Clustering and Comparing Neighborhoods in London and Paris

Coursera Capstone Project

V. Labeau

January 3rd, 2021

Introduction

In this project, we will explore, segment and cluster the neighborhoods of two large European cities: London, the capital of England, and Paris, the capital of France.

Business problem

- The goal here is to compare the neighborhoods of London and Paris and determine how similar or dissimilar they are.
- The objective is to help tourists to choose their destinations depending on the experiences that the neighbourhoods have to offer and what they would like to do.
- This model could also help people to decide about migrating to London or Paris or even relocating neighbourhoods within the cities.
- Our findings will help stakeholders make informed decisions and address any concerns they have, including the different kinds of cuisines, provision stores and what the cities have to offer.

Data Description

This section describes the processes of acquiring, cleaning, and preparing each dataset used in this project for the next stages.

Data Sources and Processing

Geographical location data needed for both London and Paris:

Neighborhood Data

- *London dataset*
  - https://en.wikipedia.org/wiki/List_of_areas_of_London, which has information about all the boroughs. Since this Wikipedia page lacks information about geographical locations, we will use *ArcGIS API* to complete the dataset with the longitude and latitude data for the boroughs of London.
- *Paris Dataset:*
  - https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e which is a JSON file data about all the boroughs in France. We will focus our study only on Paris.
  - https://opendata.paris.fr/explore/dataset/quartier_paris/download/?format=json&timezone=Europe/Berlin which contains the different neighborhoods of Paris

Venues Data:

- "Foursquare" (https://Foursquare.com): Venues data describes the venues (restaurants, cafes, parks, museums, etc.) in each neighborhood of the two cities by category. For each neighbourhood, we have chosen the radius to be 500 meters.

METHODOLOGY

The approach taken is to use the Foursquare API to explore and cluster the neighborhoods in London and Paris to find how similar or dissimilar they are neighborhoodlike.

Steps Approach

- **Step 1** Data Acquisition: Acquire and clean the data from the different sources mentioned in the Data sources section and then add the geographical coordinates
- **Step 2** Foursquare Venues: Getting location venues data using the Foursquare API.
- **Step 3** Exploratory Data Analysis:
  - Plot the prepared location data on the map for visualization purposes.
  - List down the 10 most common venues for both cities.
- **Step 4** Clustering Neighborhood:
  - Use the *k-Means* clustering algorithm to find similar neighborhoods.
  - Use the Folium library to visualize the neighborhoods in both London and Paris and their emerging clusters.
- **Step 5** Examine Clusters**:** discuss the results based on the above findings and provide a snapshot of both cities which will help travelers and immigrants in making their choice.

Data Sets

London Dataset

Neighborhood data

	Location	Londonborough	Post town	Dialcode	OS grid ref	Postcodedistrict
0	Abbey Wood	Bexley, Greenwich	LONDON	020	TQ465785	SE2
1	Acton	Ealing, Hammersmith and Fulham	LONDON	020	TQ205805	W3
1	Acton	Ealing, Hammersmith and Fulham	LONDON	020	TQ205805	W4
6	Aldgate	City	LONDON	020	TQ334813	EC3
7	Aldwych	Westminster	LONDON	020	TQ307810	WC2

	Borough	Neighborhood	Latitude	Longitude
0	Bexley, Greenwich	Abbey Wood	51.49245	0.12127
1	Ealing, Hammersmith and Fulham	Acton	51.51324	-0.26746
2	Ealing, Hammersmith and Fulham	Acton	51.48944	-0.26194
3	City	Aldgate	51.51200	-0.08058
4	Westminster	Aldwych	51.51651	-0.11968

Foursquare Venue data

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbey Wood	51.49245	0.12127	Lesnes Abbey	51.489526	0.125839	Historic Site
1	Abbey Wood	51.49245	0.12127	Sainsbury's	51.492826	0.120524	Supermarket
2	Abbey Wood	51.49245	0.12127	Lidl	51.496152	0.118417	Supermarket
3	Abbey Wood	51.49245	0.12127	Abbey Wood Railway Station (ABW)	51.490825	0.123432	Train Station
4	Abbey Wood	51.49245	0.12127	Bean @ Work	51.491172	0.120649	Coffee Shop

Data Sets

Paris Dataset

Neighborhood data

	Borough	IdNeighborhood
0	PARIS-9E-ARRONDISSEMENT	9
1	PARIS-2E-ARRONDISSEMENT	2
2	PARIS-11E-ARRONDISSEMENT	11
3	PARIS-15E-ARRONDISSEMENT	15
4	PARIS-19E-ARRONDISSEMENT	19

	IdNeighborhood	Neighborhood	Latitude	Longitude
0	12	Quinze-Vingts	48.846916	2.374402
1	6	Notre-Dame-des-Champs	48.846428	2.327357
2	14	Petit-Montrouge	48.826653	2.326437
3	19	Pont-de-Flandre	48.895556	2.384777
4	16	Muette	48.863275	2.259936

	Borough	Neighborhood	Latitude	Longitude
0	PARIS-9E-ARRONDISSEMENT	Rochechouart	48.879812	2.344861
1	PARIS-9E-ARRONDISSEMENT	Saint-Georges	48.879934	2.332850
2	PARIS-9E-ARRONDISSEMENT	Chaussée-d'Antin	48.873547	2.332269
3	PARIS-9E-ARRONDISSEMENT	Faubourg-Montmartre	48.873935	2.343253
4	PARIS-2E-ARRONDISSEMENT	Gaillon	48.869307	2.333432

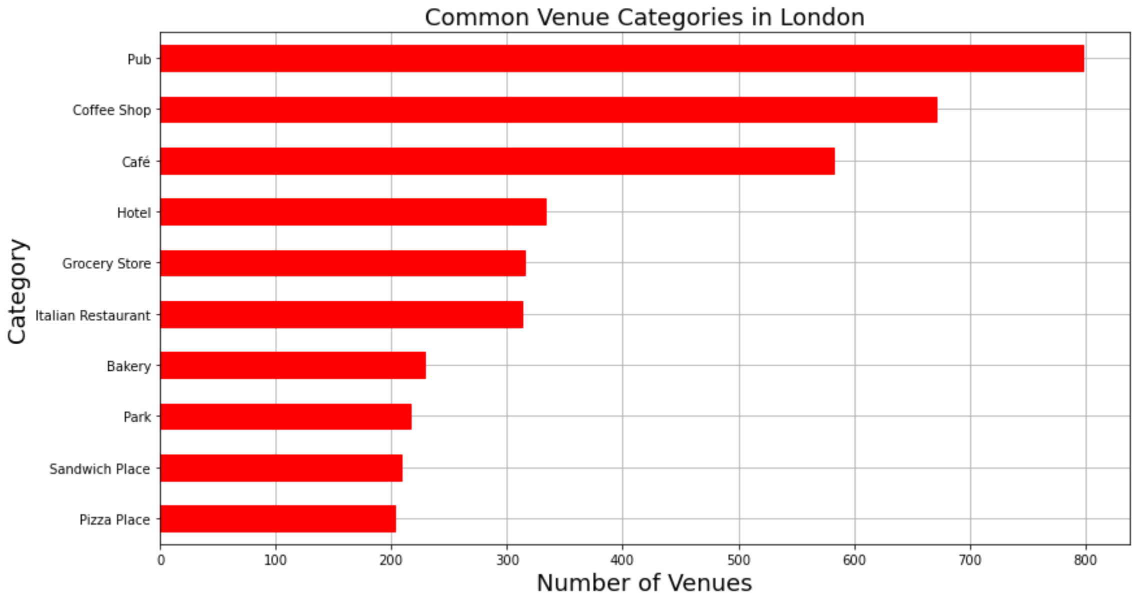
Foursquare Venue data

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rochechouart	48.879812	2.344861	Mamiche	48.880112	2.343699	Bakery
1	Rochechouart	48.879812	2.344861	Mikkeller Bar Paris	48.878663	2.345377	Beer Bar
2	Rochechouart	48.879812	2.344861	Les 36 Corneil	48.878997	2.345501	Wine Bar
3	Rochechouart	48.879812	2.344861	Le Potager de Charlotte	48.878924	2.344640	Vegetarian / Vegan Restaurant
4	Rochechouart	48.879812	2.344861	La Ferme Saint Hubert	48.878908	2.345428	Cheese Shop

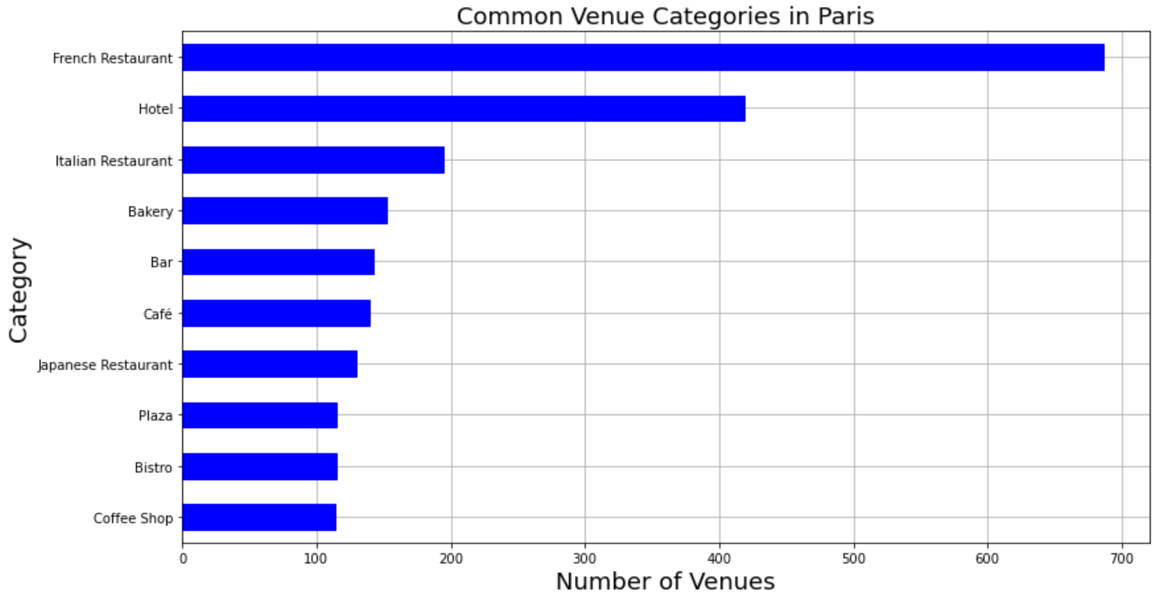
Exploratory Data Analysis

Most Common Venue categories

London



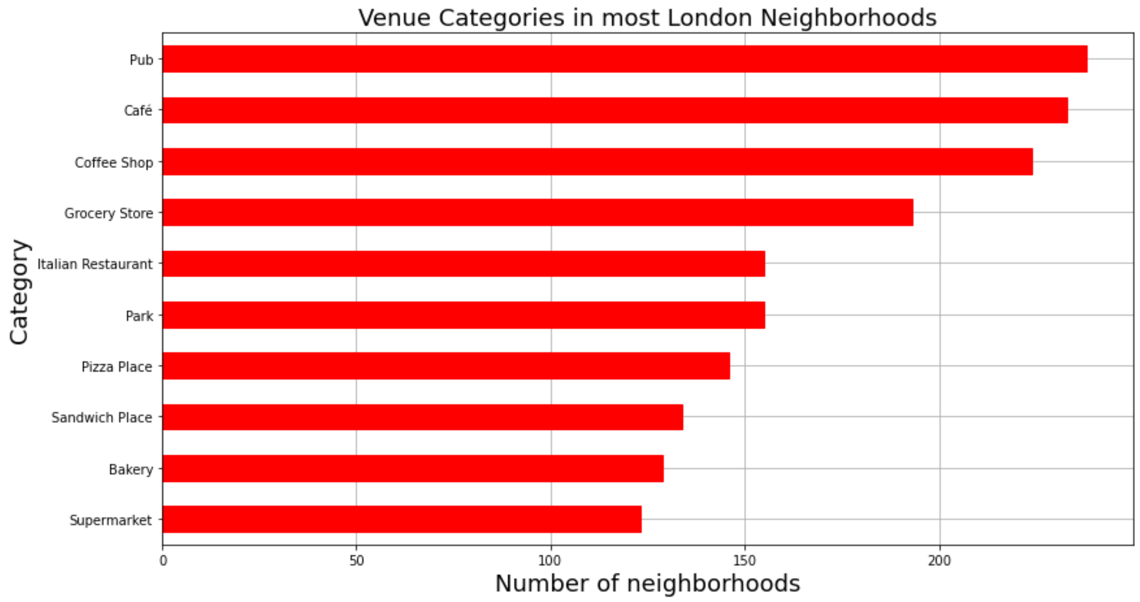
Paris



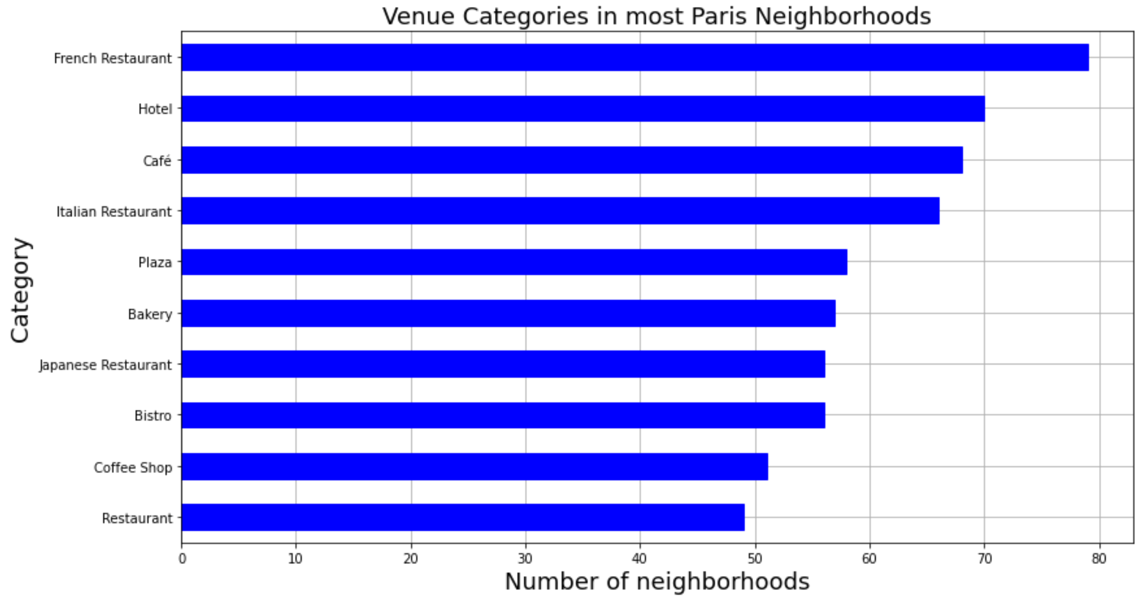
Exploratory Data Analysis

Most Widespread Venue Categories

London



Paris

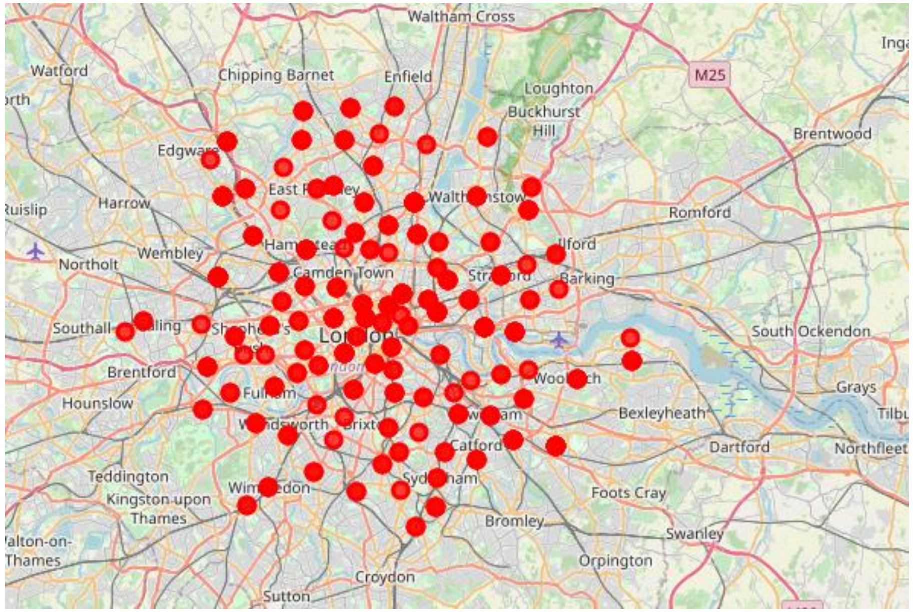


Clustering Neighborhood

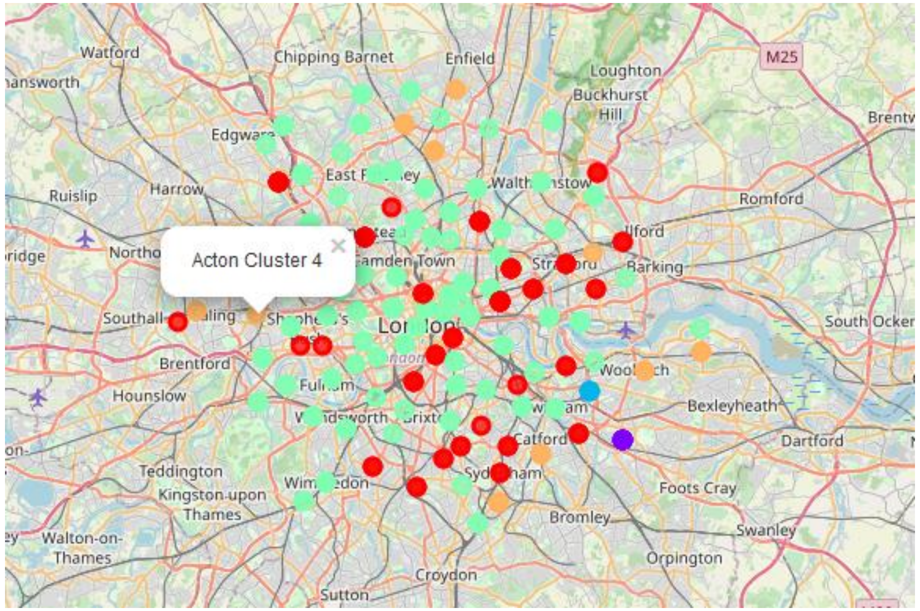
k-Means Clustering

London

Neighborhoods Map



Neighborhoods Clusters

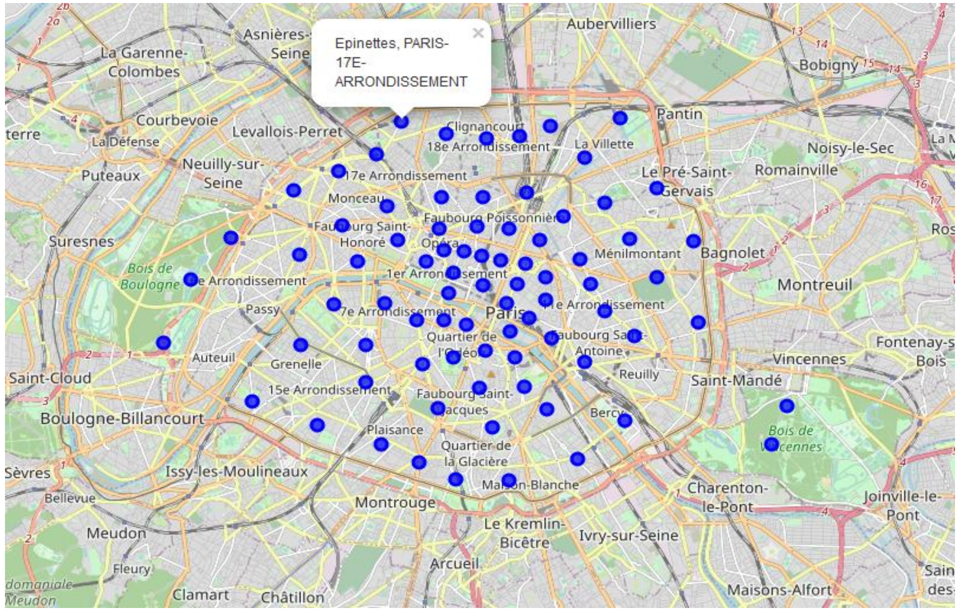


Clustering Neighborhood

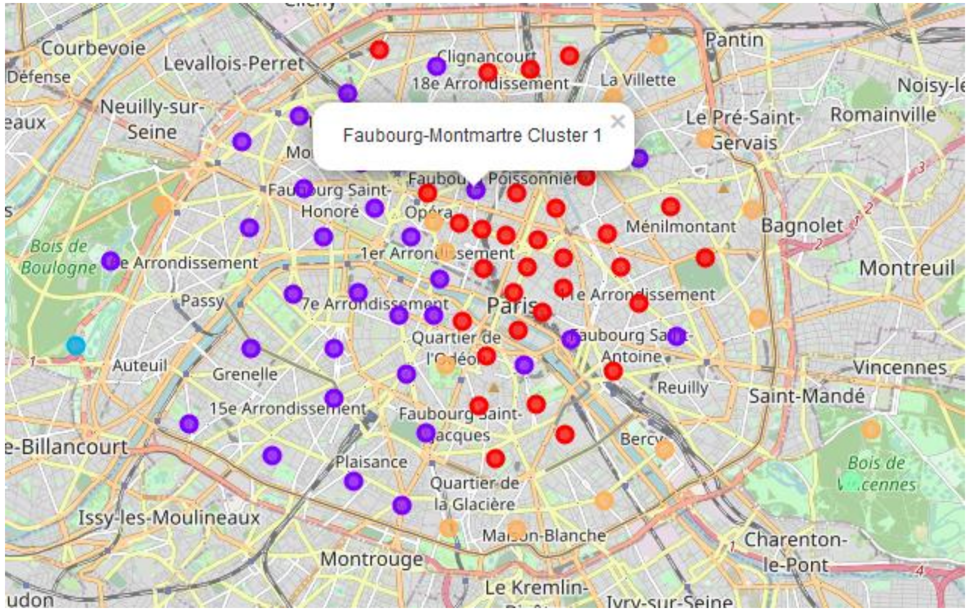
k-Means Clustering

Paris

Neighborhoods Map



Neighborhoods Clusters



Clustering Neighborhood

k-Means Clustering

London

Paris

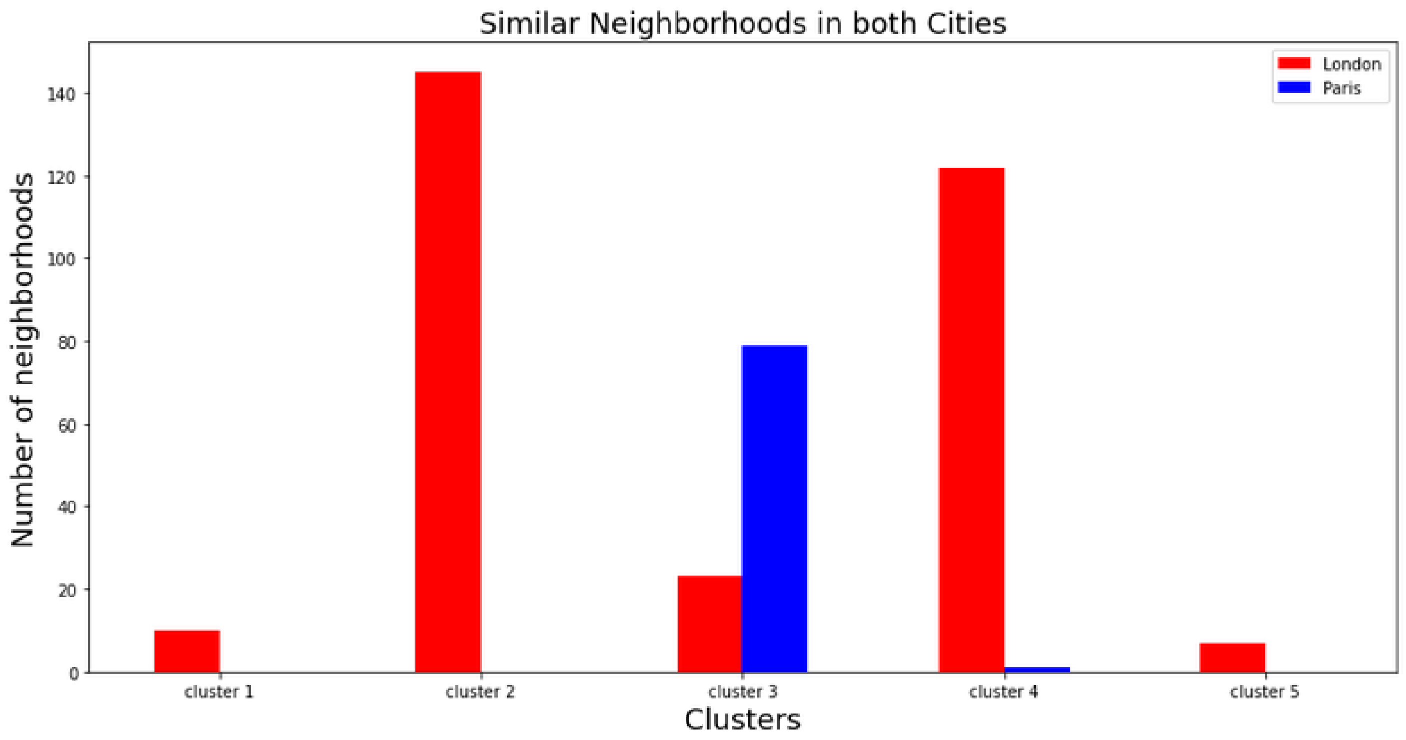
Combination of London and Paris data and applying the clustering algorithm to find the similarities between both cities.

Neighborhood	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
Winchmore Hill_London	3	Italian Restaurant	Grocery Store	Café	Supermarket	Soccer Field	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant	Farmers Market
Wood Green_London	3	Italian Restaurant	Indian Restaurant	Park	Grocery Store	Bar	Dance Studio	Event Space	Exhibit	Falafel Restaurant	Farmers Market
Woodford_London	1	Bar	BBQ Joint	Grocery Store	Pub	Seafood Restaurant	Film Studio	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant
Woodford Green_London	2	Hotel	Theater	Plaza	Garden	Monument / Landmark	Cocktail Bar	Pub	Restaurant	Tea Room	Boutique
Woodside Park_London	3	Coffee Shop	Bakery	Fast Food Restaurant	Supermarket	Pharmacy	Chinese Restaurant	Thai Restaurant	Theater	Sushi Restaurant	Turkish Restaurant
Woolwich_London	3	Child Care Service	Chinese Restaurant	Convenience Store	Indian Restaurant	Bus Stop	Middle Eastern Restaurant	Grocery Store	Fish & Chips Shop	Zoo Exhibit	Film Studio
Wormwood Scrubs_London	3	Grocery Store	Café	Pub	Gastropub	Pizza Place	Thai Restaurant	Park	Event Space	Gourmet Shop	Greek Restaurant
Amérique_Paris	2	French Restaurant	Supermarket	Pool	Bistro	Park	Bed & Breakfast	Plaza	Café	Zoo Exhibit	Falafel Restaurant
Archives_Paris	2	French Restaurant	Hotel	Coffee Shop	Clothing Store	Bar	Art Gallery	Bookstore	Bistro	Plaza	Cocktail Bar
Arsenal_Paris	2	French Restaurant	Hotel	Park	Gastropub	Plaza	Italian Restaurant	Pedestrian Plaza	Cocktail Bar	Vegetarian / Vegan Restaurant	Seafood Restaurant

RESULTS & DISCUSSION

Report the findings of our study based upon the methodology we applied to gather information.

Clustering Results



Similar Neighborhoods in both cities

- Clusters 3 and 4 shows the similarities between London and Paris. Whereas the other clusters 1, 2 and 5 show the dissimilarities.

Clustering Results

7 most common venue categories

London
Paris

Cluster 1

Venue Category	% of venues
Supermarket	27.8689
Convenience Store	16.3934
Train Station	6.55738
Platform	6.55738
Historic Site	6.55738
Coffee Shop	6.55738
Fast Food Restaurant	4.91803

Cluster 2

Venue Category	% of venues
Pub	10.7637
Coffee Shop	5.98383
Café	4.79784
Italian Restaurant	3.07278
Hotel	2.38994
Bakery	2.354
Sandwich Place	2.04852

Cluster 3

Venue Category	% of venues
French Restaurant	10.279
Hotel	8.2574
Italian Restaurant	3.75854
Coffee Shop	3.10364
Café	2.96128
Bakery	2.70501
Bar	2.22096

Cluster 4

Venue Category	% of venues
Café	8.68813
Coffee Shop	8.58429
Grocery Store	7.19972
Pub	5.08827
Pizza Place	3.32295
Park	2.83835
Italian Restaurant	2.49221

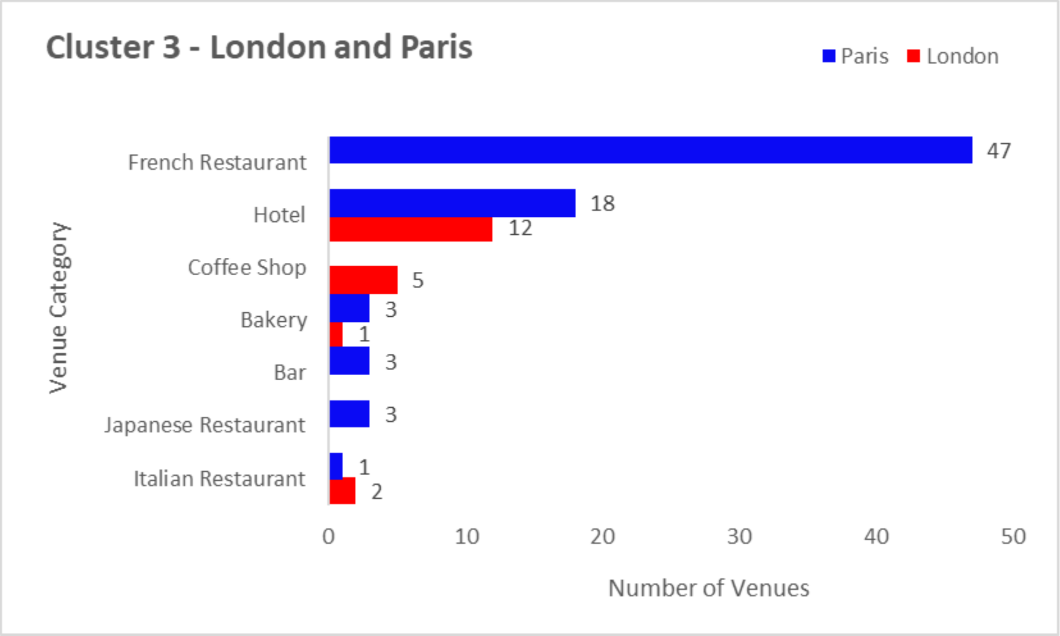
Cluster 5

Venue Category	% of venues
Historic Site	20
Bus Stop	20
Golf Course	20
Construction & Landscaping	20
Park	20

Clustering Results

Cluster 3 Detail – London vs Paris

London vs Paris



Drilling down on the cluster 3 which shows interesting data to compare both cities in term of venues.

CONCLUSION

Conclusion

In this project, the neighborhoods of London and Paris were clustered into multiple groups based on the categories (types) of the venues in these neighborhoods. The results show that there are venue categories that are more common in some cluster than others; the most common venue categories differ from one cluster to the other.

If a deeper analysis, taking more aspects into account, is performed, it might result in discovering different styles in each cluster based on the most common categories in the cluster.