

# Clustering Neighborhoods of London and Paris using Machine Learning

-

# Data Description

We require geographical location data for both London and Paris.

## **London**

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

## **Paris**

To derive our solution, We leverage JSON data available at <https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e>

## **ArcGIS API**

ArcGIS Online enables you to connect people, locations, and data using interactive maps.

## **Foursquare API Data**

Foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

# Data Collection - LONDON

```
url_london =  
"https://en.wikipedia.org/wiki/List_of_areas_of_London"  
wiki_london_url = requests.get(url_london)  
wiki_london_data = pd.read_html(wiki_london_url.text)  
wiki_london_data = wiki_london_data[1]  
wiki_london_data
```

	Location	London borough	Post town	Postcode district	Dial code	OS grid ref
0	Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
2	Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728
...	...	...	...	...	...	...
527	Woolwich	Greenwich	LONDON	SE18	020	TQ435795
528	Worcester Park	Sutton, Kingston upon Thames	WORCESTER PARK	KT4	020	TQ225655
529	Wormwood Scrubs	Hammersmith and Fulham	LONDON	W12	020	TQ225815
530	Yeading	Hillingdon	HAYES	UB4	020	TQ115825
531	Yiewsley	Hillingdon	WEST DRAYTON	UB7	020	TQ063804



# Data Collection - PARIS

```
!wget -q -O 'france-data.json'
```

```
https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e
```

```
print("Data Downloaded!")
```

```
paris_raw = pd.read_json('france-data.json')
```

```
paris_raw.head()
```

	datasetid	recordid	fields	geometry	record_timestamp
0	correspondances-code-insee-code-postal	2bf36b38314b6c39dfbcd09225f97fa532b1fc45	{'code_comm': '645', 'nom_dept': 'ESSONNE', 's...	{'type': 'Point', 'coordinates': [2.2517129721...	2016-09-21T00:29:06.175+02:00
1	correspondances-code-insee-code-postal	7ee82e74e059b443df18bb79fc5a19b1f05e5a88	{'code_comm': '133', 'nom_dept': 'SEINE-ET-MAR...	{'type': 'Point', 'coordinates': [3.0529405055...	2016-09-21T00:29:06.175+02:00
2	correspondances-code-insee-code-postal	e2cd3186f07286705ed482a10b6aebd9de633c81	{'code_comm': '378', 'nom_dept': 'ESSONNE', 's...	{'type': 'Point', 'coordinates': [2.1971816504...	2016-09-21T00:29:06.175+02:00
3	correspondances-code-insee-code-postal	868bf03527a1d0a9defe5cf4e6fa0a730d725699	{'code_comm': '243', 'nom_dept': 'SEINE-ET-MAR...	{'type': 'Point', 'coordinates': [2.7097808131...	2016-09-21T00:29:06.175+02:00
4	correspondances-code-insee-code-postal	21e809b1d4480333c8b6fe7add8f3b06f343e2c	{'code_comm': '003', 'nom_dept': 'VAL-DE-MARNE...	{'type': 'Point', 'coordinates': [2.3335102498...	2016-09-21T00:29:06.175+02:00

# Data Preprocessing

For London, We replace the spaces with underscores in the title. The *borough* column has numbers within square brackets

```
wiki_london_data.rename(columns=lambda x: x.strip().replace(" ",
"_"), inplace=True)
wiki_london_data['borough'] =
wiki_london_data['borough'].map(lambda x:
x.rstrip(']').rstrip('0123456789').rstrip('['))
```

For Paris, we break down each of the nested fields and create the dataframe

```
paris_field_data = pd.DataFrame()
for f in paris_raw.fields:
    dict_new = f
    paris_field_data = paris_field_data.append(dict_new,
ignore_index=True)

paris_field_data.head()
```

# Feature Selection & Engineering

- For both of our datasets, we need only the borough, neighbourhood, postal codes and geolocations (latitude and longitude).
- Both of our Datasets contain information related to all the cities in the country. We can narrow down and further process the data by selecting only the neighbourhoods of 'London' and 'Paris'.
- With the Help of ArcGIS API we can get the latitude and longitude of our London neighbourhood data.

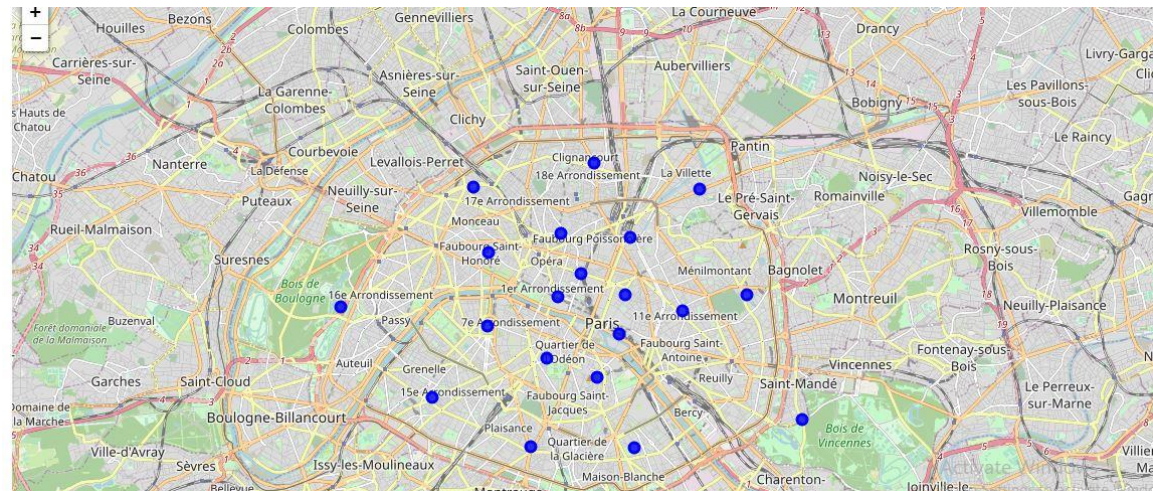
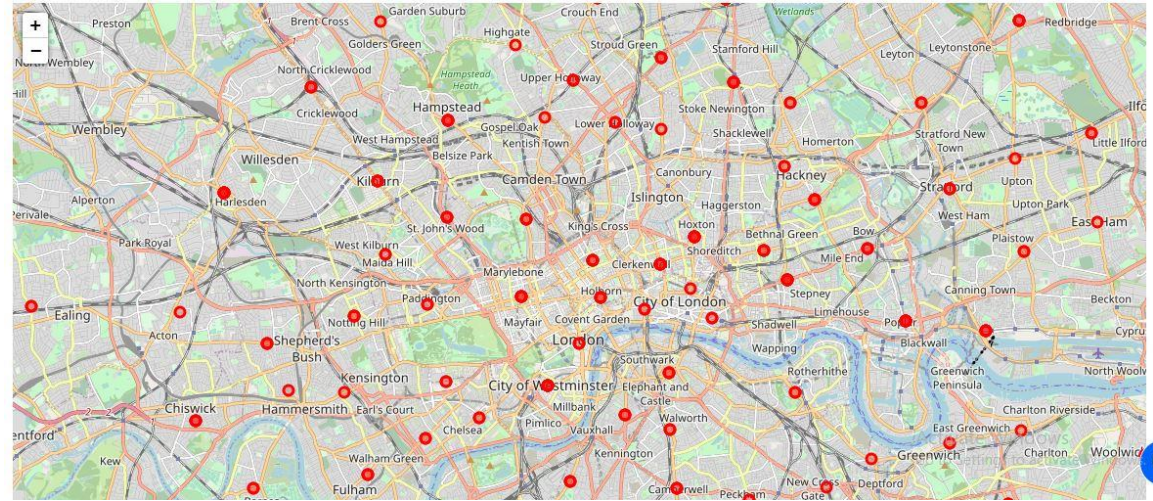
```
from arcgis.geocoding import geocode
from arcgis.gis import GIS
gis = GIS()
```

# Visualizing the Neighborhoods of London and Paris

we have visualized the neighbourhoods, we need to find out what each neighbourhood is like and what are the common venue and venue categories within a 500 m radius.

This is where Foursquare comes into play.

With the help of Foursquare we define a function which collects information pertaining to each neighbourhood including that of the name of the neighbourhood, Geo-coordinates, venue and venue categories.





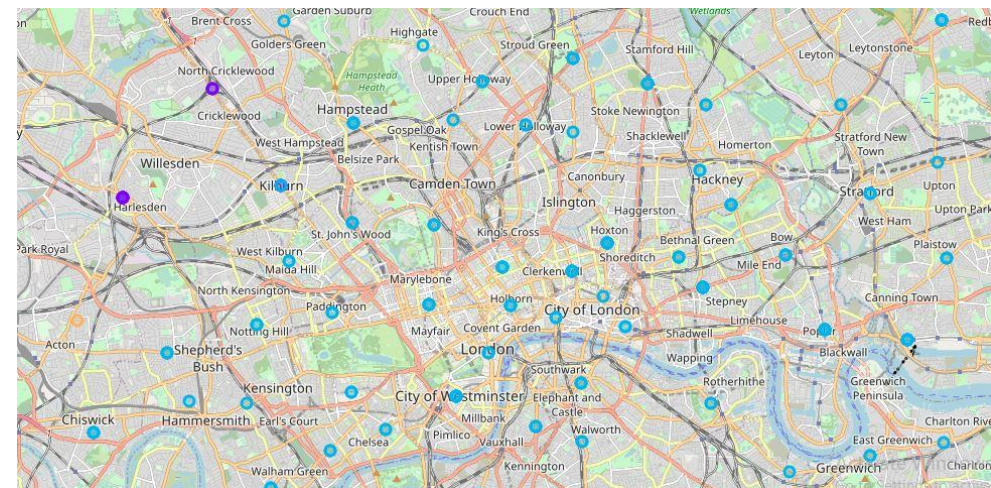
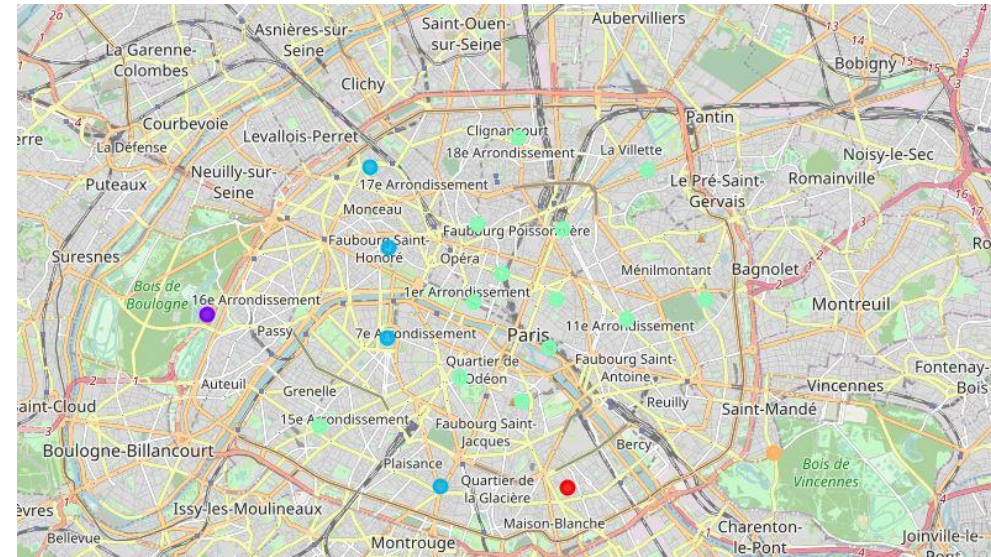
# Top Venues in the Neighborhoods

	Neighbourhood	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
0	Barnet	Coffee Shop	Café	Grocery Store	Italian Restaurant	Pub	Bus Stop	Supermarket	Sushi Restaurant	Pharmacy
1	Barnet, Brent, Camden	Bus Station	Gym / Fitness Center	Clothing Store	Supermarket	Zoo Exhibit	Fish & Chips Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant
2	Bexley	Supermarket	Historic Site	Coffee Shop	Convenience Store	Train Station	Park	Bus Stop	Construction & Landscaping	Golf Course
3	Bexley, Greenwich	Sports Club	Convenience Store	Park	Historic Site	Construction & Landscaping	Golf Course	Bus Stop	Flower Shop	Food & Drink Shop
4	Bexley, Greenwich	Supermarket	Train Station	Convenience Store	Coffee Shop	Historic Site	Fish & Chips Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant

# Visualizing the clustered Neighborhoods

Our data is processed, missing data is collected and compiled. The Model is built. All that's remaining is to see the clustered neighbourhoods on the map. Again, we use Folium package to do so.

We could examine our clusters by expanding on our code using the Cluster Labels column.



# Results and Discussion

The neighbourhoods of London are very multicultural. There are a lot of different cuisines including Indian, Italian, Turkish and Chinese. London seems to take a step further in this direction 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 too with that of the Flea markets, flower shops, fish markets, Fishing stores, clothing stores.

Paris is relatively small in size geographically. It has a wide variety of cuisines and eateries including French, Thai, Cambodian, Asian, Chinese etc. There are a lot of hangout spots including many Restaurants and Bars. Paris has a lot of Bistros. Different means of public transport in Paris which includes buses, bikes, boats or ferries. For leisure and sightseeing, there are a lot of Plazas, Trails, Parks, Historic sites, clothing shops, Art galleries and Museums.

# 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 neighbourhoods finally concluding with clustering similar neighbourhoods together.

## References

[The Battle of Neighbourhood — My London's Perspective by Dayo John](#)

[The Battle of neighborhoods! What is the best place where can I start my restaurant business in Paris? by Zakaria BOUZIANE](#)

[Foursquare API](#)

[ArcGIS API](#)