# Finding the best city to start an office in France

Christine Ock

26th, February,2021

## 1. Introduction & Business Problem

The company called ‘COV21' is looking to build a consulting business service to restaurants providing new strategy in order to cope with COVID 19 situation. They are targeting French market, because many of restaurants have been running traditionally and struggle to survive due to the pandemic. With several lockdowns, some of them had to close down. The 'COV21' has several ideas on rebranding French restaurants especially such as takeaways, menues and combining different icons while maintaining their authentic French classical image. One of the biggest stakeholders of 'COV21' is subsidised by French government meaning that the company has an up to date on government policy about COVID19 measurement. They are looking for a city where it has the most French restaurant to start the business.

In addition to restaurant industry, they have primary discussion on idea of changing business model for hotel industry, which also had huge impact during the COVID19. The company has an eligibility to get financial grant depending on the success of the business of their clients(hotels).

The problem is to find the best location in France where it has the most restaurants and hotels. Hence, they are seeking an advice from data scientist.

## 2. Data acquisition and Cleaning

The data scientist will use internet and Foursquare API as main source. The used datasets were either directly downloaded from open repository**.** First, the we will use data from https://simplemaps.com/data/fr-cities to extract latitude, longtitude and population of main 471 cities in France. We would need to clean the data in order to only extract information and rename the column of lat and log. For example,we selected data of the name of the city, latitude, longitude and population only and regenerated data frame as follow.

| | city | latitudes | longtitudes | population |
|---|---|---|---|---|
| 0 | Paris | 48.8566 | 2.3522 | 11020000.0 |
| 1 | Nice | 43.7034 | 7.2663 | 1006402.0 |
| 2 | Toulouse | 43.6045 | 1.4440 | 968638.0 |
| 3 | Marseille | 43.2964 | 5.3700 | 870018.0 |
| 4 | Rennes | 48.1147 | -1.6794 | 727357.0 |

Then, we are going to use Foursquare to retrieve all the venues of 471 cities and group by numbers by each category of the venue. We can also generate top 5 venues by each city. We will focus on the top venue in order to identify which city has the most traditional French restaurants and hotels. Based on these data acquisition and cleaning, we can start looking at the city that has the restaurant and hotel as common venue.

### 3. Methodology

Looking at the data retried from the Foursquare API in more detail, we can see that there are 5545 venues based on the limit as 100 venues and the radius 500 meters for each city from their given latitude and longitude informations. Here is a head of the list Venues name,category,latitude and longitude informations from Foursquare API.

```
print(france_venues.shape)
france_venues.head()
```

```
(5545, 7)
```

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Paris | 48.8566 | 2.3522 | Place de l'Hôtel de Ville – Esplanade de la Li... | 48.857010 | 2.351656 | Plaza |
| 1 | Paris | 48.8566 | 2.3522 | Maison Aleph | 48.857348 | 2.354873 | Pastry Shop |
| 2 | Paris | 48.8566 | 2.3522 | Berges de Seine – Rive droite | 48.855131 | 2.352289 | Pedestrian Plaza |
| 3 | Paris | 48.8566 | 2.3522 | Parc Rives de Seine | 48.855510 | 2.351419 | Park |
| 4 | Paris | 48.8566 | 2.3522 | LUSH | 48.857791 | 2.351622 | Cosmetics Shop |

In the summary of counting the venues by each each with 331 unique categories, we ran t top 5 venue category for each city and created table of top venue category by each city as below.

| | Neighborhood | 1st Most Common Venue |
|---|---|---|
| 0 | Abbeville | Hotel |
| 1 | Agde | French Restaurant |
| 2 | Agen | Multiplex |
| 3 | Aix-en-Provence | French Restaurant |
| 4 | Aix-les-Bains | Italian Restaurant |

We identify that there are some common venue category in several cities. Therefore, we use unsupervised learning K-means algorithm to cluster the cities. First, we run K-Means to cluster the cities into 5 clusters and merged table with cluster labels for each city.

| | city | latitudes | longtitudes | population | Cluster Labels | 1st Most Common Venue |
|---|---|---|---|---|---|---|
| 0 | Paris | 48.8566 | 2.3522 | 11020000.0 | 4.0 | French Restaurant |
| 1 | Nice | 43.7034 | 7.2663 | 1006402.0 | 1.0 | Hotel |
| 2 | Toulouse | 43.6045 | 1.4440 | 968638.0 | 0.0 | French Restaurant |
| 3 | Marseille | 43.2964 | 5.3700 | 870018.0 | 0.0 | French Restaurant |
| 4 | Rennes | 48.1147 | -1.6794 | 727357.0 | 4.0 | Creperie |

**4.Results**

Based on the clustering algorithm, we can see that cluster label 0 has French Restaurant as the most common venue, and cluster label 1 has hotel. The rest of cluster seems to be less irrelevant , because our goal is to find list of region that has French restaurant and hotels.

```
france_merged.loc[france_merged['Cluster Labels'] == 0, france_merged.columns[[0] + list(range(3, france_merged.shape[1]))]]
```

| | city | population | Cluster Labels | 1st Most Common Venue |
|---|---|---|---|---|
| 2 | Toulouse | 968638.0 | 0.0 | French Restaurant |
| 3 | Marseille | 870018.0 | 0.0 | French Restaurant |
| 7 | Montpellier | 607896.0 | 0.0 | French Restaurant |
| 8 | Lyon | 516092.0 | 0.0 | Bar |
| 9 | Rouen | 494362.0 | 0.0 | French Restaurant |
| ... | ... | ... | ... | ... |
| 452 | Mauriac | NaN | 0.0 | French Restaurant |
| 456 | Ambert | NaN | 0.0 | Optical Shop |
| 457 | Forcalquier | NaN | 0.0 | French Restaurant |
| 468 | Florac | NaN | 0.0 | French Restaurant |
| 470 | Castellane | NaN | 0.0 | French Restaurant |

130 rows × 4 columns

```
france_merged.loc[france_merged['Cluster Labels'] == 1, france_merged.columns[[0] + list(range(3, france_merged.shape[1]))]]
```

| | city | population | Cluster Labels | 1st Most Common Venue |
|---|---|---|---|---|
| 1 | Nice | 1006402.0 | 1.0 | Hotel |
| 13 | Brest | 300300.0 | 1.0 | Hotel |
| 37 | Belfort | 114445.0 | 1.0 | Stadium |
| 40 | Montreuil | 109897.0 | 1.0 | Hotel |
| 50 | Dunkerque | 87353.0 | 1.0 | Hotel |
| ... | ... | ... | ... | ... |
| 414 | Figeac | NaN | 1.0 | Kids Store |
| 417 | Villefranche-de-Rouergue | NaN | 1.0 | Pub |
| 420 | Apt | NaN | 1.0 | Hotel |
| 422 | Avallon | NaN | 1.0 | Hotel |
| 465 | Sainte-Menehould | NaN | 1.0 | Gym |

62 rows × 4 columns

We see that there are 130 cities that has the French restaurant and 62 cities that has the hotels as the most common venue. It is also important that we look that the city that has the high population to conclude which city to target and open the office. Based on the result, Toulouse and the Nice have the highest in each label.

## 5. Discussion

Coming back to the goal of this project, which is to find the best city to open the business, we have used two source of data in order to get latitude and longitude of main cities in France and venues in order to identify the most common venue. Using the K mean algorithm, it was easy to find 130 cities that has the French restaurant and 62 cities that has the hotels as the most common venue.

It is still difficult to target all those cities out of 471 cities, because population density is also very important. Therefore, I used population as a factor to determine which city to start with.

We can recommend Toulouse as a start the business based on the result. In terms of the hotel industry, we would recommend Nice.

We can further explore the restaurant and hotels in those cities in terms of ranking and even visualise on the map to understand the location of all those restaurants through Foursquare API. We also suggest to work with another data provide that has more local data in France.

## 6. Conclusion

In conclusion, 'COV21' should start with focusing on research about all the restaurants in Toulouse and start business in this city. For the hotel industry, the company should focus on Nice as they have the hotels as the most common venue.