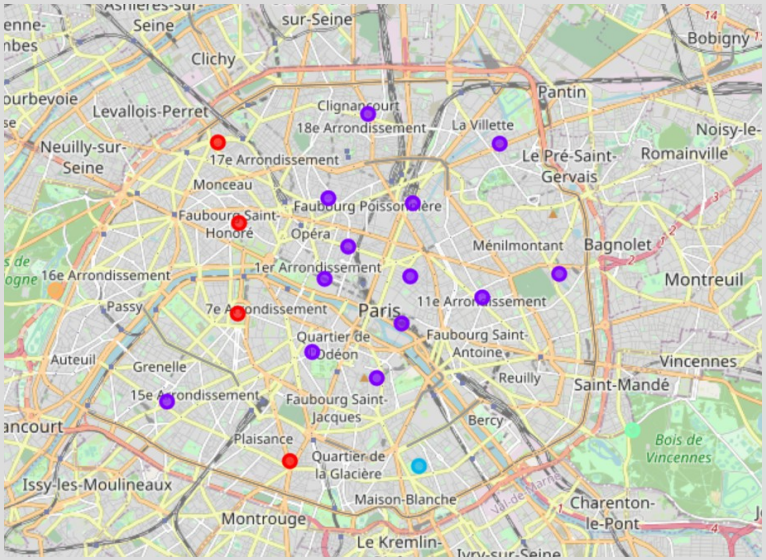
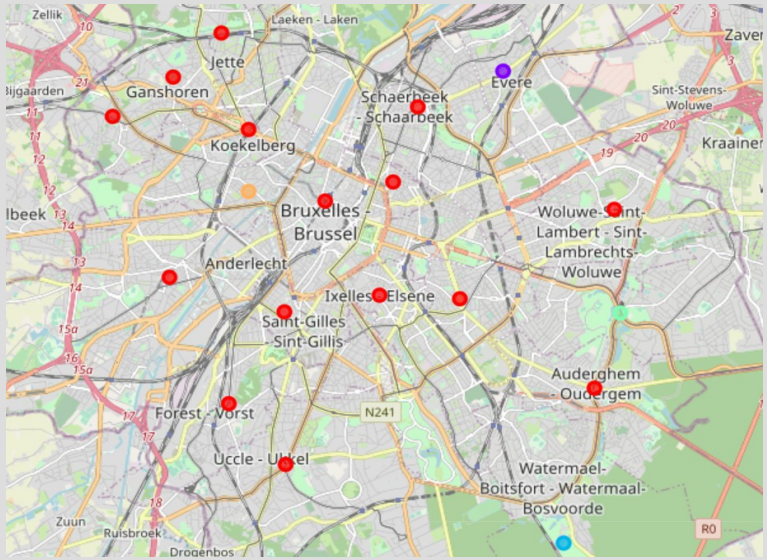
Tourism City Activities

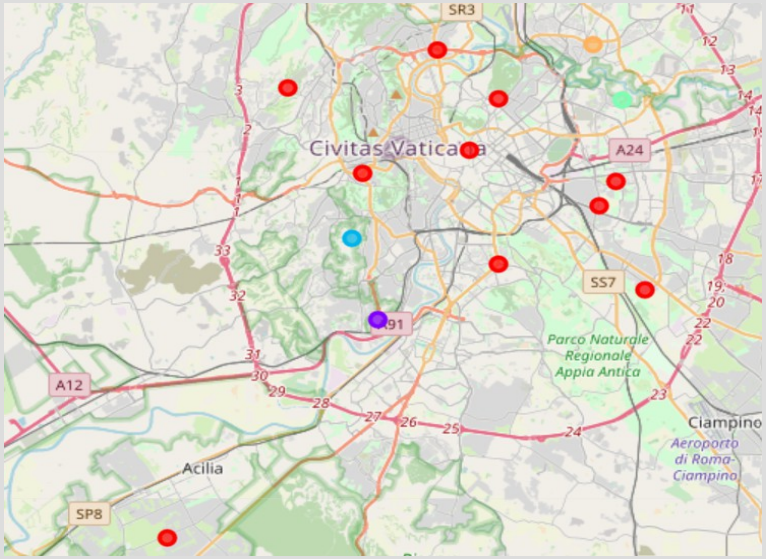
Neighborhoods Activities in different cities

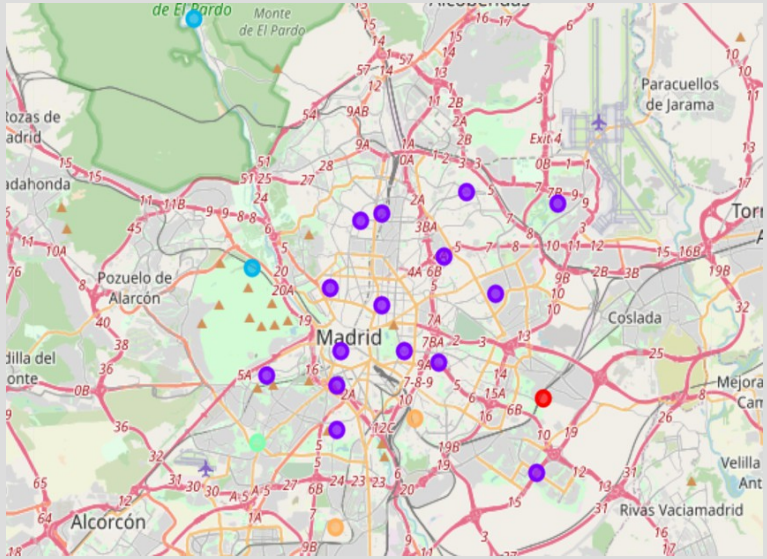
PARIS	BRUSSELS	ROME	MADRID

Neighborhood Clusters City Maps









Foursquare Venues

Foursquare Data can be used to find venues, for example :

Paris	Brussels	Rome	Madrid
Bakery	Bakery	Italian Restaurant	Bar
Bar	Bar	Basketball Stadium	Beer Garden
Beer Bar	Bookstore	Boutique	Bistro
Beer Store	Burger Joint	Café	Café
Bistro	Café	Coffee Shop	Coffee Shop
Bourse	Chocolate Shop	College Cafeteria	Comfort Food Restaurant
Brewery	Clothing Store	Concert Hall	Concert Hall
Buttes-Chaumont	Convenience Store	Cosmetics Shop	Convenience Store
Buttes-Montmartre	Cosmetics Shop	Cupcake Shop	Deli / Bodega

City, Neighborhood, Latitude, Longitude.

Paris

```
p_n=pd.read_csv('Paris.csv')
```

```
p_n.head()
```

[3]:

	Borough	Neighborhood	Latitude	Longitude
0	Paris	Louvre	48.862563	2.336443
1	Paris	Bourse	48.868279	2.342803
2	Paris	Buttes-Chaumont	48.887076	2.384821
3	Paris	Luxembourg	48.849130	2.332898
4	Paris	Passy	48.860392	2.261971

Kmeans used to cluster neighborhood by activities

Run *k*-means to cluster the neighborhood into 5 clusters.

```
# set number of clusters
kclusters = 5

n_grouped_clustering = n_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(n_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

37]: array([0, 1, 1, 1, 1, 2, 1, 1, 1, 1])

Let's create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

```
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

n_merged = p_n

# merge n_grouped with n_data to add latitude/longitude for each neighborhood
n_merged = n_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

n_merged.head() # check the last columns!
```

Kmeans Clusters

Examine Clusters in Paris

Cluster 1

In [41]:
```
n_merged.loc[n_merged['Cluster Labels'] == 0, n_merged.columns[[1] + list(range(5, n_merged.shape[1]))]]
```

Out[41]:

	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
11	Palais-Bourbon	Hotel	French Restaurant	Italian Restaurant	Plaza	Café	History Museum	Cocktail Bar	Historic Site	Japanese Restaurant	Gourmet Shop
13	Élysée	French Restaurant	Hotel	Bakery	Spa	Department Store	Cocktail Bar	Resort	Corsican Restaurant	Plaza	Italian Restaurant
15	Batignolles-Monceau	Hotel	French Restaurant	Italian Restaurant	Japanese Restaurant	Bakery	Restaurant	Bistro	Plaza	Café	Korean Restaurant
18	Observatoire	French Restaurant	Hotel	Bistro	Italian Restaurant	Bakery	Brasserie	Fast Food Restaurant	Supermarket	Sushi Restaurant	Tea Room

Cluster 2

In [42]:
```
n_merged.loc[n_merged['Cluster Labels'] == 1, n_merged.columns[[1] + list(range(5, n_merged.shape[1]))]]
```

Out[42]:

		1st Most	2nd Most	3rd Most	4th Most	5th Most	6th Most	7th Most	8th Most	9th Most	10th Most

Data Science Capstone Project

Coursera IBM Data Science Capstone Project

Victor Pagan Rubio

vpaganrubio@gmail.com