

# ART(music, museums, galleries) venues clustering in Paris

---

ALYAA SELIM



Introduction:

*Background*

Paris boasts some of the most famous and important art galleries and museums in the world and is the home of priceless iconic artworks dating back thousands of years, to modern and contemporary artists.

### *Business Problem*

The purpose of this project is to help people in exploring the best and most famous Art place in Paris. It will help people making smart and efficient decisions in saved money and time

Paris is known as the city of art and many art lovers would like to visit this project is for those people who are looking for the best Art place

This analysis seeks to present the distribution of art places referenced on Foursquare, with a summary analysis on their geographical distribution and with respect to their category.

### *Data*

#### *Data collection stage*

The main points with their location and category will be retrieved from Foursquare databases using their URL API.

also <http://tour-pedia.org/about/datasets.html> Datasets TourPedia contains two main datasets, which belong to the specific domain of tourism:

Places Reviews about places License Tourpedia is released under the Creative Commons CCZero license.

Places Places contain accommodations, restaurants, attractions and points of interest. Places were retrieved from the following social media: Facebook, Foursquare, Google Places, and Booking.

The following table shows the description of each place

These data set could help better to cluster their venues. From the Foursquare API,

.The following table shows the description of each place.

Field Description id the unique identifier of the place name name of the place (e.g. a hotel name)

Address address of the place category one among accommodation, attraction,

restaurant, poi (point of interest) location one among Rome, Amsterdam, London, Paris, Berlin, Dubai, Barcelona,

Tuscany lat Latitude

lng Longitude

Services the list of services provided by the place. It is set only if the place is an accommodation.

phone\_number national phone number associated to the place

international\_phone\_number international phone number

associated to the place website URL of the web site associated to the

place Icon picture associated to the place description description of

the place in the six languages of the OpENER project external\_urls

external URLs associated to the place.

It contains the URLs of Foursquare, Facebook, GooglePlaces and Booking (the last one is present only whether the place is an accommodation) statistics associated to the place;

they are retrieved from Foursquare and Facebook subCategory The category provided by the source. It is more specific than the field category polarity The opinion about the place Reviews The collection Reviews contains reviews on the above-described places. The following table describes the schema of each review.

Field Description Id the unique identifier of the review Text the text of the review language The language of the review source one among GooglePlaces, Foursquare, Facebook rating Rating expressed by the user. Range is between 1 and 5 Time Date of the review wordsCount Number of words of the text analysis.kaf The result of the OpeNER pipeline in KAF analysis.json The result of the OpeNER pipeline in KAF-JSON polarity The polarity of the review. It is extracted from the Polarity tagger module place.id id of the place associated to the review place.name Name of the place associated to the review place.location Location of the place associated to the review place.category Category of the place associated to the review authorName The name of the review author

### ***Target audience***

Who love Paris and art

### ***Foursquare API:***

This project would use Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

### ***Work Flow:***

Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

### ***Clustering Approach:***

we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm

### ***Libraries Which are Used to Developed 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.

## Methodology

As a database , The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

Name , categories, Neighborhood, lat, Longitude , address

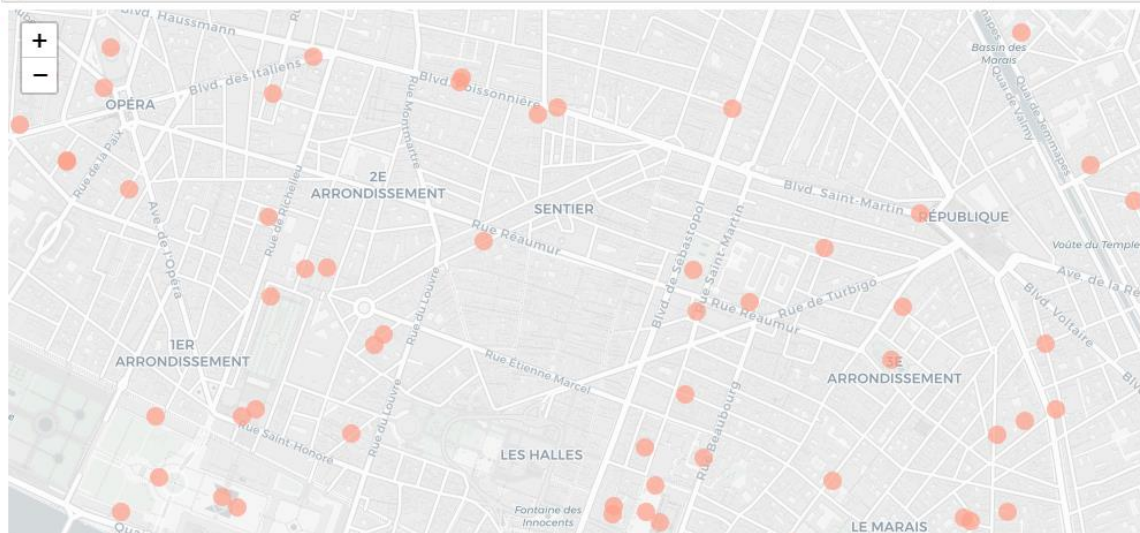
Out[37]:

	name	categories	address	lat	lng	neighborhood
0	La Gaîté Lyrique	Salle de concert	3 bis rue Papin	48.866673	2.353435	NaN
1	Centre Pompidou – Musée National d'Art Moderne	Musée d'art	Place Georges Pompidou	48.860730	2.351660	NaN
2	Galerie Daniel Templon	Galerie d'art	30 rue Beaubourg	48.862066	2.353825	NaN
3	Comedy Club	Salle de spectacle	42 boulevard de Bonne Nouvelle	48.870646	2.348376	NaN
4	Maison de Nicolas Flamel	Site historique	51 rue de Montmorency	48.863605	2.353130	NaN

I used python **folium** library to visualize geographic details of Istanbul and its boroughs and I created a map of Istanbul with boroughs superimposed on top. I used latitude and longitude values to get the visual as below:



[48]:



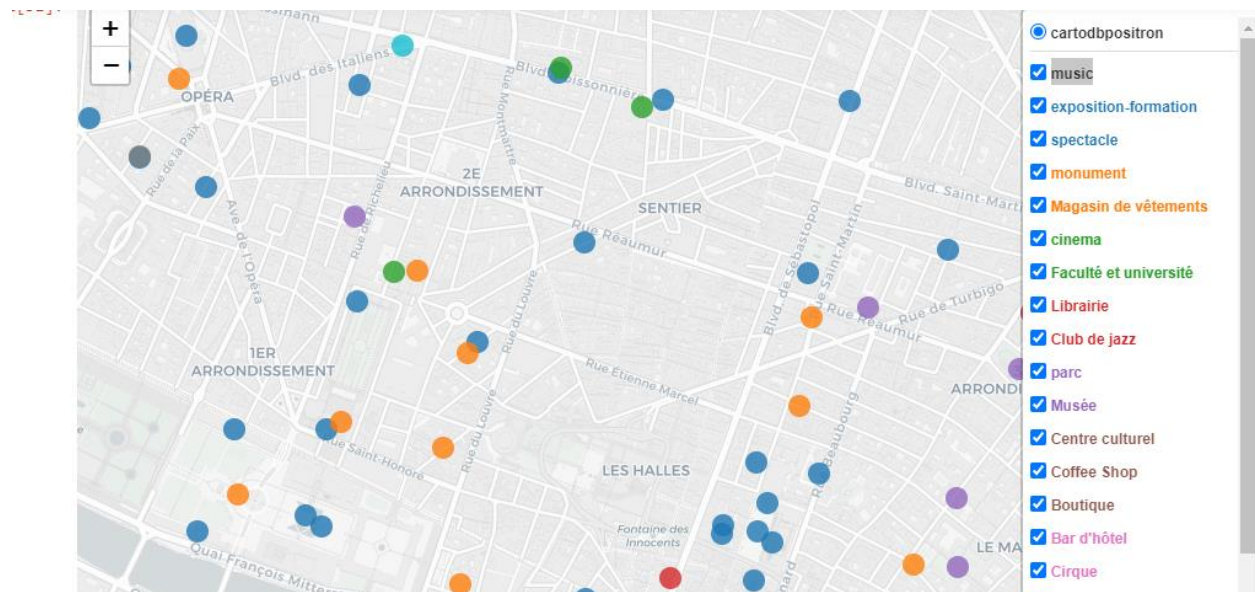
Take a look at all the different categories

In [45]: dataframe\_filtered.head()

Out[45]:

	name	categories	address	lat	lng	neighborhood	hicategories
0	La Gaîté Lyrique	music	3 bis rue Papin	48.866673	2.353435	NaN	show
1	Centre Pompidou – Musée National d'Art Moderne	exposition-formation	Place Georges Pompidou	48.860730	2.351660	NaN	exposition-formation
2	Galerie Daniel Templon	exposition-formation	30 rue Beaubourg	48.862066	2.353825	NaN	exposition-formation
3	Comedy Club	spectacle	42 boulevard de Bonne Nouvelle	48.870646	2.348376	NaN	show
4	Maison de Nicolas Flamel	monument	51 rue de Montmorency	48.863605	2.353130	NaN	monument

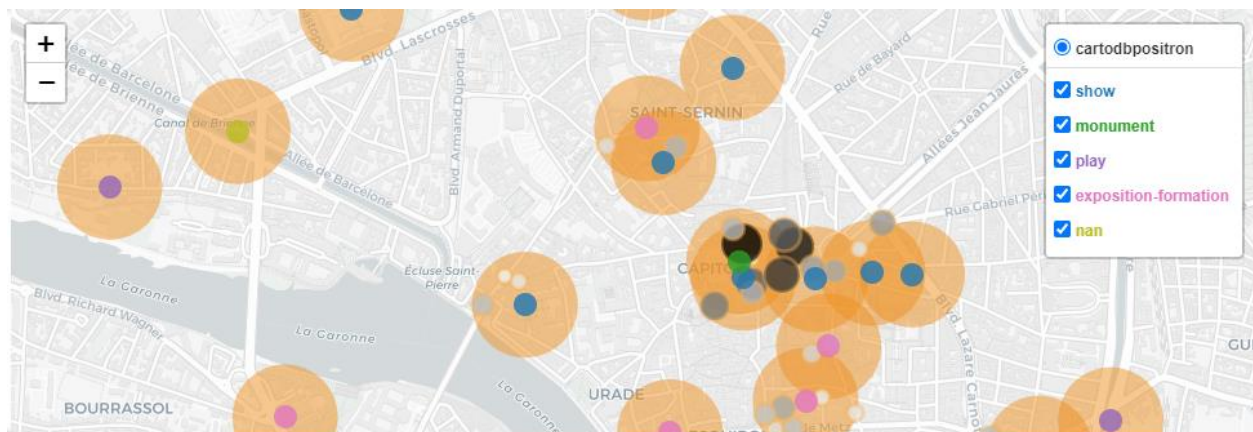
Then I do *Category color distinction*



Order to potentially add another useful value to the clustering algorithm, a pedestrian flow counts is read and add along venue locations dataset.

	address	category	id	lat	lng	location	name	subCategory
0	8 rue de Navarin	attraction	83256	48.87961034033	2.3393825451216	Paris	Hôtel Amour	Hotel Bar
1	49 Rue Pierre Charron	attraction	83321	48.869145508329	2.3021912516801	Paris	Pershing Hall	Hotel
2	99 rue Oberkampf	attraction	83355	48.8654975	2.3766887	Paris	Mécano-Bar	Bar
3	1 rue Théophile Roussel	attraction	83358	48.849622701075	2.377436673654	Paris	Le Baron Rouge	Wine Bar
4	7-15 Avenue de la Porte de la Villette	attraction	83363	48.899198480866	2.3867258494543	Paris	Glazart	Music Venue

## Exploratory visualization



#Next, let's group rows by category and by taking the mean of the frequency of occurrence of each category

I am a fan of art and history, and Paris is famous for its art galleries

Out[97]:

	subCategory	Art Gallery
0	Arcade	0
1	Art Gallery	1
2	Art Museum	0
3	Arts & Entertainment	0
4	Athletics & Sports	0

I used unsupervised learning **K-means algorithm** to cluster the boroughs. K-Means algorithm is one of the most common cluster method of unsupervised learning.

First, I will run K-Means to cluster the boroughs into **3** clusters because when I analyze the K-Means with elbow method it ensured me the 3 degree for optimum k of the K-Means.

Here is my merged table with cluster labels for each borough.

```
In [101]: # merge paris_grouped with toronto_data to add Latitude/Longitude for each subcategory
to_merged = to_merged.join(paris_venus.set_index("subCategory"), on="subCategory")

print(to_merged.shape)
to_merged.head()

(1522, 10)
```

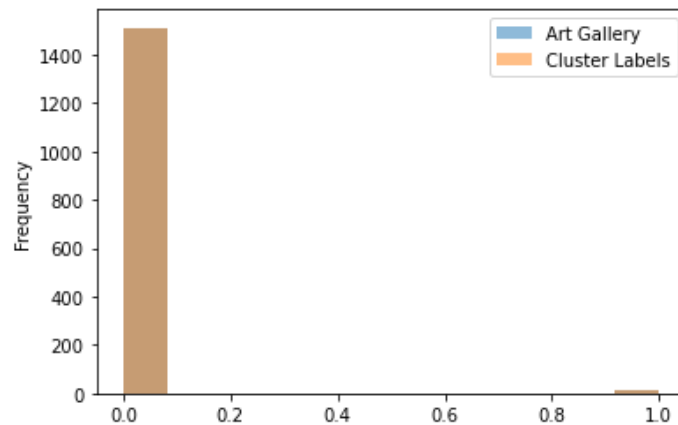
```
In [103]: to_merged
```

Out[103]:

	subCategory	Art Gallery	Cluster Labels	address	lat	lng	name
0	Arcade	0	0	10 rue des Immeubles Industriels	48.84953114483	2.3928544226342	Arcade Street
0	Arcade	0	0	Cour Bérard	48.854308646341	2.3662394285202	#Appartement4
1	Art Gallery	1	1	Parc de la Villette	48.890631905558	2.3900761307413	Pavillon Paul Delouvrier
1	Art Gallery	1	1	7 rond-point des Champs-Élysées	48.868614071405	2.3088401556015	Artcurial
1	Art Gallery	1	1	60 Rue de Bassano	48.871659203246	2.3003911972046	Espace Culturel Louis Vuitton
1	Art Gallery	1	1	2 Rue Charles Nodier	48.884372979613	2.3449059296629	Marché Saint-Pierre
1	Art Gallery	1	1	48 Rue Vieille du Temple	48.858238266037	2.358552493322	L'Espace des Blancs-Manteaux
1	Art Gallery	1	1	7bis rue René Goscinny	48.830801685139	2.3801647866044	Galerie Itinérance
1	Art Gallery	1	1	12 rue Cantagrel	48.823472446056	2.3764735307265	Cité-Refuge De L'Armée Du Salut Le Corbusier
1	Art Gallery	1	1	Berges de Seine	48.862290116178	2.2987466597392	Photoquai



```
Out[134]: <AxesSubplot:ylabel='Frequency'>
```



## Conclusion

I also performed data analysis through this information by adding the coordinates of districts and home sales price averages as static data on GitHub. In future studies, these data can also be accessed dynamically from specific platforms or packages.

Most of the art Galleries are in Cluster 1 especially Quai de Valmy,Roland Garros in lowest (close to zero) in Cluster 2 so if you from who like art the best place to visit is Quai de Valmy,Roland Garros

As a result, people are turning to big cities to start a business or work. For this reason, people can achieve better outcomes through their access to the platforms where such information is provided.

Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.