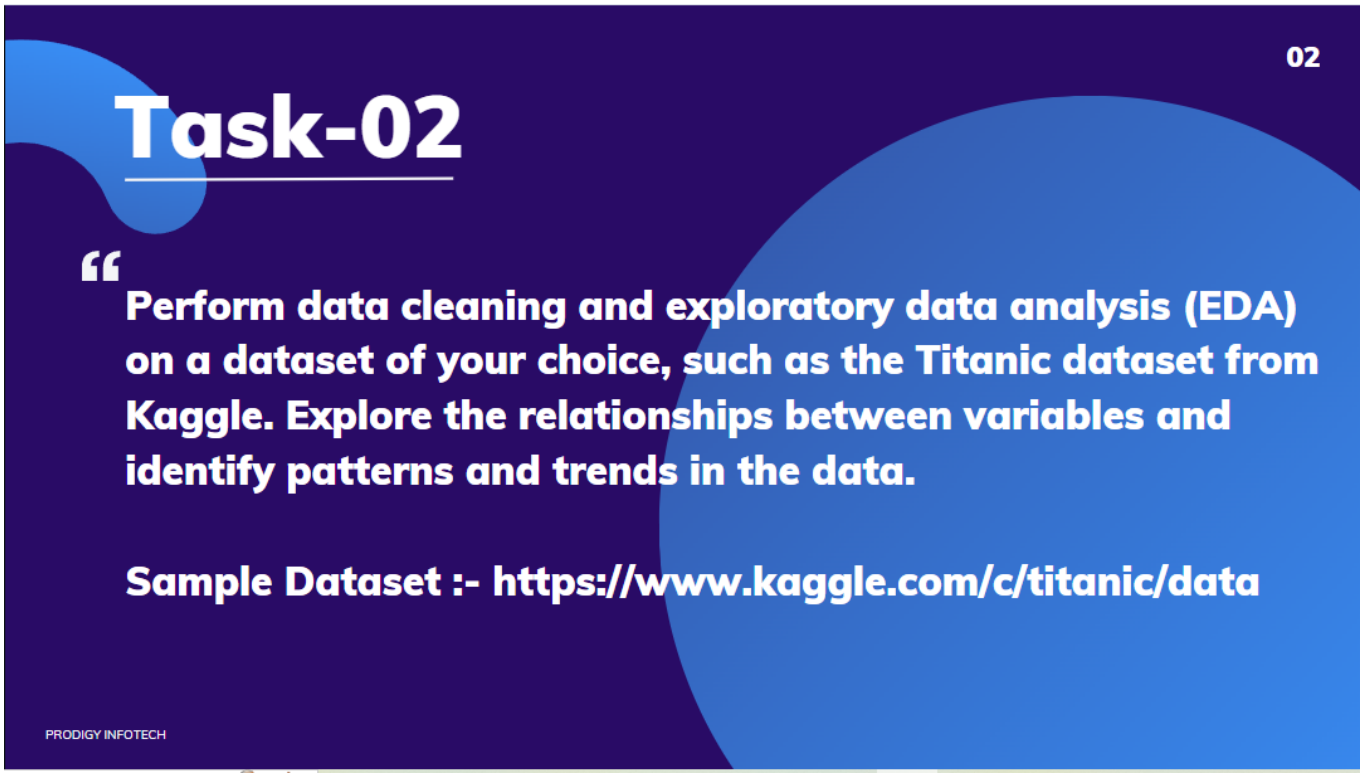



```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

```
url="/content/titanic_train.csv"
```

```
df= pd.read_csv(url)
```

survival Survival 0 = No, 1 = Yes

pclass Ticket class 1 = 1st, 2 = 2nd, 3 = 3rd

sex Sex

Age Age in years

sibsp # of siblings / spouses aboard the Titanic

parch # of parents / children aboard the Titanic

ticket Ticket number

fare Passenger fare

cabin Cabin number

embarked Port of Embarkation C = Cherbourg, Q = Queenstown, S = Southampton

```
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

```
df.tail()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	NaN	Q

```
df.shape
```

```
(891, 12)
```

```
df.info()
```

```
<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
```

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

```
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[df.isnull()]
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...
886	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
887	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
888	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
889	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
890	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

891 rows × 12 columns

```
df.dropna(axis=0,inplace=True)
```

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

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

```
df.shape
```

```
(183, 12)
```

```
df.describe()
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	183.000000	183.000000	183.000000	183.000000	183.000000	183.000000	183.000000
mean	455.366120	0.672131	1.191257	35.674426	0.464481	0.475410	78.682469
std	247.052476	0.470725	0.515187	15.643866	0.644159	0.754617	76.347843
min	2.000000	0.000000	1.000000	0.920000	0.000000	0.000000	0.000000
25%	263.500000	0.000000	1.000000	24.000000	0.000000	0.000000	29.700000
50%	457.000000	1.000000	1.000000	36.000000	0.000000	0.000000	57.000000
75%	676.000000	1.000000	1.000000	47.500000	1.000000	1.000000	90.000000
max	890.000000	1.000000	3.000000	80.000000	3.000000	4.000000	512.329200

```
df.columns
```

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

```
df.PassengerId.nunique()
```

```
183
```

```
df["Survived"].nunique()
```

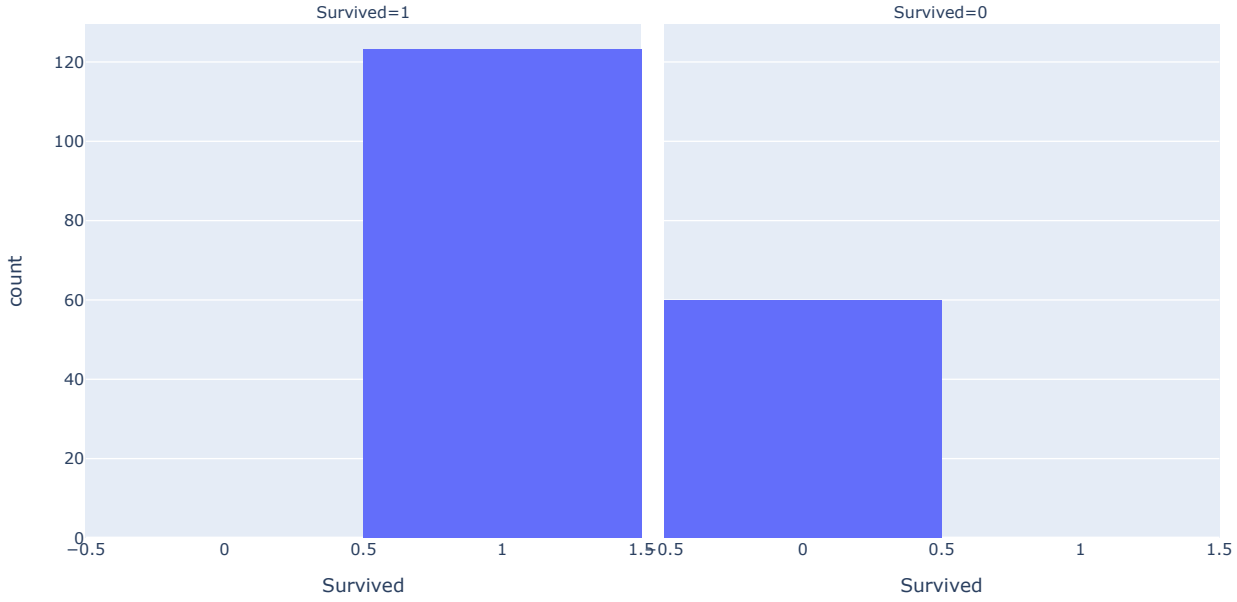
```
2
```

"In our dataset, we observed that out of the total individuals, there were 123 who survived and 60 who did not survive.

```
survived = df["Survived"].value_counts()
survived
```

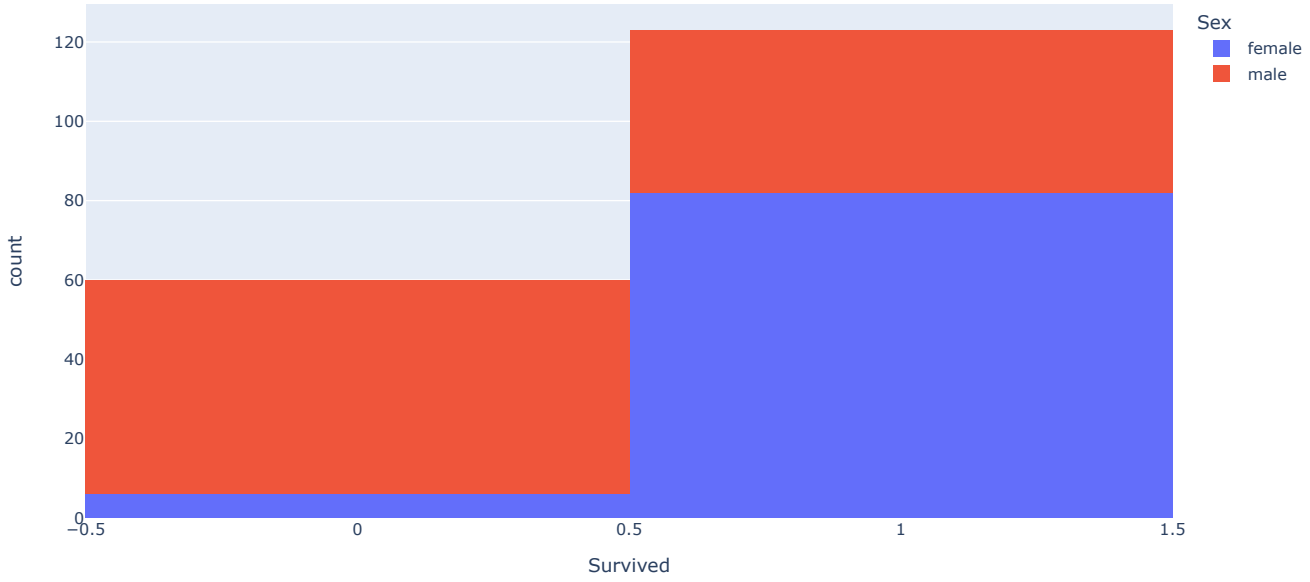
```
1    123
0     60
Name: Survived, dtype: int64
```

```
px.histogram(df,x="Survived",facet_col="Survived")
```



"Out of the 123 individuals who survived, 41 were male and 82 were female."

```
px.histogram(df,x="Survived",color="Sex")
```




```
pclass=df.Pclass.value_counts()
pclass
```

```
1    158
2     15
3     10
Name: Pclass, dtype: int64
```

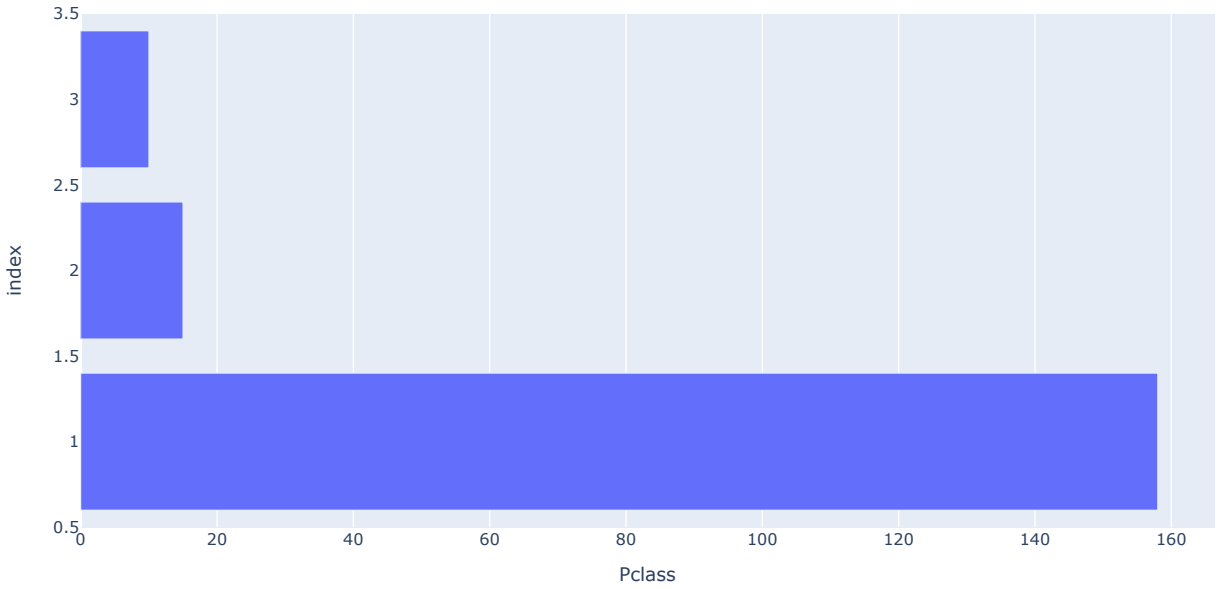
"We have data for the passenger class of individuals as follows:

158 individuals are in Class 1

15 individuals are in Class 2

10 individuals are in Class 3

```
px.bar(df["Pclass"].value_counts(),x="Pclass",hover_name="Pclass")
```

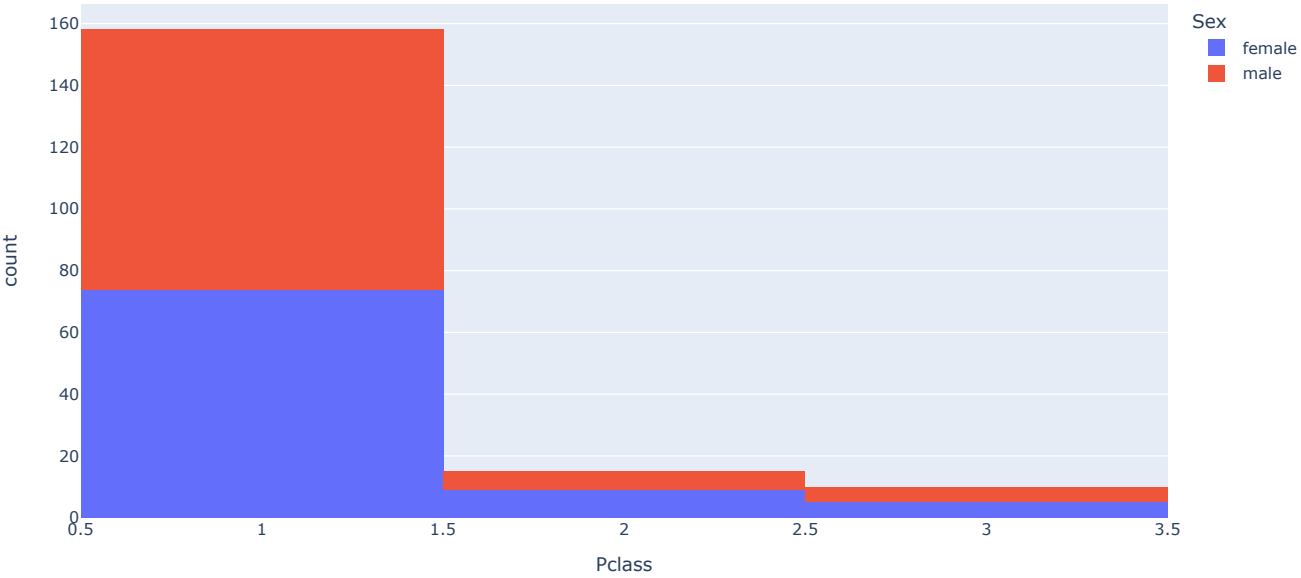


Class 1: Out of the 158 individuals, there are 84 males and 74 females.

Class 2: Out of the 15 individuals, there are 6 males and 9 females.

Class 3: Out of the 10 individuals, there are 5 males and 5 females.

```
px.histogram(df,x="Pclass",color="Sex")
```



passenger class:

Class 1:

Females: Out of the 74 females, 71 survived.

Males: Out of the 84 males, 35 survived.

Class 2:

Females: Out of the 9 females, 8 survived.

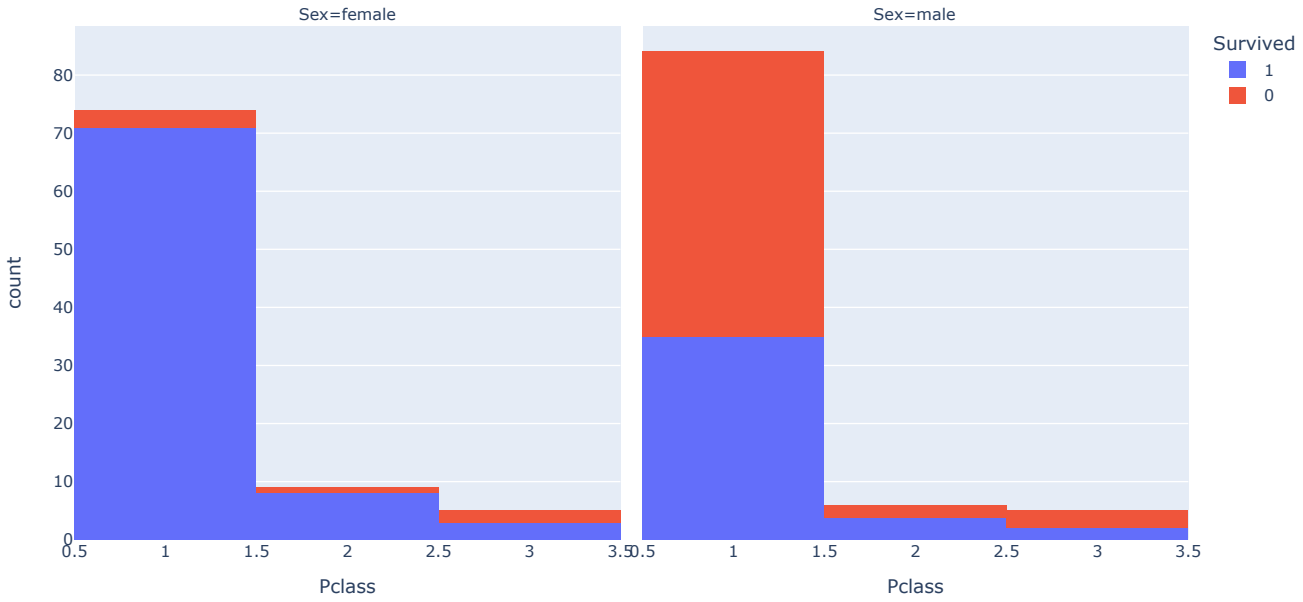
Males: Out of the 6 males, 4 survived.

Class 3:

Females: Out of the 5 females, 3 survived.

Males: Out of the 5 males, 2 survived.

```
px.histogram(df,x='Pclass',color="Survived",facet_col="Sex")
```



Females from Class 1 have a notably higher survival rate, with 71 out of 74 surviving.

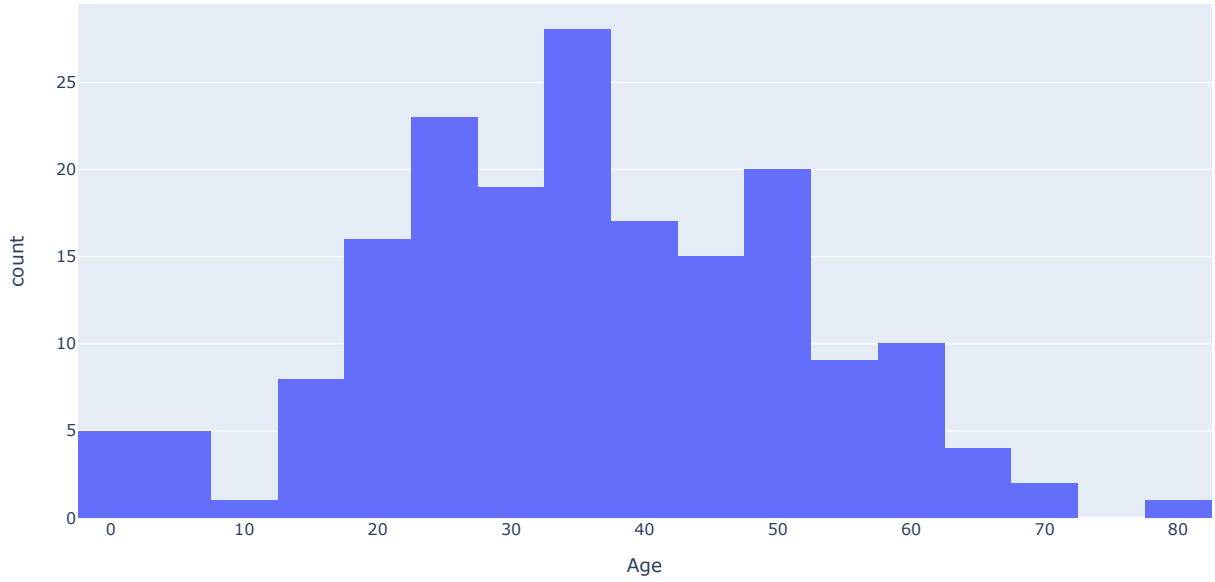
Males from Class 1 have a lower survival rate compared to females, with only 35 out of 84 surviving.

Survival rates decrease with lower passenger class.

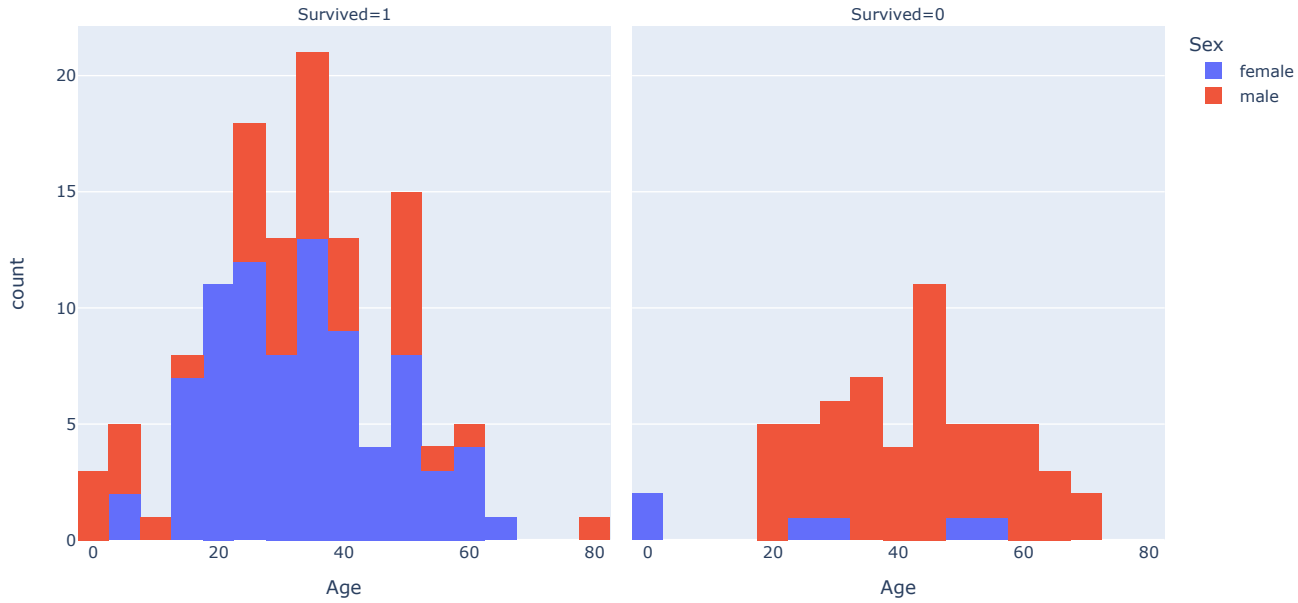
Class 3 has the lowest survival rates for both genders.

The sample sizes in Classes 2 and 3 are relatively small, which may affect the reliability of the observed survival rates.

```
px.histogram(df,x="Age")
```




```
px.histogram(df,x="Age",color="Sex",facet_col="Survived")
```




```
df.columns
```

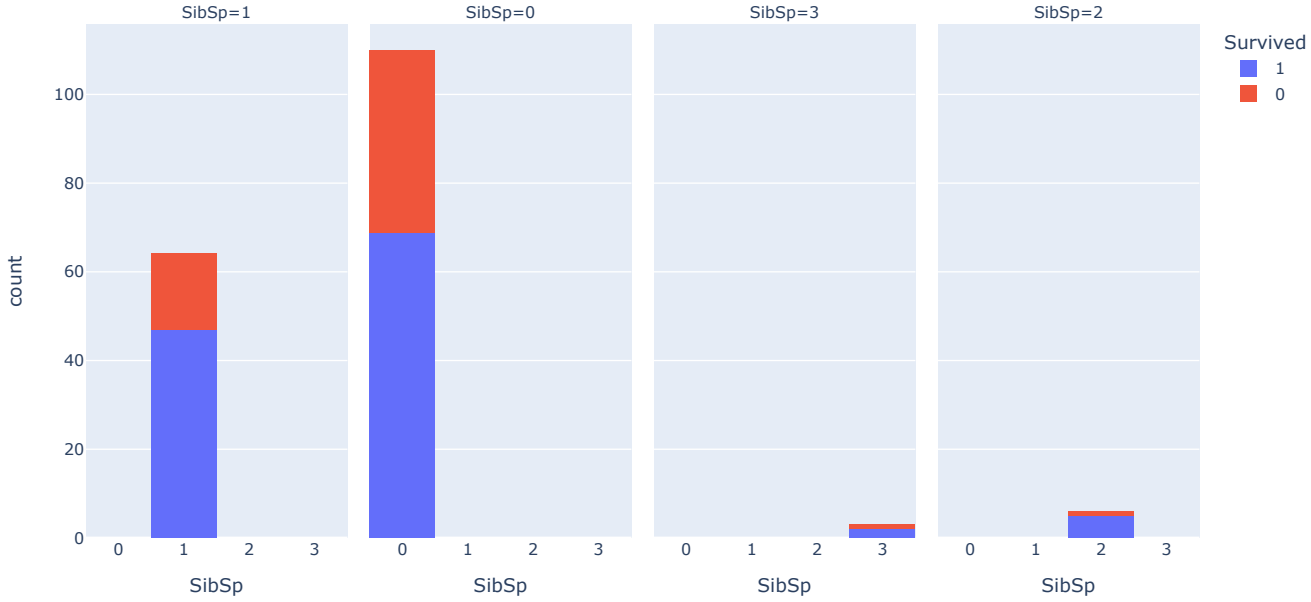
```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

```
df["SibSp"].value_counts()
```

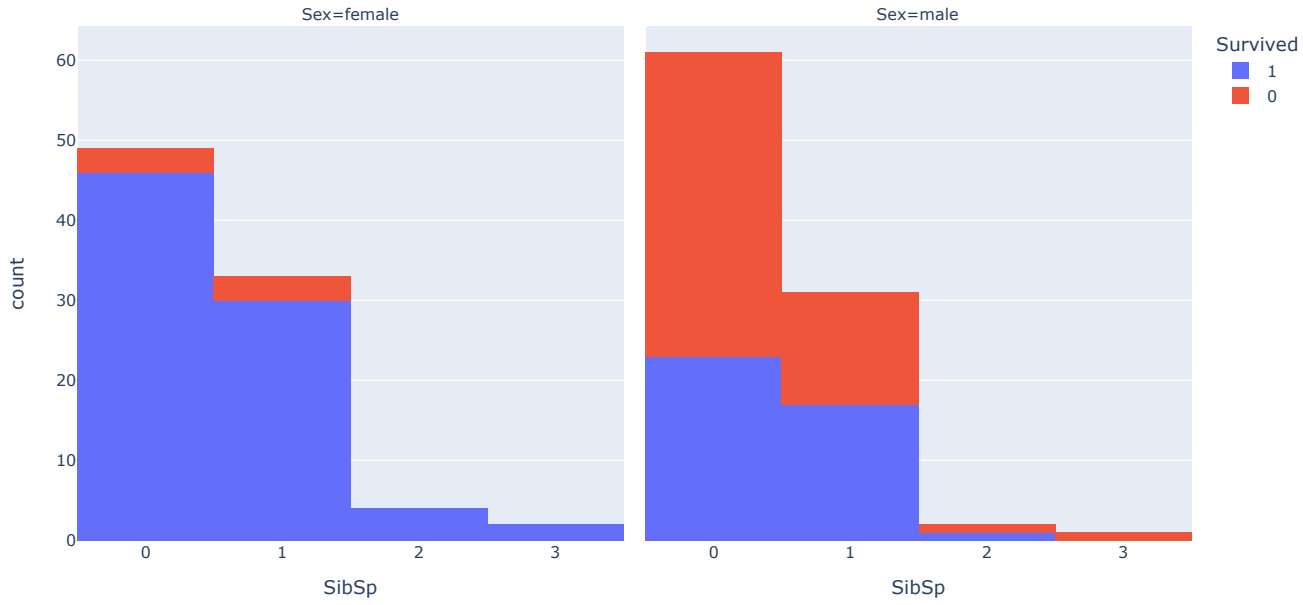
```
0    110
1     64
2      6
3      3
Name: SibSp, dtype: int64
```

```
px.bar?
```

```
px.histogram(df,x="SibSp",color="Survived",facet_col="SibSp",histfunc="count",)
```




```
px.histogram(df,x="SibSp",color="Survived",facet_col="Sex",histfunc="count",)
```




```
df.columns
```

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

```
df.Fare.describe()
```

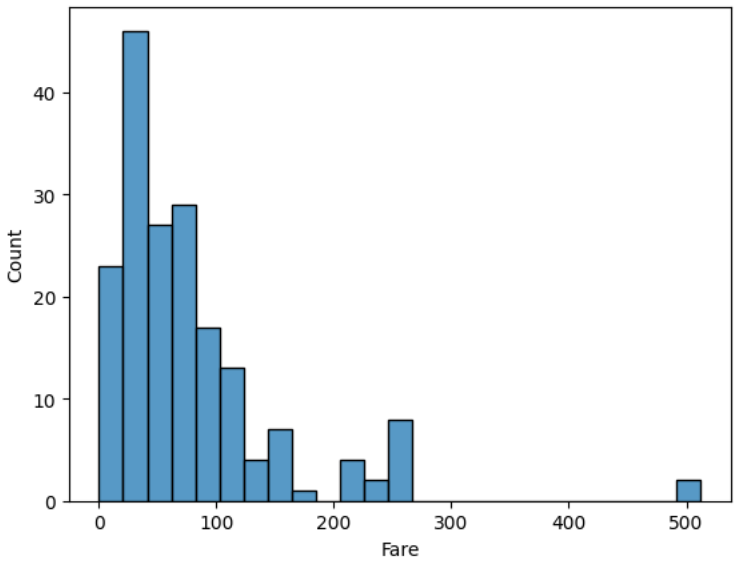
```
count    183.000000
mean      78.682469
std       76.347843
min        0.000000
25%       29.700000
50%       57.000000
75%       90.000000
max      512.329200
Name: Fare, dtype: float64
```

```
df.Ticket.describe()
```

```
count        183
unique       127
top       113760
freq           4
```

```
sns.histplot(data=df,x="Fare",)
```

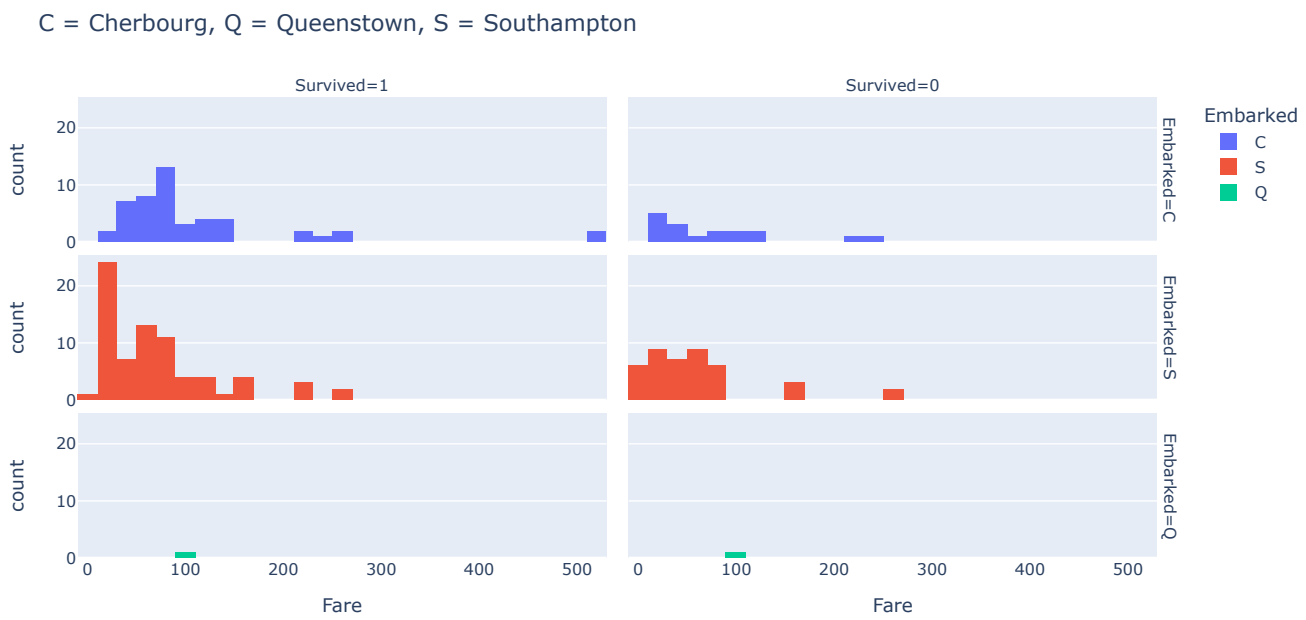
```
<Axes: xlabel='Fare', ylabel='Count'>
```




```
df.Embarked.value_counts()
```

```
S    116
C     65
Q      2
Name: Embarked, dtype: int64
```

```
px.histogram(df,x="Fare",color="Embarked",title="C = Cherbourg, Q = Queenstown, S = Southampton",facet_row="Embarked",facet_col="Survive
```




```
px.histogram(df,x="Fare",color="Pclass",facet_col="Sex",nbins=30,animation_frame="Survived",)
```