Applied Data Science Capstone

Presentation

**Battle of the Neighbourhoods**

**Paris and London**

1. **Introduction and Business Problem**

The purpose of this project is to help tourists choose their destinations depending on what experiences the neighbourhoods have to offer and what the neighbourhoods would want to have. This also helps people make decisions if they are thinking about migrating to Paris or London or even if they want to relocate between neighbourhoods within the city.

It also will help people to get awareness of the area and neighborhood before moving to a new city, state, country or place for their work or to start a new fresh life. 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 city has to offer.

* 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

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm

1. **Data Description**

We require geolocation data for both Paris and London. Postal codes in each city serve as a starting point. Using Postal codes we can find out the neighbourhoods, boroughs, venues and their most popular venue categories.

* Paris

To derive our solution, We scrape our data from *https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e*

The JSON file has data about all the neighbourhoods in France, we limit it only Paris.

1. \*postal\_code\* : Postal codes for France

2. \*nom\_comm\* : Name of Neighbourhoods in France

3. \*nom\_dept\* : Name of the boroughs, equivalent to towns in France

4. \*geo\_point\_2d\* : Tuple containing the latitude and longitude of the Neighbourhoods.

* London

To derive our solution, We scrape our data from https://en.wikipedia.org/wiki/List\_of\_areas\_of\_London

This wikipedia page has information about all the neighbourhoods, we limit it only London.

1. \*borough\* : Name of Neighbourhood

2. \*town\* : Name of borough

3. \*post\_code\* : Postal codes for London.

This wikipedia page lacks information about the geographical locations. To solve this problem we use ArcGIS API

* ArcGIS API

ArcGIS Online enables us to connect people, locations, and data using interactive maps. Work with smart, data-driven styles and intuitive analysis tools that deliver location intelligence. Share your insights with the world or specific groups. We use ArcGIS to get the geo locations of the neighbourhoods of London. The following columns are added to our initial dataset which prepares our data.

4. \*latitude\* : Latitude for Neighbourhood

5. \*longitude\* : Longitude for Neighbourhood

Based on all the information collected for both Paris and London, we have sufficient data to build our model. We cluster the neighbourhoods together based on similar venue categories. We then present our observations and findings. Using this data, our stakeholders can take the necessary decision.

* Foursquare API Data

We will need data about different venues in different neighbourhoods of that specific borough. In order to obtain information we use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighbourhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighbourhood. For each neighbourhood, we have chosen the radius to be 500 meters. 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:

1. \*Neighbourhood\* : Name of the Neighbourhood

2. \*Neighbourhood Latitude\* : Latitude of the Neighbourhood

3. \*Neighbourhood Longitude\* : Longitude of the Neighbourhood

4. \*Venue\* : Name of the Venue

5. \*Venue Latitude\* : Latitude of Venue

6. \*Venue Longitude\* : Longitude of Venue

7. \*Venue Category\* : Category of Venue

* Libraries Which are Used to Develope 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.
  + Matplotlib: Python Plotting Module.

1. **Methodology**

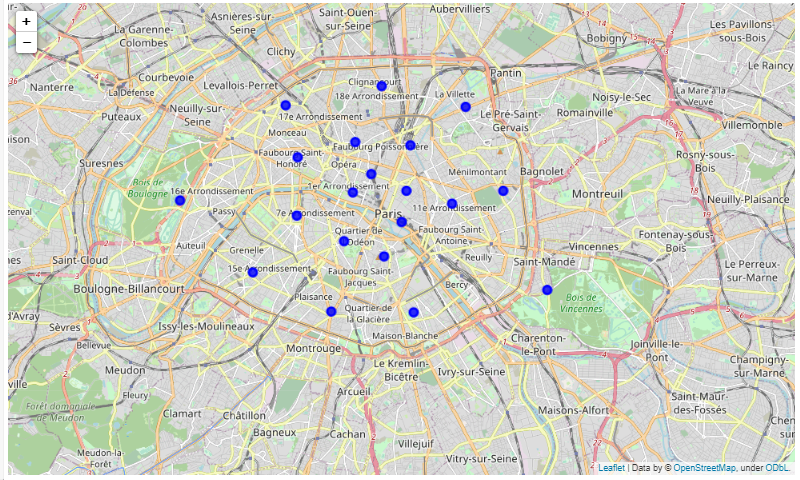


Image Neighbourhood of Paris

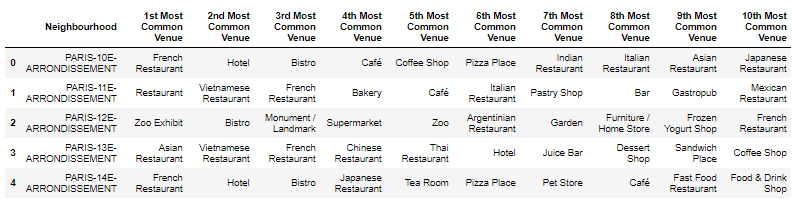


Image Most Common Venue of Paris

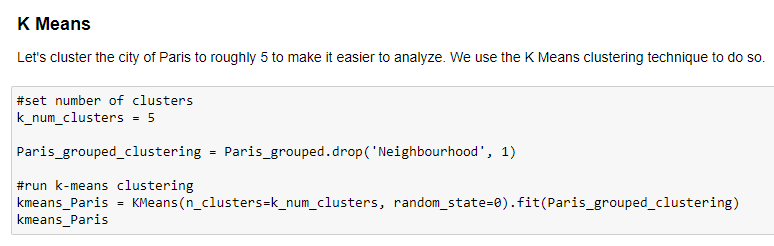


Image Model Building

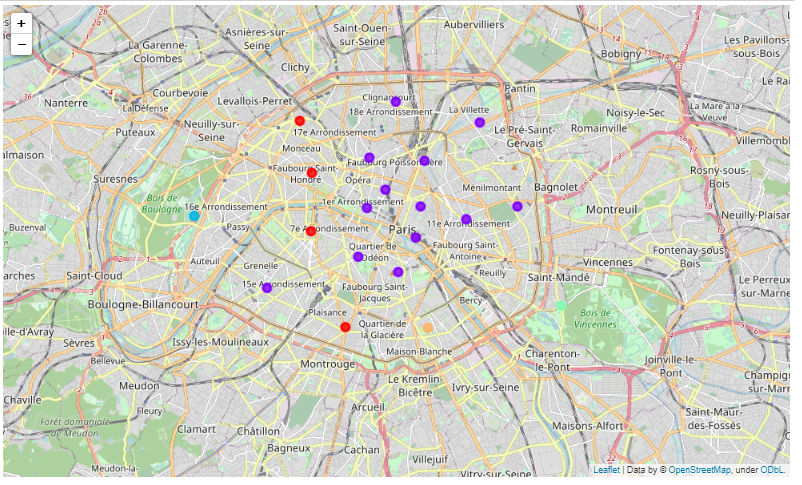


Image Clustered Neighbourhood of Paris

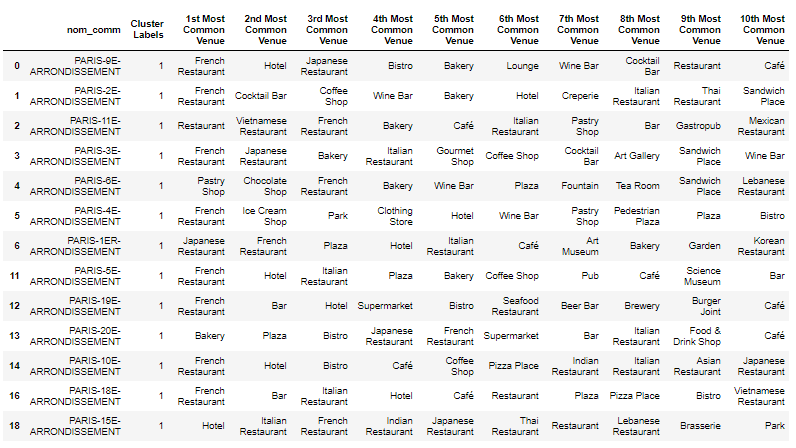


Image Cluster 1 of Paris

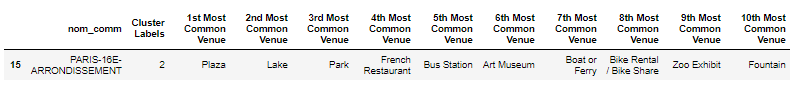


Image Cluster 2 of Paris

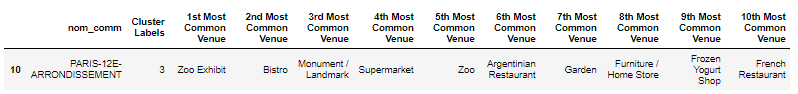


Image Cluster 3 of Paris



Image Cluster 4 of Paris



Image Cluster 5 of Paris

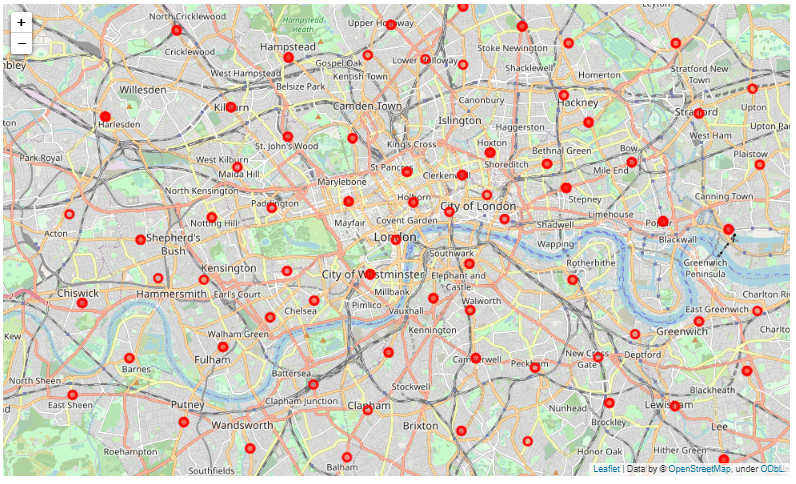


Image Neighbourhood of London



Image Most Common Venue of London

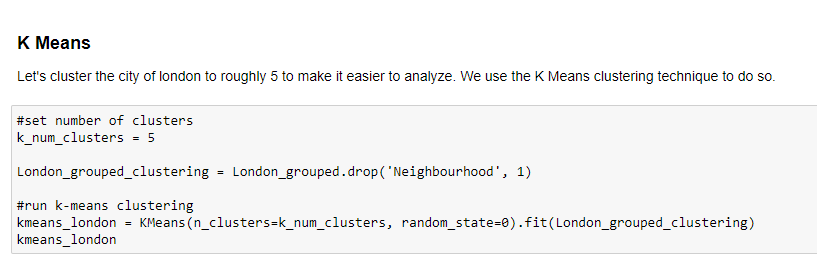


Image Model Building

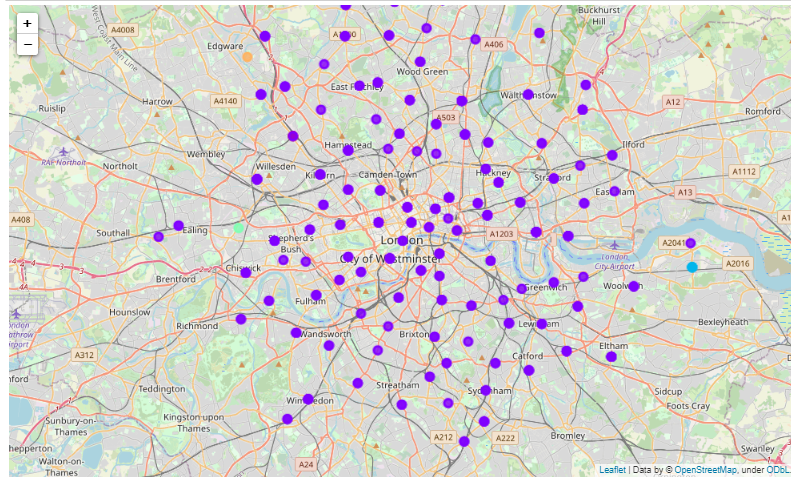


Image Clustered Neighbourhood of London

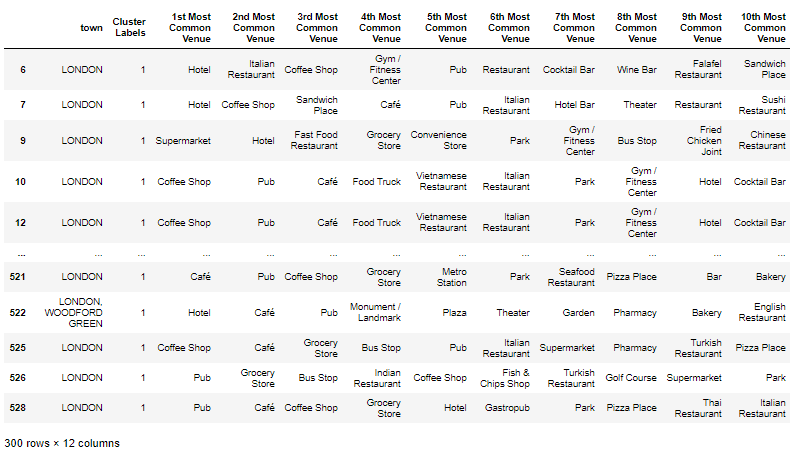


Image Cluster 1 of London



Image Cluster 2 of London

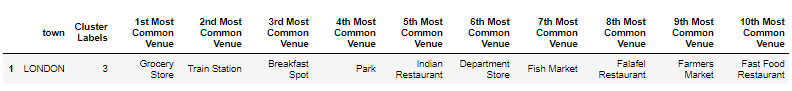


Image CLuster 3 of London

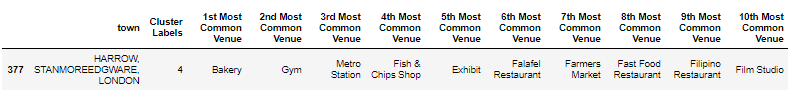


Image Cluster 4 of London



Image Cluster 5 of London

1. **Results and Discussion**

Paris is relatively smaller in size geographically. There are a lot of hangout spots including many restaurants, bars, and bistro. There are public transport in Paris such as buses, bikes, boats or ferries. For leisure and sight seeing, there are a lot of plazas, trails, parks, historic sites, clothing shops, art galleries and museums. It has a wide variety of cusines and eateries including French, Thai, Cambodian, Asian, Chinese etc. Overall, Paris seems like the relaxing vacation spot with a mix of lakes, historic spots and a wide variety of cusines to try out.

The neighbourhoods of London are very multicultural. There are a lot of different cusines including Indian, Italian, Turkish and Chinese. London seems takes a step further by having a lot of restaurants, bars, juice bars, coffee shops, fish and chips shop and breakfast spots. It has a lot of shopping options such as lea markets, flower shops, fish markets, fishing stores, clothing stores. The public transport such as buses and trains. For leisure, the neighbourhoods are set up to have lots of parks, golf courses, zoo, gyms and historic sites. Overall, London offers a multicultural, diverse and certainly an entertaining experience.

1. **Conclusion**

The purpose of this project was to explore the cities of Paris and London and see how attractive it is to potential tourists and migrants. We explored both the cities based on their postal codes and then extrapolated the common venues present in each of the neighbourhoods finally concluding with clustering similar neighbourhoods together. We could see that each of the neighbourhoods in both the cities have a wide variety of experiences to offer which is unique in it's own way.

Both Paris and London seem to offer a vacation stay or a romantic gateaway with a lot of places to explore, beautiful landscapes and a wide variety of culture. Overall, it's up to the stakeholders to decide which experience they would prefer more and which would more to their liking.