CAPSTONE PROJECT THE BATTLE OF NEIGHBORHOODS

OPENING A NEW RESTAURANT IN PARIS

TABLE OF CONTENT

1. INTRODUCTION

- According to Insee www.insee.fr, there are more than 45,000 restaurants in Paris and about 2.2 million people, therefore opening a new restaurant in Paris can be an extremely challenging task.
- According to several surveys, up to 40% of such start-ups fail in the very first year.
- What if there is a way to cluster city neighborhoods, based on their near-by restaurant similarity? What if we can visualize these clusters on a map? What if we might find what type of restaurant is the most and least popular in each location?
- Equipped with that knowledge, we might be able to make a smart choice from a huge number of restaurant types and available places.

2. DATA

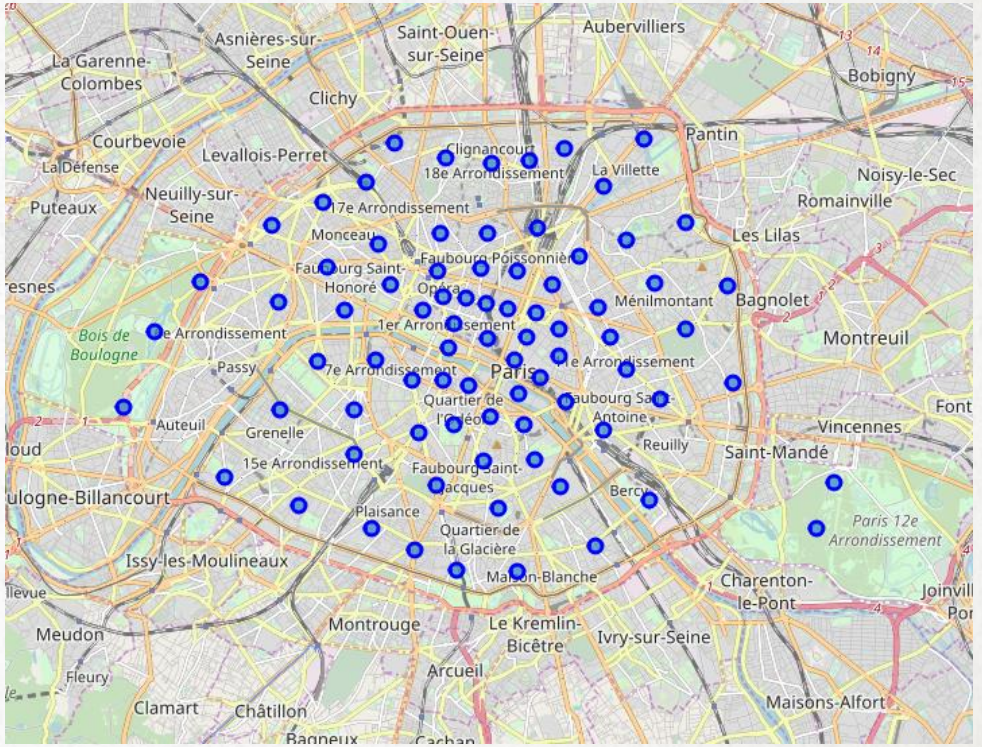
- This project will be using data from Paris open source database https://opendata.paris.fr, to collect information about boroughs and neighborhoods and merge those two separate dataframes.
- We will also use Foursquare API to collect the top 100 restaurants for each location.

3. METHODOLOGY

1. Using two tables from Paris open source database, collect information about Paris boroughs and neighborhoods.
2. Merge these two separate dataframes into one which will be used for the next steps.
3. Use the Geopy and Folium library to get the coordinates of every locations and map geospatial data on a Paris map.
4. Using Foursquare API, collect the top 100 restaurants and their categories for each location within a radius 300 meters.
5. Group collected restaurants by location and by taking the mean of the frequency of occurrence of each type, preparing them for clustering.
6. Cluster restaurants by k-means algorithm and analyze the top 10 most common restaurants in each cluster.
7. Visualize clusters on the map, thus showing the best locations for opening the chosen restaurant.

4. EXPLORING PARIS NEIGHBORHOODS

- Using Paris open source data we collected all borough and neighborhoods in Paris.
- With some cleaning and wrangling, we created a folium map to visualize Paris's neighborhoods.



5. EXPLORING PARIS RESTAURANTS

- We used Foursquare API to get top 100 restaurants in each neighborhood.
- We obtained 4791 restaurants of 123 individual types.
- We applied one-hot encoding and grouped them by taking the mean of the frequency of occurrence of each type.
- We clustered restaurants using the k-means algorithm based on their similarity.
- The k-means is an unsupervised machine learning algorithm for clustering unlabeled data.

6. RESULTS

- Cluster #1

	Neighborhood	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
count	20	20.0	20	20	20	20	20	20	20	20	20	20
unique	20	NaN	4	8	13	11	14	16	15	15	16	15
top	Père-Lachaise	NaN	French Restaurant	Italian Restaurant	Italian Restaurant	Japanese Restaurant	Bakery	Italian Restaurant	Bakery	Sushi Restaurant	Japanese Restaurant	Café

- The most common restaurant : French Restaurant
- The 10th most common restaurant : Café

6. RESULTS

- Cluster #2

	Neighborhood	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
count	9	9.0	9	9	9	9	9	9	9	9	9	9
unique	9	NaN	4	5	4	6	7	9	9	7	9	8
top	Invalides	NaN	French Restaurant	Café	French Restaurant	Bakery	Italian Restaurant	Breakfast Spot	Bakery	Bakery	Middle Eastern Restaurant	Diner

- The most common restaurant : French Restaurant
- The 10th most common restaurant : Diner

6. RESULTS

- Cluster #3

	Neighborhood	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
count	46	46.0	46	46	46	46	46	46	46	46	46	46
unique	46	NaN	4	13	12	16	20	19	24	26	27	31
top	Place-Vendôme	NaN	French Restaurant	Italian Restaurant	Italian Restaurant	Italian Restaurant	Japanese Restaurant	Japanese Restaurant	Fast Food Restaurant	Restaurant	Pizza Place	Sushi Restaurant

- The most common restaurant : French Restaurant
- The 10^{th} most common restaurant : Sushi Restaurant

6. RESULTS

- Cluster #4

	Neighborhood	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
count	2	2.0	2	2	2	2	2	2	2	2	2	2
unique	2	NaN	2	2	2	2	2	2	2	2	2	2
top	Parc-de-Montsouris	NaN	French Restaurant	Restaurant	Middle Eastern Restaurant	Pizza Place	Bakery	Cafeteria	Café	Burger Joint	French Restaurant	Diner

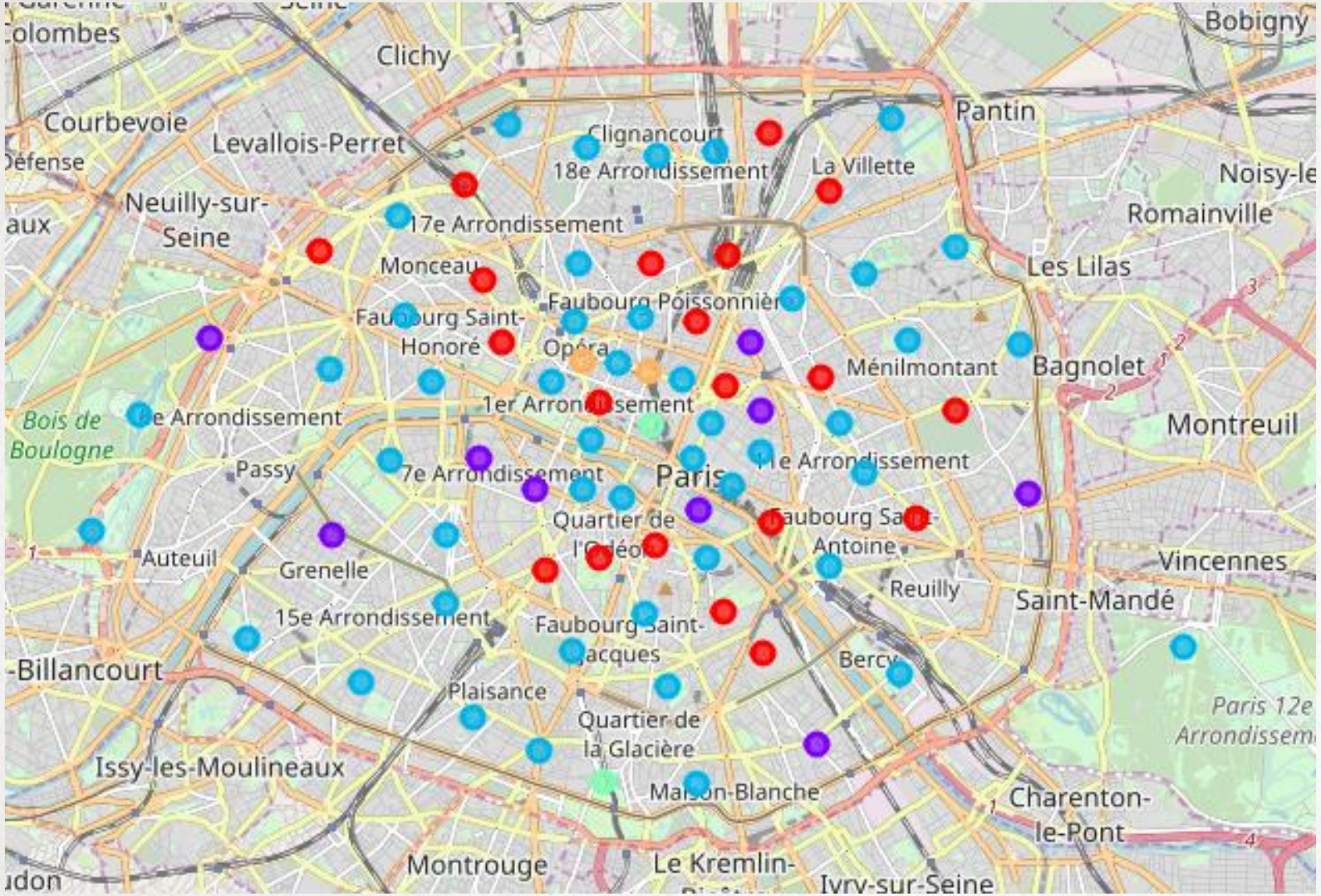
- The most common restaurant : French Restaurant
- The 10^{th} most common restaurant : Diner

6. RESULTS

- Cluster #5

	Neighborhood	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
count	2	2.0	2	2	2	2	2	2	2	2	2	2
unique	2	NaN	1	2	2	2	2	2	2	2	2	2
top	Gaillon	NaN	French Restaurant	Japanese Restaurant	Bistro	Bakery	Sandwich Place	Thai Restaurant	Salad Place	Burger Joint	Ramen Restaurant	Burger Joint

- The most common restaurant : French Restaurant
- The 10^{th} most common restaurant : Burger Joint



MAP LEGEND

Cluster 1 - Red dots
Cluster 2 - Purple dots
Cluster 3 - Blue dots
Cluster 4 - Green dots
Cluster 5 - Orange dots

7. DISCUSSION

- Analyzing the most popular restaurants in each cluster, the stakeholder should select the least popular types within the top 10 as a safe choice.
- In our recommendations, we advise selecting from the 9^{th} and 10^{th} positions as a reasonable balance between being too popular and having no customers.
- Recommendations, based on description of each cluster:
  - Cluster 1 Locations: **Japanese Restaurant** or **Café**
  - Cluster 2 Locations: **Middle Eastern Restaurant** or **Diner**
  - Cluster 3 Locations: **Pizza Place** or **Sushi Restaurant**
  - Cluster 4 Locations: **French Restaurant** or **Diner**
  - Cluster 5 Locations: **Ramen Restaurant** or **Burger Joint**

8. CONCLUSION

- In this report we established a methodology to determine what the most promising type of restaurant is , and where it should be opened.
- This type of analysis can be applied to any city of your choice that has available geospatial information.
- This type of analysis can be applied to any type of venue (shopping, clubs, etc.) that is available in Foursquare database.