

## ▼ Default title text

```
# @title Default title text
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
titanic = sns.load_dataset('titanic')
```

```
titanic = sns.load_dataset('titanic')
print(titanic.head())
```

```
survived  pclass    sex  age  sibsp  parch    fare embarked  class \
0         0        3   male  22.0    1     0   7.2500         S   Third
1         1        1  female  38.0    1     0  71.2833         C   First
2         1        3  female  26.0    0     0   7.9250         S   Third
3         1        1  female  35.0    1     0  53.1000         S   First
4         0        3   male  35.0    0     0   8.0500         S   Third

who  adult_male  deck  embark_town  alive  alone
0   man         True   NaN  Southampton    no  False
1  woman        False    C    Cherbourg   yes  False
2  woman        False   NaN  Southampton   yes   True
3  woman        False    C    Southampton   yes  False
4   man         True   NaN  Southampton    no   True
```

```
print(titanic.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 15 columns):
#   Column          Non-Null Count  Dtype
---  -
0   survived        891 non-null    int64
1   pclass          891 non-null    int64
2   sex             891 non-null    object
3   age            714 non-null    float64
4   sibsp          891 non-null    int64
5   parch          891 non-null    int64
6   fare           891 non-null    float64
7   embarked       889 non-null    object
8   class          891 non-null    category
9   who            891 non-null    object
10  adult_male      891 non-null    bool
11  deck           203 non-null    category
12  embark_town     889 non-null    object
13  alive           891 non-null    object
14  alone           891 non-null    bool
dtypes: bool(2), category(2), float64(2), int64(4), object(5)
memory usage: 80.7+ KB
None
```

```
print(titanic.describe())
```

```
count    survived    pclass    age    sibsp    parch    fare
mean     0.383838    2.308642    29.699118    0.523008    0.381594    32.204208
std      0.486592    0.836071    14.526497    1.102743    0.806057    49.693429
min      0.000000    1.000000    0.420000    0.000000    0.000000    0.000000
25%      0.000000    2.000000    20.125000    0.000000    0.000000    7.910400
50%      0.000000    3.000000    28.000000    0.000000    0.000000    14.454200
75%      1.000000    3.000000    38.000000    1.000000    0.000000    31.000000
max      1.000000    3.000000    80.000000    8.000000    6.000000    512.329200
```

```
print(titanic.isnull().sum())
```

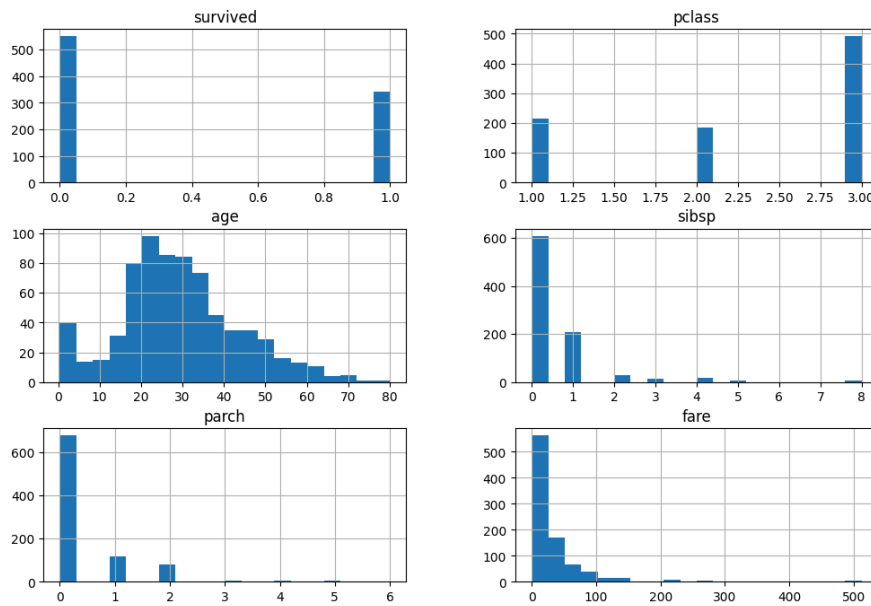
```
survived    0
pclass      0
sex         0
```

```
age          177
sibsp        0
parch        0
fare         0
embarked     2
class        0
who          0
adult_male   0
deck        688
embark_town   2
alive        0
alone        0
dtype: int64
```

```
titanic.hist(bins=20, figsize=(12, 8))
plt.suptitle('Feature Distributions')
plt.show()
```



Feature Distributions



```
sns.pairplot(titanic.select_dtypes(include=[np.number]).dropna())
plt.suptitle('Scatter Plot Matrix')
plt.show()
```

