

Finding the best city to start an office in France

Christine Ock

26th, February,2021

1. Introduction & Business Problem

The company 'SUNSHINE' is planning to start consulting business service to restaurants providing new strategy in order to cope with COVID 19 situation and looking for a city to open office in France. They are targeting French market, because many restaurants have been running traditionally and now struggling to survive due to the pandemic. With several lockdowns, some of them had to close down. The 'SUNSHINE' has several strategies on rebranding French restaurants such as takeaways, menus and combining with different icons while maintaining their authentic French classical image. One of the biggest shareholders of 'COV21' is subsidised by French government so that the company has an up to date information on government policy about COVID19 measurement.

In addition to restaurants , the company has 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 city 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, longitude and population of main 471 cities in France. Then, 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

Next, we are going to use Foursquare API to retrieve all 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 retrieved from the Foursquare API in detail, we can see that there are 5545 venues based on the search criteria, limit of 100 venues and the radius 500 meters for each city from their given latitude and longitude informations. Here is example of the list of 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 L...	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

By counting the venues by each city, we see that there are 331 unique categories. Then, we ran top 5 venue category for each city and created table of top venue category by each city as an example 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 is a 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 merge table with cluster labels for each city.

	city	latitudes	longitudes	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 and cluster label 1 has hotel as the most common venue. The rest of cluster seems to be less irrelevant , because our goal is to find list of cities that has French restaurant and hotels the most.

```
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
466	Florac	NaN	0.0	French Restaurant
470	Castellane	NaN	0.0	French Restaurant

130 rows x 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 x 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 according to the table above. It is also important that we look at the city that has the highest 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 aim of this project, which is to find the best city to start the business, we have used two sources 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 130 and 62 cities, therefore, looking at the population density is also very important. Because, the high population density can be highly related to the size of the restaurants and revenue. Therefore, we look at the city with the highest population.

We can recommend Toulouse to start with based on the result. In terms of hotel industry, we would recommend Nice.

We can further explore restaurants and hotels of 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 provider that has more local data in France.

6. Conclusion

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



