

Charger la librairie Pandas

```
In [1]: import pandas as pd
```

Lecture des données avec Pandas : Chargement du fichier

df étant le nom de l'objet de type data frame à créer
sep spécifie le caractère séparateur de colonnes
header = 0 : la ligne numéro 0 = aux noms des champs

Exercice 1

1- Lisez la base de données heart.txt

```
In [2]: df = pd.read_table("heart.txt", sep = '\t', header = 0)
```

Vérifications le type de df

L'objectif étant de prévoir le type d'opérations possibles

```
In [3]: print(type(df))
```

```
<class 'pandas.core.frame.DataFrame'>
```

2-3 Détermination du nombre d'individus et de variables

Dimensions de la table de données : nombre de lignes, nombre de colonnes

la ligne d'en-tête n'est pas comptabilisée dans le nombre de lignes

```
In [4]: dimension = df.shape
```

```
NbrLignes = df.shape[0]
```

```
NbrColonnes = df.shape[1]
```

```
In [5]: print("Dimension : ", dimension)
```

```
print("Nombre de lignes : ", NbrLignes)
```

```
print("Nombre de colonnes : ", NbrColonnes)
```

```
Dimension : (270, 13)
```

```
Nombre de lignes : 270
```

```
Nombre de colonnes : 13
```

Le nombre d'individus est égale à 270 et le nombre de variables est égale à 13

4- Afficher les six premières lignes du jeu de données

```
In [6]: print(df.head(6))
```

```
   age sexe type_douleur pression cholester sucre electro taux_max \
```

```
0 70 masculin D 130 322 A C 109
```

```
1 67 feminin C 115 564 A C 160
```

```
2 57 masculin B 124 261 A A 141
```

```
3 64 masculin B 128 263 A A 105
```

```
4 74 feminin B 120 269 A C 123
```

```
5 65 masculin D 120 177 A A 140
```

```
   angine depression pic vaisseau coeur
```

```
0 non 2.4 2 D presence
```

```
1 non 1.6 2 A absence
```

```
2 non 0.3 1 A presence
```

```
3 oui 0.2 2 B absence
```

```
4 oui 0.2 1 B absence
```

```
5 non 0.4 1 A absence
```

```
In [7]: display(df)
```

```
   age sexe type_douleur pression cholester sucre electro taux_max \
```

```
0 70 masculin D 130 322 A C 109 non 2.4 2 D presence
```

```
1 67 feminin C 115 564 A C 160 non 1.6 2 A absence
```

```
2 57 masculin B 124 261 A A 141 non 0.3 1 A presence
```

```
3 64 masculin B 128 263 A A 105 oui 0.2 2 B absence
```

```
4 74 feminin B 120 269 A C 123 oui 0.2 1 B absence
```

```
... ... ... ... ... ... ... ... ... ... ... ... ...
```

```
265 52 masculin C 172 199 B A 162 non 0.5 1 A absence
```

```
266 44 masculin B 120 263 A A 173 non 0.0 1 A absence
```

```
267 56 feminin B 140 294 A C 153 non 1.3 2 A absence
```

```
268 57 masculin D 140 192 A A 148 non 0.4 2 A absence
```

```
269 67 masculin D 160 286 A C 108 oui 1.5 2 D presence
```

270 rows × 13 columns

5- Énumération des colonnes

```
In [8]: print(df.columns)
```

```
Index(['age', 'sexe', 'type_douleur', 'pression', 'cholester', 'sucre', 'electro', 'taux_max', 'angine', 'depression', 'pic', 'vaisseau', 'coeur'],
```

```
dtype='object')
```

Type de chaque variable

```
In [9]: print(df.dtypes)
```

```
age          int64
```

```
sexe         object
```

```
type_douleur object
```

```
pression      int64
```

```
cholester     int64
```

```
sucre         object
```

```
electro        object
```

```
taux_max      int64
```

```
angine         object
```

```
depression    float64
```

```
pic           int64
```

```
vaisseau      object
```

```
coeur          object
```

```
dtype: object
```

Informations sur les données

```
In [10]: print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 270 entries, 0 to 269
```

```
Data columns (total 13 columns):
```

```
# Column Non-Null Count Dtype
```

```
--
```

```
0 age    270 non-null int64
```

```
1 sexe   270 non-null object
```

```
2 type_douleur 270 non-null object
```

```
3 pression 270 non-null int64
```

```
4 cholester 270 non-null int64
```

```
5 sucre    270 non-null int64
```

```
6 electro   270 non-null object
```

```
7 taux_max 270 non-null int64
```

```
8 angine   270 non-null object
```

```
9 depression 270 non-null float64
```

```
10 pic     270 non-null int64
```

```
11 vaisseau 270 non-null object
```

```
12 coeur    270 non-null object
```

```
dtypes: float64(1), int64(5), object(7)
```

```
memory usage: 27.5+ KB
```

```
None
```

La base de données (structure df) comprend 270 individus (lignes) et 13 variables (colonnes). Les variables sont 'age', 'pression', 'cholester', 'taux_max', 'depression', 'pic' sont des variables quantitatives. Les variables 'sexe', 'type_douleur', 'sucre', 'electro', 'angine', 'vaisseau', 'coeur' sont des variables qualitatives.

6- Considérez un sous-ensemble df1 de 8 variables

```
In [11]: df1=df.iloc[:,0:8]
```

```
dimension = df1.shape
```

```
NbrLignes1 = df1.shape[0]
```

```
NbrColonnes1 = df1.shape[1]
```

```
print("Dimension : ", dimension)
```

```
print("Nombre de lignes : ", NbrLignes1)
```

```
print("Nombre de colonnes : ", NbrColonnes1)
```

```
Dimension : (270, 8)
```

```
Nombre de lignes : 270
```

```
Nombre de colonnes : 8
```

7- Énumération des colonnes de df1

```
In [12]: print(df1.columns)
```

```
Index(['age', 'sexe', 'type_douleur', 'pression', 'cholester', 'sucre', 'electro', 'taux_max'],
```

```
dtype='object')
```

La structure df1 comprend 270 individus (lignes) et 8 variables (colonnes). Les variables sont 'age', 'pression', 'cholester', 'taux_max', 'depression', 'pic' sont des variables quantitatives.

8- Considérez un sous-ensemble df1 de 8 variables

On peut également réduire le nombre d'individus à considérer

```
In [13]: df2=df1.iloc[0:100]
```

```
dimension2 = df2.shape
```

```
NbrLignes2 = df2.shape[0]
```

```
NbrColonnes2 = df2.shape[1]
```

```
print("Dimension : ", dimension2)
```

```
print("Nombre de lignes : ", NbrLignes2)
```

```
print("Nombre de colonnes : ", NbrColonnes2)
```

```
Dimension : (10, 8)
```

```
Nombre de lignes : 10
```

```
Nombre de colonnes : 8
```

Exercice 2: Analyse de la variable type_douleur

1- Vérifiez les modalités de la variable

```
In [14]: Modalites=df['type_douleur'].unique()
```

```
print("Les modalités sont: ", Modalites)
```

```
Les modalités sont: ['D' 'C' 'B' 'A']
```

Les modalités de la variable type_douleur sont: D, C, B, A

Attention les sorties des scripts ne constituent pas une réponse aux questions posées. Vous devez formuler des phrases

2- Calcul des effectifs

La fonction value_counts calcule les effectifs par modalité de la variable

```
In [15]: E1 = df['type_douleur'].value_counts()
```

```
print(E1)
```

```
D 129
```

```
C 79
```

```
B 42
```

```
A 20
```

```
Name: type_douleur, dtype: int64
```

Les effectifs des différentes modalités sont les suivants: A = 20, B = 42, C = 79 et D = 129

3- Calcul des fréquences

```
In [16]: E2 = 100*df['type_douleur'].value_counts()/df1.shape[0]
```

```
print("Effectif :")
```

```
D 47.77
```

```
C 25.555556
```

```
B 15.555556
```

```
A 7.407407
```

```
Name: type_douleur, dtype: float64 %
```

Les fréquences des différentes modalités sont les suivantes: A = 7.40, B = 15.55, C = 29.25 et D = 47.77

4- Représentez les graphiques de la variable type_douleur

Importation de la librairie

(a) Diagramme en batons

Permet de représenter la distribution d'une variable statistique qualitative.

```
In [18]: E2.plot.bar(figsize=(4,3))
```

```
Out[18]: <AxesSubplot: >
```


(b) Diagramme circulaire

```
In [19]: E1.plot.pie(subplots=True, figsize=(4, 4))
```

```
array(
```

```
Out[19]: <Figure>
```


5- Représentez le diagramme en boîte de la variable age

```
In [20]: df.boxplot(column='age', figsize=(6,3), vert=False)
```

```
Out[20]: <AxesSubplot:
```