

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

Traitement de données

Le Titanic

Christophe Viroulaud

Seconde - SNT

Donn 01

1. Histoire

2. Découvrir la structure des données

3. Exploitation des données

4. Filtrer les données

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

Histoire

Le RMS Titanic est un paquebot transatlantique britannique qui 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.

La coque du Titanic était pourvue de seize compartiments étanches servant à protéger le navire en cas de voies d'eau ou d'avaries importantes, ce qui lui donna la réputation de paquebot « insubmersible » et conduisit les médias contemporains à le présenter comme l'un des navires les plus sûrs.

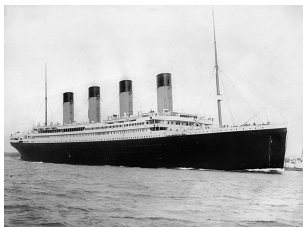


FIGURE 1 – Le Titanic

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

Le 14 avril 1912, quatre jours après le commencement de son voyage inaugural, il heurte un iceberg à 23 h 40 (heure locale) et coule le 15 avril 1912 à 2 h 20 au large de Terre-Neuve.

L'épave du Titanic est localisée le 1er septembre 1985 par le professeur Robert Ballard. Elle gît à 3 843 mètres de profondeur à 650 km au sud-est de Terre-Neuve¹.

Les données étudiées par la suite, recensent les informations des passagers du navire.

1. Source : <https://fr.wikipedia.org/wiki/Titanic>

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

1. Histoire
2. Découvrir la structure des données
3. Exploitation des données
4. Filtrer les données

Les données sont contenues dans un fichier *csv* (*Comma Separated Values*).

Activité 1 :

1. Sur le site <https://cviroulaud.github.io>, télécharger le dossier compressé *titanic.zip* et en extraire le fichier **titanic.csv**
2. Ouvrir le fichier avec **le bloc-notes de Windows**. Observer le symbole utilisé pour séparer les données.

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

Activité 1 :

3. Ouvrir le fichier à l'aide du logiciel **LibreOffice Calc**. Le logiciel repère les symboles et classe les données dans des colonnes.
4. La première ligne présente les **descripteurs**. Que représentent ces termes ?
5. Pour le descripteur *sexe*, à quoi correspond la valeur *1* ?
6. Pour le descripteur *tarif*, à quoi correspond la valeur indiquée pour chaque objet ?

À retenir

Un fichier **csv (Comma Separated Values)** rassemble des informations sur un thème. Les données sont classées selon des **descripteurs**. Chaque ligne du fichier représente un **enregistrement**.

- ▶ Un descripteur représente une caractéristique des données contenues dans le fichier.
- ▶ Le 1 représente les hommes et le 2 les femmes.
- ▶ Le prix du billet payé par chaque passager est donné dans la colonne *tarif*. Il est donné en Livre Sterling. On peut remarquer que certains passagers n'ont pas payé.

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

1. Histoire
2. Découvrir la structure des données
3. Exploitation des données
4. Filtrer les données

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

Le tableur LibreOffice propose des fonctions de calculs automatiques. Une méthode possible consiste à :

1. se placer dans une case vide,
2. entrer le signe = puis la fonction désirée. Par exemple, **MOYENNE** (ou **AVERAGE** en anglais),
3. entrer la plage de calcul ; exemple : **A2:A10** pour prendre les cellules A2 à A10.

Activité 2 : À l'aide des fonctions usuelles du tableur, déterminer :

1. l'âge moyen des passagers,
2. le tarif moyen payé,
3. le tarif le plus élevé qui a été payé.

Avant de regarder la correction



- ▶ Prendre le temps de réfléchir,
- ▶ Analyser les messages d'erreur,
- ▶ Demander au professeur.

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

```
1 = MOYENNE(E2 : E1310)
```

Code 1 – Âge moyen

L'âge moyen est 29,90 ans.

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

```
1 = MOYENNE(F2 : F1310)
```

Code 2 – Tarif moyen

Le tarif moyen est 33,36£.

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

1 = MAX(F2 : F1310)

Code 3 – Tarif maximum

Le tarif maximum est 512£.

1. Histoire
2. Découvrir la structure des données
3. Exploitation des données
4. Filtrer les données

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

L'outil **Données/Autofiltre** permet de réaliser des filtres sur les valeurs de chaque descripteur.

Activité 3 :

1. Activer l'outil **Données/Autofiltre**,
2. Trier les tarifs par ordre croissant. Construire et compléter le tableau suivant :

Tarif payé	[0; 50[[50; 100[...	[500; 550[
Effectif				

3. Dans le tableur, créer une nouvelle feuille, recopier le tableau et construire une représentation graphique correspondante aux données.

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

Avant de regarder la correction



- ▶ Prendre le temps de réfléchir,
- ▶ Analyser les messages d'erreur,
- ▶ Demander au professeur.

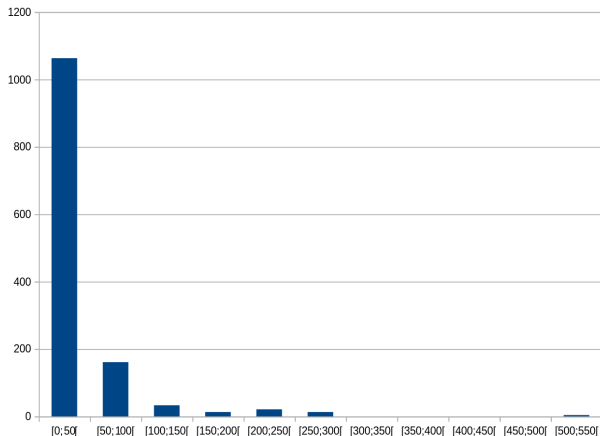
- ▶ $[0; 50[\rightarrow 1063$
- ▶ $[50; 100[\rightarrow 161$
- ▶ $[100; 150[\rightarrow 33$
- ▶ $[150; 200[\rightarrow 13$
- ▶ $[200; 250[\rightarrow 21$
- ▶ $[250; 300[\rightarrow 13$
- ▶ $[300; 350[\rightarrow 0$
- ▶ $[350; 400[\rightarrow 0$
- ▶ $[400; 450[\rightarrow 0$
- ▶ $[450; 500[\rightarrow 0$
- ▶ $[500; 550[\rightarrow 4$

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données



À retenir

L'icône **Insérer un diagramme** permet de construire des représentations graphiques dans LibreOffice.

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

Activité 4 :

1. Déterminer la moyenne des données du descripteur *survie*. Quelle interprétation peut-on en donner ?

Avant de regarder la correction



- ▶ Prendre le temps de réfléchir,
- ▶ Analyser les messages d'erreur,
- ▶ Demander au professeur.

```
1 = MOYENNE(B2 : B1310)
```

Code 4 – Survie

La moyenne de la colonne *survie* est 0,38. Cela signifie que moins de la moitié des passagers ont survécu.

Il n'y avait pas suffisamment de places dans les canots de sauvetage du Titanic pour tous les passagers et les membres de l'équipage (et certains canots sont partis à peine remplis). On souhaite examiner l'influence de la classe sociale des passagers sur l'obtention d'une place sur un canot de sauvetage.

Pour calculer la fréquence de survie d'une classe, il faut, pour une classe donnée, calculer le rapport entre le nombre de survivants et le nombre total de passagers de cette classe. Pour calculer le nombre de survivants d'une classe, il peut être intéressant d'utiliser la fonction **SOMME**.

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

Activité 5 :

1. Trier les données du descripteur *classe* par ordre croissant à l'aide de l'autofiltre.
2. Calculer la fréquence de survivants pour chaque classe.

Avant de regarder la correction



- ▶ Prendre le temps de réfléchir,
- ▶ Analyser les messages d'erreur,
- ▶ Demander au professeur.

Histoire

Découvrir la
structure des
données

Exploitation des
données

Filtrer les données

1 = SOMME(B2:B324)/323

Code 5 – Calcul pour la classe 1

- ▶ classe 1 : 0,62
- ▶ classe 2 : 0,43
- ▶ classe 3 : 0,26