A Tale of Two Cities

Clustering Neighborhoods of London and Paris using Machine Learning

1. Introduction

A Tale of Two Cities, a novel written by Charles Dickens was set in London and Paris, which takes place during the French Revolution. These cities were both happening then and now. A lot has changed over the years, and we now take a look at how the cities have grown.

London and Paris are quite a popular tourist and vacation destinations for people all around the world. They are diverse and multicultural and offer a wide variety of experiences that are widely sought after. We try to group the neighbourhoods of London and Paris respectively and draw insights to what they look like now.

2. Business Problem

The aim is to help tourists choose their destinations depending on the experiences that the neighbourhoods have to offer and what they would want to have. This model also helps people 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.

3. Data Description

We require geographical location data for both London and Paris. 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.

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 to London.

borough: Name of Neighborhood

town: Name of the borough

post_code: Postal codes for London.

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

ArcGIS API

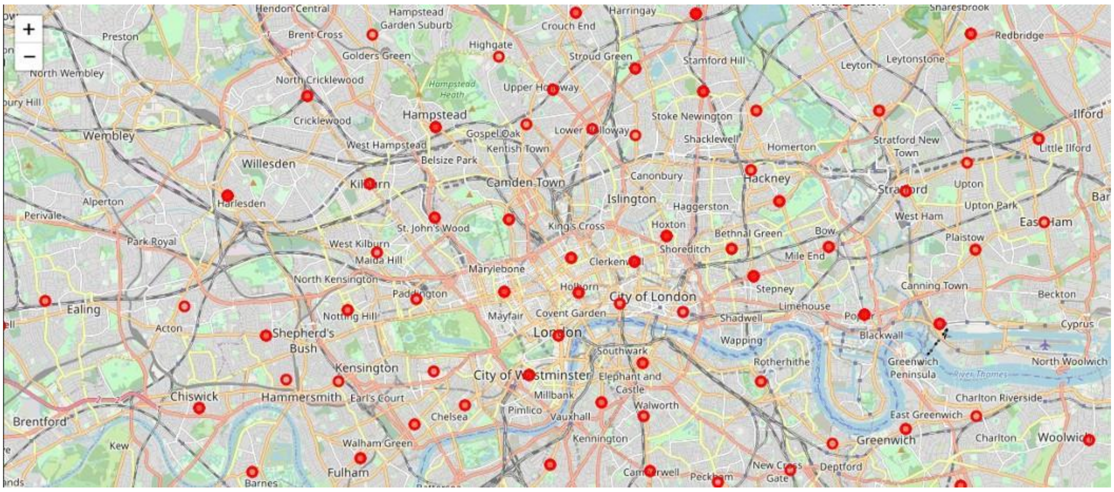
ArcGIS Online enables you to connect people, locations, and data using interactive maps. Work with smart, data-driven styles and intuitive analysis tools that deliver location intelligence. Share your insights with the world or specific groups.

More specifically, we use ArcGIS to get the geographical locations of the neighbourhoods of London. Adding the following columns to our initial data set prepares our data.
latitude: Latitude for Neighborhood
longitude: Longitude for Neighborhood

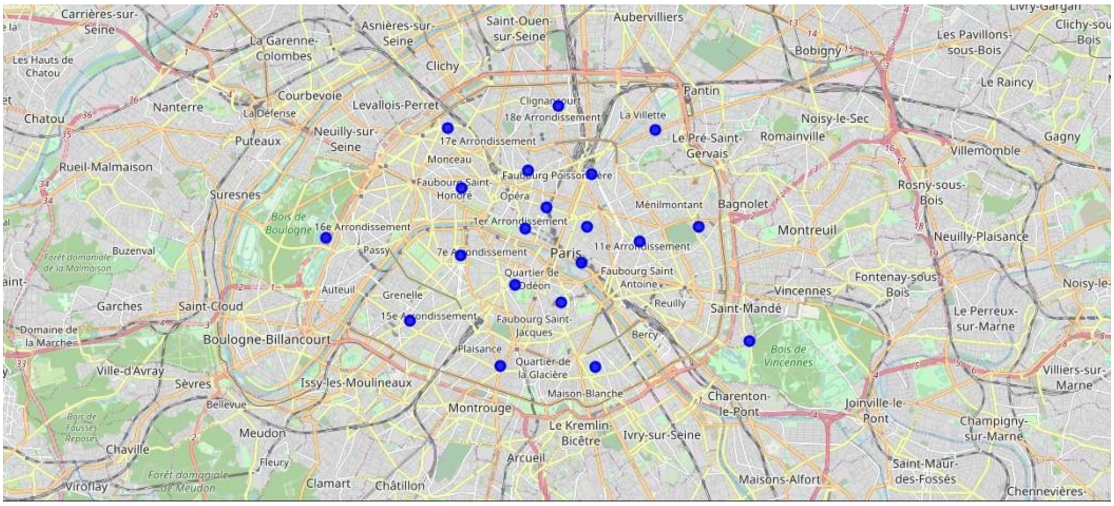
Paris
To derive our solution, We leverage JSON data available at https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e
The JSON file has data about all the neighbourhoods in France; we limit it to Paris.
postal_code: Postal codes for France
nom_comm: Name of Neighborhoods in France
nom_dept: Name of the boroughs, equivalent to towns in France
geo_point_2d: Tuple containing the latitude and longitude of the Neighborhoods.

Foursquare API Data
We will need data about different venues in different neighbourhoods of that specific borough. To gain that information, we will use "Foursquare" location 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 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:
Neighbourhood: Name of the Neighborhood
Neighbourhood Latitude: Latitude of the Neighborhood
Neighbourhood Longitude: Longitude of the Neighborhood
Venue: Name of the Venue
Venue Latitude: Latitude of Venue
Venue Longitude: Longitude of Venue
Venue Category: Category of Venue
Based on all the information collected for both London and Paris, 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.

4. Visualizing the Neighborhoods of London and Paris

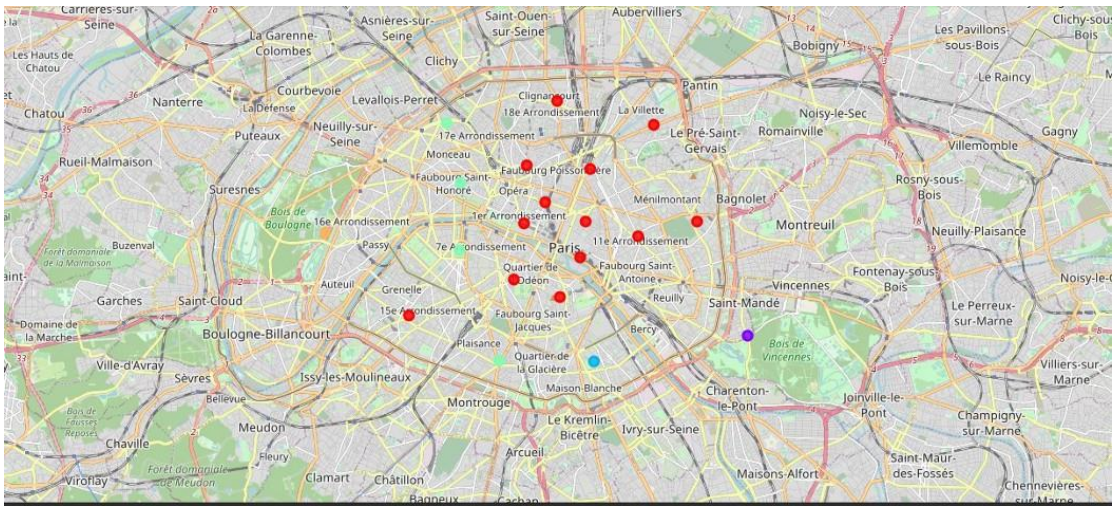


Hendon Central
Hampstead Garden Suburb
Harringay
Preston
Brent Cross
Highgate
Crouch End
Redbridge
Golders Green
Stroud Green
Stamford Hill
Leytonstone
Leyton
North Wembley
North Cricklewood
Upper Holloway
Cricklewood
Hampstead
Gospel Oak
Lower Holloway
Stoke Newington
Ilford
Wembley
West Hampstead
Kentish Town
Shacklewell
Homerton
Stratford New Town
Little Ilford
Belsize Park
Willesden
Camden Town
Canonbury
Hackney
Kilburn
Islington
Stratford
Upton
Harlesden
Alperton
Haggerston
Upton Park
Perivale
St. John's Wood
King's Cross
Hoxton
Bethnal Green
Bow
West Ham
East Ham
Park Royal
West Kilburn
Maida Hill
Shoreditch
Plaistow
Mile End
Marylebone
Clerkenwell
North Kensington
Stepney
Canning Town
Paddington
Holborn
City of London
Beckton
Ealing
Notting Hill
Mayfair
Covent Garden
Limehouse
Shadwell
Poplar
Acton
Shepherd's Bush
London
Blackwall
Cyprus
Wapping
Southwark
Rotherhithe
Greenwich Peninsula
North Woolwich
Kensington
City of Westminster
Elephant and Castle
Chiswick
Hammersmith
Earl's Court
Millbank
Charlton Riverside
Walworth
Brentford
Chelsea
Pimlico
Vauxhall
East Greenwich
Kennington
Greenwich
Charlton
Woolwich
Kew
Walham Green
Fulham
Camberwell
Peckham
New Cross Gate
Deptford
Barnes



Carrières-sur-Seine
Asnières-sur-Seine
Saint-Ouen-sur-Seine
Aubervilliers
Livry-Gargan
Clichy-sous-Bois
La Garenne-Colombes
Les Pavillons-sous-Bois
Les Hauts de Chatou
Bobigny
Clichy
Le Raincy
Courbevoie
Pantin
Nanterre
Levallois-Perret
Clignancourt
Chatou
La Défense
18e Arrondissement
La Villette
Noisy-le-Sec
Neuilly-sur-Seine
Puteaux
17e Arrondissement
Le Pré-Saint-Gervais
Romainville
Gagny
Monceau
Villemomble
Rueil-Malmaison
Faubourg Poissonnière
Faubourg Saint-Honoré
Rosny-sous-Bois
Suresnes
Opéra
Ménilmontant
Bagnolet
Bois de Boulogne
16e Arrondissement
1er Arrondissement
Montreuil
Neuilly-Plaisance
Buzenval
11e Arrondissement
Passy
7e Arrondissement
Paris
Forêt domaniale de la Malmaison
Faubourg Saint-Antoine
Quartier de l'Odéon
Auteuil
Vincennes
Fontenay-sous-Bois
Grenelle
Noisy-le-
Reuilly
Garches
Saint-Cloud
Saint-Mandé
15e Arrondissement
Le Perreux-sur-Marne
Faubourg Saint-Jacques
Domaine de la Marche
Boulogne-Billancourt
Bercy
Plaisance
Bois de Vincennes
Quartier de la Glacière
Ville-d'Avray
Villiers-sur-Marne
Sèvres
Issy-les-Moulineaux
Maison-Blanche
Charenton-le-Pont
Bellevue
Joinville-le-Pont
Montrouge
Bois de Fausses Reposes
Champigny-sur-Marne
Meudon
Le Kremlin-Bicêtre
Ivry-sur-Seine
Chaville
Arcueil
Fleury
Saint-Maur-des-Fossés
Forêt domaniale de Meudon
Viroflay
Clamart
Châtillon
Maisons-Alfort
Chennevières-

Cricklewood
Hampstead
Gospel Oak
Lower Holloway
Stoke Newington
Ilford
Wembley
West Hampstead
Kentish Town
Shacklewell
Homerton
Stratford New Town
Little Ilford
Belsize Park
Willesden
Camden Town
Canonbury
Hackney
Kilburn
Islington
Stratford
Upton
Harlesden
Alperton
Haggerston
Upton Park
Perivale
St. John's Wood
King's Cross
Hoxton
Bethnal Green
Bow
West Ham
East Ham
Park Royal
West Kilburn
Maida Hill
Shoreditch
Plaistow
Mile End
Marylebone
Clerkenwell
North Kensington
Stepney
Canning Town
Paddington
Holborn
City of London
Beckton
Ealing
Notting Hill
Mayfair
Covent Garden
Limehouse
Shadwell
Poplar
Acton
Shepherd's Bush
London
Blackwall
Cyprus
Wapping
Southwark
Rotherhithe
Greenwich Peninsula
North Woolwich
Kensington
City of Westminster
Elephant and Castle
Chiswick
Hammersmith
Earl's Court
Millbank
Charlton Riverside
Walworth
Brentford
Chelsea
Pimlico
Vauxhall
East Greenwich
Kennington
Greenwich
Charlton
Woolwich
Kew
Walham Green
Fulham
Camberwell
Peckham
New Cross Gate
Deptford
Barnes
Battersea
Isleworth
North Sheen
Stockwell
Blackheath
Shooters Hill
Richmond
East Sheen
Putney
Clapham Junction
Clapham
Nunhead
Lewisham
St Margarets
Wandsworth
Brixton
Brockley
Lee
Well Hall
Eltham
Roehampton
Honor Oak
Hither Green



5. Conclusion

The purpose of this project was to explore the cities of London and Paris 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 neighborhoods finally concluding with clustering similar neighborhoods together.

We could see that each of the neighborhoods in both the cities have a wide variety of experiences to offer which is unique in its own way. The cultural diversity is quite evident which also gives the feeling of a sense of inclusion.

Both Paris and London seem to offer a vacation stay or a romantic getaway with a lot of places to explore, beautiful landscapes, amazing food 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.