

```
import pandas as pd
df=pd.read_excel("titanic-passengers.xlsx")
```

```
print(df.isnull())
print(df.isnull().sum())
```

↗

	PassengerId	Survived	Pclass	Name	...	Ticket	Fare	Cabin	Embarked
0	False	False	False	False	...	False	False	True	False
1	False	False	False	False	...	False	False	False	False
2	False	False	False	False	...	False	False	True	False
3	False	False	False	False	...	False	False	True	False
4	False	False	False	False	...	False	False	False	False
..
886	False	False	False	False	...	False	False	True	False
887	False	False	False	False	...	False	False	True	False
888	False	False	False	False	...	False	False	True	False
889	False	False	False	False	...	False	False	True	False
890	False	False	False	False	...	False	False	True	False

[891 rows x 12 columns]
PassengerId 0
Survived 0
Pclass 0
Name 0

Enregistrement... X

Parch 0
Ticket 0
Fare 0
Cabin 687
Embarked 2
dtype: int64

```
df1=df.copy()
df.dropna(axis=0,how='any',thresh=None,inplace=False)
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
1	76	No	3	Moen, Mr. Sigurd ..	male	25.0	0	0	348123	7

```
df.dropna(axis=1,how='any')
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticke
0	343	No	2	Collander, Mr. Erik Gustaf	male	28.000000	0	0	24874
1	76	No	3	Moen, Mr. Sigurd Hansen	male	25.000000	0	0	34812
2	641	No	3	Jensen, Mr. Hans Peder	male	20.000000	0	0	35005
3	568	No	3	Palsson, Mrs. Nils (Alma Cornelia Berglund)	female	29.000000	0	4	34990
			1	Davidson, Mr. Thornton	male	31.000000	1	0	F.C 1275

Enregistrement...✕

```
df.dropna(axis=1,how='all')
```

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticke
-------------	----------	--------	------	-----	-----	-------	-------	-------

```
df1['Age'].fillna(df1['Age'].mean(),inplace=False)
```

0	28.000000
1	25.000000
2	20.000000
3	29.000000
4	31.000000
...	
886	14.000000
887	22.000000
888	30.000000
889	29.699118
890	19.000000

Name: Age, Length: 891, dtype: float64

Completed

```
df3=df1[df['Sex']=='male']
df3.head(20)
```

Enregistrement...

×

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
1	0	3	Mr. Braund	male	22	0	0	3101274

```
df2=df[df['Age']>=20]
df2.head(20)
df2.value_counts()
```

PassengerId	Survived	Pclass	Name
891	0	3	Dooley, Mr. Patrick
309	0	2	Abelson, Mr. Samuel
316	1	3	Nilsson, Miss. Helmina Josefina
315	0	2	Hart, Mr. Benjamin
314	0	3	Hendekovic, Mr. Ignjac
609	1	2	Laroche, Mrs. Joseph (Juliette Marie Louise Lafargue)
610	1	1	Shutes, Miss. Elizabeth W
611	0	3	Andersson, Mrs. Anders Johan (Alfrida Konstantia Brogr
615	0	3	Brocklebank, Mr. William Alfred
1	0	3	Braund, Mr. Owen Harris

Length: 548, dtype: int64

Lovell, Mr.

3101274

```
df1.isnull().sum()
```

PassengerId	0
Name	0
Sex	0
Age	0
SibSp	0
Parch	0
Ticket	0
Fare	0
Cabin	687
Embarked	2

dtype: int64

Frland

3101274

```
df['Cabin'].value_counts()
```

C23 C25 C27	4
G6	4
B96 B98	4
E101	3
F33	3
..	
C86	1
C50	1
C111	1
A34	1
A31	1

Name: Cabin, Length: 147, dtype: int64

```
df['Cabin'].fillna('G6',inplace=True)
df['Cabin'].value_counts()
```

```

G6          691
C23 C25 C27    4
B96 B98        4
E101         3
F33          3
...
C86          1
C50          1
C111         1
A34          1
A31          1
Name: Cabin, Length: 147, dtype: int64

```

```

#methode 1:
from sklearn.preprocessing import LabelEncoder
df['Survived'].value_counts()
encoder=LabelEncoder()
df['Survived']=encoder.fit_transform(df['Survived'])
df['Survived'].value_counts()

```

```

0    549
1    342
Name: Survived, dtype: int64

```

Enregistrement...



{"No":100,"Yes":200})

```

100    549
200    342
Name: Survived, dtype: int64

```

```

onehot=pd.get_dummies(df['Sex'])
onehot.head()

```

	female	male
0	0	1
1	0	1
2	0	1
3	1	0
4	0	1

```

new_data=df3.drop('Sex',axis=1)
new_data.head()

```

	PassengerId	Survived	Pclass	Name	Age	SibSp	Parch	Ticket	Fare	Ca
0	343	100	2	Collander, Mr. Erik Gustaf	28.0	0	0	248740	13.0000	I

```
new_data=new_data.join(onehot)
new_data.head()
```

	PassengerId	Survived	Pclass	Name	Age	SibSp	Parch	Ticket	Fare	Ca
0	343	100	2	Collander, Mr. Erik Gustaf	28.0	0	0	248740	13.0000	I
1	76	100	3	Moen, Mr. Sigurd Hansen	25.0	0	0	348123	7.6500	F
2	641	100	3	Jensen, Mr. ...	20.0	0	0	350050	7.8542	I

Enregistrement...

×