

TP 1 - DATASCIENCE

Nom : SILINOu NICOLAS SHEDDY

Gitlab: [SILINOu NICOLAS / TP1 Datascience](#) · [GitLab \(esiea.fr\)](#)

Manipulation de données

```
import pandas as pd
dadd=pd.read_csv('nodes-addresses.csv', low_memory=False)
doff=pd.read_csv('nodes-officers.csv', low_memory=False)
dent=pd.read_csv('nodes-entities.csv', low_memory=False)
drel=pd.read_csv('relationships.csv', low_memory=False)
```

✓ 12.6s

Exercice I avec Pandas

1. Combien de pays ont été déclarés dans les offshore leaks ?

```
pays_uniques = dadd[dadd["sourceID"] == "Offshore Leaks"]["countries"].nunique()
print(f"Nombre de pays déclarés dans les offshore leaks est : {pays_uniques}")
```

✓ 0.0s

Nombre de pays déclarés dans les offshore leaks est : 172

2. Dans combien d'entrée le mot "Ireland" est présent ?

Le mot 'Ireland' est présent dans 981 entrées

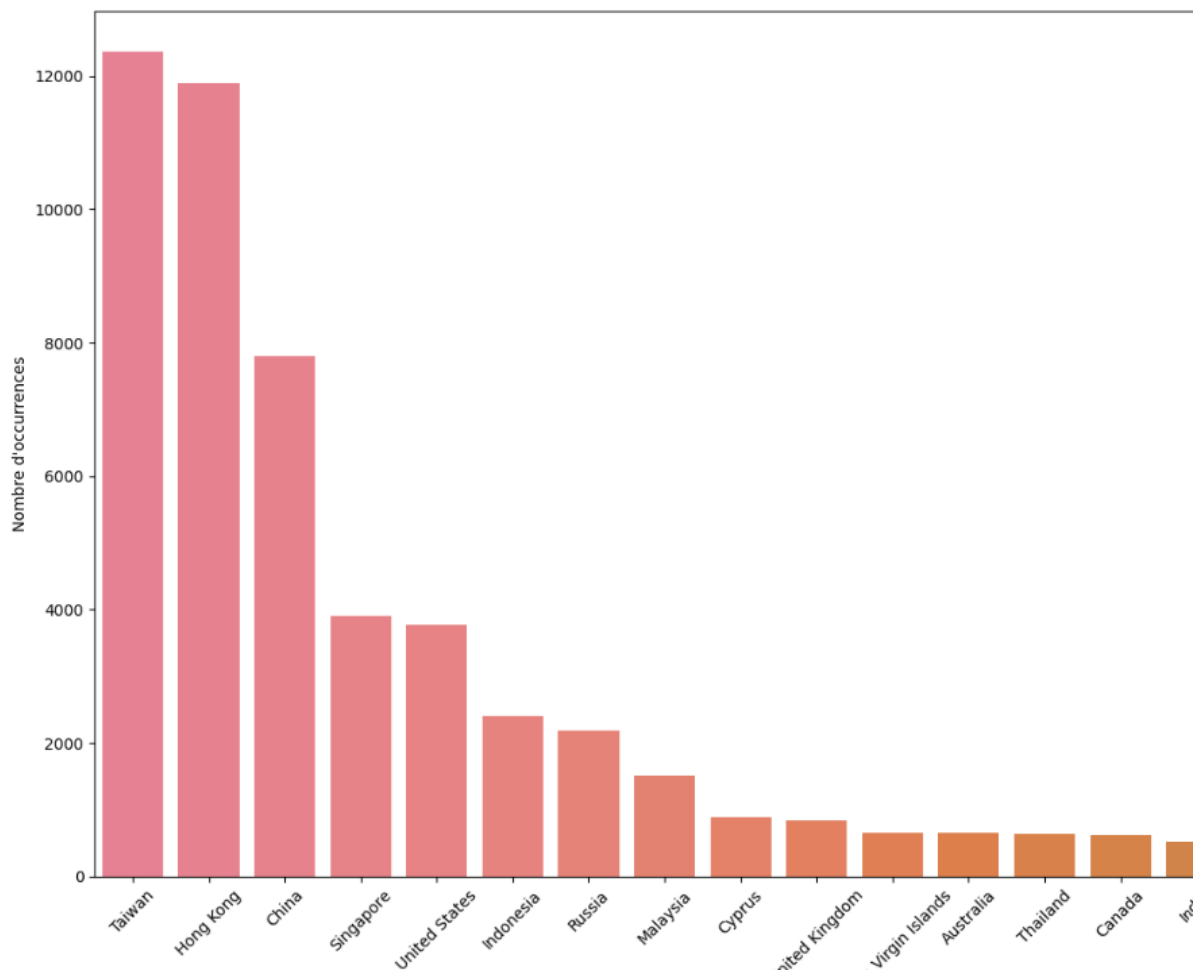
3. Dans combien de montage apparaît "EL PORTADOR" ?

'EL PORTADOR' apparaît dans 9325 montages

EXERCICE 2

4. Faites un histogramme des pays qui ont été déclarés dans les offshore leaks

(j'ai mis un fichier « output.png » sur le gitlab qui contient le graphe complet)



5. Est-on en présence d'une loi normale si non quel est le top 20 des pays les plus déclarés ?

Nous ne sommes pas en présence d'une normale car la courbe ne suit pas la forme d'une cloche. De plus, d'après le calcul ci-dessous, p-valeur > 0.05.

```
from scipy import stats

# Générer des données normalement distribuées pour l'exemple
data = stats.norm.rvs(size=1000)

# Test de Shapiro-Wilk
shapiro_test = stats.shapiro(data)

print(f"Statistique de test: {shapiro_test[0]}, p-valeur: {shapiro_test[1]}")
```

✓ 0.0s

Pyth

Statistique de test: 0.9983358974503368, p-valeur: 0.4523520788588303

	Country	Count
0	Taiwan	12361
1	Hong Kong	11894
2	China	7806
3	Singapore	3909
4	United States	3766
5	Indonesia	2400
6	Russia	2188
7	Malaysia	1506
8	Cyprus	896
9	United Kingdom	843
10	British Virgin Islands	658
11	Australia	653
12	Thailand	634
13	Canada	624
14	India	514
15	Japan	467
16	Switzerland	380
17	Not identified	356
18	Philippines	355
19	United Arab Emirates	354

EXERCICE 3

6. Combien de service providers ont été déclarés dans les offshore leaks et quelle est la proportion de ces service providers vis-à-vis de des officers ?

```

Nombre de service providers uniques : 2
Nombre total d'officers : 746949
Proportion de service providers par rapport aux officers : 2.677558976583408e-06

```