In [12]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [13]:
```
df = pd.read_csv("tested.csv")
```

In [14]:
```
df.head()
```

Out[14]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298

In [15]:
```
df.tail()
```

Out[15]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
413	1305	0	3	Spector, Mr. Woolf	male	NaN	0	0	A. 323
414	1306	1	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	1775
415	1307	0	3	Saether, Mr. Simon Sivertsen	male	38.5	0	0	SOTO O. 310126
416	1308	0	3	Ware, Mr. Frederick	male	NaN	0	0	3593
417	1309	0	3	Peter, Master. Michael J	male	NaN	1	1	266

In [16]:
```
df.shape
```

Out[16]:
```
(418, 12)
```

In [69]:
```
df.drop("PassengerId", axis = 1, inplace = True)
```

In [70]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 11 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   Survived  418 non-null    int64  
 1   Pclass    418 non-null    int64  
 2   Name      418 non-null    object 
 3   Sex       418 non-null    object 
 4   Age       418 non-null    float64
 5   SibSp     418 non-null    int64  
 6   Parch     418 non-null    int64  
 7   Ticket    418 non-null    object 
 8   Fare      417 non-null    float64
 9   Cabin     418 non-null    object 
 10  Embarked  418 non-null    object 
dtypes: float64(2), int64(4), object(5)
memory usage: 36.1+ KB
```

In [71]:
```
df.describe()
```

Out[71]:

	Survived	Pclass	Age	SibSp	Parch	Fare
count	418.000000	418.000000	418.000000	418.000000	418.000000	417.000000
mean	0.363636	2.265550	30.272590	0.447368	0.392344	35.627188
std	0.481622	0.841838	12.634534	0.896760	0.981429	55.907576
min	0.000000	1.000000	0.170000	0.000000	0.000000	0.000000
25%	0.000000	1.000000	23.000000	0.000000	0.000000	7.895800
50%	0.000000	3.000000	30.272590	0.000000	0.000000	14.454200
75%	1.000000	3.000000	35.750000	1.000000	0.000000	31.500000
max	1.000000	3.000000	76.000000	8.000000	9.000000	512.329200

In [72]:
```
df.dtypes
```

Out[72]:
```
Survived      int64
Pclass        int64
Name         object
Sex          object
Age         float64
SibSp         int64
Parch         int64
Ticket       object
Fare        float64
Cabin        object
Embarked     object
dtype: object
```

In [73]:
```
df.isnull().sum()
```

Out[73]:
```
Survived    0
Pclass      0
Name        0
Sex         0
Age         0
SibSp       0
Parch       0
Ticket      0
Fare        1
Cabin       0
Embarked    0
dtype: int64
```

In [74]:
```
df["Age"].mean()
```

Out[74]:
```
np.float64(30.272590361445783)
```

In [75]:
```
df["Age"] = df["Age"].fillna(df["Age"].mean())
```

In [76]:
```
df["Cabin"] = df["Cabin"].fillna("Unknown")
```

In [77]:
```
df.isnull().sum()
```

Out[77]:
```
Survived    0
Pclass      0
Name        0
Sex         0
Age         0
SibSp       0
Parch       0
Ticket      0
Fare        1
Cabin       0
Embarked    0
dtype: int64
```

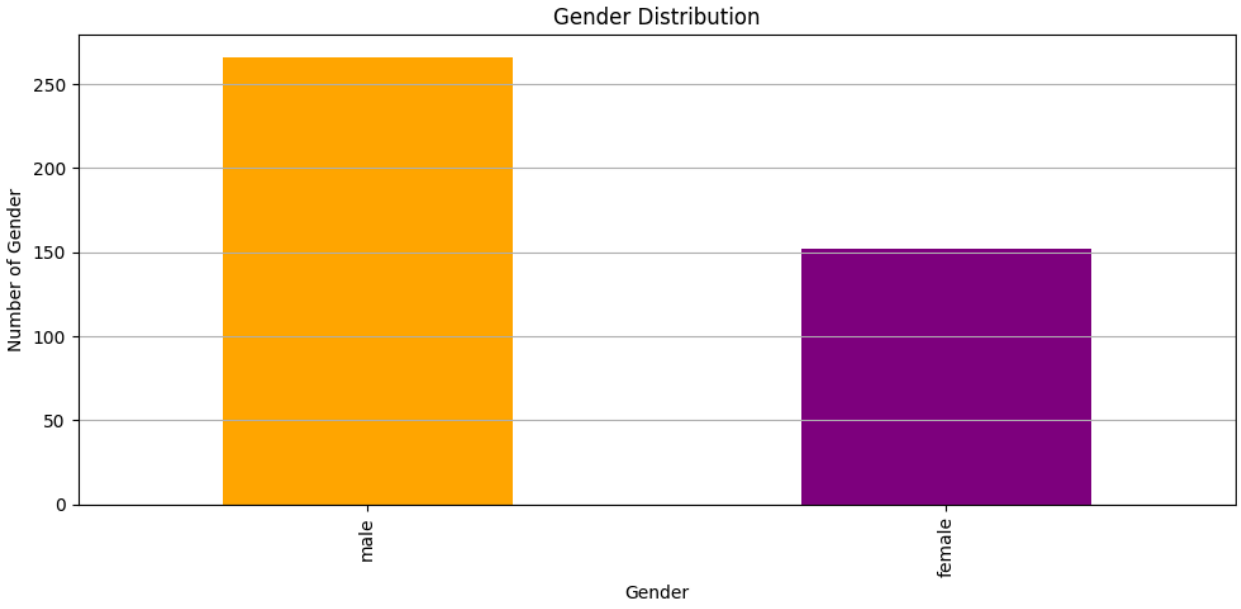
In [78]:
```
df.duplicated().sum()
```

Out[78]:
```
np.int64(0)
```

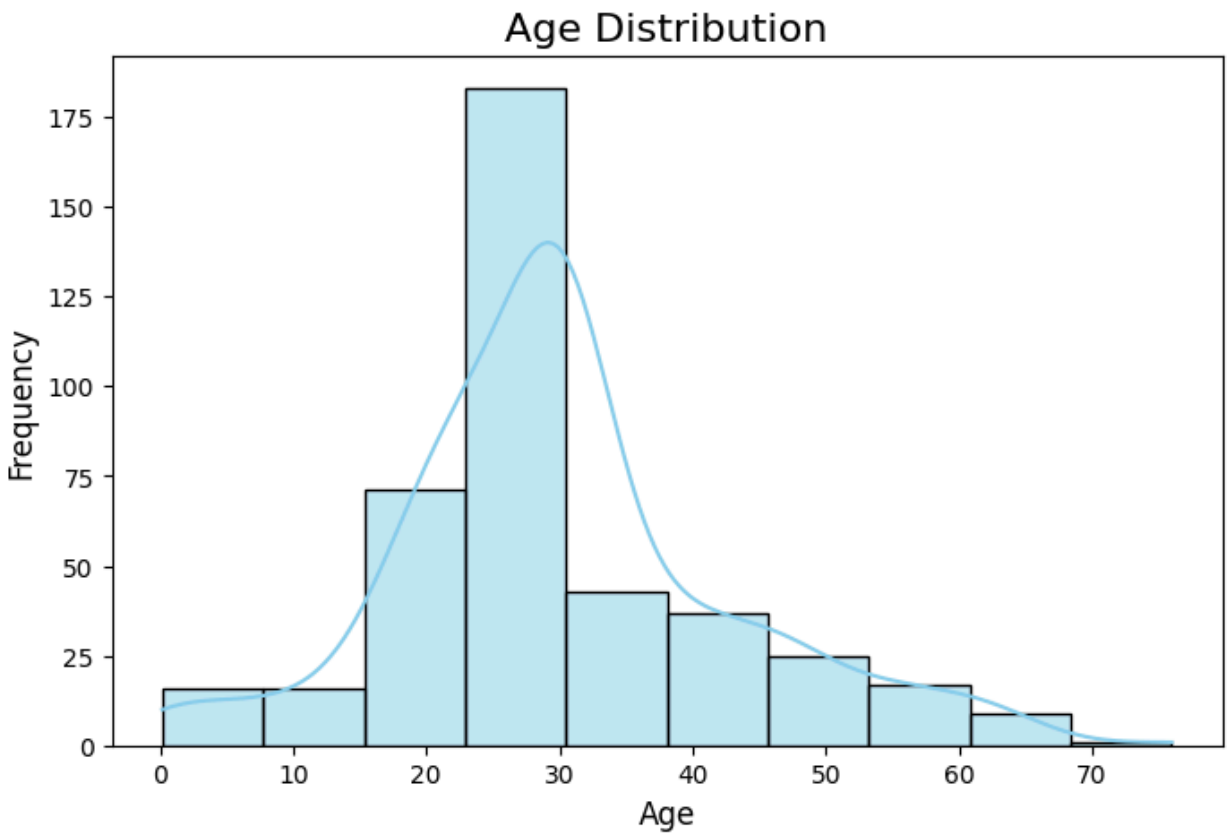
In [79]:
```
df.columns
```

Out[79]:
```
Index(['Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp', 'Parch', 'Ticke
t',
       'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

In [88]:
```
plt.figure(figsize = (10,5))
df["Sex"].value_counts().plot(kind = "bar", color = ["Orange","Purple"])
plt.xlabel("Gender")
plt.ylabel("Number of Gender")df
plt.title("Gender Distribution")
plt.tight_layout()
plt.grid(axis = "y")
plt.show()
```



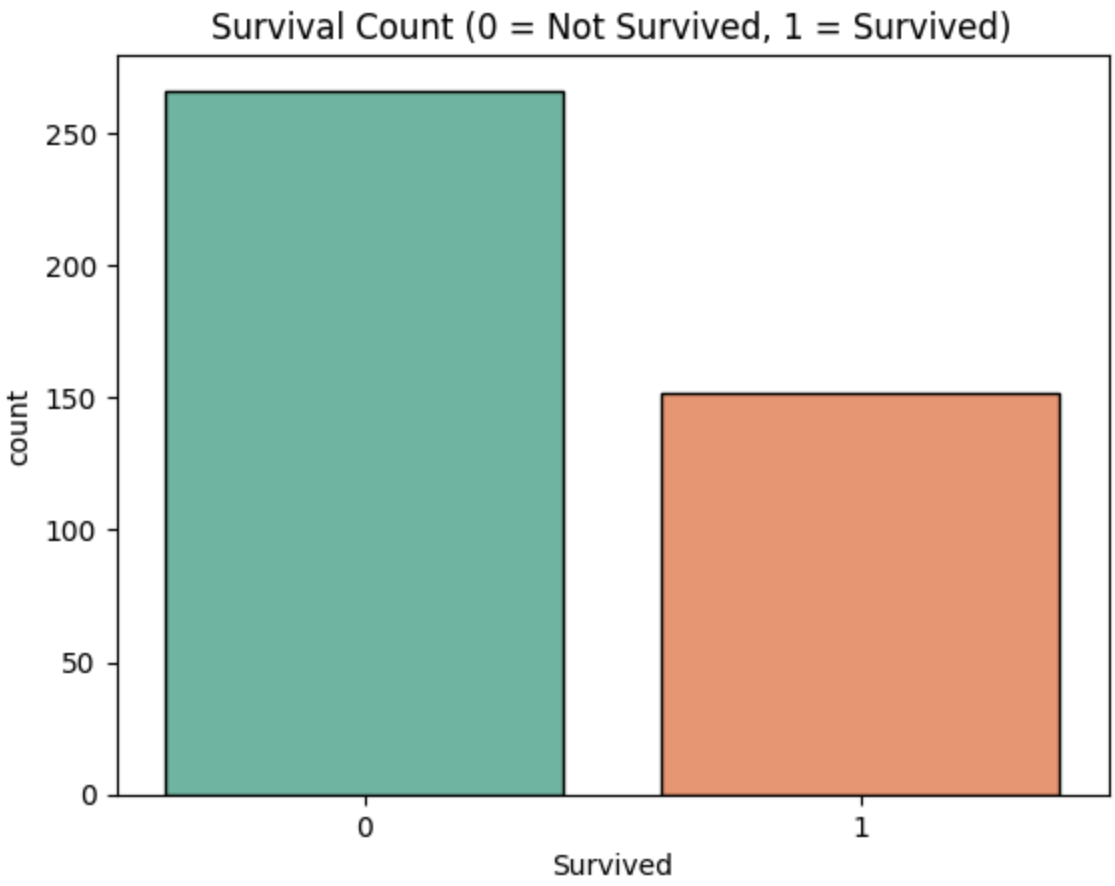
In [95]:
```
plt.figure(figsize=(8,5))  # set size
sns.histplot(data=df, x="Age", bins=10, kde=True, color="skyblue", edgecolor="

# Add labels and title
plt.title("Age Distribution", fontsize=16)
plt.xlabel("Age", fontsize=12)
plt.ylabel("Frequency", fontsize=12)

plt.show()
```



In [131…

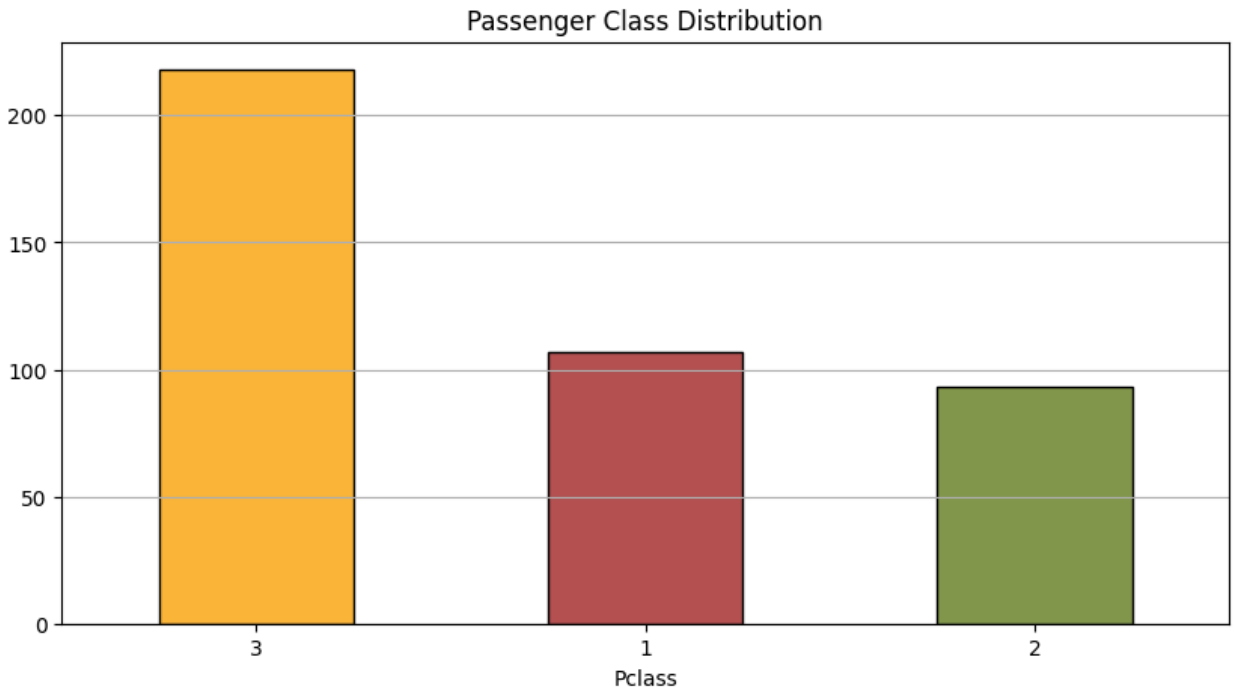
```
sns.countplot(data=df, x="Survived", palette="Set2", edgecolor = "black")
plt.title("Survival Count (0 = Not Survived, 1 = Survived)")
plt.show()
```

```
C:\Users\shash\AppData\Local\Temp\ipykernel_18028\902830758.py:1: FutureWarnin
g:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same e
ffect.

  sns.countplot(data=df, x="Survived", palette="Set2", edgecolor = "black")
```



In [128…

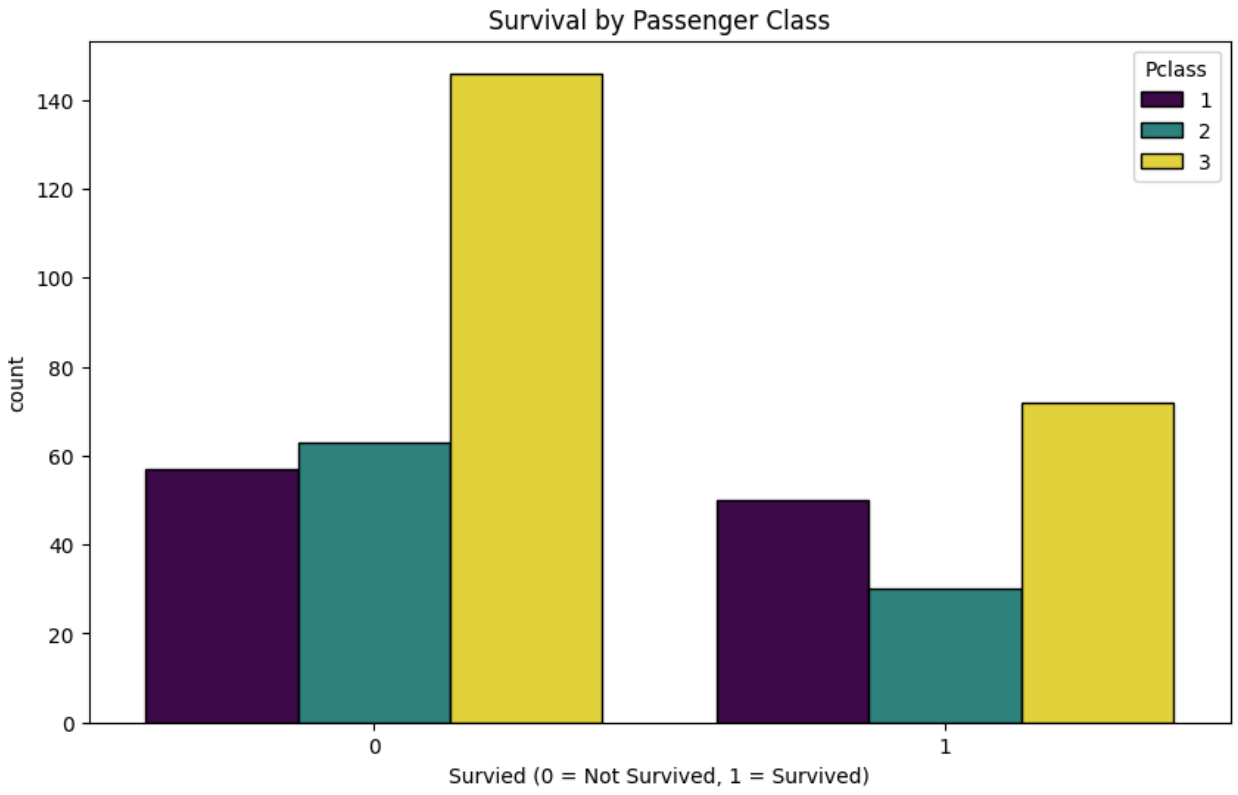
```
plt.figure(figsize=(10,5))
df["Pclass"].value_counts().plot(kind = "bar", color = ["#FCB53B","#B45253","#
plt.xticks(rotation=0)
plt.title("Passenger Class Distribution")
plt.grid(axis = "y")
plt.show()
```



Most passengers were in 3rd class.

In [129…

```
plt.figure(figsize = (10,6))
sns.countplot(data = df , x = "Survived", hue = "Pclass", palette="viridis", e
plt.title("Survival by Passenger Class")
plt.xlabel("Survied (0 = Not Survived, 1 = Survived)")
plt.show()
```

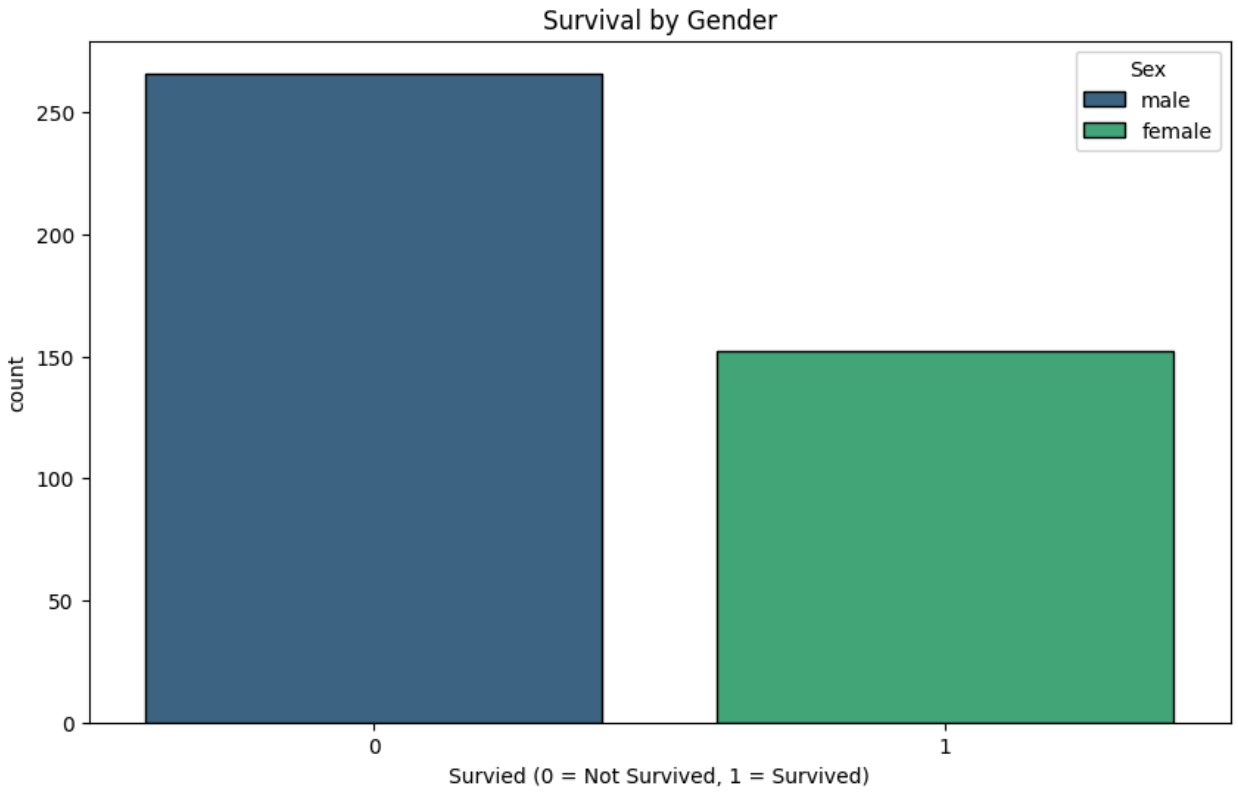


Higher survival rate in 1st class.

In [122…
```
df.columns
```

Out[122…
```
Index(['Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp', 'Parch', 'Ticke
t',
       'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

In [130…
```
plt.figure(figsize = (10,6))
sns.countplot(data = df , x = "Survived", hue = "Sex", palette="viridis", edge
plt.title("Survival by Gender")
plt.xlabel("Survied (0 = Not Survived, 1 = Survived)")
plt.show()
```

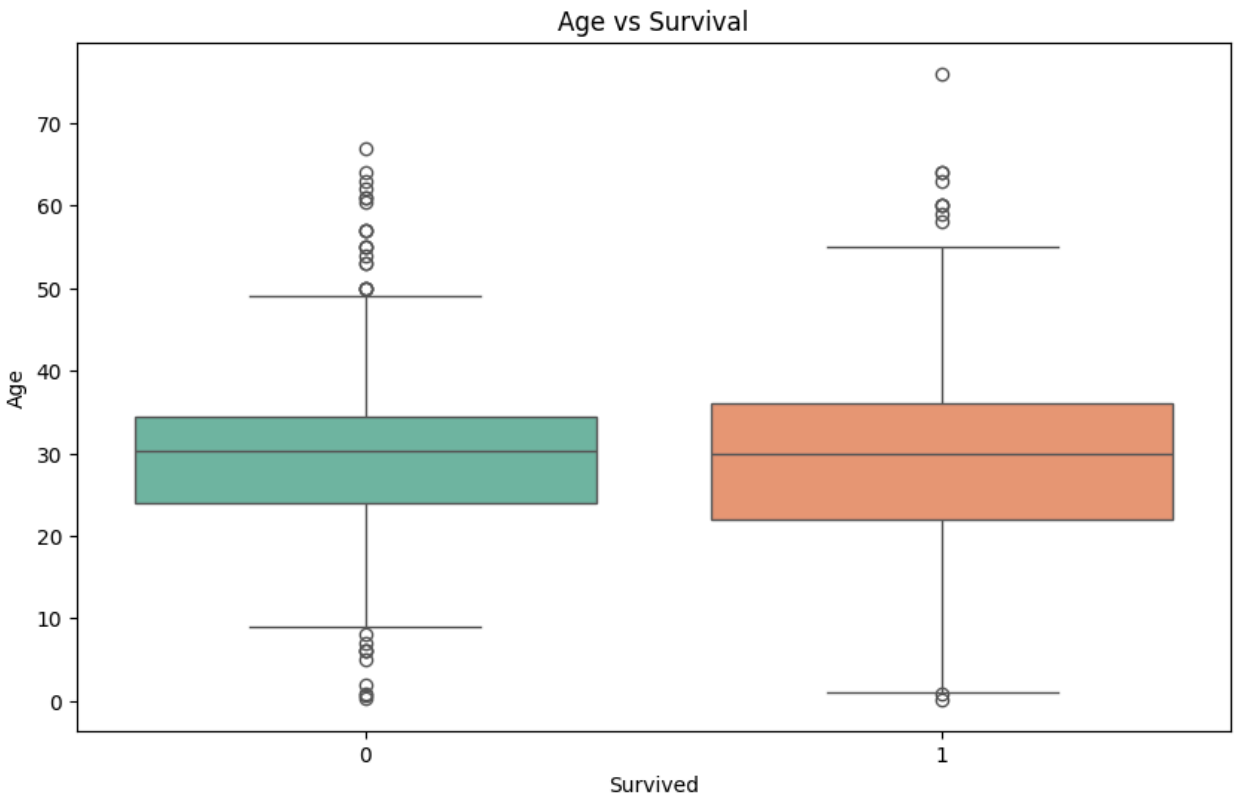


More women survived compared to men.

In [133…

```
plt.figure(figsize = (10,6))
sns.boxplot(data=df, x="Survived", y="Age", palette="Set2")
plt.title("Age vs Survival")
plt.show()
```

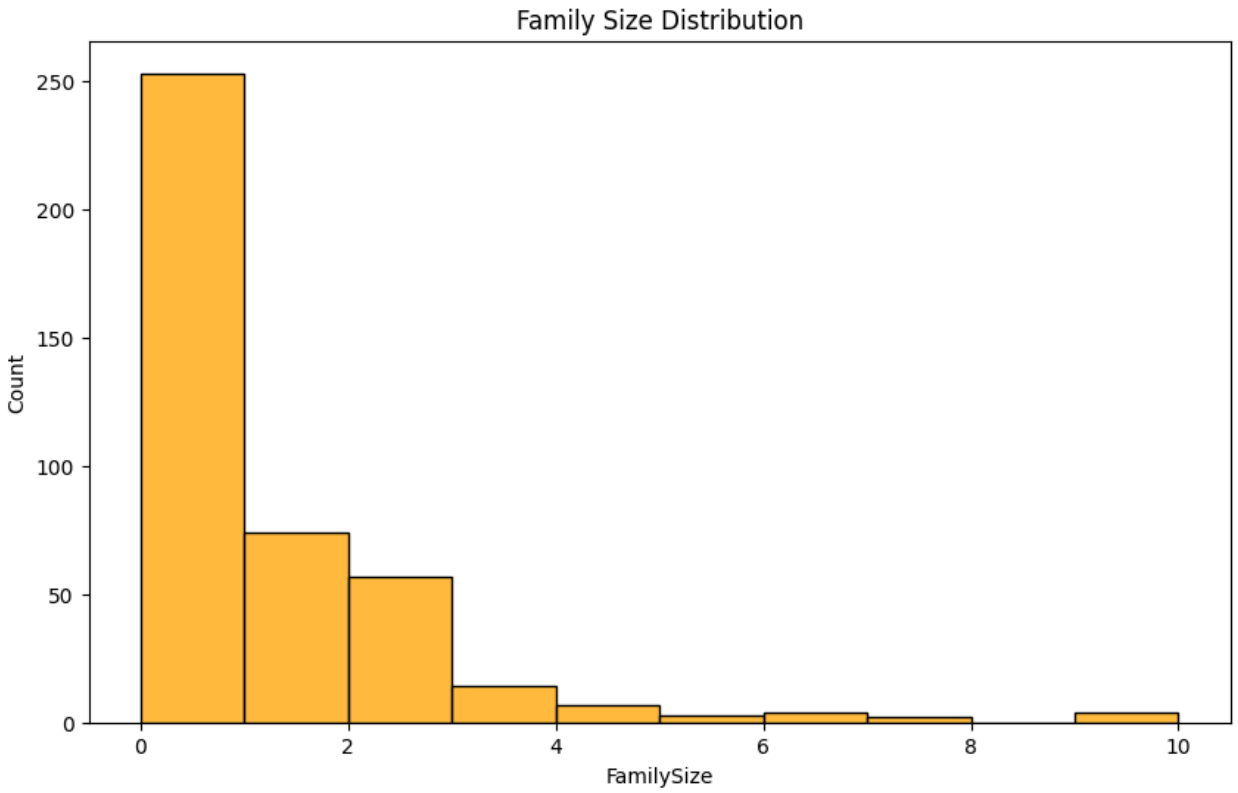
```
C:\Users\shash\AppData\Local\Temp\ipykernel_18028\405743674.py:2: FutureWarnin
g:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same e
ffect.

  sns.boxplot(data=df, x="Survived", y="Age", palette="Set2")
```



Children had higher survival chances.

In [138…

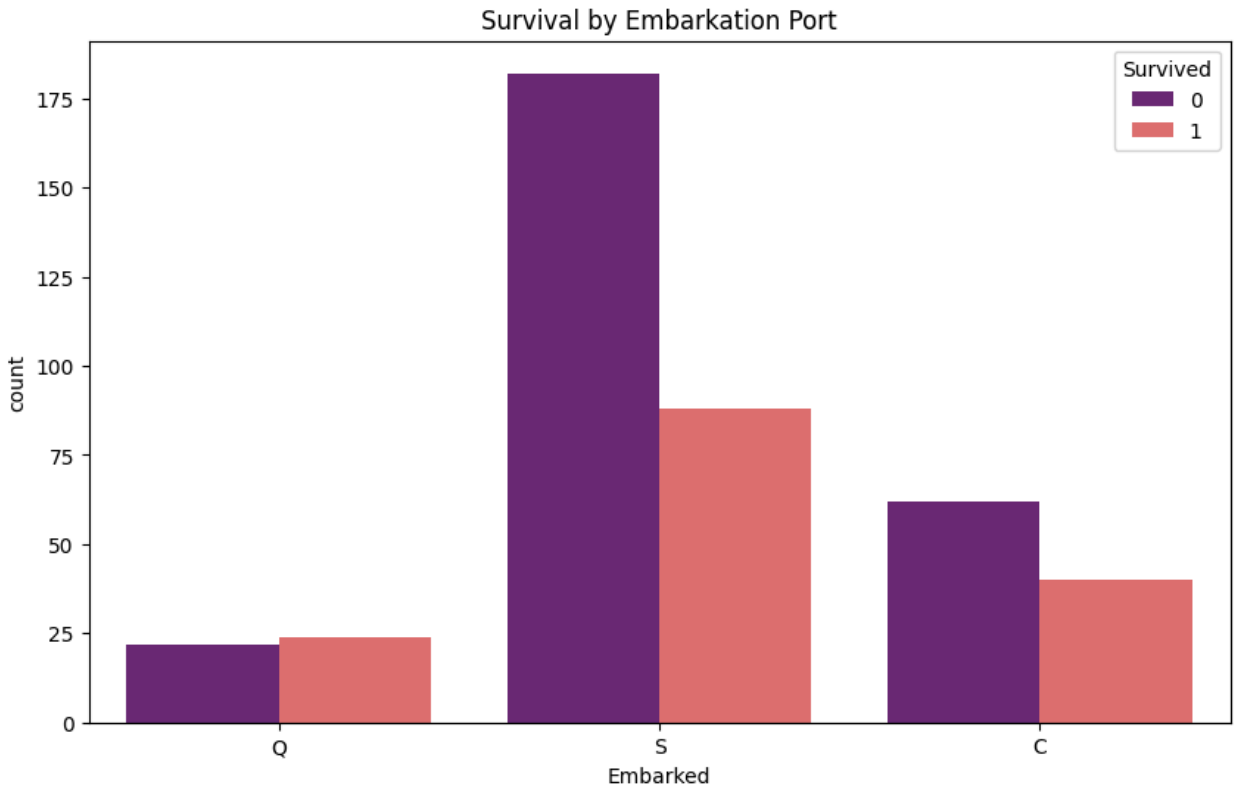
```
plt.figure(figsize = (10,6))
df["FamilySize"] = df["SibSp"] + df["Parch"]

sns.histplot(data=df, x="FamilySize", bins=10, kde=False, color="orange")
plt.title("Family Size Distribution")
plt.show()
```



Many passengers were traveling alone.

In [139…

```
plt.figure(figsize = (10,6))
sns.countplot(data=df, x="Embarked", hue="Survived", palette="magma")
plt.title("Survival by Embarkation Port")
plt.show()
```


Survival by Embarkation Port
Survived
0
1
count
175
150
125
100
75
50
25
0
Q
S
C
Embarked

In [ ]: