ideally next to each to other- and then expand from there once they become financially ready.

This company wants to know which are the two best districts among the twenty that will maximize their revenue for the first year while minimizing their operating costs.

These districts must have:

- A minimum number of restaurants and supermarkets.
- Plenty of offices, residential areas… mostly urban.

2. Data Description

The information about all the districts of Paris was found on this website : https://fr.geneawiki.com/index.php/Liste_des_quartiers_de_Paris, it has a table containing relevant information about the districts names and postal codes along with the names of the neighborhoods that each district has. Here is a sneak peak of the table mentioned above.

Code INSEE [1]	Code Postal	Arrondissements		Quartiers
75101	75001	I	Le Louvre	01 - Saint-Germain-l'Auxerrois 02 - Les Halles 03 - Le Palais-Royal 04 - La Place-Vendôme
75102	75002	II	La Bourse	05 - Gaillon 06 - Vivienne 07 - Le Mail 08 - Bonne-Nouvelle
75103	75003	III	Le Temple	09 - Les Arts-et-Métiers 10 - Les Enfants-Rouges 11 - Les Archives 12 - Saint-Avoye
75104	75004	IV	L'Hôtel-de-Ville	13 - Saint-Merri 14 - Saint-Gervais 15 - L'Arsenal 16 - Notre-Dame

This data was combined with geospatial data from Google Maps. I used Google Maps because the number of districts is not that big, 20 districts in total. The search was relatively easy in comparison with API calls to Google Places. You can find the csv file containing the geospatial data (latitudes, longitudes, and postal codes) in the Github repository assigned to this project. A column was added to file which is the population density in ha/km² in each district. This information is going to be very important for our analysis later in this project.

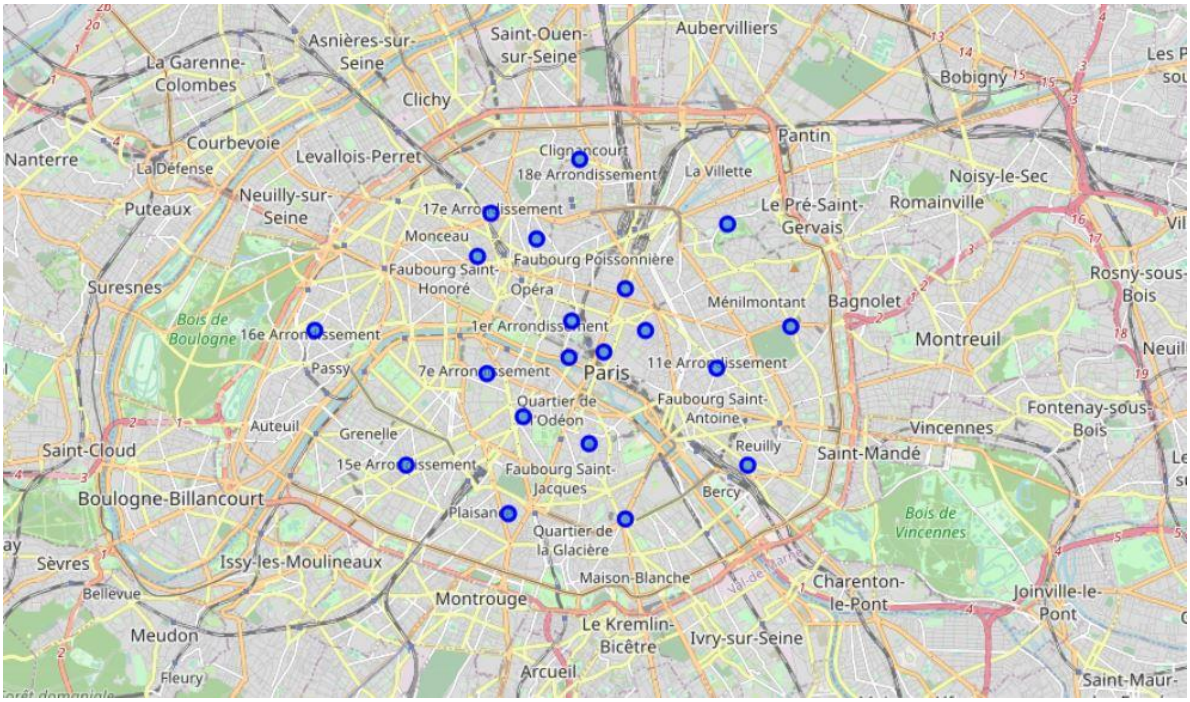
The location data that will be used in this analysis is the *Foursquare* data. We will manipulate the huge amount of data that Foursquare has about the venues. We will be using API calls to their server to identify the location of venues that best match the company's objectives and business plan.

3. Methodology

In the first section of the analysis, I imported the data from the website mentioned in the previous section, the goal was to clean the data and make it easy to identify and manipulate. Then I merged it with the data from the csv file that has the geospatial coordinates of the districts to get a final complete dataframe ready for analysis. You can find the code used for this step in the Github repository assigned to this project. Here is how it looks like:

	Postal Code	District	Neighborhoods	Latitude	Longitude	Population Density ha/km²
0	75001	Le Louvre	01 - Saint-Germain-l'Auxerrois 02 - Les Halles...	48.8592	2.3417	9 041
1	75002	La Bourse	05 - Gaillon 06 - Vivienne 07 - Le Mail 08 - B...	48.8655	2.3426	21 006
2	75003	Le Temple	09 - Les Arts-et-Métiers 10 - Les Enfants-Roug...	48.8637	2.3615	29 956
3	75004	L'Hôtel-de-Ville	13 - Saint-Merri 14 - Saint-Gervais 15 - L'Ars...	48.8601	2.3507	16 966
4	75005	Le Panthéon	17 - Saint-Victor 18 - Le Jardin-des-Plantes 1...	48.8448	2.3471	23 359

This dataframe was used to produce a map describing the locations of all the districts of Paris with the help of the Folium library. The map looks like this, where the blue dots represent the centers of each district in Paris.



The Foursquare location data was then manipulated to explore the venues in each districts of Paris. Because of the small startup budget of the company, a limit of **50 venues** per each district was set and a radius of **500 m** (1 km in diameter from end to end) was chosen to account for the max distance that an agent can travel and at the same time ensure the fastest delivery for the customer. At the outset of the API calls, **937 venues** were extracted and **162 unique categories.** The districts are then grouped by taking the mean frequency of occurrence of each category. Using this information, I created a dataframe that has the 10 most common venues for each district. Here is how it looks like:

	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	L'Enclos-Saint-Laurent	Cocktail Bar	Hotel	French Restaurant	Seafood Restaurant	Bar	Breakfast Spot	Pizza Place	Bakery	Coffee Shop	Bistro
1	L'Hôtel-de-Ville	French Restaurant	Art Gallery	Ice Cream Shop	Bakery	Burger Joint	Pub	Bar	Plaza	Gay Bar	Park
2	L'Observatoire	French Restaurant	Hotel	Bar	Bakery	Italian Restaurant	Bistro	Garden	Bookstore	Brasserie	Southwestern French Restaurant
3	L'Opéra	French Restaurant	Hotel	Cocktail Bar	Italian Restaurant	Wine Bar	Lounge	Bistro	Café	Museum	Juice Bar
4	L'Élysée	Hotel	Italian Restaurant	Café	French Restaurant	Bar	Thai Restaurant	Theater	Bakery	Wine Shop	Modern European Restaurant

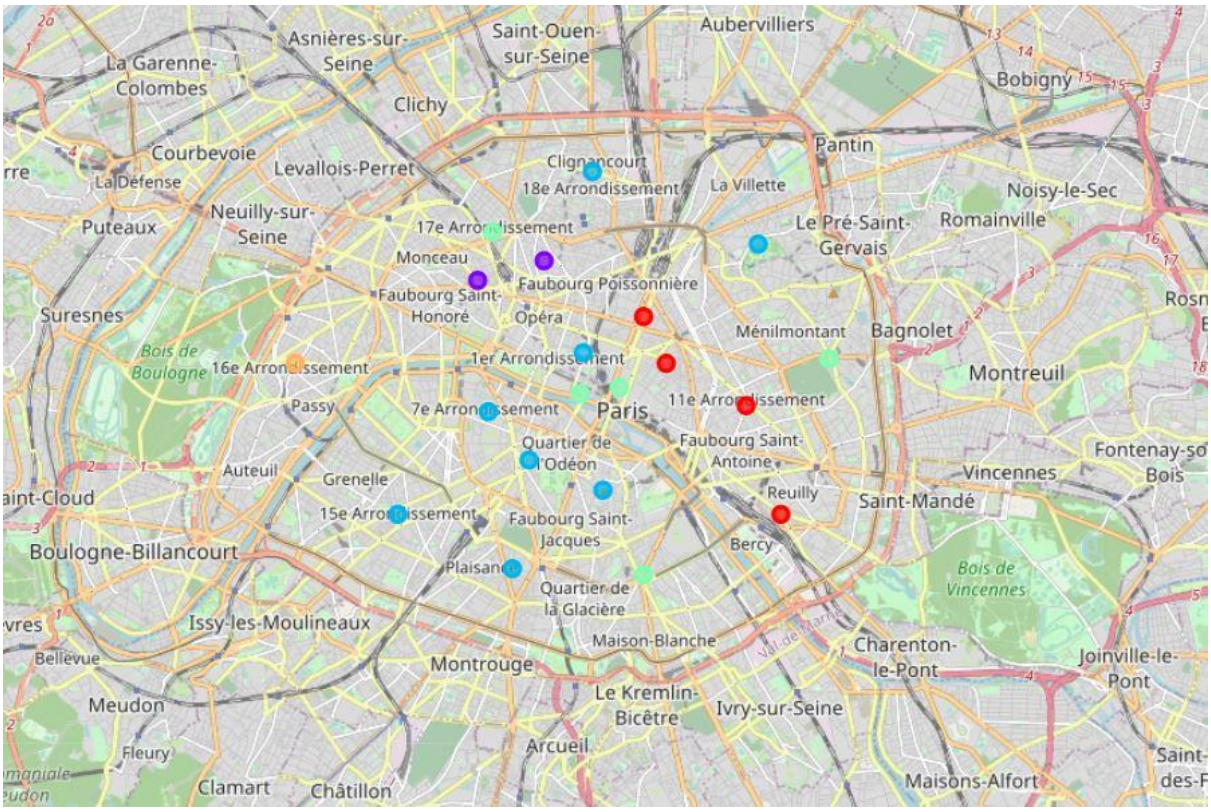
The machine learning algorithm that will be used for this project is KMeans, it is a clustering algorithm which belongs to the family of unsupervised learning algorithms. It learns by itself the labels to give to the data it has. This algorithm is chosen because the goal is to find out which are the districts that form the same cluster. In other words, the districts that look like

each other in terms of the venues they share. This information is very important to us because it will help us solve the problem by identifying the districts that offer the best business opportunities for our food delivery company.

4. Results

The results of the KMeans analysis were very promising. The algorithm produced 5 clusters labeled from 0 to 4. These cluster labels were added to the original dataframe and then were put on a folium map with a color distinguishing each cluster. The results are shown in the figures below.

	Postal Code	District	Neighborhoods	Latitude	Longitude	Population Density ha/km²	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
0	75001	Le Louvre	01 - Saint-Germain-l'Auxerrois 02 - Les Halles...	48.8592	2.3417	9 041	3	Plaza	Hotel	French Restaurant	Bar	Tea Room	Theater	Wine Bar
1	75002	La Bourse	05 - Gaillon 06 - Vivienne 07 - Le Mail 08 - B...	48.8655	2.3426	21 006	2	French Restaurant	Hotel	Bistro	Italian Restaurant	Restaurant	Spa	Clothing Store
2	75003	Le Temple	09 - Les Arts-et-Métiers 10 - Les Enfants-Roug...	48.8637	2.3615	29 956	0	Japanese Restaurant	Italian Restaurant	Vietnamese Restaurant	Cocktail Bar	Sandwich Place	Art Gallery	Burger Joint
3	75004	L'Hôtel-de-Ville	13 - Saint-Merri 14 - Saint-Gervais 15 - L'Ars...	48.8601	2.3507	16 966	3	French Restaurant	Art Gallery	Ice Cream Shop	Bakery	Burger Joint	Pub	Bar
4	75005	Le Panthéon	17 - Saint-Victor 18 - Le Jardin-des-Plantes 1...	48.8448	2.3471	23 359	2	French Restaurant	Italian Restaurant	Pub	Wine Bar	Coffee Shop	Plaza	Bakery



After examining the contents of the five clusters, I found the following results:

- **Cluster 1**

	District	Population Density ha/km²	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Le Temple	29 956	0	Japanese Restaurant	Italian Restaurant	Vietnamese Restaurant	Cocktail Bar	Sandwich Place	Art Gallery	Burger Joint	Bistro	Café	Bookstore
9	L'Enclos-Saint-Laurent	31 754	0	Cocktail Bar	Hotel	French Restaurant	Seafood Restaurant	Bar	Breakfast Spot	Pizza Place	Bakery	Coffee Shop	Bistro
10	Popincourt	40 827	0	French Restaurant	Cocktail Bar	Italian Restaurant	Bistro	Café	Wine Bar	Pastry Shop	Bar	Gluten-free Restaurant	Pizza Place
11	Reuilly	22 345	0	Hotel	Bistro	Supermarket	Chinese Restaurant	French Restaurant	Japanese Restaurant	Bar	Cheese Shop	Middle Eastern Restaurant	Soup Place

- **Cluster 2**

	District	Population Density ha/km²	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
7	L'Élysée	9 457	1	Hotel	Italian Restaurant	Café	French Restaurant	Bar	Thai Restaurant	Theater	Bakery	Wine Shop	Modern European Restaurant
8	L'Opéra	27 251	1	French Restaurant	Hotel	Cocktail Bar	Italian Restaurant	Wine Bar	Lounge	Bistro	Café	Museum	Juice Bar

- **Cluster 3**

	District	Population Density ha/km²	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	La Bourse	21 006	2	French Restaurant	Hotel	Bistro	Italian Restaurant	Restaurant	Spa	Clothing Store	Bakery	Pizza Place	Plaza
4	Le Panthéon	23 359	2	French Restaurant	Italian Restaurant	Pub	Wine Bar	Coffee Shop	Plaza	Bakery	Bar	Creperie	Hotel
5	Le Luxembourg	19 734	2	French Restaurant	Bakery	Hotel	Wine Bar	Art Gallery	Chocolate Shop	Dessert Shop	Cocktail Bar	Coffee Shop	Cupcake Shop
6	Le Palais-Bourbon	13 235	2	French Restaurant	Hotel	Bistro	Italian Restaurant	Coffee Shop	Art Museum	Park	Chocolate Shop	Bakery	Pizza Place
13	L'Observatoire	24 821	2	French Restaurant	Hotel	Bar	Bakery	Italian Restaurant	Bistro	Garden	Bookstore	Brasserie	Southwestern French Restaurant
14	Vaugirard	27 712	2	French Restaurant	Hotel	Italian Restaurant	Coffee Shop	Japanese Restaurant	Supermarket	Korean Restaurant	Lebanese Restaurant	Bar	Ethiopian Restaurant
17	La Butte-Montmartre	32 875	2	French Restaurant	Bar	Pizza Place	Café	Wine Bar	Italian Restaurant	Gastropub	Restaurant	Deli / Bodega	Coffee Shop
18	Les Buttes-Chaumont	27 342	2	French Restaurant	Italian Restaurant	Bar	Pool	Restaurant	Beer Garden	Dessert Shop	Concert Hall	Park	Scenic Lookout

- **Cluster 4**

	District	Population Density ha/km²	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Le Louvre	9 041	3	Plaza	Hotel	French Restaurant	Bar	Tea Room	Theater	Wine Bar	Clothing Store	Restaurant	Church
3	L'Hôtel-de-Ville	16 966	3	French Restaurant	Art Gallery	Ice Cream Shop	Bakery	Burger Joint	Pub	Bar	Plaza	Gay Bar	Park
12	Les Gobelins	25 625	3	French Restaurant	Thai Restaurant	Hotel	Vietnamese Restaurant	Italian Restaurant	Asian Restaurant	Bakery	Indian Restaurant	Bar	Cambodian Restaurant
16	Les Batignolles-Monceau	29 724	3	Wine Bar	French Restaurant	Bar	Thai Restaurant	Coffee Shop	Italian Restaurant	Korean Restaurant	Pizza Place	Bookstore	Restaurant
19	Ménilmontant	32 702	3	French Restaurant	Bar	Bakery	Japanese Restaurant	Bistro	Plaza	Park	Food & Drink Shop	Italian Restaurant	Café

- **Cluster 5**

	District	Population Density ha/km²	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
15	Passy	20 921	4	French Restaurant	Bakery	Café	Grocery Store	Supermarket	Garden	Sandwich Place	Chinese Restaurant	Restaurant	Asian Restaurant

5. Discussion

After looking at the clusters, it appears that cluster 3 and cluster 4 are the best suited for the company's needs because they offer the biggest number of business opportunities. French restaurants and hotels are the most common venues in these clusters. However, as the company starting budget is limited, suggesting this big number of districts in confusing, that's why further analysis is needed. The company's busines plan is to start with only two districts and then expand from there once their financial capability increases. Hence, an important factor which was overlooked during this analysis comes into play. It is the "Population density ha/km²". In other words, choosing only two districts: cluster 3 and cluster 4 comes down to the ratio of the number of people who live in each district versus the area in km², the bigger this number the more a great number of people live in small confined area, which is optimal for the company objectives as it creates more business opportunities and reduces operating costs (number of employees, total distance covered…).

After taking a closer look at the clusters, two districts pop out, **L'Observatoire** district with a population density of 24 821 and **Vaugirard** district with 27 712. **These two districts are the best match for the company's needs, because they have diversified venues, are close to each other and have a great population density.**

6. Conclusion

The goal of the analysis that was done in this report is to determine the two best districts of Paris that have the most upside for our food deliver company. They want to find out which are the districts that offer the best business opportunities. Based on the data collected, and the machine learning approach used (KMeans) it was concluded that, **L'Observatoire district and Vaugirard district are the best match for the company**.