

## IBM Applied Data Science Capstone

# *Opening a New Hotel in Paris, France*

By: Vipul Sindha

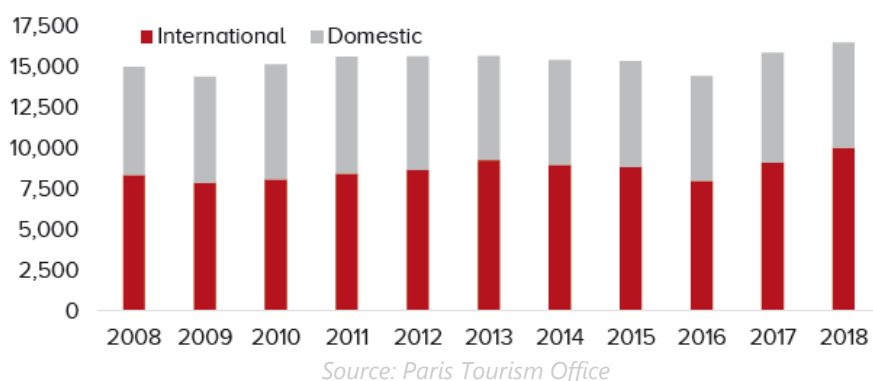
May 2020



## Introduction

Business environment in France has experienced positive change since start of 2019 with a decrease in unemployment rate and an increase in purchasing power, following the 'gilets jaunes' disruption. The Economist Intelligence Unit forecasts modest GDP growth to pick up to around 1.5% for the foreseeable future. In terms of tourism, Paris enjoys an exceptionally well-balanced mix of business and leisure demand, which allows the City of Lights to have both a broad seasonality and strong average rate. Paris has remained, alongside London, one of the two most desirable destinations for hotel investment for more than a decade.

Visitation to Paris is robust and has remained broadly static at 15-16.5 million over the past ten years. 2018 saw a further increase in visitation thanks to the international market (accounting for approximately 60%), especially with the return of US visitors – the number one source country – and thanks to city-wide events such as the 2018 Ryder Cup.



Paris is still one of the most attractive markets in Europe and, once again, achieved the highest price per room in HVI. The general resilience and strong fundamentals of the Paris market have led us to consider that a continued improvement in performance can be expected for the rest of 2020, despite a rocky start to the year. The organization of large events such as the 2023 Rugby World Cup and the 2024 Summer Olympic Games will certainly help maintain, and even improve, Paris's strong performance.



## Business Problem

Aspects considered by planners when a Hotel is to be built include feasibility of the site in terms of the traveler's demand, market conditions, property prices, land availability, adequate connectivity with airports, train stations; and size, access, and topography of the site, as well as availability of utilities, zoning laws, and land use in the immediate area. Economic conditions of the area, the sociology of the region, and local commercial competition and attitudes determine the size of Hotel that can be supported and the kind of accommodation acceptable to a given guest.

The objective of this capstone project is to analyze and select the best locations in Paris to open a new Hotel. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question:

In the city of Paris, if a property developer is looking to open a new Hotel, where would you recommend that they open it?

## Target Audience of this project

This project is particularly useful to property developers and investors looking to open or invest in new Hotels in the capital city of France i.e. Paris.

Paris's hotel investment market has remained liquid and sought-after, despite the recent turmoil. French REIT Gecina sold its entire portfolio of five hotels to Angelo Gordon and EQ Group for €181 million (€222,000 per room) in June 2019. In 2018, the most significant single hotel transaction was The Westin Paris – Vendôme, sold by Singapore sovereign wealth fund GIC to Henderson Park for €550 million (€1.3 million per room, the highest value per key European transaction in 2018). For the latest value trends, please refer to our annual *European Hotel Valuation Index*, which showed that Paris maintains by a significant margin the highest value per room compared to other European markets.

Property	Sale Date	Rooms	Estimated Price (€)	Per Room (€)
Staycity Courbevoie (forward sale)	Jul 2019	216	73,000,000	338,000
Gecina Hotels Portfolio (5 Hotels)	Jun 2019	814	181,000,000	222,000
Mama Shelter La Defense	Apr 2019	211	Undisclosed	
ibis Paris Tour Montparnasse 15eme	Jan 2019	45	Undisclosed	
ibis Styles Paris Montmartre Nord	Jan 2019	46	Undisclosed	
Phileas Hotel Gare Saint Lazare	Dec 2018	39	5,300,000	136,000
Hotel Lancaster	Nov 2018	56	Undisclosed	
Westin Paris	Oct 2018	428	550,000,000	1,285,000
66 Champs-Elysees - 128 Boetie	Jun 2018	109	80,000,000	734,000
Motel One Porte Dorée (forward sale)	Jan 2018	255	Undisclosed	

Source: HVS Research

## Data Section

To solve the problem, we will need the following data:

- List of neighborhoods in Paris. This defines the scope of this project, which is confined to the city of Paris.
- Latitude and longitude coordinates of those neighborhoods. This is required to plot the map and to get the venue data.
- Venue data, particularly data related to Hotels. We will use this data to perform clustering on the neighborhoods.

### Sources of data and methods of extraction

This Wikipedia page ([https://en.wikipedia.org/wiki/Arrondissements\\_of\\_Paris](https://en.wikipedia.org/wiki/Arrondissements_of_Paris)) contains a list of neighborhoods in Paris, with a total of 20 neighborhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and Pandas packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest databases of 105+ million places and is used by over 150,000 developers. Foursquare APIs provide many categories of the venue data, we are particularly interested in the Hotels category to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.

## Methodology

Firstly, we need to get the list of neighborhoods in the city of Paris. Fortunately, the list is available in the Wikipedia page ([https://en.wikipedia.org/wiki/Arrondissements\\_of\\_Paris](https://en.wikipedia.org/wiki/Arrondissements_of_Paris)). We will do web scraping using Python requests and Pandas packages to extract the list of neighborhoods data. However, this is just a list of names. We need to get the geographical coordinates in the form of latitude and longitude to be able to use Foursquare API. To do so, we will use the wonderful Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the data into a pandas Data Frame and then visualize the neighborhoods in a map using Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the city of Paris.

Next, we will use Foursquare API to get the top 100 venues that are within a radius of 2000 meters. We need to register a Foursquare Developer Account to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighborhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each neighborhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyze each neighborhood by grouping the rows by neighborhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analyzing the Hotels data, we will filter the “Hotel” as venue category for the neighborhoods.

Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighborhoods into 3 clusters based on their frequency of occurrence for “Hotel”. The results will allow us to identify which neighborhoods have higher concentration of Hotels while which neighborhoods have fewer number of Hotels. Based on the occurrence of Hotels in different neighborhoods, it will help us to answer the question as to which neighborhoods are most suitable to open new Hotel.

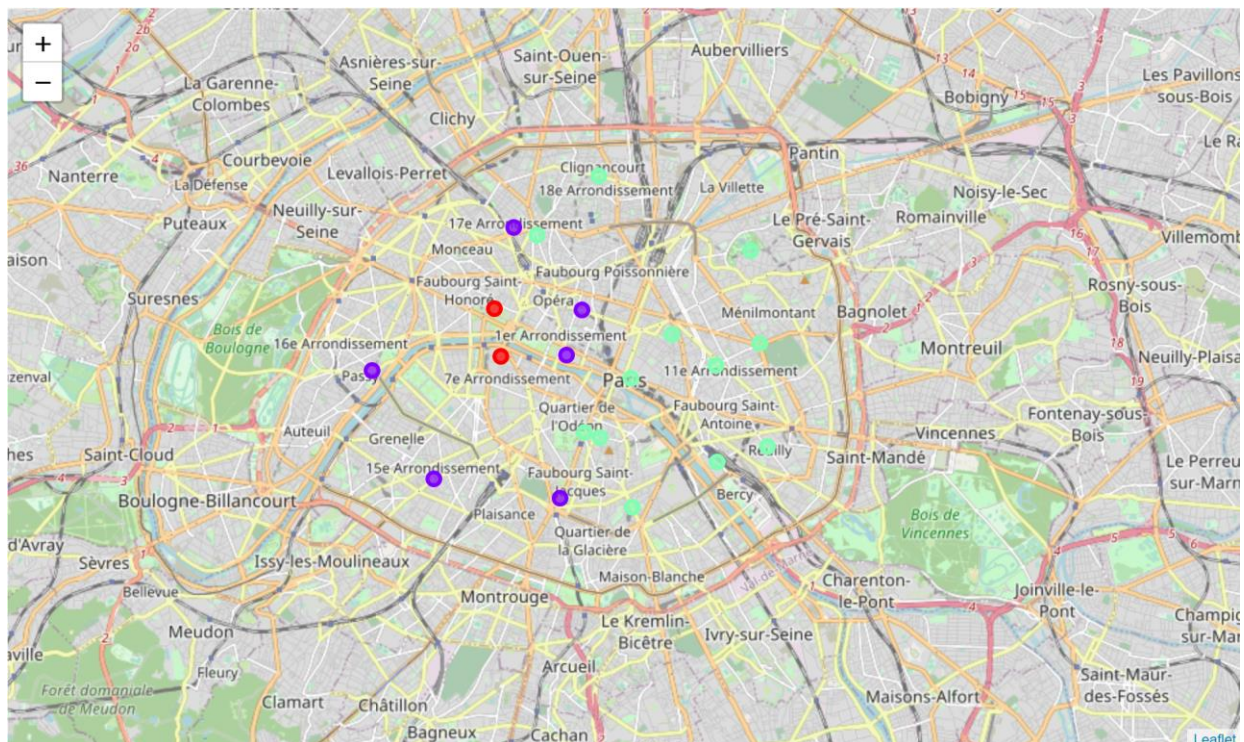


## Results

The results from the k-means clustering show that we can categorize the neighborhoods into 3 clusters based on the frequency of occurrence for “Hotels”:

- Cluster 0: Neighborhoods with moderate number of Hotels
- Cluster 1: Neighborhoods with high concentration of Hotels
- Cluster 2: Neighborhoods with low number to no existence of Hotels

The results of the clustering are visualized in the map below with cluster 0 in purple color, cluster 1 in mint green color, and cluster 2 in red color.



## Discussion

As observations noted from the map in the Results section, most of the Hotels are concentrated in the Central and Northern side of the river Seine in Paris city, with the highest number in cluster 1 and moderate number in cluster 0. On the other hand, cluster 2 has exceptionally low number to no Hotels in the neighborhoods. This represents a great opportunity and high potential areas to open new Hotel as there is extraordinarily little to no competition from Hotels. Meanwhile, Hotels in cluster 1 are likely suffering from intense competition due to oversupply and high concentration of Hotels. From another perspective, the results also show that the oversupply of Hotels mostly happened in the central and north – east area of the city, with the south – west areas still have very few Hotels. Therefore, this project recommends property developers to capitalize on these findings to open new Hotel in neighborhoods in cluster 2 with little to no competition. Property developers with unique selling offerings to guests to stand out from the competition can also open new Hotels in neighborhoods in cluster 0 with moderate competition. Lastly, property developers are advised to avoid neighborhoods in cluster 1 which already have high concentration of Hotels and suffering from intense competition.

## Limitations and Suggestions for Future Research

In this project, we only consider one factor i.e. frequency of occurrence of Hotels, there are other factors such as population density, nearby attractions, nightlife, nearby restaurants, prices per room, average expenditure of travelers, etc., that could influence the location decision of a new Hotel. However, to the best knowledge of this research such data analysis of the neighborhoods is not required for this project. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new Hotel. In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Future research could make use of paid account to bypass these limitations and obtain more results.

## **Conclusion**

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. property developers and investors regarding the best locations to open a new Hotel.

To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighborhoods in cluster 2 are the most preferred locations to open a new Hotel.

The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new hotel.



## References

Arrondissements of Paris. Retrieved from *Wikipedia*.

[https://en.wikipedia.org/wiki/https://en.wikipedia.org/wiki/Arrondissements\\_of\\_Paris](https://en.wikipedia.org/wiki/https://en.wikipedia.org/wiki/Arrondissements_of_Paris)

Foursquare Developers Documentation. *Foursquare*. Retrieved from

<https://developer.foursquare.com/docs>

Paris Hotel Market Report Retrieved from

<https://www.hvs.com/article/8579-paris-hotel-market-a-rocky-start-to-the-year>

French hotel industry locations inside Retrieved from

<https://tophotel.news/french-hotel-industry-sees-7-6-increase-in-foreign-guest-arrivals-in-2018/>

# Appendix

## Cluster 0:

- Batignolles-Monceau
- Bourse
- Passy
- Louvre
- Vaugirard
- Observatoire

## Cluster 1:

- Temple
- Reuilly
- Popincourt
- Panthéon
- Opéra
- Ménilmontant
- Hôtel-de-Ville
- Gobelins
- Entrepôt
- Buttes-Chaumont
- Butte-Montmartre
- Luxembourg

## Cluster 2:

- Palais-Bourbon
- Élysée