1. **Data where you describe the data that will be used to solve the problem and the source of the data**

We will use the following data :

1. List of London Borough & their GPS coordinates : Longitude - Latitude that i have available in an excel format from a previous project. :
   1. <https://github.com/Philreb/coursera_capstone_project/blob/main/london_coordinates2.xlsx>
2. Venue data that will be extracted from Foursquare API and will be used for the clustering of the neighbourhood

**Methodologie :**

1. Visualize the Borough of London
   1. Download and read coordinates data
   2. Use **Folium package** to create a map of London with Borough superimposed on top
2. Explore neighborhood
   1. **Foursquare API**
   2. Define Foursquare credentials and version
   3. Explore the neighbourhood in our dataframe and extract the top 50 venues in a 5000 radius.
   4. Convert the venues list in a new dataframe
   5. Add the venues to the london map
3. Analyse neighborhood
   1. Proceed a one hot encoding
   2. Group by rows
   3. Define top venues
4. Cluster the neighborhood
   1. Create the clusters with with **k-means**
   2. Visualize the clusters by adding them to the map
5. Examine cluster & conclusion
   1. Identify the cluster with “pub” as most common venues

---------------------------------------------------------------------------------------------------------------------------