

Atelier Données – Formation SNT – Mardi 7 mai 2019

Titanic « Les femmes et les enfants d'abord ? »

Partie A – Avec le tableur.

A l'aide du fichier CSV proposé, répondre à l'ensemble des questions suivantes :

- 1) Identifier les différents descripteurs de la collection.
Préciser le type et le format de donnée d'un objet.
- 2) Combien de personnes étaient âgées de 10 ans ?
- 3) Quel est le prix du billet le plus élevé ?
- 4) Combien de femmes en 1^{ère} classe voyageaient sur le Titanic ? Quel âge pour la plus jeune ? pour la plus âgée ?
- 5) « Les femmes et les enfants d'abord ? »

En considérant que l'on est enfant lorsqu'on a moins de 18 ans, déterminer le nombre de femmes et d'enfants qui ont survécu.

Que penser alors de l'adage "les femmes et les enfants d'abord" ?

- 6) Quel est le tarif moyen du billet des personnes qui voyageaient en 1^{ère} classe ?
- 7) Quelle est la moyenne d'âge des personnes qui voyageaient en 3^{ème} classe ?
- 8) Indiquer selon le genre des personnes de chaque classe, la fréquence du descripteur « Survivant. »

Partie B – Avec le module PANDAS de Python

Dans Edupython, installer (avec conda), la librairie pandas. Exécuter ensuite les lignes suivantes :

```
import pandas
```

```
data=pandas.read_csv('titanic.csv', sep=';')
```

L'objet data contient ainsi l'ensemble des données.

- 1) Exécuter successivement dans la console, les instructions suivantes et commenter les résultats :
`data.head()` ; `data.shape` ; `data.dtypes`.
- 2) Exécuter la commande `data.describe()`
Quel est le prix moyen du billet ?
- 3) Voici un panel de fonctions de la librairie pandas :
`data['nom du descripteur'].mean()`
`groupe1=data.groupby(['nom du descripteur 1','nom du descripteur 2']).mean()`
`groupe2=data.dropna()`
`groupe3=data.query('sexe==2 or age<18')`

Reprendre l'ensemble des questions de la partie A.