

Contrôle continu N°2

Données : heart.txt

Résultat attendu : Rendre un fichier de script « contrle2.py »

- que fait le code ci-dessous ?

```
df = pandas.read_table("heart.txt",sep = '\t',header = 0)
```

Elle permet de charger la base de données « heart.txt ».

- Afficher les dimensions du jeu de données : nombre de lignes, nombre de colonnes

```
#Afficher la dimension
```

```
df.shape
```

```
print(df.shape)
```

#2) nombre de ligne

```
df.lines
```

```
# afficher les colonnes
```

```
df.columns
```

```
df.shape
```

```
print(df.shape)
```

- afficher les dernières lignes du jeu de données

```
df.under()
```

```
print(df.under())
```

- énumérer des colonnes

```
df.dtypes
```

- afficher le type de chaque colonnes

```
print(type(df))
```

- afficher les statistiques descriptives des variables quantitatives

```
df['age'][0:]
```

- afficher les statistiques descriptives des variables qualitatives

```
df['typedouleur'][0:]
```

8. accéder à la colonne sexe

```
df.loc['sexe']
```

9. accéder aux colonnes sexe et sucre

```
df[['sexe','sucre']]
```

10. afficher les premières valeurs du sexe

```
df['sexe'].head()
```

11. calculer de manière explicite : le nombre de valeurs, la moyenne, ecart-type, le minimum, le maximum, les quartiles de la variable « typedouleur

```
df['age'].mean()
```

```
df['age'].std()
```

```
df['age'].min()
```

```
df['age'].max()
```

```
taille=df['age'].describe()
```

```
taille['25%']
```

```
taille['50%']
```

```
taille['75%']
```

12. afficher les 12 premiers ages

```
df['age'][1:12]
```

```
df.iloc[1:14,1:1]
```

```
#13-tri
```

```
df['typedouleur'].sort_values()
```

13. Trier le jeux de données par age croissant et type de douleur decroissant

```
df['typedouleur'].sort_values()
```

14. Utilisation de iloc

- a. 5 premières valeurs de toutes les colonnes

```
df.iloc[0:5,:]
```

- b. 5 premières lignes et deux premières colonnes

```
df.iloc[1:5,1:2]
```

- c. 5 premières lignes et colonnes 0, 1 et 4

```
df.iloc[1:5,[0,2,4]]
```

15. Utilisation de loc

- a. liste des individus présentant une douleur de type A

```
df.loc[df['typedouleur']=="A",:]
```

- b. liste des individus présentant une douleur de type A ou B

```
df.loc[df['typedouleur'].isin(['A','B']),:]
```

- c. liste des individus présentant une douleur de type A et angine == oui

```
df.loc[df['typedouleur']=="A"] & (df['angine']=="oui"),:]
```

16. liste des personnes de moins de 45 ans, de sexe masculin, présentant une maladie cardiaque

```
df.loc[(df['age'] < 45) & (df['sexe'] == "masculin") & (df['coeur'] == "presence"),:]
```

17. fréquences selon sexe et cœur

```
pandas.crosstab(df['sexe'],df['cœur'], normalize='all')
```

18. fréquences relative par ligne selon sexe et cœur

```
pandas.crosstab(df['sexe'],df['cœur'], normalize='index')
```