A SUMMER IN PARIS: PLACES YOU CAN'T AFFORD TO MISS

1. INTRODUCTION / BUSINESS PROBLEM

Tens of thousands of people visit the metropolitan city of Paris every year for tourism. Some of the famous sites they visit are monuments and museums such as "LeTour Eiffel", "Le Musee de Louvre", etc. When they do visit these sites, which do not often have other social facilities and relaxation centers, they sometimes miss out on some of the amazing ancillary facilities located near to these touristic sites such as the wide variety of restaurants, gift shops, etc. Some of the visitors do not even know the rating of the touristic sites and how they are spread out in the geographical region. In this project, I intend to inform visitors of the many amazing social amenities and ancillary services located close to the top 20 monuments/museums in Paris. This will enable them to better plan their trips, know which places to visit, and the proximity of these places to each other.

2. DESCRIPTION OF THE DATA

Foursquare location data will be used to explore the sites near the monuments/museums. Foursquare Labs Inc. commonly known as Foursquare is an American technology company which has a very reliable location platform. Their technology and data powers apps such as Apple's Maps, Uber, Twitter, Microsoft, Samsung, and over 100,000 other developers. Foursquare data such as the touristic site name, location, etc will be used. Data on the list of visitors per attraction published by the Paris office of Tourism on Wikipedia as shown below will also be used.

List of visitors per attraction [edit]

The 20 top Paris museums and monuments - (2007/2006 figures from the Paris Office of Tourism)[26]

Rank	Change 07/06	Museums and Monuments	2007	2006	Variation 07/06
1	=	Notre Dame de Paris	13,650,000	13,650,000	—
2	=	Basilique du Sacré-Cœur	10,500,000	10,500,000	—
3	=	The Louvre	8,260,000	8,348,000	-1.1%
4	=	Eiffel Tower	6,797,410	6,695,135	1.5%
5	=	Pompidou Centre	5,509,425	5,133,506	7.3%
6	+1	Musée d'Orsay	3,166,509	3,009,203	5.2%
7	-1	Cité des Sciences et de l'Industrie	3,030,628	3,055,000	-0.8%
8	=	Chapel of Our Lady of the Miraculous Medal	2,000,000	2,000,000	—
9	+1	Arc de Triomphe	1,543,295	1,330,738	16.0%
10	+2	Musée du Quai Branly	1,379,623	952,770	44.8%
11	-2	Muséum d'Histoire Naturelle	1,372,804	1,344,344	2.1%
12	-1	Musée de l'Armée	1,188,728	1,130,841	5.1%
13	=	Sainte Chapelle	866,982	833,392	4.0%
14	+3	Musée Grévin	762,000	682,000	11.7%

3. METHODOLOGY

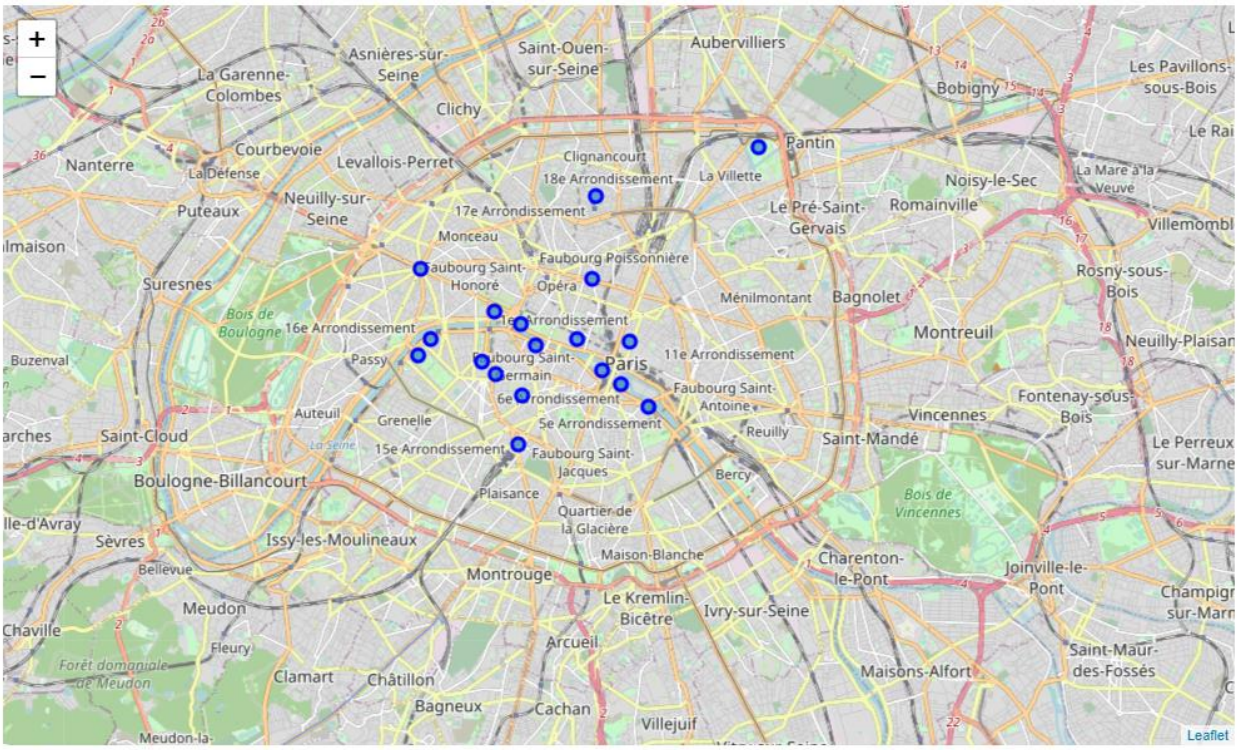
In this section, the data analysis process which led to the results is described. Firstly, data from Wikipedia about the top 20 museums and monuments in Paris was scrapped and used to create a data frame. The python pandas library was used to clean and prepare the data, removing unnecessary information or data that could not be handled in the data frame. This resulted in the nice data frame shown below.

	Rank	Change 07/06	Museums and Monuments	Visitors in 2007	Visitors in 2006	Variation 07/06
0	1	=	Notre Dame de Paris	13650000	13650000	—
1	2	=	Basilique du Sacré-Cœur	10500000	10500000	—
2	3	=	The Louvre	8260000	8348000	-1.1%
3	4	=	Eiffel Tower	6797410	6695135	1.5%
4	5	=	Pompidou Centre	5509425	5133506	7.3%

Then, the nominatim function from the geopy library was used to add geospatial data to the data frame, that is the latitude and the longitude to produce the new data frame shown below.

	Rank	Change 07/06	Museums and Monuments	Visitors in 2007	Visitors in 2006	Variation 07/06	Latitude	Longitude
0	1	=	Notre Dame de Paris	13650000	13650000	—	48.852937	2.350050
1	2	=	Basilique du Sacré-Cœur	10500000	10500000	—	48.886806	2.343015
2	3	=	The Louvre	8260000	8348000	-1.1%	48.861147	2.338028
3	4	=	Eiffel Tower	6797410	6695135	1.5%	48.858260	2.294499
4	5	=	Pompidou Centre	5509425	5133506	7.3%	48.860592	2.352474

Subsequently, using the folium package and the data frame, a map was created, highlighting the museums and monuments.



Foursquare data was then retrieved for a maximum of 25 sites at a radius of 500 m near each monument or museum, and a total number of 485 sites were retrieved, as shown in the data frame below.

```
Total number of sites is 485
```

	Site	Site Latitude	Site Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Notre Dame de Paris	48.852937	2.35005	Shakespeare & Company	48.852568	2.347096	Bookstore
1	Notre Dame de Paris	48.852937	2.35005	Comme chai Toi	48.851749	2.349319	French Restaurant
2	Notre Dame de Paris	48.852937	2.35005	Tours de la Cathédrale Notre-Dame de Paris	48.853230	2.349207	Scenic Lookout
3	Notre Dame de Paris	48.852937	2.35005	Sola	48.851569	2.348391	Japanese Restaurant
4	Notre Dame de Paris	48.852937	2.35005	Au Vieux Paris d'Arcole	48.854196	2.350312	French Restaurant

To find clusters of site categories in the different touristic sites, The data frame with the site venues associated to the monuments or museums was first of all transformed by one-hot encoding (0/1), as shown below.

	Site	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	Asian Restaurant	Auvergne Restaurant	Bagel Shop	Bakery	...	Tourist Information Center	Toy / Game Store	Track	Tram Station	Trattoria/Osteria
0	Notre Dame de Paris	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
1	Notre Dame de Paris	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
2	Notre Dame de Paris	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
3	Notre Dame de Paris	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
4	Notre Dame de Paris	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0

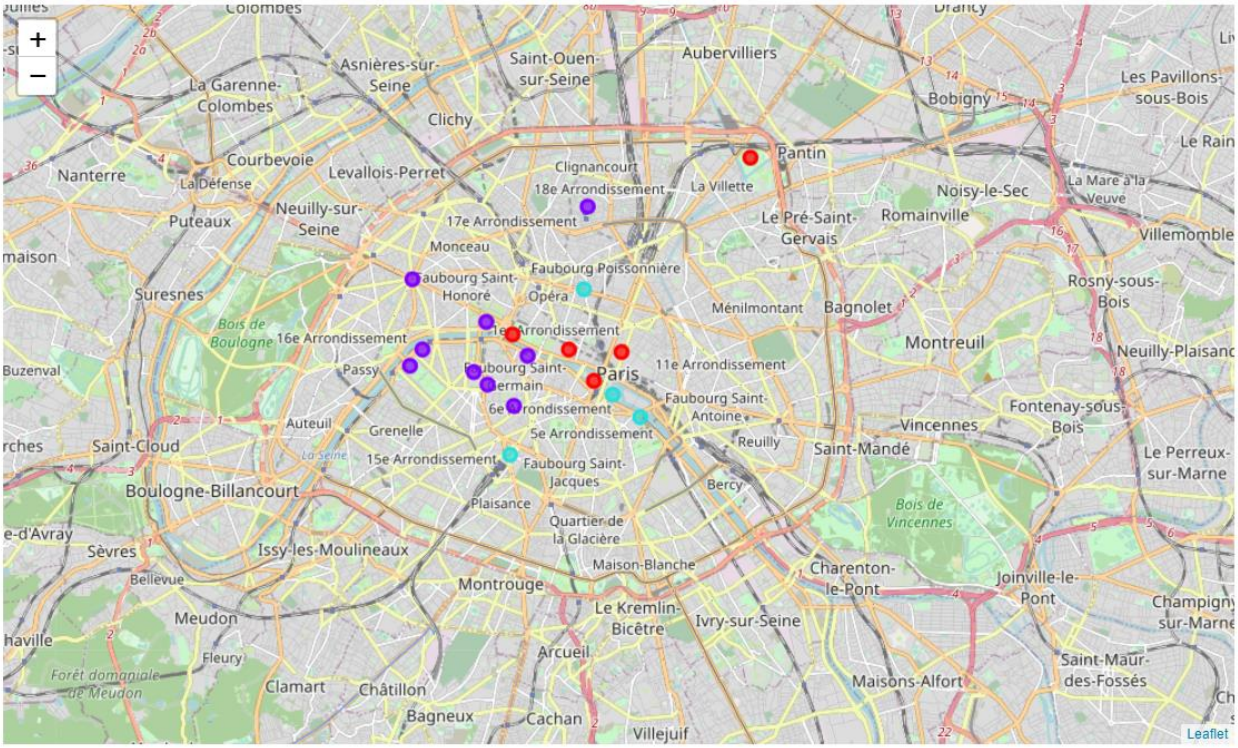
Next, grouping was used to show the frequency.

	Site	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	Asian Restaurant	Auvergne Restaurant	Bagel Shop	Bakery	...	Tourist Information Center	Toy / Game Store	Track	Tram Station	Trattoria/O
0	Arc de Triomphe	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	...	0.04	0.00	0.00	0.00	
1	Basilique du Sacré-Cœur	0.00	0.00	0.00	0.04	0.04	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	
2	Chapel of Our Lady of the Miraculous Medal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	
3	Cité des Sciences et de l'Industrie	0.04	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	...	0.00	0.00	0.00	0.04	
4	Eiffel Tower	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.04	...	0.00	0.00	0.04	0.00	
5	Institut du Monde Arabe	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	
6	Musée Grévin	0.00	0.04	0.04	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	
7	Musée Rodin	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	
8	Musée d'Orsay	0.00	0.00	0.00	0.04	0.08	0.04	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	
9	Musée de l'Armée	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	
10	Musée de l'Orangerie	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	
11	Musée du Quai Branly	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	

With these data, an unsupervised machine learning algorithm was simulated, more specifically, a k-means clustering algorithm from the scikit-learn package with a k value of 4.

4. RESULTS AND DISCUSSION

A visualization of the clusters can be seen in the map below. You will see 18 bubbles representing the touristic sites. Two of the sites do not appear on the map. ***Muséum national d'histoire naturelle*** for example has been spread out into 4 locations, which could not be located by foursquare in the analysis.



The different cluster data frames are shown below. The clusters show the 10 most common venues near each museum/monument.

Cluster 0:

	Site	Cluster Labels	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
10	Muséum d'Histoire Naturelle	0	Garden	Science Museum	Playground	Military Base	French Restaurant	Plaza	Chinese Restaurant	Concert Hall	Corsican Restaurant	Comedy Club

Cluster 1

	Site	Cluster Labels	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
2	The Louvre	1	Plaza	French Restaurant	Historic Site	Hotel	Cocktail Bar	Church	Chinese Restaurant	Restaurant	Cheese Shop	Shoe Store
4	Pompidou Centre	1	Coffee Shop	Restaurant	Lebanese Restaurant	French Restaurant	Liquor Store	Clothing Store	Cosmetics Shop	Chocolate Shop	Burger Joint	Bubble Tea Shop
6	Cité des Sciences et de l'Industrie	1	Music Venue	Rock Club	American Restaurant	Steakhouse	Performing Arts Venue	Multiplex	Movie Theater	Pizza Place	Plaza	Concert Hall
12	Sainte Chapelle	1	Plaza	Falafel Restaurant	Wine Bar	Hotel	Restaurant	Indie Movie Theater	Burger Joint	Moroccan Restaurant	Bookstore	Lebanese Restaurant
16	Musée de l'Orangerie	1	Hotel	Art Museum	Bookstore	Hotel Bar	Plaza	Fountain	Tea Room	Dessert Shop	Scandinavian Restaurant	Perfume Shop
19	Panthéon	1	Italian Restaurant	Plaza	Hotel	Sandwich Place	Fountain	Sicilian Restaurant	Monument / Landmark	Café	Restaurant	Pizza Place

Cluster 2:

	Site	Cluster Labels	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
1	Basilique du Sacré-Cœur	2	French Restaurant	Plaza	Middle Eastern Restaurant	Pizza Place	Bar	Chocolate Shop	History Museum	Bistro	Scenic Lookout	Gift Shop
3	Eiffel Tower	2	French Restaurant	Hotel	Italian Restaurant	Garden	Beach Bar	Library	Falafel Restaurant	Monument / Landmark	Plaza	Restaurant
5	Musée d'Orsay	2	French Restaurant	Hotel	Garden	Art Museum	Bookstore	Historic Site	Fountain	Food Truck	Exhibit	Pizza Place
7	Chapel of Our Lady of the Miraculous Medal	2	French Restaurant	Italian Restaurant	Coffee Shop	Hotel	Garden	Chocolate Shop	Cupcake Shop	Peruvian Restaurant	Bistro	Tailor Shop
8	Arc de Triomphe	2	Hotel	French Restaurant	Italian Restaurant	Jewelry Store	Steakhouse	Molecular Gastronomy Restaurant	Moroccan Restaurant	Movie Theater	Cocktail Bar	Roof Deck

Cluster 3:

	Change 07/06	Cluster Labels	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
0	=	3	French Restaurant	Bookstore	Bakery	Yoga Studio	Ice Cream Shop	Japanese Restaurant	Lebanese Restaurant	Falafel Restaurant	Park	Pastry Shop
13	+3	3	French Restaurant	Hotel	Wine Bar	Steakhouse	Ice Cream Shop	Women's Store	Fish & Chips Shop	Corsican Restaurant	New American Restaurant	Concert Hall
14	-1	3	French Restaurant	Ice Cream Shop	Wine Bar	Museum	Indie Movie Theater	Boat or Ferry	Historic Site	Japanese Restaurant	Garden	Escape Room
18	+3	3	Japanese Restaurant	Creperie	French Restaurant	Chinese Restaurant	Caribbean Restaurant	Museum	Argentinian Restaurant	Scenic Lookout	Italian Restaurant	Auvergne Restaurant

From the cluster data frames, it can be seen the clusters have been grouped according to the frequency of the most common venues.

5. CONCLUSION

The goal presented at the outset of this project was achieved: tourists can see in the results the most common venues near the touristic centres. Updated information on the ranking of the touristic sites was not available. It could have given an updated picture of the touristic scenery in Paris. Nonetheless, when tourists do visit Paris, they will better know which nearby venues to look out for at the touristic centres. This is just one example of the fantastic data science use cases one can realize by applying technology which is available for free today.