The Battle of the Neighborhoods - Report

# Introduction & Business Problem

The city of Paris is the most populated agglomeration in France and one if not the most attractive in Europe. Paris is a lively, diverse, dynamic city and is the financial as well as political capital of France.

It is a multicultural city which provides a lot of opportunities as well as a favorable business environment.

The city has attracted many players in the market. It is a global hub of business and commerce. The city is a major center for banking and financial (one of the most important in Europe), retail trade, world trade, transport, tourism, real estate, new media, traditional media, advertising, legal services, accounting, insurance, theater, fashion and arts in Europe. It also means that the market is very competitive. As it is a very developed city, the cost of business is also one of the highest.

Thus, any new commercial enterprise or expansion must be carefully analyzed. The information drawn from the analysis will allow a good understanding the business environment that helps strategically target the market. This analysis will help in risk reduction. And the return on investment will be reasonable, though.

## Political Division of Paris

The City of Paris like any big city is divided in many, many ways. The city itself was divided into 20 different Faubourgs and later into **Arrondissments** (districts) for the purpose of taxation and [policing](https://www.mtholyoke.edu/courses/rschwart/hist255-s01/thenardier/index.html), but there were social divisions as well that were just as concrete. These divisions were influenced by the differing ideals between social groups. Understanding these divisions is important to both the political and social history of Paris.

You will find in this report a map showing Paris arrondissements and a map presenting Paris neighborhoods. Paris arrondissements and districts are the administrative divisions that share Paris in 20 municipals arrondissements. Within each of its boroughs, the neighborhoods are making the charm and the reputation of the city of Paris.

The twenty arrondissements of Paris all have their own character through their tourist sites, places to go out, restaurants, and homes. Each Parisian arrondissement has four administrative districts.

Each of Paris' 20 administrative districts (or arrondissements) are officially divided into 4 quartiers.

## Business Problem

City of Paris is famous for its excellent cuisine. Its food culture includes an array of international cuisines influenced by the city's immigrant history. Chinese restaurants have become so popular in France now it seems that there is one on every corner, not only in major cities but also in smaller cities.

Starting a Chinese restaurant can be a great business opportunity, but you need to distinguish yourself from others to enjoy long-term success.

If you plan a real restaurant that can demand higher prices for fresh fish or fresh foods, delivered daily from Japan or China, focus on neighborhoods and outlets that already attract a sophisticated Chinese or others client. If you plan a cheap buffet restaurant, points to the masses looking for affordable high-traffic locations with large shopping centers and other local points of interest.

My client wants to open their business in Paris area, not necessary in the center of Paris. So according to his need, I focus on that borough during my analysis. We define potential neighborhood based on the number of Chinese bars which are operating right in each neighborhood. Paris has full potential but also is a very challenging district to open a business because of high competition. New Chinese bar should be open in an area that inadequate neighborhood in this way the bar can attract more customers. Therefore, this analysis necessary to ensure that we have enough customers and that we are not so close to other Chinese places, and if we are close to others, to provide the best food in order to attract as many customers as possible.

# Data acquisition and cleaning

As specify earlier, Paris has a total of 20 boroughs and 80 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 20 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

The first dataset consists of the 20 boroughs with their respective geographical coordinates and are available at:

* <https://opendata.paris.fr/explore/dataset/arrondissements/table/>

The second dataset consists of the 80 neighborhoods with their respective geographical coordinate, the boroughs to witch the belong and are available at:

* <https://opendata.paris.fr/explore/embed/dataset/quartier_paris/table/>

Merging those two datasets, we finally have our dataset with all 20 boroughs and 4 neighborhoods in each.



Figure 1 - Boroughs, Neighborhoods of Paris

**Why using geographical data?**

* Paris city geographical coordinates data will be utilized as input for the *Foursquare API*[[1]](#footnote-1), that will be leveraged to provision venues information for each neighborhood.
* We will use the Foursquare API to explore neighborhoods in Paris.

In addition, Chinese category Id **4bf58dd8d48988d145941735[[2]](#footnote-2)** is used for retrieving data from Foursquare API.

# Methodology

In this project, I will use the basic methodology. Given the dataset with Boroughs, Neighborhoods and their respective geographical coordinates, we will use the *Foursquare API* to explore neighborhoods of Paris. After that, explore function to get Chinese restaurant categories in each neighborhood. Finally, I will use K-Means[[3]](#footnote-3) clustering to cluster and find the best places(venues) to build a Chinese restaurant in Paris.

For the sake of simplicity, we will limit our research to boroughs like **Louvre, Bourse, Temple, Hôtel-de-Ville, Panthéon, Luxembourg and Palais-Bourbon.**





Results returns by *Foursquare API* give us information’s about neighborhood (geographical coordinates), venue (geographical coordinates and different category).

* The most important thing here is the **hierarchical list of categories applied to venues.**

For our research perimeter, we find more than 19 unique categories.



We then use those 19 unique categories to One-Hot[[4]](#footnote-4) encode each Venue category values.

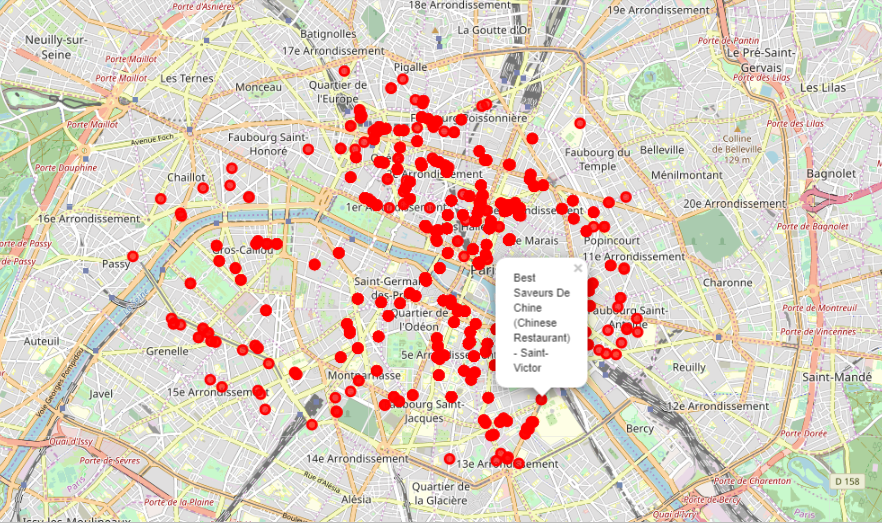


Figure 2 - Chinese restaurants in Paris (Louvre, Bourse, Temple, Hôtel-de-Ville, Panthéon, Luxembourg and Palais-Bourbon).

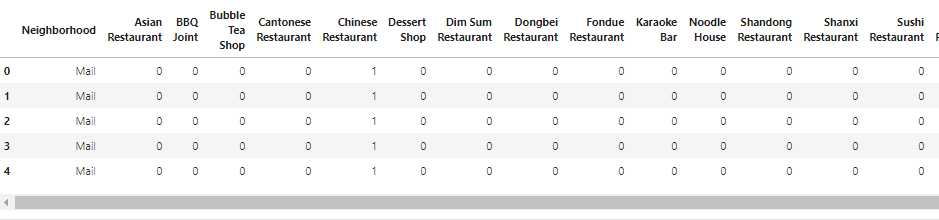


Figure 3 -One-Hot Encoding Venue Categories by Neighborhood.

We then use this feature to group the neighborhoods into clusters. *K-means* clustering algorithm with 5 clusters will be used to complete this task. *Folium* library will also be used to visualize the neighborhoods in Paris and its emerging clusters.



Figure 4 - 7 of 10 Most Common Venue

Applying *K-means*, each borough, neighborhood as well as their respective Latitude and Longitude will be grouped in each cluster. A plot will be made to visualize all those results.

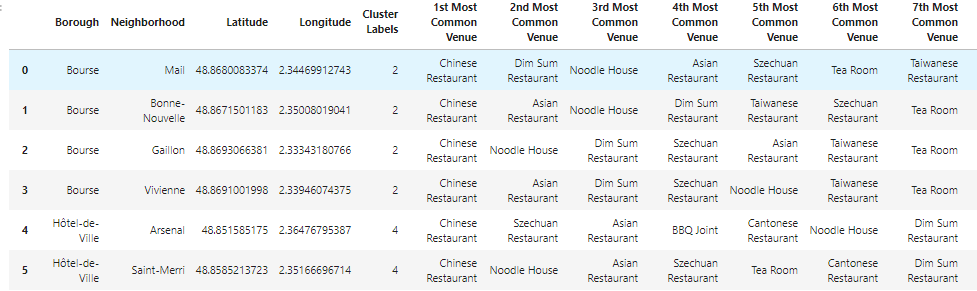


Figure 5- Borough, Neighborhood grouped in clusters.

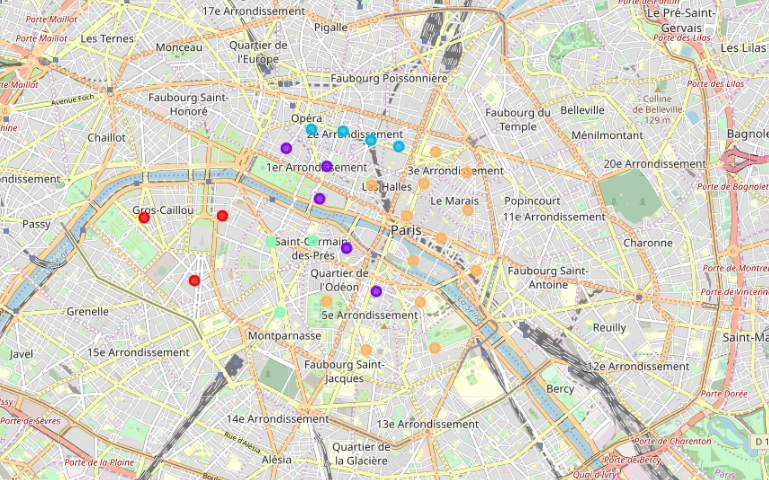


Figure 6 - Results with 5 clusters

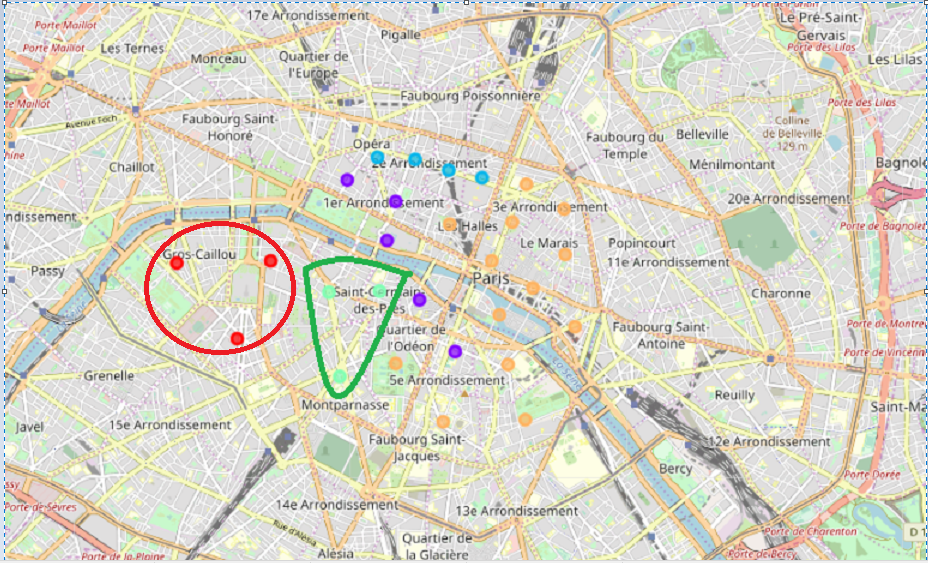
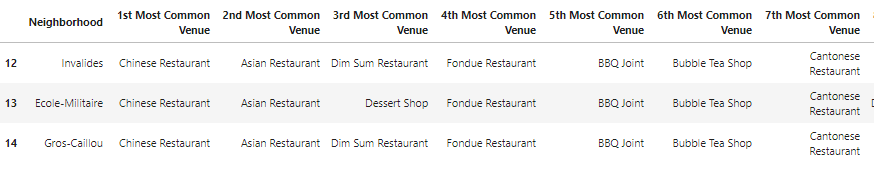


Figure 7 - Selected locations

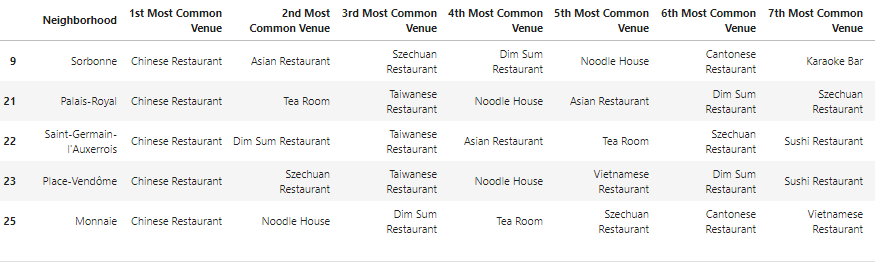
# Results analysis

*K-means* cluster our data into 5 clusters and return those items.

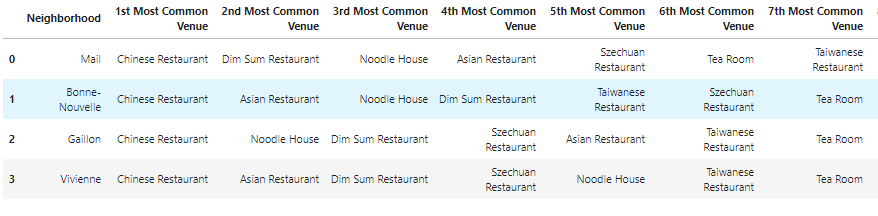
## Cluster 0



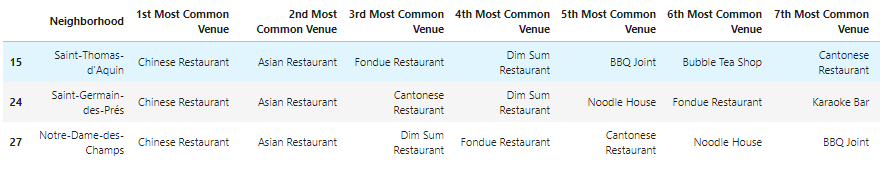
## Cluster 1



## Cluster 2



## Cluster 3



## Cluster 4



Based on dataframe analysis above

* **Cluster 0 (Invalides, Ecole-Militaire, Gros-Caillou**) and
* **Cluster 3 (Saint-Thomas d'Acquin, Saint-Germain-des-Près, Notre-Dame-Des-Champs)**

areas are the best places to open a new Chinese restaurant business.

# Discussion

In this section, I will focus on what we have noticed and the recommendation that I can make based on previous results.

* This analysis is performed on one hand on limited data and on other on reduced perimeter. We have focused our research around Louvre, Bourse, Temple, Hôtel-de-Ville, Panthéon, Luxembourg and Palais-Bourbon. This may be right or not, but if good amount of data were available there is scope to come up with better results.
* There is high competition around Notre-Dame, Jardins-des-Plantes and Val-de-Grace so it is very risky to open business in these areas.
* Sorbonne has also potential where it’s closes to University and high schools.
* We can more detailed analysis by adding some factors such as transportation, demographics of inhabitants.

Finally, *Foursquare API* proved to be a good source of data but frustrating at times. Despite having a Developer account, I regularly exceeded my hourly limit locking me out for the day.

# Conclusion

Although the goals of this project were met there is a lot to do to improve it.

In this study, I analyzed the relationship between business development and some useful machine learning technics such as clustering to group a set of Chinese restaurants in such a way that those restaurants in the same group (called a **cluster**) are more similar (in some sense) to each other than to those in other groups (clusters).

I also performed data analysis through this information by adding geographical coordinates of districts and some Chinese restaurants as static data on GitHub. In future studies, these data can also be accessed dynamically from specific platforms or packages.

I ended the study by visualizing the data and clustering information on the Paris map. In future studies, web or telephone applications can be carried out to direct investors.

With these analyses and information’s, my client can choose the place where open his Chinese restaurant.

# References

* [Forsquare API](https://developer.foursquare.com/) <https://developer.foursquare.com/>
* Coursera <https://www.coursera.org/professional-certificates/ibm-data-science>
* Paris <https://www.wikiwand.com/en/Paris>
* Arrondissements of Paris <https://www.wikiwand.com/en/Arrondissements_of_Paris>
* Paris Neighborhoods <https://www.parisinsidersguide.com/paris-neighborhoods.html>

1. The **Foursquare API** provides location-based experiences with diverse information about venues, users, photos, and check-ins. [↑](#footnote-ref-1)
2. **Foursquare API** categories list are available at https://developer.foursquare.com/docs/resources/categories [↑](#footnote-ref-2)
3. https://www.wikiwand.com/en/K-means\_clustering [↑](#footnote-ref-3)
4. https://www.wikiwand.com/en/One-hot [↑](#footnote-ref-4)