In [20]:

```
import pandas as pd

# Create a dataframe from csv
df = pd.read_csv('ClasseurResultat_LIENS.csv', delimiter=';')

# User list comprehension to create a list of lists from Dataframe rows
list_of_rows = [list(row) for row in df.values]
```

In [27]:

```
liste_pseudo = []
for i in range(len(list_of_rows)):
    liste_pseudo.append(list_of_rows[i][0])
```

In [63]:

```
liste_pseudo_unique = set(liste_pseudo)
liste_pseudo_unique = list(liste_pseudo_unique)
```

In [64]:

```
liste_pseudo_unique
```

Out[64]:

```
['KalfaOnogo',
 'Gladiat474',
 'simonot_annick',
 'aloifiafia_z',
 'doudou_2_0_',
 'jorgembneto',
 'PelegrinLea',
 'Emir15822794',
 'Loic_MJ',
 '...']
```

In [31]:

```
# Create a dataframe from csv
df = pd.read_csv('dico_unifie_NOEUDS.csv', delimiter=';')

# User list comprehension to create a list of lists from Dataframe rows
list_of_rows = [list(row) for row in df.values]
```

In [131]:

```
# Create a dataframe from csv
df = pd.read_csv('liste.csv')

# User list comprehension to create a list of lists from Dataframe rows
list_of_rows2 = [list(row) for row in df.values]
```

In [132]:

```
liste_pseudo_unique_2 = []
for i in range(0, len(list_of_rows2)):
    liste_pseudo_unique_2.append(list_of_rows2[i][0])
```

In [133]:

```
liste_pseudo_unique_2
```

Out[133]:

```
['EmmanuelMacron',
```

```
 'MLP_officiel',
 'nicolassarkozy',
 'fhollande',
 'jlmelenchon',
 'anne_hidalgo',
 'manuelvalls',
 'alainjuppe',
 'najatvb',
 'ephilippe_lh',
 'benoithamon',
 'chtaubira',
 'nk_m',
 'royalsegolene',
 'bayrou',
 'Lagarde',
 'fleurpellerin',
 'FrancoisFillon',
 'BrunoLeMaire',
 'montebourg',
 'CecileDuflot',
 'MarionMarechal',
 '...'
]
```

In [135]:

```
liste_a_garder = liste_pseudo_unique + liste_pseudo_unique_2
```

In [136]:

```
liste_a_garder
```

Out[136]:

```
['KalfaOnogo',
 'Gladiat474',
 'simonot_annick',
 'aloifiafia_z',
 'doudou_2_0_',
 'jorgembneto',
 'PelegrinLea',
 'Emir15822794',
 'Loic_MJ',
 'Mathys9360',
 'PinsonMichel1',
 'RomaneCaille',
 'MustafaDuman_',
 'nab_chat',
 '...']
```

In [69]:

```
# On peut trier les csv, on commence avec les noeuds
```

In [77]:

```
# Create a dataframe from csv
df = pd.read_csv('dico_unifie_NOEUDS.csv', delimiter=',')

# User list comprehension to create a list of lists from Dataframe rows
list_of_rows = [list(row) for row in df.values]
```

In [137]:

```
liste_a_garder_casse = []
for i in range(len(liste_a_garder)):
    liste_a_garder_casse.append(liste_a_garder[i].lower())
```

In [115]:

```
len(liste_a_garder_casse)
```

Out[115]:

```
486
```

In [138]:

```
liste_noeud = []
for i in range(len(list_of_rows)):
    if list_of_rows[i][0].lower() in liste_a_garder_casse:
        liste_noeud.append([list_of_rows[i][0] , list_of_rows[i][1]])
```

In [139]:

```
len(liste_noeud)
```

Out[139]:

```
153
```

In [140]:

```
list_of_rows
```

Out[140]:

```
[['EmmanuelMacron', '62808'],
 ['lemondefr', '68472'],
 ['Le_Figaro', '421a04'],
 ['BarackObama', '50102'],
 ['JoeBiden', '38524'],
 ['le_Parisien', '42546'],
 ['RFI', '32639'],
 ['BFMTV', '36270'],
 ['France24_fr', '30509'],
 ['CNEWS', '32668'],
 ['franceinfo', '34471'],
 ['jeune_afrique', '23211'],
 ['afpfr', '30669'],
 ['realDonaldTrump', '24374'],
 ['NetflixFR', '26365'],
 ['Elysee', '23174'],
 ['MichelleObama', '22406'],
 ...]
```

In [141]:

```
liste_noeud
```

Out[141]:

```
[['EmmanuelMacron', '62808'],
 ['lemondefr', '68472'],
 ['Le_Figaro', '421a04'],
 ['le_Parisien', '42546'],
 ['RFI', '32639'],
 ['BFMTV', '36270'],
 ['France24_fr', '30509'],
 ['CNEWS', '32668'],
 ['franceinfo', '34471'],
 ['jeune_afrique', '23211'],
 ['afpfr', '30669'],...]
```

In [142]:

```
df = pd.DataFrame(liste_noeud)

# saving the dataframe
df.to_csv('ln.csv')
```

In [143]:

```
# Create a dataframe from csv
df = pd.read_csv('ClasseurResultat_LIENS.csv', delimiter=';')

# User list comprehension to create a list of lists from Dataframe rows
list_of_rows = [list(row) for row in df.values]
```

In [169]:

```
liste_noeud_tmp = []
for i in range (len(liste_noeud)):
        liste_noeud_tmp.append(liste_noeud[i][0])
```

In [170]:

```
nouvelle_liste = []
for i in range (len(list_of_rows)):
    if list_of_rows[i][1] in liste_noeud_tmp:
        nouvelle_liste.append([list_of_rows[i][0], list_of_rows[i][1]])
```

In [172]:

```
len(nouvelle_liste)
```

Out[172]:

```
4048
```

In [173]:

```
df = pd.DataFrame(nouvelle_liste)

# saving the dataframe
df.to_csv('LIENS_TRIES.csv')
```

In [ ]:

```

```