

Analyse de quelques données des panama papers

- [1 Récupération des données](#)
- [2 Fichier d'arêtes](#)
- [3 Fichier d'intermédiaires](#)
- [4 Recoupement des informations](#)
- [5 Carte](#)
 - [5.1 Basemap](#)
 - [5.2 Folium](#)

1 Récupération des données

Commencez par télécharger les deux fichiers suivants :

- [panama-all_edges.csv](#)
- [panama-Intermediaries](#)

Placez les dans votre dossier de travail.

Pour lire ces fichiers CSV, vous pouvez utiliser le code suivant, qui renvoie une liste (contenant autant d'éléments qu'il y a de lignes dans le fichier csv):

```
def read_csv(filename):
    import csv
    lines = []
    with open(filename) as csv_file:
        csv_reader = csv.reader(csv_file, delimiter=',')
        for row in csv_reader:
            if row:
                lines.append(row)
    return lines

edges = read_csv("panama-all_edges.csv")
inter = read_csv("panama-Intermediaries.csv")
```

2 Fichier d'arêtes

edges contient des triplets, qui décrivent les arcs orientés d'un graphe. Ces arcs sont tagués par un nom de relation.

Par exemple, la ligne :

```
['11000002', 'intermediary of', '10188436']
```

signifie que l'entité 11000002 (c'est un sommet du graphe) a joué un rôle d'intermédiaire pour l'entité 10188436. Le tag de l'arc est intermediary of.

Combien y a-t-il de tags (réellement) différents dans ce fichier ?

3 Fichier d'intermédiaires

On va s'intéresser plus particulièrement aux entités qui ont joué un rôle d'intermédiaire. Elles sont toutes référencées dans le second fichier. Par exemple, 11000002 est détaillée sur la deuxième ligne de données du fichier d'intermédiaires :

```
'CORFIDUCIA ANSTALT', '10004', '', 'The Panama Papers data is current through 2015',
'LIE', 'Liechtenstein', 'ACTIVE', '11000002', 'Panama Papers'
```

La colonne 0 contient le nom de l'entité (personne, société...), la colonne 7 son indentifiant 11000002 et la colonne 4 le code du pays de résidence (ISO 3166-1 alpha-3 *a priori*).

Combien de pays différents sont présents dans le fichier des intermédiaires ? Constatez que certaines données sont manquantes, ou que parfois, plusieurs pays sont mentionnés.

4 Recoupement des informations

Notre objectif est de calculer un score par pays qui est d'autant plus grand qu'il héberge un grand nombre d'intermédiaires. Attention, un intermédiaire particulier a pu jouer son rôle d'intermédiaire plusieurs fois. S'il a joué 10 fois le rôle d'intermédiaire, on peut considérer qu'il faut augmenter le score du pays de 1 (puisque'on a trouvé 1 intermédiaire de plus) ou de 10 (puisque'il y a 10 rôles d'intermédiaires). Les deux solutions sont acceptables, essayez de faire les deux.

Calculez le score de chaque pays.

Quels sont les pays qui ont les scores les plus élevés ?

5 Carte

Téléchargez le fichier suivant, il contient les coordonnées de la plupart des pays du globe :

- [Country List ISO 3166 Codes Latitude Longitude.csv](#)

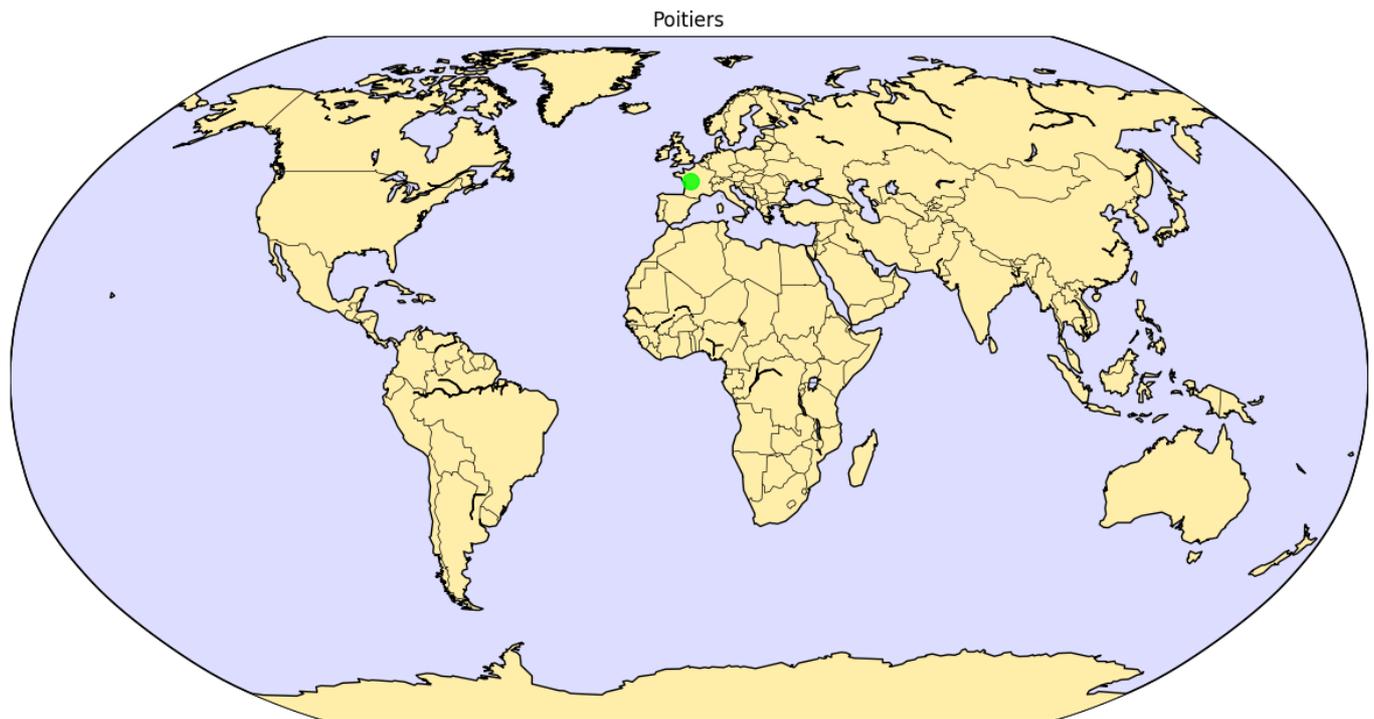
Créez la structure de données adaptée, qui vous permet d'obtenir la latitude et la longitude d'un pays à partir de son code ISO 3166.

Pour tracer des cartes en Python, vous pouvez utiliser les modules Basemap ou Folium. Un exemple d'utilisation de ces deux modules est donné un peu plus loin (testez-les pour savoir quel module choisir).

Sur une carte du monde, mettez en évidence les scores calculés pour chaque pays (changement de couleur, ou taille du marqueur).

5.1 Basemap

Le programme suivant produit cette carte :



```
from mpl_toolkits.basemap import Basemap
import matplotlib.pyplot as plt

# FOND DE CARTE
plt.figure(figsize=(15,15))
carte = Basemap(projection='robin',lon_0=0,resolution='c')
carte.drawcoastlines()
carte.drawcountries()
#m.drawstates()
carte.fillcontinents(color='#ffeeaa',lake_color='#ddddff')
carte.drawmapboundary(fill_color='#ddddff')

# Ajout d'un marqueur
# Convertit long/lat en coordonnées sur le fond de cartes
xpt,ypt = carte(0.3404, 46.5802)
# Ajout d'un point vert
carte.plot(xpt,ypt, 'ro', markersize=10, alpha=0.8, color=(0,1,0))

# Affichage de la carte
plt.title("Poitiers")
plt.show()
```

5.2 Folium

Les cartes Folium sont interactives (et il faut les enregistrer pour les ouvrir ensuite dans un navigateur) :



```
import folium
from folium import plugins

carte = folium.Map(location=[46.580224,0.340375], tiles='cartodbpositron', zoom_start=3)

loc = [46.5802, 0.3404]
folium.RegularPolygonMarker(loc, color='black',
                             opacity=0, weight=0, fill_color='#f00',
                             fill_opacity=0.4, number_of_sides=6,
                             rotation=0, radius=10, popup="",
                             tooltip="Poitiers").add_to(carte)

carte.save("carte.html")
```