# Theatres and Entertainment venues

# Data Analysis of Paris

# 1. Introduction/Business Understanding

## 1.1 Description of the problem

The business problem we are currently helping clients to solve is: how a choose a district to open a new theatre in Paris.

## 1.2 Discussion of the background

Whatever 'golden era' means to you, the roaring twenties, the swinging sixites.... Paris theatres attract visitor worldwide not only because of their high-quality shows not also because of their ultra-rich history. Discovering a play or musical in a historical Parisian monument is a great alternative to the Parisian monuments/museums overrated tourist traps. Visitors might discover a theatre

commissioned by Queen Marie Antoinette, a modern theatre-cinema which hosted Orson Welles and Picasso; We could go on and on!

For those interested in learning about the French way of life, there's truly no better way to do so than by experiencing it first-hand, alongside a Parisian audience: think shows about Edith Piaf (complete with her repertoire), Molière's and Rostand's most famous masterpieces, and modern and sometimes risque, quintessentially Parisian vaudevilles. Theatre in Paris is the answer to all your burning questions about the French!

Our clients, investors in entertainment -related industry, would like to choose one among the 20 districts to open a new theatre.

## Questions that can be asked using the above-mentioned datasets¶
- •       How many theatres and entertainment venues in Paris?
- •       How are they located in Paris?
- •       Which areas have large number of theatres?
- •       Do existed theatres or entertainment venues impact our choice?
- •       What is the best area if open a new theatre?

## Data requirements
For this project we need the following data:

a. List of Theatres and entertainment venues in Paris that contains list Locality and district information.
Data source:
https://en.wikipedia.org/wiki/List_of_theatres_and_entertainment_venues_in_Paris
Description: This data set contains the required information.

b. 20 district geometry information with latitude and longitude.
Data source:   http://www.cartesfrance.fr
Description: we will use this data set to explore district locality of Paris. No direct listed information available, we got this information and formed a csv file.

c. Nearby places in each locality of Paris.
Data source: Fousquare API
Description: By using this api we will get all the venues in each district.

# 3. Methodology

## 3.1 Data Preparation

### 3.1.1 List of theatres and entrainment venues in Paris from Wikipedia:

We first used List_of_theatres_and_entertainment_venues_in_Paris from Wikipedia to scrap the table to create a data-frame. For this, we used pandas to transform the data in the table on the Wikipedia page into a dataframe containing 138 names of the theatres, District (Arrt), Present use. We started as below:

| | Name | Address | Arrt | Opened | Seats | Present use | Notes |
|---|---|---|---|---|---|---|---|
| 0 | AccorHotels Arena | NaN | 12th | 1984 | 17000 | concerts | formerly Palais omnisports de Paris-Bercy and ... |
| 1 | Aktéon | NaN | 11th | 1986 | 60 | theatre (plays), concerts | NaN |
| 2 | Alhambra | NaN | 11th | 2008 | 530 | music hall | formerly Théâtre Art Déco des Cheminots |
| 3 | Artistic-Athévains | NaN | 11th | 1913 | 220 | theatre (plays) | formerly Folies Artistic, Artistic Concert, Ar... |
| 4 | Ateliers Berthier | NaN | 6th | 2003 | 500 | theatre (plays) | second stage of the Théâtre de l'Odéon |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 133 | Le Trabendo | 211, avenue Jean Jaurès | 19th | NaN | 700 | concerts | NaN |
| 134 | Le Trianon | NaN | 18th | 1895 | 1000 | concerts | formerly Trianon-Concert,Trianon-Théâtre, Théâ... |
| 135 | Les Trois Baudets | NaN | 18th | NaN | 250 | concerts, chanson | NaN |
| 136 | Vingtième Théâtre | NaN | 20th | NaN | 245 | general use | NaN |
| 137 | Le Zèbre de Belleville | NaN | 11th | 1939 | 199 | cabaret, concerts | NaN |

### 3.1.2 Getting Coordinates of the 20 Districts:

Next objective is to get the coordinates of these 20 districts using information found from http://www.cartesfrance.fr , there wasn't any synthetic information, got this information manually and formed a csv file.

|    | Arrt | Latitude | Longitude |
|----|------|----------|-----------|
| 0  | 1th  | 48.864   | 2.331     |
| 1  | 2th  | 48.867   | 2.34      |
| 2  | 3th  | 48.864   | 2.362     |
| 3  | 4th  | 48.856   | 2.356     |
| 4  | 5th  | 48.846   | 2.345     |
| 5  | 6th  | 48.851   | 2.333     |
| 6  | 7th  | 48.857   | 2.32      |
| 7  | 8th  | 48.878   | 2.318     |
| 8  | 9th  | 48.872   | 2.34      |
| 9  | 10th | 48.872   | 2.357     |
| 10 | 11th | 48.858   | 2.38      |
| 11 | 12th | 48.841   | 2.388     |
| 12 | 13th | 48.832   | 2.356     |
| 13 | 14th | 48.833   | 2.327     |
| 14 | 15th | 48.841   | 2.3       |
| 15 | 16th | 48.863   | 2.276     |
| 16 | 17th | 48.884   | 2.322     |
| 17 | 18th | 48.892   | 2.345     |
| 18 | 19th | 48.883   | 2.382     |
| 19 | 20th | 48.865   | 2.399     |

### 3.1.3 Folium

We used python folium library to visualize geographic details of Paris and its 20 districts and I created a map of Paris. I used latitude and longitude values to get the visual as below:
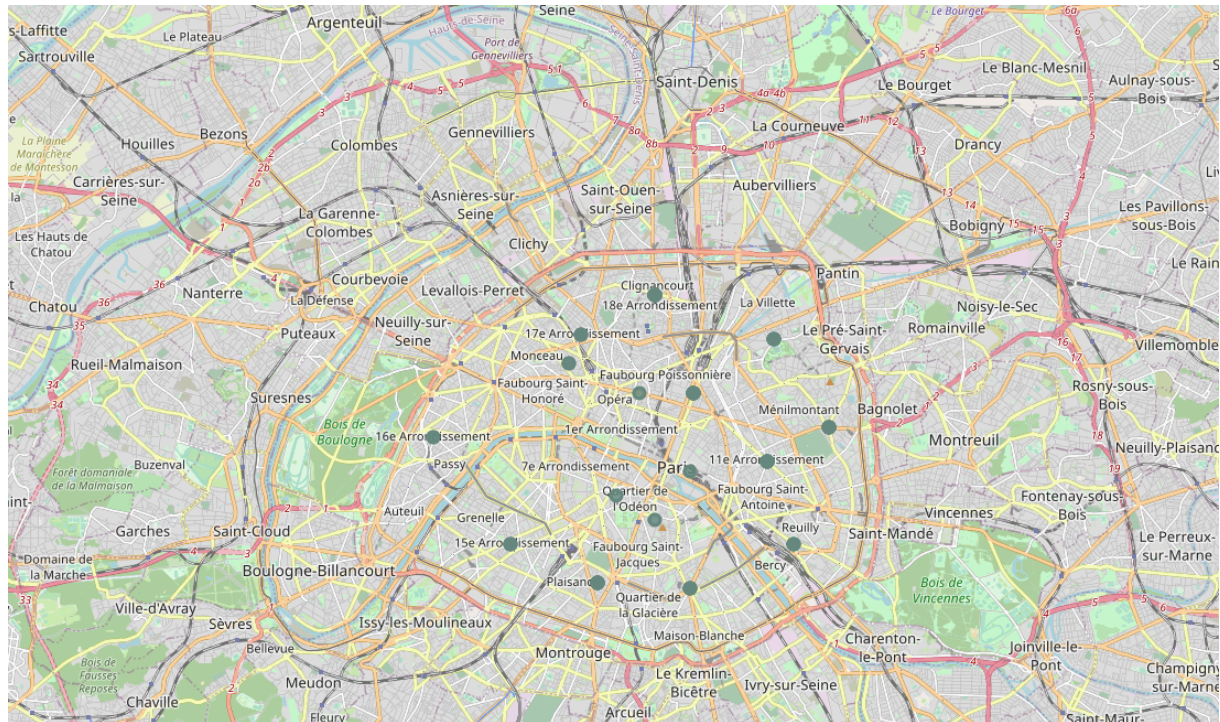
## 3.2. Exploratory Data Analysis:

### 3.2.1 Create a data-frame with pandas one hot encoding for the Present use.

| | Arrt | Asian theatre | cabaret | cabaret, concerts | chanson | comedy | concerts | concerts, chanson | dance, theatre (plays) | general | ... | music hall | music, dance and drama | night club on boat | one-man shows | opera | theatre (plays) | theatre (plays), concerts | theatre (plays), dance | theatre (plays), general | theatre (plays), musicals |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 12th | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 12th | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 12th | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 11th | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 4 | 11th | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 73 | 15th | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 74 | 15th | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 75 | 5th | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 76 | 16th | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 77 | 16th | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | Arrt | Asian theatre | cabaret | cabaret, concerts | chanson | comedy | concerts | concerts, chanson | dance, theatre (plays) | general | ... | music hall | music, dance and drama | night club on boat | one-man shows | opera | theatre (plays) | theatre (plays), concerts | theatre (plays), dance | theatre (plays), general | theatre (plays), musicals |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 10th | 0.000000 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.111111 | 0.0 | 0.0 | 0.111111 | ... | 0.000000 | 0.0 | 0.00 | 0.111111 | 0.000000 | 0.555556 | 0.000000 | 0.000000 | 0.0 | 0.111111 |
| 1 | 11th | 0.090909 | 0.0 | 0.090909 | 0.0 | 0.0 | 0.181818 | 0.0 | 0.0 | 0.090909 | ... | 0.090909 | 0.0 | 0.00 | 0.090909 | 0.000000 | 0.090909 | 0.090909 | 0.090909 | 0.0 | 0.000000 |
| 2 | 12th | 0.000000 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.333333 | 0.0 | 0.0 | 0.333333 | ... | 0.000000 | 0.0 | 0.00 | 0.000000 | 0.333333 | 0.000000 | 0.000000 | 0.000000 | 0.0 | 0.000000 |
| 3 | 13th | 0.000000 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.000000 | ... | 0.000000 | 0.0 | 0.25 | 0.000000 | 0.000000 | 0.500000 | 0.000000 | 0.250000 | 0.0 | 0.000000 |
| 4 | 14th | 0.000000 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.111111 | ... | 0.111111 | 0.0 | 0.00 | 0.000000 | 0.000000 | 0.777778 | 0.000000 | 0.000000 | 0.0 | 0.000000 |

### 3.2.2 Output each district ( Arrt ) along with the top 5 most common venues:

| | Arrt | theatre (plays) |
|---|------|-----------------|
| 0 | 10th | 0.555556 |
| 1 | 11th | 0.090909 |
| 2 | 12th | 0.000000 |
| 3 | 13th | 0.500000 |
| 4 | 14th | 0.777778 |

3.2.3 Use clustering (KMeans)

These prescriptive analytics to help investors decide a location to open a theatre.

K vs Error
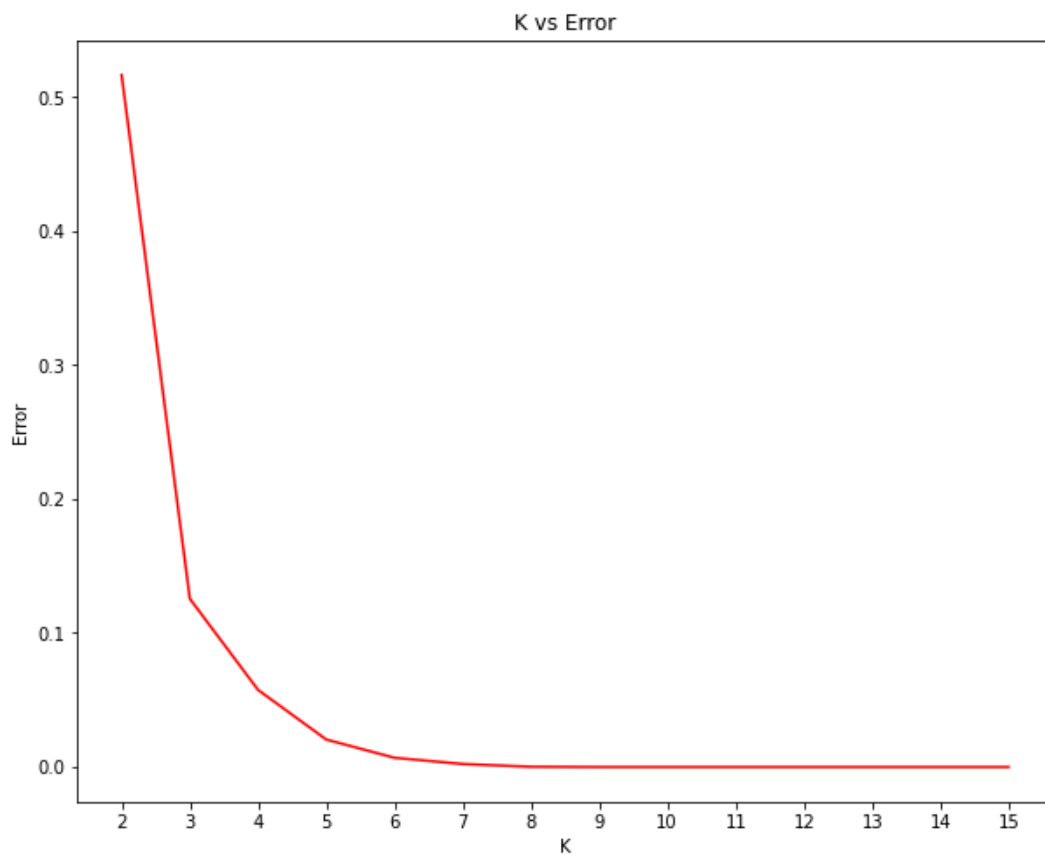


Finally, we try to cluster these 20 districts based on the present use and use K-Means clustering. So, our expectation would be based on the similarities of present use, these districts will be clustered. We have used the code below:

```
# sort the results by Cluster Labels
print(to_merged.shape)
to_merged.sort_values(["Cluster Labels"], inplace=True)
to_merged
```

(78, 7)

]:

| | Arrt | theatre (plays) | Cluster Labels | Name | Present use | Latitude | Longitude |
|---|---|---|---|---|---|---|---|
| 0 | 10th | 0.555556 | 0 | Théâtre des Bouffes du Nord | theatre (plays) | 48.872 | 2.357 |
| 3 | 13th | 0.500000 | 0 | Théâtre 13 | theatre (plays) | 48.832 | 2.356 |
| 3 | 13th | 0.500000 | 0 | Théâtre du Lierre | theatre (plays), dance | 48.832 | 2.356 |
| 5 | 15th | 0.333333 | 0 | Palais des Sports | general | 48.841 | 2.3 |
| 5 | 15th | 0.333333 | 0 | Théâtre Silvia-Monfort | theatre (plays), general | 48.841 | 2.3 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 4 | 14th | 0.777778 | 2 | Théâtre le Guichet Montparnasse | theatre (plays) | 48.833 | 2.327 |
| 4 | 14th | 0.777778 | 2 | Théâtre Montparnasse | theatre (plays) | 48.833 | 2.327 |
| 4 | 14th | 0.777778 | 2 | Théâtre Rive Gauche | theatre (plays) | 48.833 | 2.327 |
| 4 | 14th | 0.777778 | 2 | La Comédie Italienne | theatre (plays) | 48.833 | 2.327 |
| 4 | 14th | 0.777778 | 2 | Théâtre de la Gaîté-Montparnasse | theatre (plays) | 48.833 | 2.327 |

78 rows × 7 columns

We can represent these clusters in a leaflet map using Folium library as below:
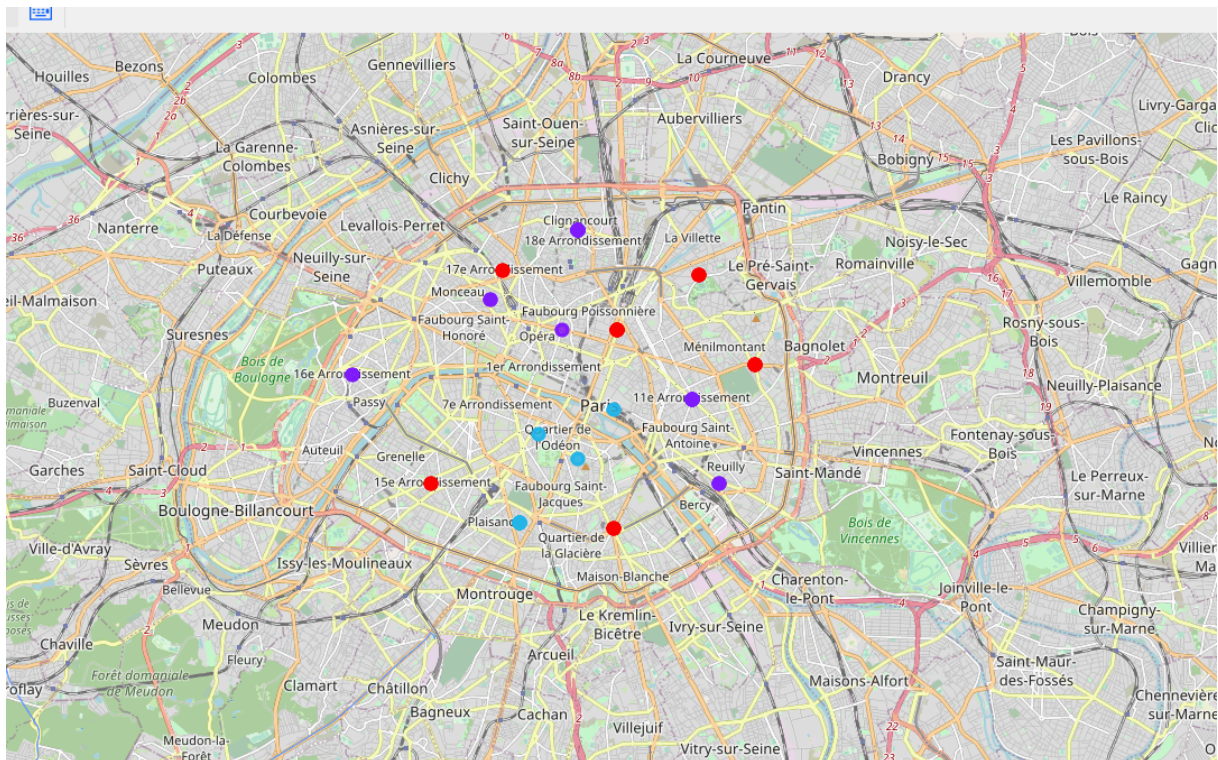
# 4. Results

We got a glimpse of the Theatres and entrainments venues in Paris  and were able to find out some interesting insights which might be useful to investors with business interests. Let's summarize our findings:

- 138 Theatres and entrainments venues in Paris.
- 26 Play Theatres in Paris.
- Play Theatres are widely located in all Paris , there's no a dominated district.
- 13th and 15th have the least Play Theatres, which might be a wise choice for our investors.

## 5. Discussion Section

The major purpose of this project, is to suggest a better district in Paris to open a play theatre. Social presence in society in terms of like-minded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

This Capstone project can be continued for making it more precise in terms to the connectivity of each theatres and find out how businesses impact each other (cafés, restaurants, stores and theatres) and build a better social-based urban ecosystem.

## *6. Conclusion*

*Working through this Capstone project, I found that there are many real-life problems or scenarios where data can be used to find solutions to those problems. Like seen in the example above, data was used to analysis entertainment industry in the 20 districts in Paris, to better fit people's needs as well as those of investors. The results can help to decide about the district that fit the most his needs.*

*I feel rewarded with the efforts and believe this course with all the topics covered is well worthy of appreciation.*

*I have made use of some frequently used python libraries to scrap web-data, use Foursquare API to explore the districts information in Paris and saw the results of segmentation of districts using Folium leaflet map.*

*This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools.*

***Libraries Which are Used to Develope the Project:***

Pandas: For creating and manipulating dataframes.

Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

XML: To separate data from presentation and XML stores data in plain text format.

Geocoder: To retrieve Location Data.

Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.


GitHub Link of Complete Project :  https://github.com/daily92000/Capstone-Project---The-Battle-of-Neighborhoods

Jupyter Notebook Link for code : https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/9ccb95cf-a50f-4097-8655-84d60c799d9a?projectid=8ad0418b-1fa1-4959-9976-24365fe6b44f&context=cpdaas