

LES DONNÉES STRUCTURÉES E02

Merci à S.Colomban pour cette activité

Définition n°1. Le format CSV

Les données doivent être structurées, c'est-à-dire organisées, pour permettre leur tri, leur exploitation, leur échange, leur stockage.

Le sigle CSV désigne un format libre qui signifie Comma-Separated Values. Il s'agit d'un fichier informatique de données structurées, dont les valeurs sont généralement séparées par des virgules ou des points-virgules (comma signifie virgule en anglais).

Les tableurs tels qu'Excel (Microsoft) ou Calc (LibreOffice) et des bases de données telles que MySQL et Oracle sont capables d'importer et exporter des fichiers CSV. Les langages tels que php et python sont capables d'importer et de travailler sur les fichiers CSV.

Situation

Le RMS Titanic est un paquebot transatlantique britannique qui a fait naufrage dans l'océan Atlantique Nord en 1912 à la suite d'une collision avec un iceberg, lors de son voyage inaugural de Southampton à New York.

Cet événement l'une des plus grandes catastrophes maritimes survenues en temps de paix.

Le fichier [S2-titanic.csv](#) présente des données sur plus de 1300 passagers du Titanic.

EXERCICE N°1 Un peu de vocabulaire

- 1) Télécharger le fichier [S2-titanic.csv](#).
- 2) Ouvrir ce fichier avec notepad++.
- 3) La première ligne du fichier correspond aux descripteurs, c'est-à-dire aux titres des colonnes de classement :
 - Donner le nombre de descripteurs du fichier.
 - Citer les descripteurs.
 - Quel symbole est le séparateur ?

EXERCICE N°2

- 1) Ouvrir le fichier [S2-titanic.csv](#) avec LibreOffice.
- 2) A l'ouverture cocher uniquement le bon séparateur.
- 3)
 - 3.a) Pour le descripteur "sexe", à quoi correspond la valeur 1?
 - 3.b) Sachant que Andrews, Mr. Thomas Jr n'a pas survécu au naufrage, pour le descripteur "survie", à quoi correspond la valeur 1?
 - 3.c) Choisir le 8ème objet du fichier (c'est-à-dire le 8ème passager) et rédiger une phrase le décrivant.
- 4) En insérant une commande de la forme = **MOYENNE(...)** dans une cellule bien choisie, déterminer le prix moyen d'un billet à bord du Titanic. .
- 5) Créer un AutoFiltre sur la première ligne puis :
 - Déterminer le tarif le plus élevé
 - Déterminer le nombre de femmes présentes à bord
 - Trier les données afin de compléter le tableau suivant :

Tarif (£)	[0;50[[50;100[[100;150[[150;200[[200;250[[250;300[
Effectif	161	33	13	21

Tarif (£)	[300;350[[350;400[[400;450[[450;500[[500;550[
Effectif	0	0	0	0

EXERCICE N°3

- 1) Lancer le logiciel Edupython.
- 2) Saisir le programme suivant, le sauvegarder sous le nom csv-python.py puis l'exécuter :

```
1 import csv
2 fichier_csv = open("S2-titanic.csv", "r")
3 csv_passagers= csv.reader(fichier_csv, delimiter=";")
4 numero = 0
5 for ligne in csv_passagers:
6     if numero == 0:
7         print("Descripteurs :")
8         print(ligne)
9     else:
10        print("Données :")
11        print(ligne)
12        numero = numero+1
13 print("total :",numero, "passagers")
14 fichier_csv.close()
```

on importe la bibliothèque (ou module) csv
on ouvre le fichier S2-titanic.csv
le délimiteur est le ";"

on parcourt toutes les lignes du fichier
la ligne numéro 0 : les descripteurs

les autres lignes : des données

on a compté tous les passagers
on ferme le fichier S2-titanic.csv

S. COLOMBAN – Lycée Ella Fitzgerald

- 3) Il peut-être intéressant d'avoir des informations sur un seul passager.
- 3.a) Modifier le programme précédent comme ci-dessous, le sauver sous le nom :
csv-passager.py

puis l'exécuter :

```
1 import csv
2 fichier_csv = open("S2-titanic.csv", "r")
3 csv_passagers= csv.reader(fichier_csv, delimiter=";")
4 numero = 0
5 for ligne in csv_passagers:
6     if numero == 800:
7         print("passager numéro :", numero)
8         print(ligne)
9         break #inutile de continuer le for
10        numero=numero+1
11 fichier_csv.close()
```

S. COLOMBAN – Lycée Ella Fitzgerald

- 3.b) Modifier le programme précédent afin d'obtenir des informations sur le 1197^e passager.

Chaque ligne de données est de la forme : ['3', '1', 'Sheerlinck, Mr. Jan Baptist', '1', '29', '10']
Cette ligne peut être analysée ainsi :

Rang dans la ligne	0	1	2	3	4	5
Descripteur	classe	survie	nom	sexe	age	tarif
Type de variable	int	int	chaîne de caractères	int	int	float
Valeur	3	1	Sheerlinck, Mr. Jan Baptist	1	29	10

- Pour obtenir le **tarif** de cette ligne, il suffit d'écrire `float(ligne[5])`
- Pour obtenir le **nom** de cette ligne, il suffit d'écrire `ligne[2]`
- Pour obtenir l'**age** de cette ligne, il suffit d'écrire ...
- Pour obtenir la **classe** de cette ligne, il suffit d'écrire ...

S. COLOMBAN – Lycée Ella Fitzgerald

- 4) Le programme suivant compte nombre de passagers de 3^eme classe.

```
1 import csv
2 fichier_csv = open("S2-titanic.csv", "r")
3 csv_passagers= csv.reader(fichier_csv, delimiter=";")
4 numero = 0
5 compteur = 0
6 for ligne in csv_passagers:
7     if numero >= 1:
8         if (int(ligne[0]) == 3) :
9             compteur = compteur +1
10        numero=numero+1
11 fichier_csv.close()
12 print("Nombre de passagers de 3ème classe : ", compteur)
```

S. COLOMBAN – Lycée Ella Fitzgerald

- 4.a) Le modifier afin qu'il compte le nombre de femmes.
- 4.b) Le modifier à nouveau afin qu'il calcule les ages du plus jeune passager et du plus vieux passager.
- 4.c) Le modifier à nouveau afin qu'il calcule le tarif moyen d'un billet.