# Neighborhood and Venue Data Analysis of Paris

Monil Tushar Shah

June 9, 2020

## Introduction

Paris is the capital of France and the most populous city in the country. Being the City of Light it is the world's most visited capital. The city is proud of its many monuments from the iconic Eiffel Tower to the lofty Notre-Dame cathedral and the majestic Arc de Triomphe. No doubt this is Europe's most enchanting capital. There are around 80 neighborhoods in paris. Each neighborhood may have different popular venues, and thus if we can find similiar neighborhoods and their popular venue categories, it can aid businesses in making investment decisions.

The target audience for this project are Entrpreneurs, Investors, or Organizations who are planning to expand their business in the city of Paris.

## Data Description

First, we have to find out all the neighborhoods of Paris city. We will download the relevant data from trusted sources like the french government website and other open data sources provided by local authorities of Paris city. Further, we would need location data (postal codes, coordinates) of each of these neighborhoods, which we will obtain using the Geopy Library. Lastly, to find the nearby venues of each neighborhood using their location coordinates we will utilize the Foursquare API.

## Methodology

The main aim of this project is to find the most popular and common venues in each neighborhood of Paris and cluster them based on similar categories of venues. Doing so will help us recognizing similar neighborhoods, which can help Businesses/Organizations and Entrepreneurs make smart and data-driven investment decisions in Paris that are profitable for their business ventures.

For example, let's assume an investor wants to start a French restaurant in Paris. The criteria used to select a site are the following:

- Neighborhoods with fewer competitors are preferred.
- Neighborhoods, where supermarkets are easily accessible for food supplies are preferable.

We can suggest suitable neighborhood based on the criteria above.

I utilized the K-Means Algorithm to find out the Clusters of Neighborhood with similar Venues. K-Means clustering is a method of vector quantization, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

I have used used elbow method to find the optimal number of K. After analyzing using elbw method, the optimal K value I got was 3. So 3 clusters were formed whose results We will discuss in the result section of this report.

# Result

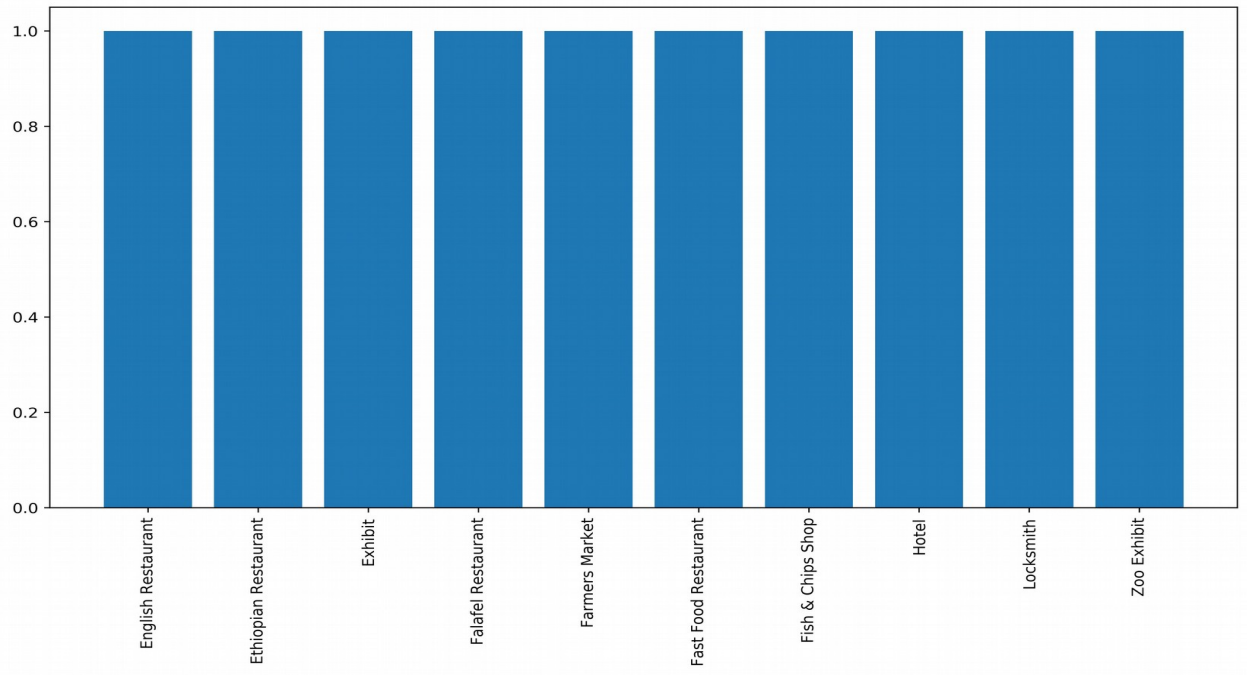
As discussed in the earlier sections of this report, our result consists of 3 different clusters on the basis of similarity in venues. As you can see in the below results, we were able to find the top 10 most common venues for each of the Neighborhoods. Not only that but we were also able to figure out popular venues in each cluster. I also created bar-graphs depicting the popular venue categories in each cluster v/s their count.

## Cluster-1

Cluster-1 is the smallest with only one neighborhood. It does contain any valuable information for our analysis.

| | Neighborhood | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 45 | Picpus | Locksmith | Hotel | Zoo Exhibit | Fish & Chips Shop | English Restaurant | Ethiopian Restaurant | Exhibit | Falafel Restaurant | Farmers Market | Fast Food Restaurant |

Bar-graph of venues in cluster-1 v/s their count:

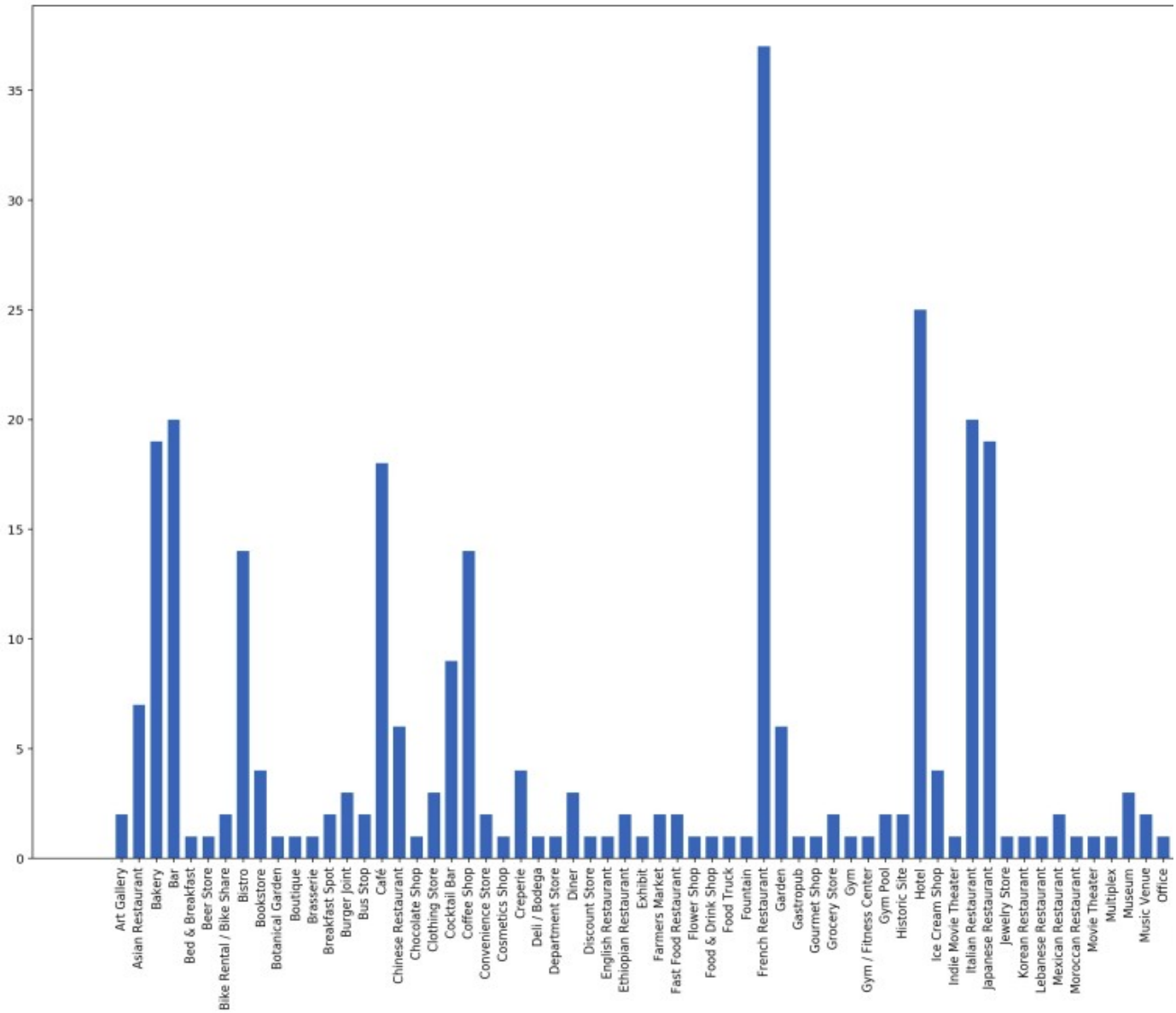



## Cluster-2

Top 10 most common venues for each of the Neighborhoods in Cluster-2.

| | Neighborhood | 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 |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | Palais-Royal | Japanese Restaurant | French Restaurant | Hotel | Italian Restaurant | Plaza | Historic Site | Wine Bar | Bookstore | Pastry Shop |
| 4 | Gaillon | Japanese Restaurant | Hotel | French Restaurant | Chocolate Shop | Wine Bar | Italian Restaurant | Korean Restaurant | Theater | Jewelry Store |
| 5 | Bonne-Nouvelle | French Restaurant | Cocktail Bar | Wine Bar | Bakery | Italian Restaurant | Hotel | Japanese Restaurant | Coffee Shop | Restaurant |
| 6 | Mail | French Restaurant | Cocktail Bar | Wine Bar | Bakery | Hotel | Bar | Creperie | Coffee Shop | Boutique |
| 7 | Vivienne | French Restaurant | Japanese Restaurant | Hotel | Wine Bar | Coffee Shop | Bistro | Bookstore | Bakery | Salad Place |
| 8 | Archives | French Restaurant | Hotel | Coffee Shop | Art Gallery | Italian Restaurant | Bar | Burger Joint | Bistro | Tapas Restaurant |
| 9 | Arts-et-Métiers | French Restaurant | Cocktail Bar | Hotel | Wine Bar | Italian Restaurant | Restaurant | Moroccan Restaurant | Bar | Japanese Restaurant |
| 10 | Enfants-Rouges | Hotel | French Restaurant | Wine Bar | Café | Japanese Restaurant | Sandwich Place | Vietnamese Restaurant | Bakery | Italian Restaurant |
| 11 | Sainte-Avoie | French Restaurant | Art Gallery | Café | Hotel | Coffee Shop | Chinese Restaurant | Restaurant | Bakery | Vietnamese Restaurant |

Bar-graph of venues in cluster-2 v/s their count:

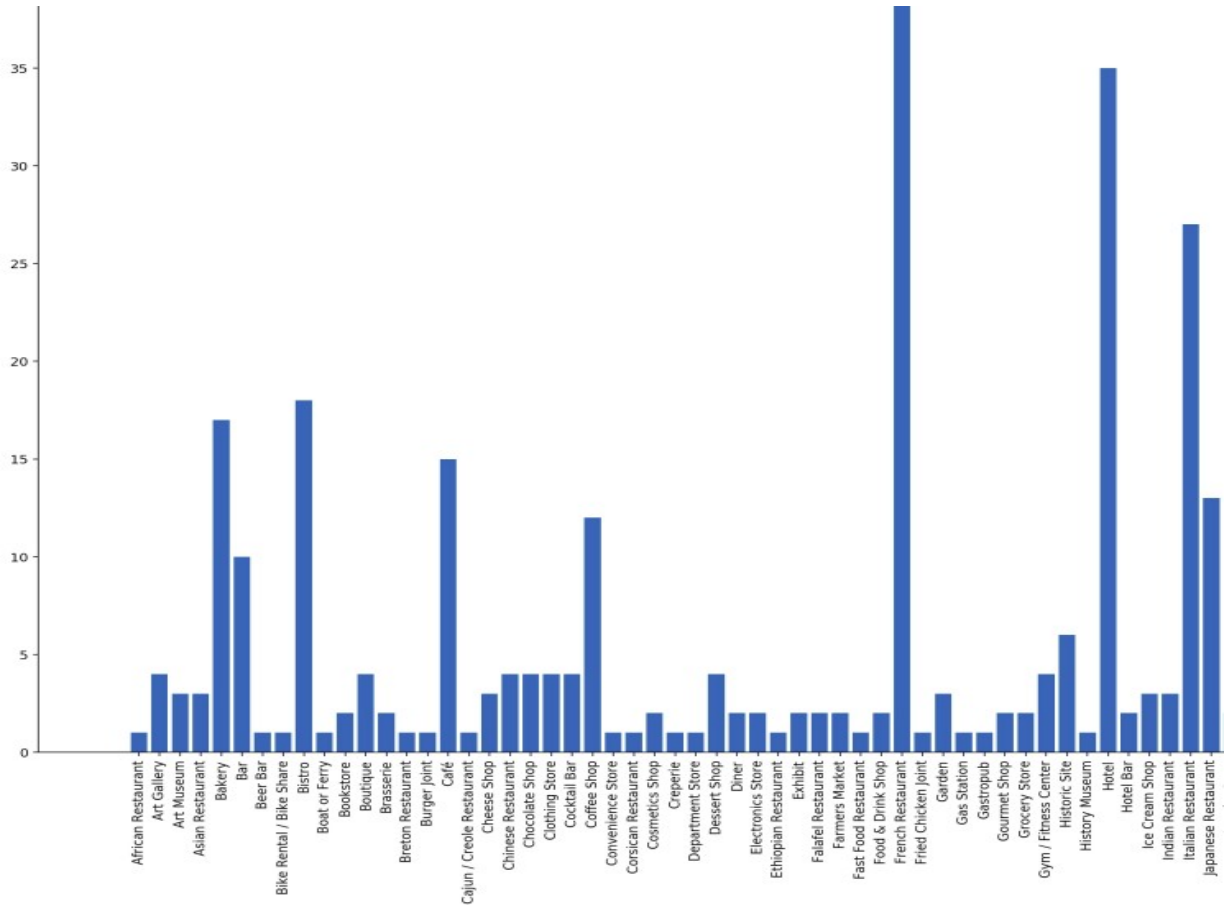


## Cluster-3

Top 10 most common venues for each of the Neighborhoods in Cluster-3.

| Neighborhood | 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 |
|---|---|---|---|---|---|---|---|---|---|---|
| Halles | French Restaurant | Bakery | Hotel | Ice Cream Shop | Art Gallery | Cocktail Bar | Pedestrian Plaza | Cheese Shop | Coffee Shop | Plaza |
| Place-Vendôme | French Restaurant | Hotel | Jewelry Store | Chocolate Shop | Japanese Restaurant | Boutique | Dessert Shop | Bookstore | Hotel Bar | Italian Restaurant |
| Saint-Germain-l'Auxerrois | French Restaurant | Hotel | Plaza | Art Museum | Historic Site | Coffee Shop | Italian Restaurant | Cosmetics Shop | Exhibit | Clothing Store |
| Arsenal | French Restaurant | Hotel | Plaza | Seafood Restaurant | Park | Tapas Restaurant | Pedestrian Plaza | Cocktail Bar | Gastropub | Vegetarian / Vegan Restaurant |
| Saint-Merri | French Restaurant | Ice Cream Shop | Art Gallery | Café | Plaza | Historic Site | Italian Restaurant | Coffee Shop | Gourmet Shop | Pedestrian Plaza |
| Saint-Victor | French Restaurant | Hotel | Bakery | Italian Restaurant | Bar | Plaza | Japanese Restaurant | Bistro | Museum | Pub |
| Notre-Dame-des-Champs | French Restaurant | Hotel | Bakery | Japanese Restaurant | Italian Restaurant | Bistro | Steakhouse | Creperie | Cosmetics Shop | Chocolate Shop |
| Saint-Germain-des-Prés | French Restaurant | Italian Restaurant | Hotel | Plaza | Wine Bar | Café | Sandwich Place | Historic Site | Boutique | Pastry Shop |
| Saint-Thomas-d'Aquin | French Restaurant | Hotel | Italian Restaurant | Art Gallery | Café | Historic Site | Bakery | Cheese Shop | Tailor Shop | Bistro |

Bar-graph of venues in cluster-3 v/s their count:



## Discussion

One of the drawbacks of the K-Means algorithm is because the initial centroids are chosen randomly, it will likely give different results each time it is run. Furthermore, different values of k should be experimented with to find different results.

K-Means is sensitive to outliers and works well only with the spherical shape of clusters. To overcome this issue, a clustering algorithm like K-Medoids can be used which is robust to outliers and reduces noise as compared to K-Means. K-Medoids has it's own pros and cons as well. You can use clustering algorithms other than the two mentioned here and experiment to see which algorithm performs well on your data.

Also, Paris being the most populous city in the country, additional factors like population density and median household income of neighborhoods can be taken into consideration to achieve better results from clustering.

As a further application, a platform can be developed which provides such information and visualizations achieved in this project to provide services to Investors, Businesses, and the City administration department.

## Conclusion

To conclude, from the above clusters it can be derived that Cluster-1 does not provide any valuable information. Cluster-2 seems more popular for Restaurants, Cafes, Wine Bar, Bistros, and SuperMarkets. Cluster-3 also being popular for restaurants includes Vegetarian/Vegan restaurants in contrast to Cluster-2 (having none) and other common venues like Parks, Plaza, Clothing Stores, Gym and Fitness centers, and Historic Sites.

Also, city administrators can manage the city more effectively by performing similar data analysis from time to time.