

TD 6 : Le module pandas

Répondez aux questions suivantes :

1. Importez les librairies pandas et numpy

```
import numpy as np
import pandas as pd
```

2. Lisez le fichier de données cities.csv

```
df = pd.read_csv('cities.csv')
```

3. Affichez sa taille et regardez quelques premières lignes

```
df.shape
df.head()
```

4. Vous remarquez une première colonne inutile, utilisez la méthode drop des dataframes pour supprimer cette colonne

```
df.drop(columns=['unnamed'], inplace=True)
ou
df = df.drop(columns=['unnamed'])
```

5. Appelez la méthode info des dataframes

```
df.info()
```

6. Remarquez une colonne entièrement vide, utilisez la méthode dropna des dataframes pour supprimer en place les colonnes qui ont toutes leurs valeurs manquantes

```
df.dropna(axis=1, how='all', inplace=True)
```

7. Vérifiez que vous avez bien enlevé les colonnes vides

```
df.info()
```

8. Affichez la ligne d'index 78, que remarquez-vous ?

```
df.iloc[78]
On remarque que la ligne est vide.
```

9. Utilisez la méthode dropna des dataframes pour supprimer en place les lignes qui ont toutes leurs valeurs manquantes (et pas uniquement la ligne d'index 78)

```
df.dropna(axis=0, how='all', inplace=True)
```

10. Quelle est la population totale des villes dans le dataset ?

```
df['population'].sum()
```

11. Quelle est la population totale des villes aux Maroc ?

```
df2 = df[df['country']=='Morocco']  
df2['population'].sum()
```

12. Quelle ville a la plus petite population ?

```
df[df['population']==df['population'].min()]
```

13. Quelle ville a la plus grande population ?

```
df[df['population']==df['population'].max()]
```

14. Quelle est le nombre des villes aux Maroc ?

```
df2 = df[df['country']=='Morocco']  
df2['city'].count()
```

15. Quelle est capital du Maroc ?

```
df[(df['country']=='Morocco') & (df['capital']=='primary')]
```

16. Donner la liste des villes triée par leur population dans ordre décroissant ?

```
df.sort_values('population', ascending=False)
```