

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

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

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 211
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briags	female	38.0	1	0	PC 175

Next steps:

[Generate code with df](#)[New interactive sheet](#)

```
df.info()
df.isnull().sum()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   PassengerId     891 non-null    int64
1   Survived        891 non-null    int64
2   Pclass          891 non-null    int64
3   Name            891 non-null    object
4   Sex             891 non-null    object
5   Age            714 non-null    float64
6   SibSp           891 non-null    int64
7   Parch           891 non-null    int64
8   Ticket          891 non-null    object
9   Fare           891 non-null    float64
10  Cabin           204 non-null    object
11  Embarked        889 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB

```

	0
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

```

df['Age'].fillna(df['Age'].mean(), inplace=True)
df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)
df.drop(columns=['Cabin'], inplace=True)
df.isnull().sum()

```

```
/tmp/ipykernel_251/3262979027.py:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series consisting of zero rows.
The behavior will change in pandas 3.0. This inplace method will never work b
```

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.

```
df['Age'].fillna(df['Age'].mean(), inplace=True)
```

```
/tmp/ipykernel_251/3262979027.py:2: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series consisting of zero rows.
The behavior will change in pandas 3.0. This inplace method will never work b
```

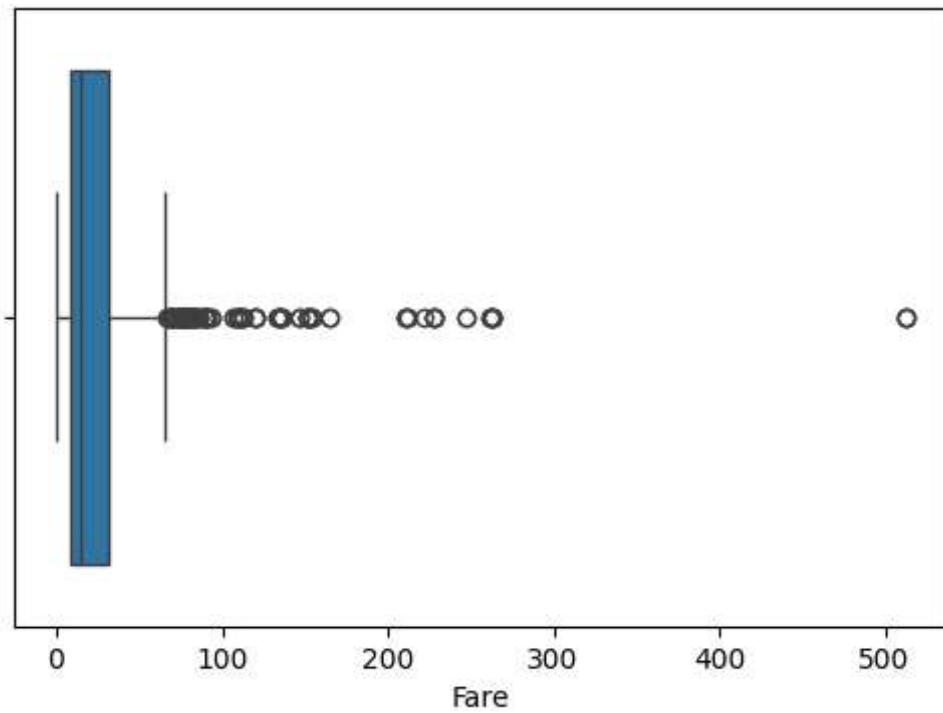
For example, when doing 'df[col].method(value, inplace=True)', try using 'df.

```
df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)
```

	0
PassengerId	0
Survived	0
Pclass	0
Name	0
Sex	0
Age	0
SibSp	0
Parch	0
Ticket	0
Fare	0
Embarked	0

dtype: int64

```
plt.figure(figsize=(6,4))
sns.boxplot(x=df['Fare'])
plt.show()
```



```
Q1 = df['Fare'].quantile(0.25)
Q3 = df['Fare'].quantile(0.75)

IQR = Q3 - Q1

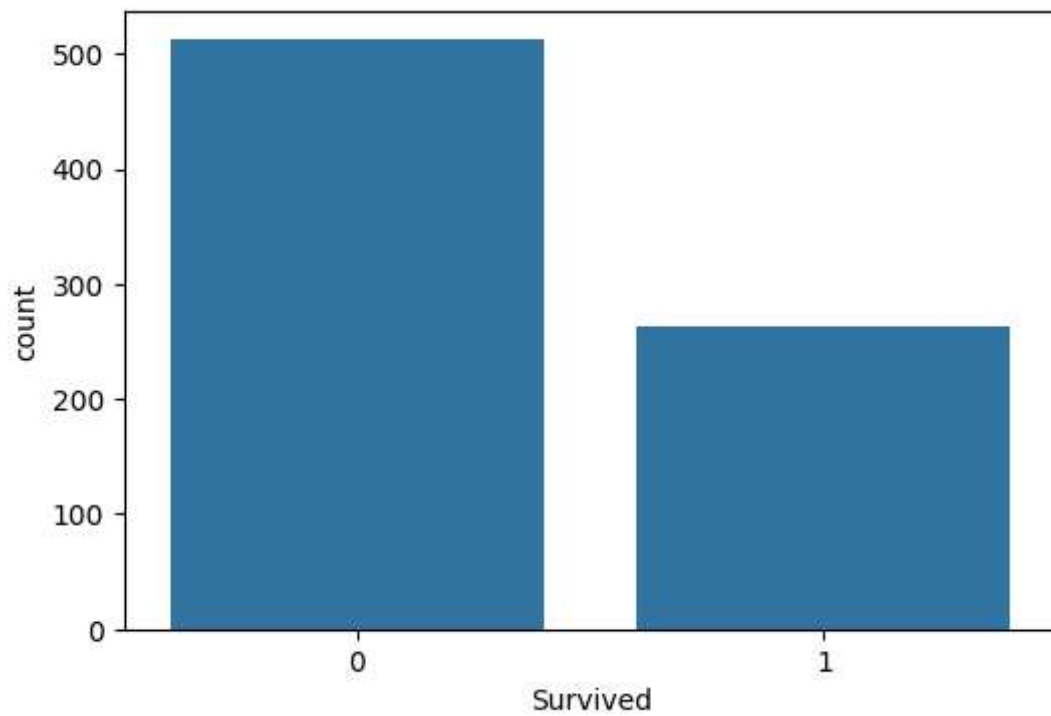
lower = Q1 - 1.5 * IQR
upper = Q3 + 1.5 * IQR

df = df[(df['Fare'] >= lower) & (df['Fare'] <= upper)]

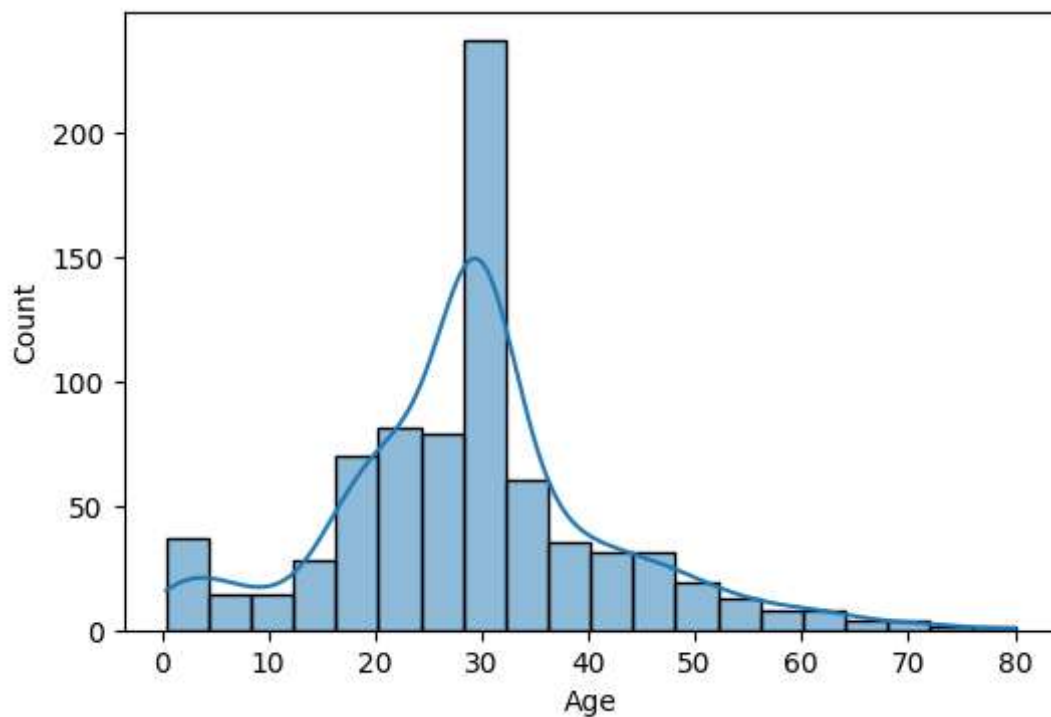
df.shape

(775, 11)
```

```
plt.figure(figsize=(6,4))
sns.countplot(x='Survived', data=df)
plt.show()
```



```
plt.figure(figsize=(6,4))  
sns.histplot(df['Age'], bins=20, kde=True)  
plt.show()
```



```
plt.figure(figsize=(6,4))  
sns.countplot(x='Sex', hue='Survived', data=df)  
plt.show()
```

