

Another wine bar in Paris

Coursera Captsone Project

Tom Hohweiller, PhD.

2021

A client problem

- A client wants to open a wine bar in Paris
- In an area with existing restaurants for a dynamic neighbourhood
- But not too many of them

Business understanding

- Get a location in Paris for a wine-related venue
- In an area with already a good medium density of restaurants

Planning the study

Four steps ahead to find the best location:

1. Extract every venues in each *arrondissement* to have a better view of the business situation
2. Study the number of restaurants as well as the number per 1000 residents (for wine bars or shops also)
3. After selecting a *arrondissement*, refine the venues and clustering them to identify small area with prominent restaurants
4. Finally, choosing the best area will be our best location in Paris

Data - Data requirements

Data needed

Two type of data - for each *arrondissement* and venues

- Coordinates of each *arrondissement* - **Webscrapping wikipedia webpage**
- Venues in Paris for each *arrondissement* - **Using the Foursquare API**
 - Name of the venue
 - Coordinates
 - Venue category
 - Postal code

DataFrame

Each dataframes will be usefull, the followed ones will be constructed:

- 1- DataSet of each *arrondissement*
- 2- DataSet of all the venues (for each *arrondissement*)
- 3- DataSet for restaurants and wine bars and shops
- 4- DataSet for the selected *arrondissement*

Steps to take

1. Extract every venues in each *arrondissement* to have a better view of the business situation
2. Study the number of restaurants as well as the number per 1000 residents (for wine bars or shops also)
3. After selecting a *arrondissement*, refine the venues and clustering them to identify small area with prominent restaurants
4. Finally, choosing the best area will be our best location in Paris

WebScraping and venues

Step 0

Webscrapping wikipedia pages is done using beautifulsoup

Step 1

Calling the Foursquare API make gathering venues information really easy

These two step created the following dataframes, with the dataframe of each *arrondissement* (left) and data frame of each venues (right).

| | Arr | Latitude | Longitude | Population |
|----|-----|-----------|-----------|------------|
| 1 | 1 | 48.860000 | 2.341944 | 16385 |
| 2 | 2 | 48.866944 | 2.340556 | 21042 |
| 3 | 3 | 48.863889 | 2.361667 | 34389 |
| 4 | 4 | 48.866111 | 2.355556 | 28370 |
| 5 | 5 | 48.846111 | 2.344722 | 50631 |
| 6 | 6 | 48.850556 | 2.332778 | 41976 |
| 7 | 7 | 48.856944 | 2.320000 | 52193 |
| 8 | 8 | 48.877778 | 2.317778 | 37368 |
| 9 | 9 | 48.872500 | 2.340278 | 60071 |
| 10 | 10 | 48.871944 | 2.357500 | 90836 |
| 11 | 11 | 48.858333 | 2.379222 | 147470 |
| 12 | 12 | 48.841111 | 2.368556 | 141287 |
| 13 | 13 | 48.832222 | 2.355556 | 183399 |
| 14 | 14 | 48.830556 | 2.326667 | 159941 |
| 15 | 15 | 48.841389 | 2.302278 | 235178 |
| 16 | 16 | 48.862778 | 2.276111 | 168554 |
| 17 | 17 | 48.864444 | 2.321944 | 158737 |
| 18 | 18 | 48.892222 | 2.344444 | 106131 |
| 19 | 19 | 48.862778 | 2.381944 | 188066 |
| 20 | 20 | 48.865000 | 2.399167 | 106739 |

| | Arr | Arr Latitude | Arr Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category | Venue Id |
|---|-----|--------------|---------------|--|----------------|-----------------|--------------------|--------------------------|
| 0 | 1 | 48.86 | 2.341944 | Place du Louvre | 48.859541 | 2.340822 | Plaza | 48979c8ae4b05a85da0714f |
| 1 | 1 | 48.86 | 2.341944 | Cour Carrée du Louvre | 48.860360 | 2.339543 | Pedestrian Plaza | 4c079d740ed3c920b6be797d |
| 2 | 1 | 48.86 | 2.341944 | Eglise Saint-Germain-l'Auxerrois (Eglise Saint-) | 48.860620 | 2.341306 | Church | 4ad0da0f064a620173421e3 |
| 3 | 1 | 48.86 | 2.341944 | Boutique yam'Tcha | 48.861710 | 2.342380 | Chinese Restaurant | 5498b44c490e7a7ca5e5ed84 |
| 4 | 1 | 48.86 | 2.341944 | La Vénus de Milo (Vénus de Milo) | 48.859043 | 2.337234 | Exhibit | 5664eb745c3e31e7088e96d |

Restaurants and wine bars/shops

Step 2

Creating a statistic dataframe with the number of restaurants and wine bars/shops, as well as the number of venues per 1000 residents is easy. Creating bar plots, representing each value for each *arrondissement*

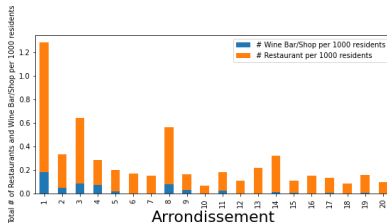
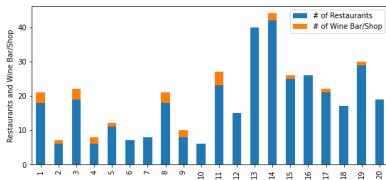


Figure: Bar plots of the total number of restaurants and wine bars/shops (left). Number of restaurants and wine bars/shops per 1000 residents (right).

Best *arrondissement*

Step 3

- Taking into account the fact that the client wants a area with existing restaurant (but not too many). The **third *arrondissement*** will be chosen.
- However, take surrounded *arrondissement* is a good solution for broader research
- Looking at the following *arrondissement*: 1st, 2nd, 3rd, 4th, 10th and 11th

Creating this dataframe, for all venues in the third *arrondissement*

| | Arr | Arr Latitude | Arr Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category | Venue id | Postal Code |
|-----|-----|--------------|---------------|---------------------------|----------------|-----------------|----------------|--------------------------|-------------|
| 200 | 3 | 48.863889 | 2.361667 | Mmmozza | 48.863910 | 2.360591 | Sandwich Place | 4d974096a2c654814aa6d353 | 75003 |
| 201 | 3 | 48.863889 | 2.361667 | Chez Alain Miam Miam | 48.862369 | 2.361950 | Sandwich Place | 5b546a4a82a750002c940e7f | 75003 |
| 202 | 3 | 48.863889 | 2.361667 | Marché des Enfants Rouges | 48.862806 | 2.361996 | Farmers Market | 4b75734cf964a5202c0d2ee3 | 75003 |
| 203 | 3 | 48.863889 | 2.361667 | Candelaria | 48.863032 | 2.364059 | Cocktail Bar | 4d77b39caf63cbff3997be0f | 75003 |
| 204 | 3 | 48.863889 | 2.361667 | Le Barav | 48.865166 | 2.363155 | Wine Bar | 4b68a117f964a520c8832be3 | 75003 |

Figure: Dataframe of the third *arrondissement* with additional information.

Best *arrondissement*

Venues in these areas are depicted in the following map

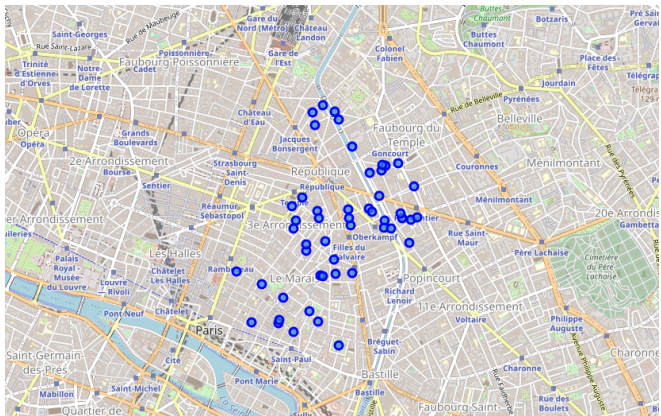


Figure: Spatial repartition of the venues of the third arrondissement and around.

Best *arrondissement*

Step 3.5

From the previous map, four area can be found. Using *k-means*, they can be identify

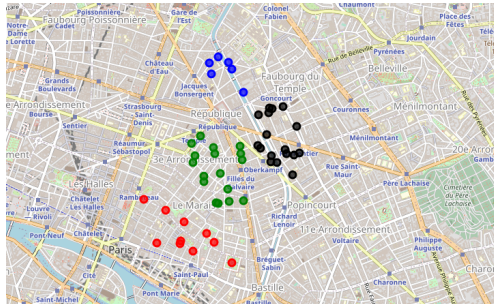


Figure: Spatial repartition of the venues of the third arrondissement and around - color coded with respect with cluster label.

Step 4

From these clusters, the top five venues category can be extracted. Which are the following:

| | Cluster Label | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 0 | 0 | Bakery | Bookstore | Asian Restaurant | Canal | Cheese Shop |
| 1 | 1 | Clothing Store | Art Gallery | Plaza | Israeli Restaurant | Falafel Restaurant |
| 2 | 2 | French Restaurant | Restaurant | Italian Restaurant | Bakery | Wine Bar |
| 3 | 3 | Art Gallery | Sandwich Place | Cocktail Bar | Coffee Shop | Vietnamese Restaurant |

Figure: Top 5 venues category for each clusters.

Selecting the cluster label 2 was done:

- Existing restaurants
- Other clusters have day-related venues
- Bonus point: close to the Republic place in Paris (lot of people)

Best area

Finally, the best location would be in the following circle:

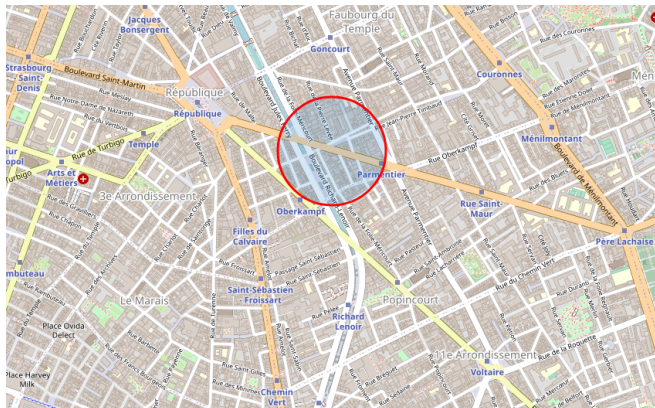


Figure: Best location for a wine bar/shop in Paris.

Conclusion

- This study was conducted using several techniques:
 - webscrapping
 - getting information using API
 - clustering
- It was successful at deciding to find the best location for opening a wine bar/shop

Thank you for your reading!
Hope it was pleasant
Have a nice day!