Picking the right location for a new fast food restaurant in Paris

Business Problem

- For this project I have created a hypothetical business problem. I would like to start a fast food restaurant, and therefore I am looking to identify which areas in Paris would be the best to do so. Hence, I could use data science techniques to analyse paris boroughs environements to determine which boroughs are the most suitable. The ideal location will depend on the target market (families, young adults…).
- This analysis could be useful for business entrepreneur looking to open a new fast food restaurant.

Business Problem: rational

- According to Statistita, there are around 180,000 restaurants in France including 20% located in Paris.
- The French restaurant industry sales increased by 1.6% to reach EUR50.4 billion in 2018.
- The fast food segment represents 37% of restaurants in France and generates around EUR19 billion of sales.
- The fast-food average ticket increased by 1.4%.
- Based on the figures displayed by the Paris Chamber of Commerce, 23% Parisian restaurants are fast-food restaurant only topped by traditional French restaurants.
- It is the fastest growing segment with a growth of 11% (in terms of number of restaurants) between 2014 and 2017.

Data

- The list of Paris metro stations
- The annual trafic for each metro stations
- The Foursquare API

Data Preparation

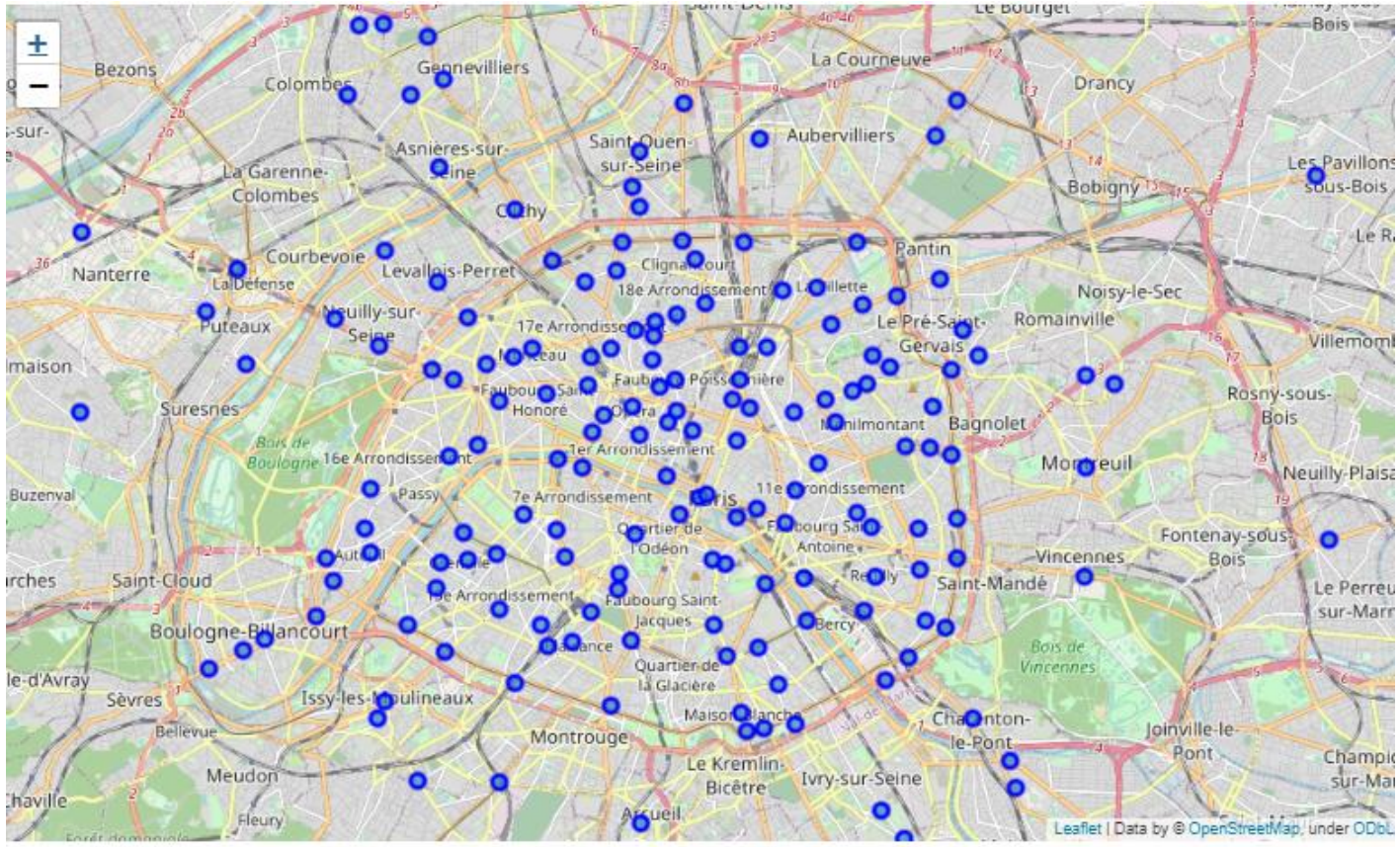
- Datasets downloaded from the Autonomous Parisian Transportation Administration:
  - Public transport stations' traffic
  - Metro stations' coordinates.
- Data Preparation:
  - First we need to exclude the stations belonging to other public transport network (bus, RER etc…) from the public transport stations' traffic dataset.
  - We decided we also segmented and sorted stations into 5 bins based on their annual traffics , using the pandas' cut function. The output of this operation is the dataframe below:
  - We also have to format the coordinates. After performing these tasks, we merge the datasets into a dataframe.

Data Preparation

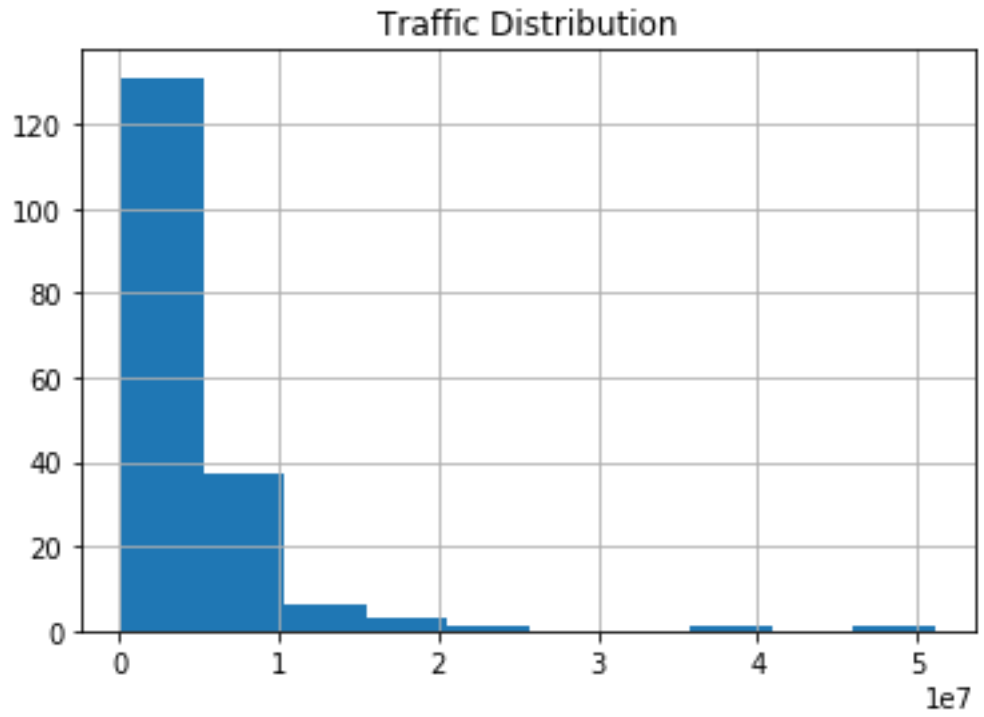
- Formatted dataframe:

Station	Description	Coordinates	Rang	Réseau	Trafic	Ville	Arrondissement pour Paris	Traffic_Cat	Latitude	Longitude
AVRON	63 BOULEVARD ALSACE-LORRAINE - 94058	48.8500655011,2.49939528589	256	Métro	1871024	Paris	11.0	(0, 2500000]	48.8500655011	2.49939528589
CHATEAU DE VINCENNES	PISTE GARE ROUTIERE - 75112	48.8442170813,2.44079723454	50	Métro	6353285	Vincennes	NaN	(5000000, 7500000]	48.8442170813	2.44079723454
ECOLE VETERINAIRE DE MAISONS-ALFORT	31-35 AVENUE DU GENERAL LECLERC - 94046	48.8147969334,2.42270643461	157	Métro	3193857	Maisons - Alfort	NaN	(2500000, 5000000]	48.8147969334	2.42270643461
MARX DORMOY	AVENUE FRANCOIS MITTERRAND - 91027	48.7036035904,2.37143564263	151	Métro	3359946	Paris	18.0	(2500000, 5000000]	48.7036035904	2.37143564263
SIMPLON	46 BOULEVARD ORNANO - 75118	48.8948064764,2.34715016514	218	Métro	2366858	Paris	18.0	(0, 2500000]	48.8948064764	2.34715016514

Metro Station Locations



Traffic Distribution



Venues

- We used the Foursquare API to get the top 100 venues within 500 meters of each metro station.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	AVRON	48.850066	2.499395	Monceau Fleurs	48.849502	2.498391	Flower Shop
2	AVRON	48.850066	2.499395	Onela Perreux-sur-Marne	48.849233	2.496895	Home Service
3	AVRON	48.850066	2.499395	Arrêt Raymond Poincarré [116]	48.850198	2.494422	Bus Stop
4	AVRON	48.850066	2.499395	Mille Et Un Vin	48.848123	2.494047	Wine Shop
5	CHATEAU DE VINCENNES	48.844217	2.440797	Tamarin	48.845311	2.438471	Thai Restaurant

- After a onhot encondin, we add the traffic categories and obtain the following dataframe:

	(0, 2500000]	(2500000, 5000000]	(5000000, 7500000]	(7500000, 10000000]	(10000000, 55000000]	Accessories Store	Afghan Restaurant	African Restaurant	Alsatian Restaurant	American Restaurant	...	Vietnamese Restaurant	Water Park	Waterfall	W
0	1	0	0	0	0	0.0	0.0	0.000000	0.0	0.000000	...	0.000000	0.0	0.0	0.000
1	0	0	1	0	0	0.0	0.0	0.000000	0.0	0.000000	...	0.000000	0.0	0.0	0.000
2	0	1	0	0	0	0.0	0.0	0.000000	0.0	0.000000	...	0.000000	0.0	0.0	0.000
3	0	1	0	0	0	0.0	0.0	0.000000	0.0	0.000000	...	0.000000	0.0	0.0	0.000
4	1	0	0	0	0	0.0	0.0	0.023256	0.0	0.023256	...	0.023256	0.0	0.0	0.023

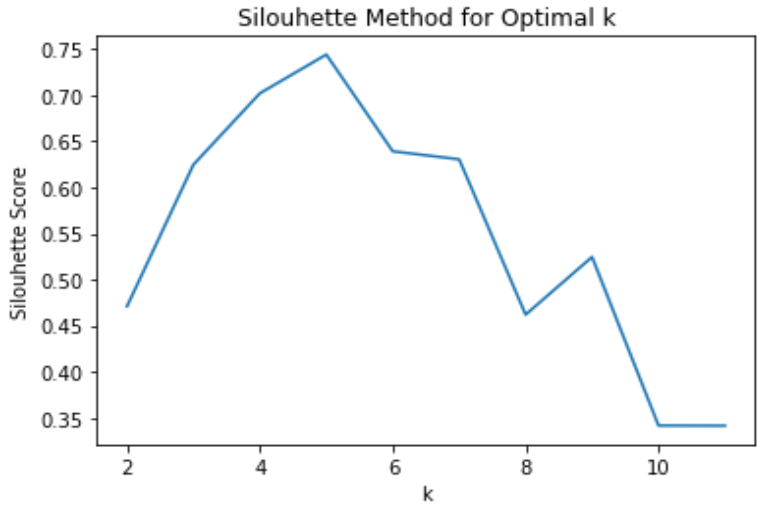
Clustering: k-means

- Now that we have the dataset that will be used to train our model, we needed to choose a clustering algorithm. We decided to use the K-means algorithm which is an unsupervised learning. It is easy to implement and it is quite computationally fast.

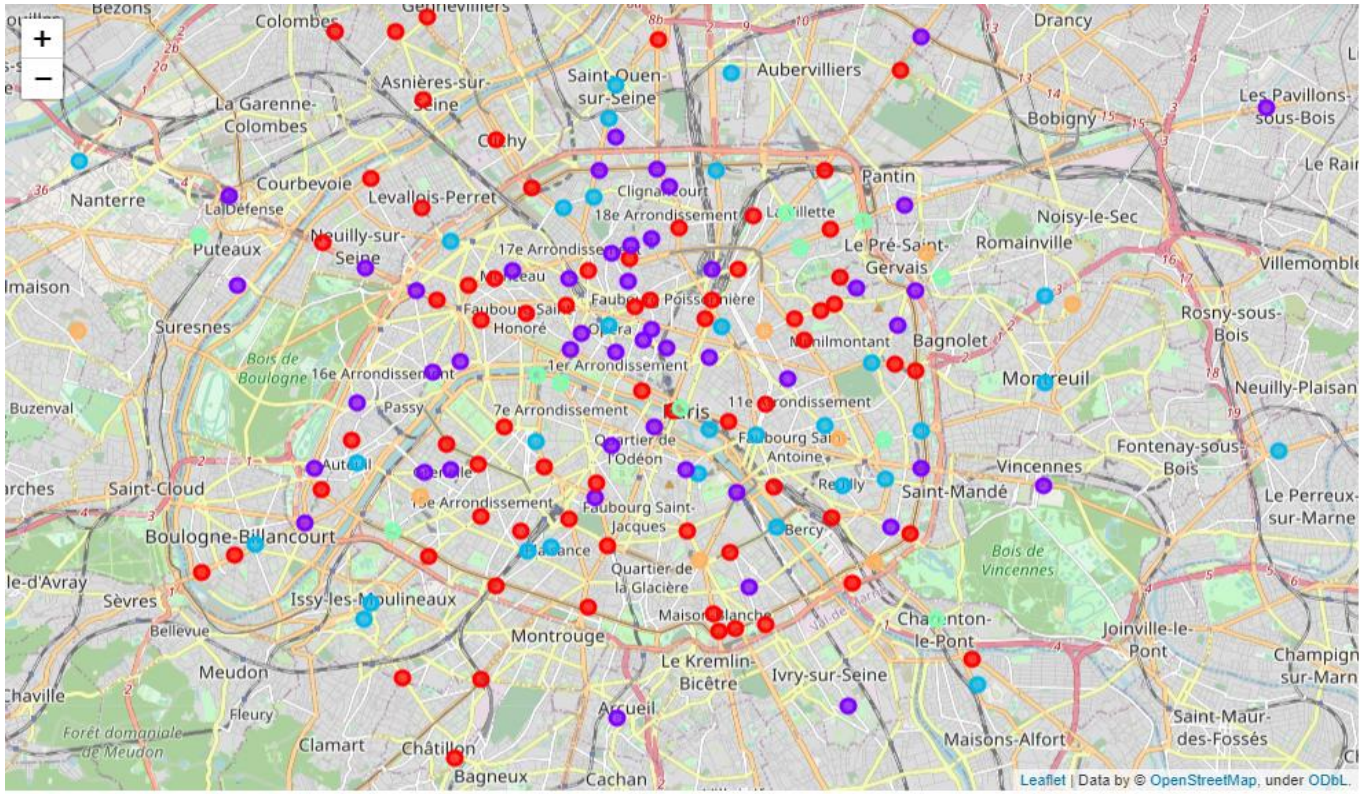
Clustering: Finding the Optimal Number of Clusters

- The first step is to decide how many clusters we want build. For this purpose we use the silouhette analysis which can be used to study the separation distance between the resulting clusters. The graph below displays the silouhette scores for each number of clusters.

- In this graph, we can see that the optimal cluster number is five since it has the highest silouhette score.



Clustering: Output



Markers color: Cluster 0: red, CLuster 1: purple, Cluster 2: blue, Cluster 3: green, Cluster 4: orange.

Results

- Cluster 0, cluster 1 are the biggest clusters in terms of metro stations counts.
  - They are surrounded by a lot of restaurants, the most fequent venue being italian restaurants. We can also see a competition from Japanese restaurants which appeared in the top four venues as well as from pizza places, vietnamese restaurants and sandwich places. In terms of traffic cluster 1 consist of bigger stations with 20% of them having more than 7.5 million tavellers. La Defense which is the main business district and the third biggest metro station falls within this cluster.
- Cluster 2 is quite similar, but consist of stations with less traffic.
- Cluster 3 consists of 12 stations. Three stations could be interesting spots because they are located on the Seine banks, in central Paris. Three stations are located in the canal de l'Ourcq neighborhood which is in the process of gentrification with a young population.
- Cluster 4 is quite interesting with most of its stations located in the eastern part of Paris. Place d'Italie could be a good location since it is big a transport hub.

Discussion

- Based on this analysis, we recommend to consider cluster 3 or 4 to start a new fast food restaurants. These clusters gather metro stations located in areas where the competition is less intense than in other clusters.
- This analysis can be improved with additional data such as demography, population youth, offices locations etc.

Conclusion

- In this report, we conducted an analysis aiming for determining the best postential locations for a new fast food restaurant. In order to do so, we performed data cleaning/preparation using common python libraries such as pandas, matplotlib and numpy. In order to implement the k-means algorithm, we used the scikit learn package. The result of the analysis can serve as a basis for business decision making.

- Jupyter notebook related to this analysis.
  - https://github.com/Kadrik87/IBM-Capstone Project/blob/master/Fast%20Food%20Restaurant.ipynb