Dataminig TP1 :
Data clean

Ressources:

- book.csv
- university_towns.txt
- olympics.csv

PART 1 (book.csv)

Q1: Exporter book.csv dans un data frame nommé df puis afficher les 5 premières lignes

Q2: Afficher les propriétés du data frame

Q3: Nous allons restreindre l'étude à seulement 7 colonnes pour cela nous allons supprimer 7 colonnes que nous jugeons inutiles pour le moment :

```
to_drop = ['Edition Statement',
           'Corporate Author',
           'Corporate Contributors',
           'Former owner',
           'Engraver',
           'Contributors',
           'Issuance type',
           'Shelfmarks']
```

Q4: Changer l'index pourqu'il prenne la valeur de l'identifiant “Identifier”

Q5: Afficher les 25 premières valeurs de la colonne “Date of Publication”

Q6: Ecrire une fonction qui permet de nettoyer la colonne des dates

Q7: Ecrire une fonction qui permet de nettoyer la colonne des auteurs

Q8: Ecrire une fonction qui permet de nettoyer la colonne des titres

PART 2 (university_tows.txt)

Q1: Exporter book.csv dans un data frame nommé df puis afficher les 5 premières lignes

Q2: Nous voyons que nous avons des noms d'états périodiques suivis des villes universitaires de cet état: StateA TownA1 TownA2 StateB TownB1 TownB2 .... Si nous regardons la façon dont les noms d'états sont écrits dans le fichier, nous verrons qu'ils ont tous la sous-chaîne «[edit]» en eux.

Exporter book.csv dans un data frame avec les colonnes “State” et City” puis afficher les 5 premières lignes

Q3: Ecrire une fonction qui permet de nettoyer les datas

PART 3 (olympics.csv)

Q1 : exporter 'olympics.csv' sous forme de data frame nommé olympics_df puis afficher les 5 premières lignes

Q2 : Exporter 'olympics.csv' sous forme de data frame nommé olympics_df en supprimant la première ligne puis afficher les 5 premières lignes

Q3 : Changer le nom des colonnes en utilisant le dictionnaire new_names :

```
new_names =  {'Unnamed: 0': 'Country',
              '? Summer': 'Summer Olympics',
              '01 !': 'Gold',
              '02 !': 'Silver',
              '03 !': 'Bronze',
              '? Winter': 'Winter Olympics',
              '01 !.1': 'Gold.1',
              '02 !.1': 'Silver.1',
              '03 !.1': 'Bronze.1',
              '? Games': '# Games',
              '01 !.2': 'Gold.2',
              '02 !.2': 'Silver.2',
              '03 !.2': 'Bronze.2'}
```

Bravo, vous avez terminé