

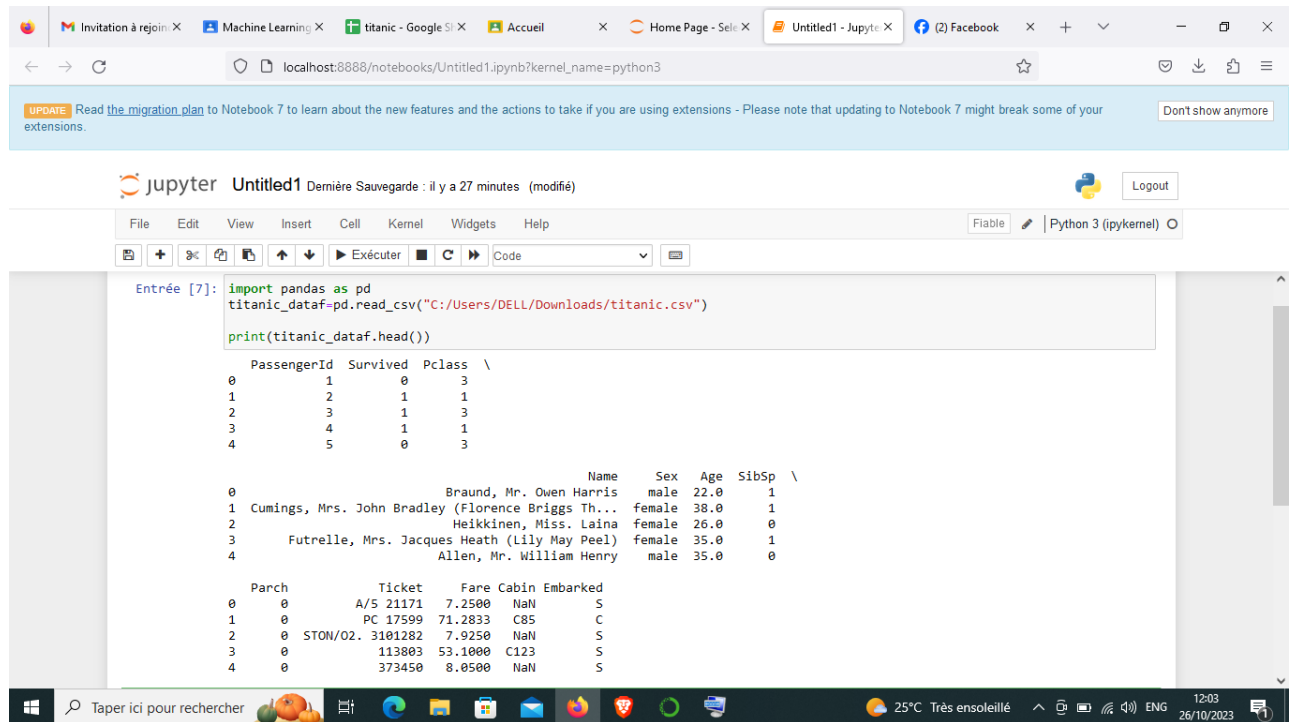
COMPTE RENDU DU TP2

Realisé par Barry Alpha Mamadou Douah

Année Universitaire 2023-2024

Manipulation de données stockés au format csv avec pandas

1. Charger le jeu de données Titanic sous forme d'un DataFrame et l'afficher.
2. Afficher les statistiques descriptives de ce jeu de données.

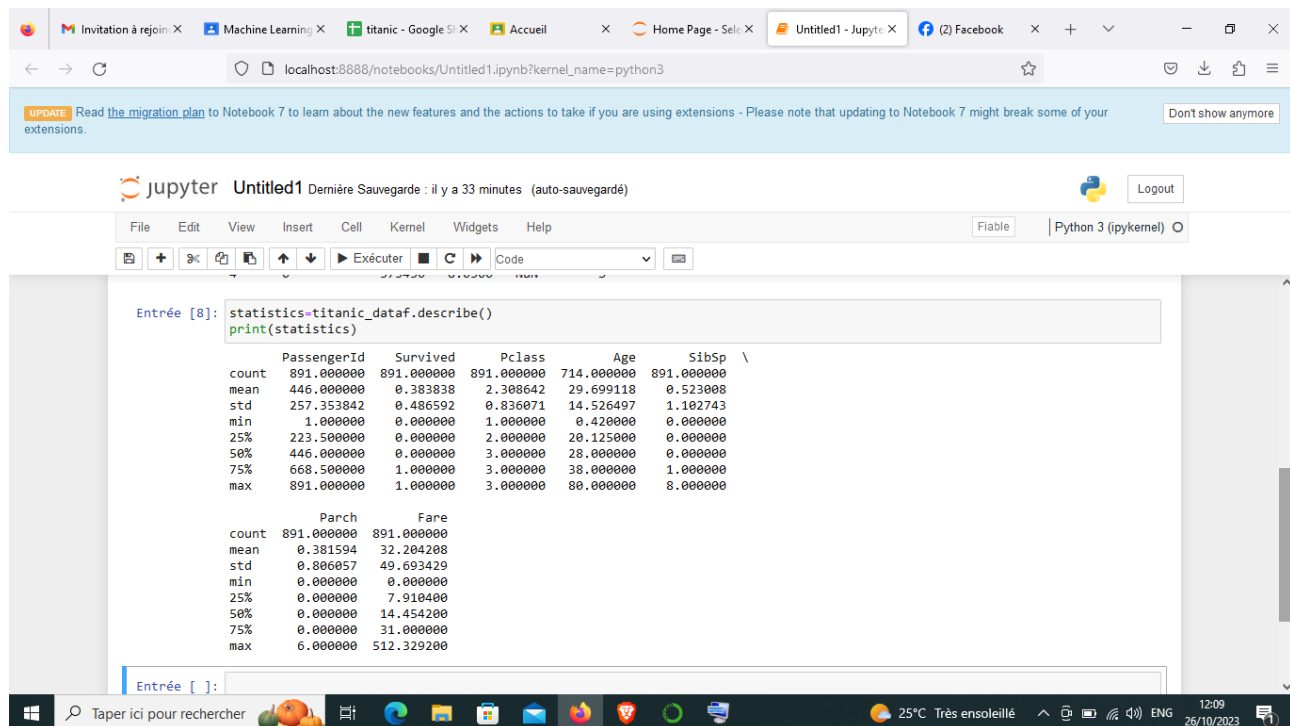


The screenshot shows a Jupyter Notebook interface with the following code in the first cell:

```
Entrée [7]: import pandas as pd
titanic_dataaf=pd.read_csv("C:/Users/DELL/Downloads/titanic.csv")
print(titanic_dataaf.head())
```

The output displays the first five rows of the dataset, showing columns: PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, and Embarked.

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S



The screenshot shows a Jupyter Notebook interface with the following code in the second cell:

```
Entrée [8]: statistics=titanic_dataaf.describe()
print(statistics)
```

The output displays the descriptive statistics for the dataset, showing columns: PassengerId, Survived, Pclass, Age, SibSp, Parch, Fare, and Embarked.

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Embarked
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.523008	32.204208	0.523008
std	257.353842	0.486592	0.836071	14.526497	1.102743	1.102743	49.693429	1.102743
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400	0.000000
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200	0.000000
75%	668.500000	1.000000	3.000000	38.000000	1.000000	1.000000	31.000000	1.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	8.000000	512.329200	8.000000

3. Afficher le nombre de passagers ainsi que le nombre d'attributs du jeu de données Titanic.
4. Afficher les étiquettes des attributs.

The screenshot shows a Jupyter Notebook interface with a browser window at the top displaying tabs for 'Invitation à rejoindre', 'Machine Learning', 'titanic - Google Slides', 'Accueil', 'Home Page - Selection', 'Untitled1 - Jupyter', and '(2) Facebook'. The address bar shows 'localhost:8888/notebooks/Untitled1.ipynb?kernel_name=python3'. A blue banner at the top contains an update message about migrating to Notebook 7. The notebook title is 'Untitled1' with a save status 'Dernière Sauvegarde : il y a 43 minutes (modifié)'. The menu bar includes 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Widgets', and 'Help'. The toolbar shows icons for file operations and execution. The main area displays a table of statistics for the Titanic dataset:

	mean	std	min	25%	50%	75%	max
	0.381594	0.806057	0.000000	0.000000	0.000000	0.000000	0.000000
	32.204208	49.693429	0.000000	7.910400	14.454200	31.000000	512.329200

Below the table, a code cell (Entrée [9]) contains the following Python code:

```
passenger_count=len(titanic_data)
attribute_count=len(titanic_data.columns)
print("nombre de passagers:",passenger_count)
print("nombre d'attributs:",attribute_count)
```

The output of the code cell shows:

```
nombre de passagers: 891
nombre d'attributs: 12
```

The bottom of the image shows a Windows taskbar with a search bar and various application icons.

The screenshot shows a Jupyter Notebook interface with a browser window at the top displaying tabs for 'Invitation à rejoindre', 'Machine Learning', 'titanic - Google Slides', 'Accueil', 'Home Page - Selection', 'Untitled1 - Jupyter', and '(2) Facebook'. The address bar shows 'localhost:8888/notebooks/Untitled1.ipynb?kernel_name=python3'. A blue banner at the top contains an update message about migrating to Notebook 7. The notebook title is 'Untitled1' with a save status 'Dernière Sauvegarde : il y a une heure (modifié)'. The menu bar includes 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Widgets', and 'Help'. The toolbar shows icons for file operations and execution. The main area displays a table of statistics for the Titanic dataset:

	25%	50%	75%	max
	0.000000	0.000000	0.000000	0.000000
	7.910400	14.454200	31.000000	512.329200

Below the table, a code cell (Entrée [9]) contains the following Python code:

```
passenger_count=len(titanic_data)
attribute_count=len(titanic_data.columns)
print("nombre de passagers:",passenger_count)
print("nombre d'attributs:",attribute_count)
```

The output of the code cell shows:

```
nombre de passagers: 891
nombre d'attributs: 12
```

Below this, another code cell (Entrée [10]) contains the following Python code:

```
titanic_data.columns
```

The output of the code cell shows:

```
Out[10]: Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
         'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
         dtype='object')
```

The bottom of the image shows a Windows taskbar with a search bar and various application icons.

5. Afficher toutes les informations relatives aux 10 premiers passagers.
6. Afficher les identifiants ainsi que les noms des différents passagers.

Invitation à rejoindre Machine Learning titanic - Google Sheets Accueil Home Page - Selection Untitled1 - Jupyter (2) Facebook

localhost:8888/notebooks/Untitled1.ipynb?kernel_name=python3

UPDATE Read the [migration plan](#) to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions. Don't show anymore

jupyter Untitled1 Dernière Sauvegarde : il y a une heure (auto-sauvegardé) Logout

File Edit View Insert Cell Kernel Widgets Help

Entrée [11]: titanic_dataf.head(10)

Out[11]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C

Entrée []:

Invitation à rejoindre Machine Learning titanic - Google Sheets Accueil Home Page - Selection Untitled1 - Jupyter (2) Facebook

localhost:8888/notebooks/Untitled1.ipynb?kernel_name=python3

UPDATE Read the [migration plan](#) to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions. Don't show anymore

jupyter Untitled1 Dernière Sauvegarde : il y a une heure (auto-sauvegardé) Logout

File Edit View Insert Cell Kernel Widgets Help

Notebook enregistré

Entrée [17]:

```
passenger_id=titanic_dataf[['PassengerId','Name']]
print(passenger_id)
```

PassengerId	Name
0	1 Braund, Mr. Owen Harris
1	2 Cumings, Mrs. John Bradley (Florence Briggs Th...
2	3 Heikinen, Miss. Laina
3	4 Futrelle, Mrs. Jacques Heath (Lily May Peel)
4	5 Allen, Mr. William Henry
...	...
886	887 Montvila, Rev. Juozas
887	888 Graham, Miss. Margaret Edith
888	889 Johnston, Miss. Catherine Helen "Carrie"
889	890 Behr, Mr. Karl Howell
890	891 Dooley, Mr. Patrick

[891 rows x 2 columns]

Entrée []:

7. Afficher les différentes valeurs de la colonne Pclass.
8. Calculer et afficher le nombre de passagers de sexe féminin et le nombre de passagers de sexe masculin.
9. Calculer et afficher le nombre de passagers par classe.
10. Calculer et afficher le nombre de personnes qui ont survécu ainsi que le pourcentage.
11. Calculer et afficher le nombre de femmes ayant survécu au naufrage, ainsi que le pourcentage (ensuite les hommes).
12. Calculer et afficher l'âge moyen des hommes.

The screenshot shows a Jupyter Notebook interface with the following code and output:

```
Entrée [24]: femmes_survivantes = titanic_data[titanic_data['Sex'] == 'female' & (titanic_data['Survived'] == 1)]
nombre_femmes_survivantes = len(femmes_survivantes)
pourcentage_femmes_survivantes = (nombre_femmes_survivantes / len(titanic_data[titanic_data['Sex'] == 'female'])) * 100
print("Nombre de femmes ayant survécu au naufrage :", nombre_femmes_survivantes)
print("Pourcentage de survie des femmes :", pourcentage_femmes_survivantes)

Nombre de femmes ayant survécu au naufrage : 233
Pourcentage de survie des femmes : 74.20382165605095

Entrée [25]: hommes_survivantes = titanic_data[titanic_data['Sex'] == 'male' & (titanic_data['Survived'] == 1)]
nombre_hommes_survivantes = len(hommes_survivantes)
pourcentage_hommes_survivantes = (nombre_hommes_survivantes / len(titanic_data[titanic_data['Sex'] == 'male'])) * 100
print("Nombre d'hommes ayant survécu au naufrage :", nombre_hommes_survivantes)
print("Pourcentage de survie des hommes :", pourcentage_hommes_survivantes)

Nombre d'hommes ayant survécu au naufrage : 109
Pourcentage de survie des hommes : 18.890814558058924

Entrée [26]: age_moyen_hommes = titanic_data[titanic_data['Sex'] == 'male']['Age'].mean()
print("Âge moyen des hommes :", age_moyen_hommes)

Âge moyen des hommes : 30.72664459161148
```

The screenshot shows a Jupyter Notebook interface with the following code and output:

```
Entrée [18]: titanic_data['Pclass'].unique()

Out[18]: array([3, 1, 2], dtype=int64)

Entrée [21]: nombre_femmes = (titanic_data['Sex'] == 'female').sum()
nombre_hommes = (titanic_data['Sex'] == 'male').sum()
print("Nombre de passagers de sexe féminin :", nombre_femmes)
print("Nombre de passagers de sexe masculin :", nombre_hommes)

Nombre de passagers de sexe féminin : 314
Nombre de passagers de sexe masculin : 577

Entrée [22]: # 9. Calculer et afficher le nombre de passagers par classe
passagers_par_classe = titanic_data.groupby('Pclass')['Pclass'].count()
print("Nombre de passagers par classe :")
print(passagers_par_classe)

Nombre de passagers par classe :
Pclass
1    216
2    184
3    491
Name: Pclass, dtype: int64
```

13. Donner le code qui permet de créer une liste L contenant n chaînes ("bébé", "enfant", "adulte", "vieux") selon l'âge du passager sachant que n est le nombre de passager de notre jeu de données.

— Si $\text{age} \leq 3$: "bébé"

— Si $\text{age} > 3$ et $\text{age} \geq 15$: "enfant"

— Si $\text{age} > 15$ et $\text{age} \leq 60$: "adulte"

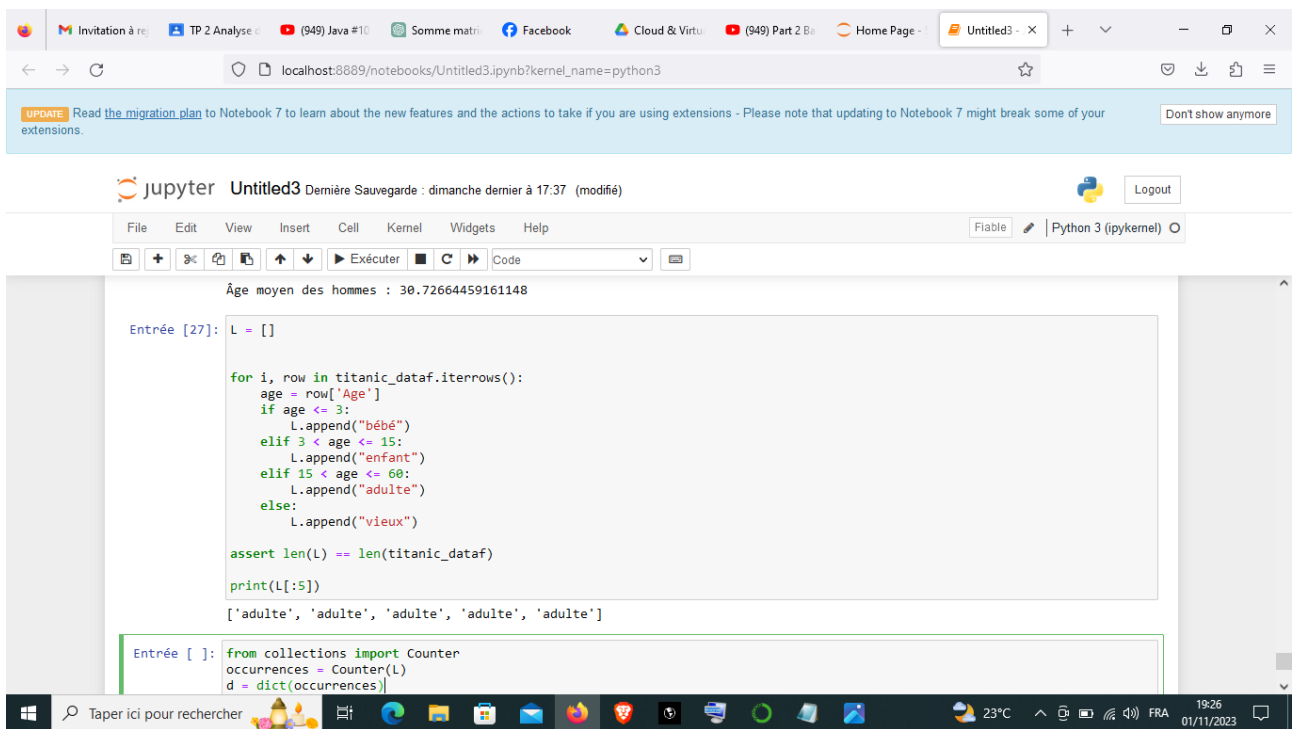
— Sinon "Vieux".

14. Créer à partir de la liste L, un dictionnaire d contenant comme clefs les chaînes "bébé", "enfant",

"adulte" et "vieux". Les valeurs sont le nombre d'occurrence de la chaîne dans la liste L.

15. Afficher le nombre de valeurs manquantes par attribut.

16. Calculer le pourcentage de valeurs manquantes sur l'attribut Age.



The screenshot shows a Jupyter Notebook with two code cells. The first cell contains a loop that iterates over the 'titanic_data' dataset, categorizing passengers into 'bébé', 'enfant', 'adulte', or 'vieux' based on their age. The second cell contains code to count the occurrences of each category using a Counter and to print the result as a dictionary.

```
Entrée [27]: L = []

for i, row in titanic_data.iterrows():
    age = row['Age']
    if age <= 3:
        L.append("bébé")
    elif 3 < age <= 15:
        L.append("enfant")
    elif 15 < age <= 60:
        L.append("adulte")
    else:
        L.append("vieux")

assert len(L) == len(titanic_data)
print(L[:5])

['adulte', 'adulte', 'adulte', 'adulte', 'adulte']
```

```
Entrée [ ]: from collections import Counter
occurrences = Counter(L)
d = dict(occurrences)
```

Invitation à re... TP 2 Analyse... (949) Java #10... Somme mat... Facebook... Cloud & Virt... (949) Part 2 B... Home Page... Untitled3 - X

localhost:8889/notebooks/Untitled3.ipynb?kernel_name=python3

UPDATE Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions. Don't show anymore

jupyter Untitled3 Dernière Sauvegarde : dimanche dernier à 17:37 (modifié) Logout

File Edit View Insert Cell Kernel Widgets Help

Exécuter Code

```
Entrée [28]: from collections import Counter
occurrences = Counter(L)
d = dict(occurrences)
print("Dictionnaire d :")
print(d)

Dictionnaire d :
{'adulte': 609, 'vieux': 199, 'bébé': 30, 'enfant': 53}

Entrée [29]: nombre_valeurs_manquantes_par_attribut = titanic_dataf.isnull().sum()
print("Nombre de valeurs manquantes par attribut :")
print(nombre_valeurs_manquantes_par_attribut)

Nombre de valeurs manquantes par attribut :
PassengerId      0
Survived          0
Pclass           0
Name             0
Sex              0
Age             177
SibSp            0
Parch           0
Ticket           0
Fare             0
Cabin           687
Embarked         2
dtype: int64
```

Taper ici pour rechercher

Maxi... 19:27 01/11/2023

Invitation à re... TP 2 Analyse... (949) Java #10... Somme mat... Facebook... Cloud & Virt... (949) Part 2 B... Home Page... Untitled3 - X

localhost:8889/notebooks/Untitled3.ipynb?kernel_name=python3

UPDATE Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions. Don't show anymore

jupyter Untitled3 Dernière Sauvegarde : dimanche dernier à 17:37 (modifié) Logout

File Edit View Insert Cell Kernel Widgets Help

Exécuter Code

```
Survived          0
Pclass           0
Name             0
Sex              0
Age             177
SibSp            0
Parch           0
Ticket           0
Fare             0
Cabin           687
Embarked         2
dtype: int64

Entrée [30]: pourcentage_valeurs_manquantes_age = (titanic_dataf['Age'].isnull().sum() / len(titanic_dataf)) * 100
print("Pourcentage de valeurs manquantes sur l'attribut Age :", pourcentage_valeurs_manquantes_age)

Pourcentage de valeurs manquantes sur l'attribut Age : 19.865319865319865

Entrée [ ]:
```

Taper ici pour rechercher

23°C 19:28 01/11/2023