```
import pandas as pd

df = pd.read_csv('Titanic-Dataset.csv')
df.head()
```

	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	Heikkinen, 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

Étapes suivantes : Afficher les graphiques recommandés | New interactive sheet

```
dfbis = df[['Pclass','Sex','Age','Fare','Survived']]
dfbis.head()
dfbis.describe()
```

	Pclass	Age	Fare	Survived
count	891.000000	714.000000	891.000000	891.000000
mean	2.308642	29.699118	32.204208	0.383838
std	0.836071	14.526497	49.693429	0.486592
min	1.000000	0.420000	0.000000	0.000000
25%	2.000000	20.125000	7.910400	0.000000
50%	3.000000	28.000000	14.454200	0.000000
75%	3.000000	38.000000	31.000000	1.000000
max	3.000000	80.000000	512.329200	1.000000

```
target = df['Survived']
inputs = dfbis.drop('Survived', axis=1)
```

```
dumnies = pd.get_dummies(inputs.Sex,dtype=int)
dumnies.head()
inputs = pd.concat([inputs,dumnies],axis=1).drop('Sex', axis=1)
inputs.head()
```

	Pclass	Age	Fare	female	male
0	3	22.0	7.2500	0	1
1	1	38.0	71.2833	1	0
2	3	26.0	7.9250	1	0
3	1	35.0	53.1000	1	0
4	3	35.0	8.0500	0	1

Étapes suivantes : Afficher les graphiques recommandés | New interactive sheet

```
inputs.isna().sum()
```

	0
Pclass	0
Age	177
Fare	0
female	0
male	0

dtype: int64

```
inputs.Age = inputs.Age.fillna(inputs.Age.mean())
```

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(inputs, target, test_size=0.3, train_size=0.7)
```

```
from sklearn.naive_bayes import GaussianNB
model = GaussianNB()

model.fit(X_train, y_train)
```

```
GaussianNB()
```

```
model.score(X_test,y_test)
```

```
0.7835820895522388
```

+ Code + Texte

```
print(X_test[:10])
model.predict(X_test[:10])
```

```
      Pclass        Age      Fare  female  male
823        3  27.000000   12.4750       1     0
428        3  29.699118    7.7500       0     1
291        1  19.000000   91.0792       1     0
689        1  15.000000  211.3375       1     0
564        3  29.699118    8.0500       1     0
323        2  22.000000   29.0000       1     0
180        3  29.699118   69.5500       1     0
756        3  28.000000    7.7958       0     1
293        3  24.000000    8.8500       1     0
612        3  29.699118   15.5000       1     0
array([1, 0, 1, 1, 1, 1, 1, 0, 1, 1])
```

```
import pickle

with open('titanic_nbayes_model.pkl', 'wb') as f:
    pickle.dump(model, f)
```