12/27/2020

Clustering and Comparing Neighborhoods in London and Paris

IBM Data Science Professional Certificate

Veronique Labeau
COURSERA

Table of Contents

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.

Both London and Paris are found at the heart of two great European nations. They are quite popular as vacation destinations for people all around the world. They are diverse and multicultural, and offer a wide variety of experiences that are particularly sought after.

While London is the capital of England, it is also the largest city within the country. The city stands on River Thames in South East England, with its history stretching back to the Roman times.

On the other hand, Paris, the capital of France, is in the north-central part of the nation. Like London, the city also stands along a river, commonly known as the Seine River. Paris has a rich European history and is regarded to be a global center for culture, fashion, art, and gastronomy.

Business problem

London vs Paris, which is your favorite destination? Picking a favorite city when it comes to London vs Paris can be such a tricky task. It is always a struggle since both cities are famous for their soul-refreshing experiences, as well as iconic attractions. Most tourists find it hard to pick one-holiday destination between the two.

Our goal here is to compare the neighborhoods of the two cities 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 will also help people to make decisions if they are thinking about migrating to London or Paris or even if they wish to relocate neighbourhoods within the city. 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.

Data Description

This section describes the processes of acquiring, cleaning, and preparing each dataset used in this project for the next stages. We need geographical location data for both London and Paris to segment and explore their neighborhoods. For both cities, we will essentially need a dataset that contains the boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhoods.

The data below will be used for this analysis. Once downloaded and cleaned up, the data will be combined into one table:

- Borough
- Neighborhood
- Latitude
- Longitude

Data 1: Paris boroughs and neighborhoods

Paris has a total of 20 boroughs (called *arrondissements* in French) and 80 neighborhoods (called *quartiers* in French). To explore, analyze and segment neighborhoods, the longitude and latitude of each of them, as well as the boroughs, will be added using the links to the following dataset:

- *Paris Boroughs*: 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.
- *Paris Neighborhoods*: https://opendata.paris.fr/explore/dataset/quartier_paris/download/?format=json&timezone=Europe/Berlin which contains the different neighborhoods of Paris.

	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

Data 2: London boroughs

London has a total of 32 boroughs. To explore, analyze and segment them, their longitude and latitude will be added using the link to the following dataset:

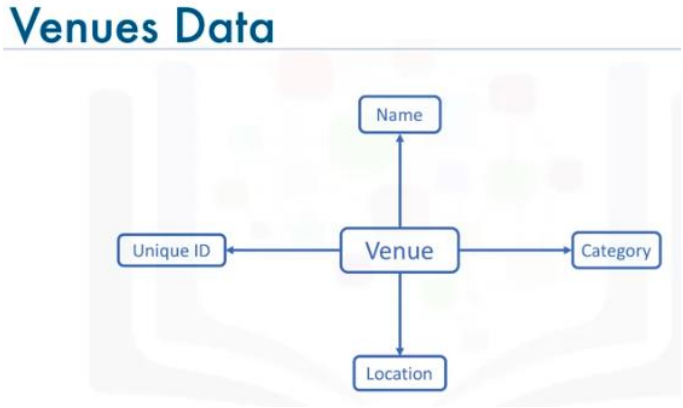
- We will scrape London data from the Wikipedia page https://en.wikipedia.org/wiki/List_of_areas_of_London, which has information about all the boroughs. For our study, we will filter the dataset only on the boroughs with London in Post-town column.
- *Locations coordinates*: Since this Wikipedia page 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.

	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

Data 3: Venues data

The 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.



To gain that information, we will use "Foursquare" location information (https://Foursquare.com). 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 and returned via a JSON file.