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

```
df=pd.read_csv('/content/drive/MyDrive/Sem6_DSBDA/DataViz1n2/titanic.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 21171
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450

Next steps: View recommended plots

```
df.describe()
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.00
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.20
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.69
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.00
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.91
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.45
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.00
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.32

```
df.isnull()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Far
0	False	False	False	False	False	False	False	False	False	Fals
1	False	False	False	False	False	False	False	False	False	Fals
2	False	False	False	False	False	False	False	False	False	Fals
3	False	False	False	False	False	False	False	False	False	Fals
4	False	False	False	False	False	False	False	False	False	Fals
...	...	...	...	...	...	...	...	...	...	..
886	False	False	False	False	False	False	False	False	False	Fals
887	False	False	False	False	False	False	False	False	False	Fals
888	False	False	False	False	False	True	False	False	False	Fals
889	False	False	False	False	False	False	False	False	False	Fals
890	False	False	False	False	False	False	False	False	False	Fals

891 rows × 12 columns

```
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['Embarked']=df['Embarked'].fillna('S')
```

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

```
df['Cabin_new'] = df['Cabin'].fillna(0)
df['Cabin_new'] = df['Cabin'].notnull().astype(int)
```

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

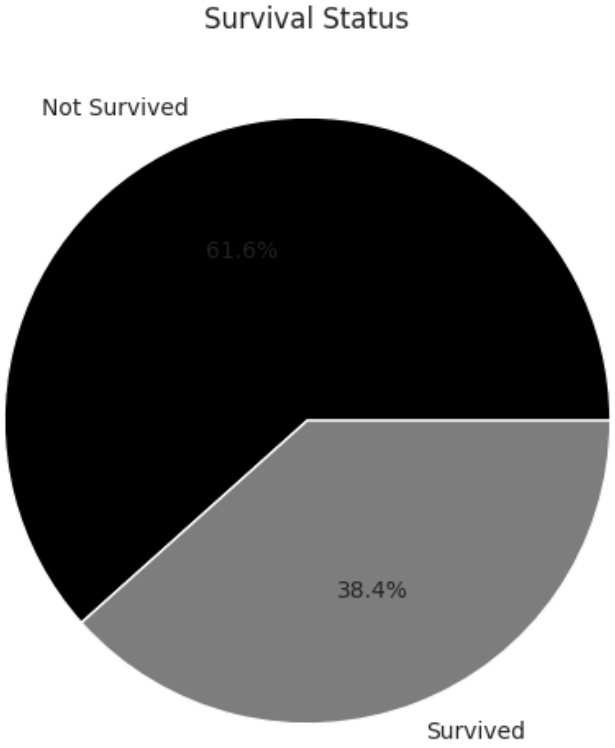
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age              0
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         0
Cabin_new        0
dtype: int64
```

```
df.drop(columns=['Cabin'], inplace=True)
```

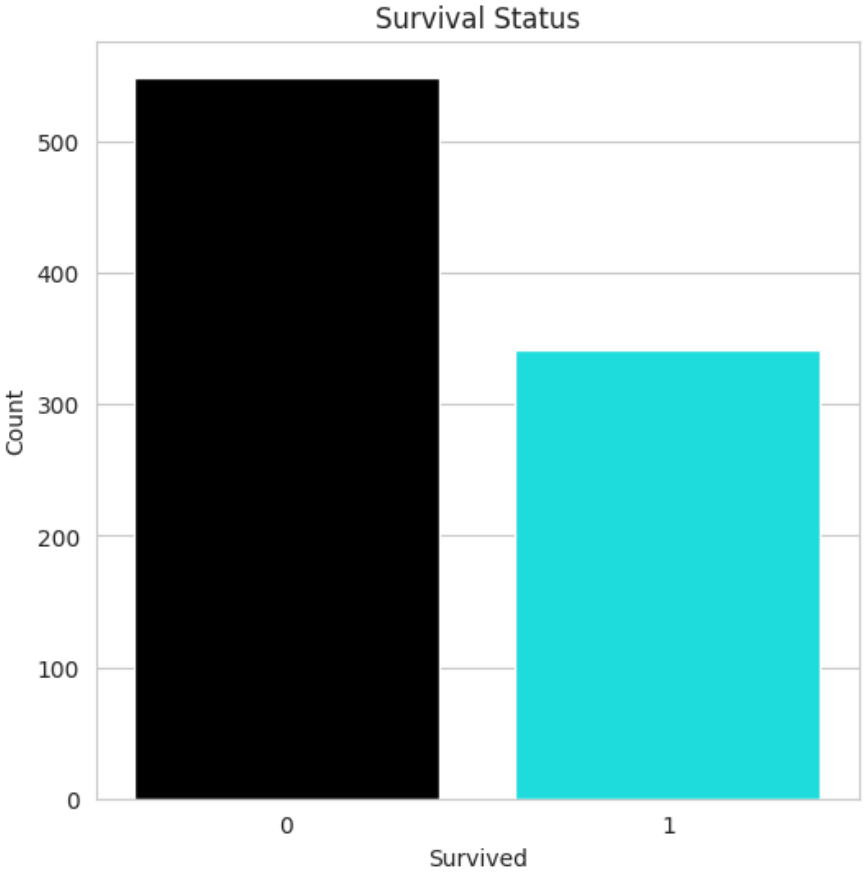
```
survived = df['Survived'].value_counts()
survived
```

```
0    549
1    342
Name: Survived, dtype: int64
```

```
# pieplot
plt.figure(figsize=(6, 6))
plt.pie(survived, labels=['Not Survived', 'Survived'],
autopct='%1.1f%%', colors=['black', 'grey'])
plt.title('Survival Status')
plt.show()
```




```
#barplot or countplot
sns.set_style("whitegrid")
plt.figure(figsize=(6, 6))
sns.countplot(x='Survived', hue='Survived', data=df, palette=['black', 'cyan'], legend=False)
plt.title('Survival Status')
plt.xlabel('Survived')
plt.ylabel('Count')
plt.show()
```



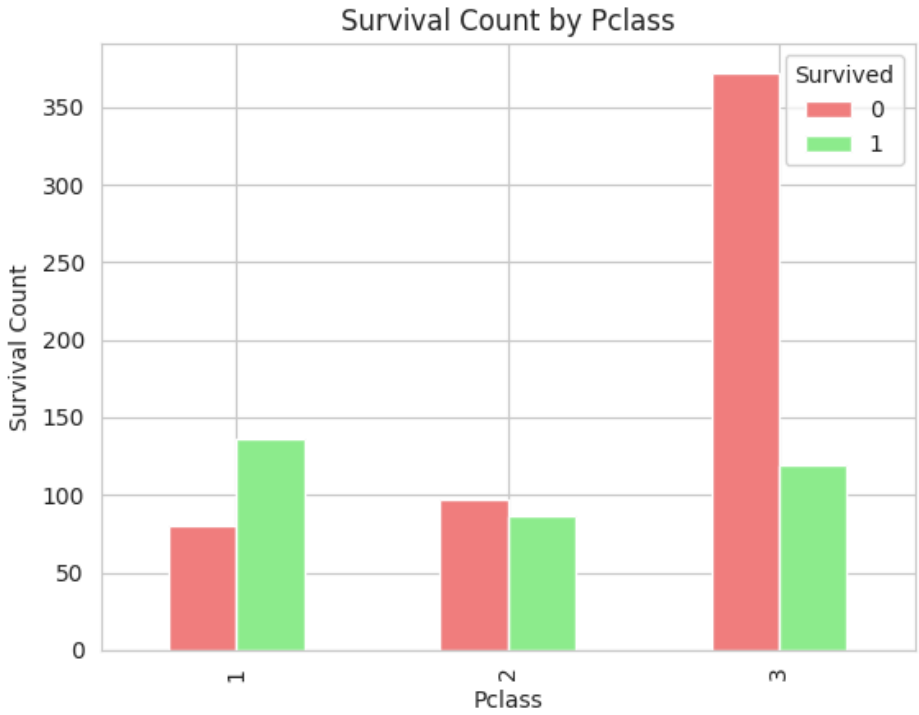

```
survival_count_byPclass = df.groupby(['Pclass', 'Survived']).size().unstack(fill_value=0)
survival_count_byPclass
```

Survived	0	1
Pclass		
1	80	136
2	97	87
3	372	119

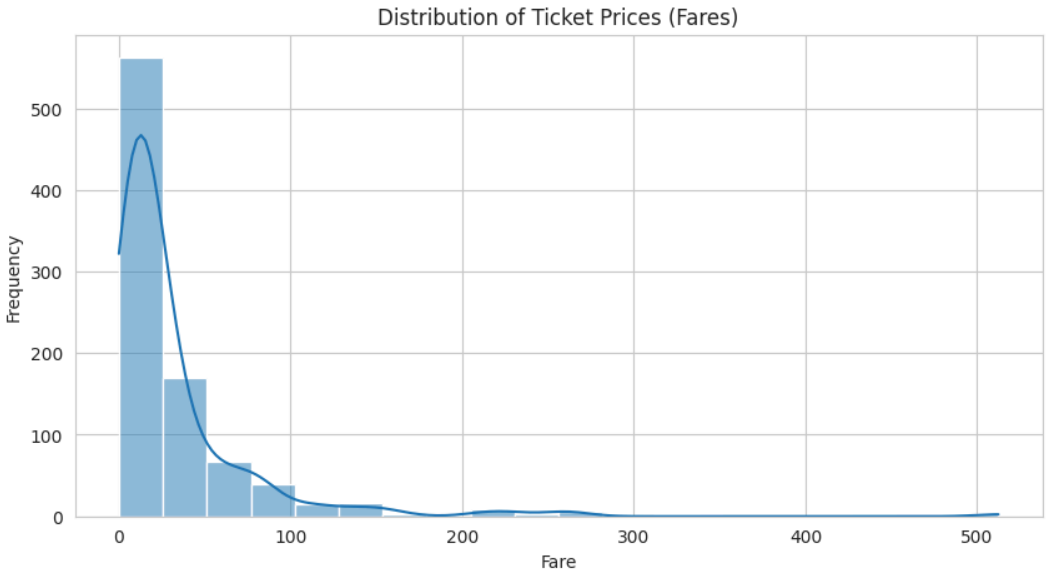
Next steps: View recommended plots

```
survival_count_byPclass.plot(kind='bar', color=['lightcoral', 'lightgreen'])
plt.xlabel('Pclass')
plt.ylabel('Survival Count')

plt.title('Survival Count by Pclass')
plt.show()
```

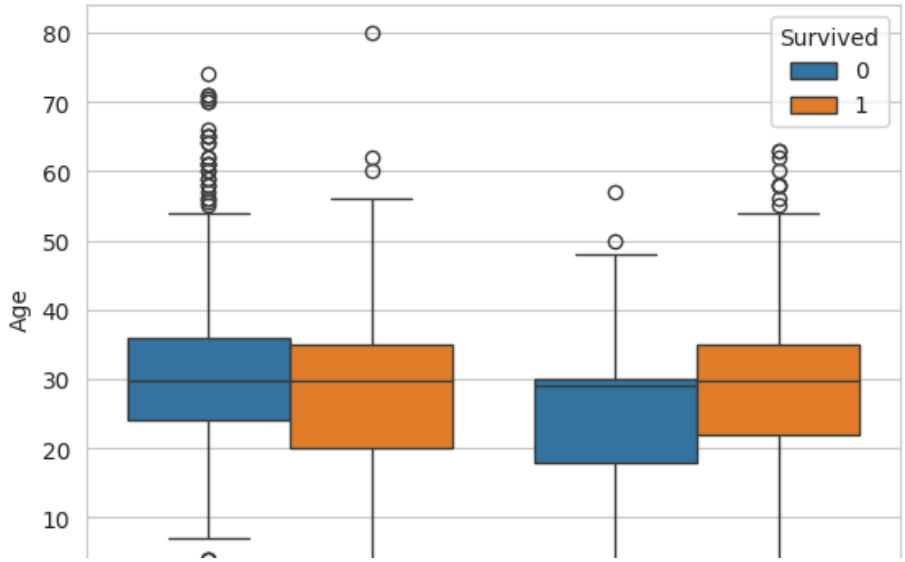



```
#histplot for tickect fare
plt.figure(figsize=(10, 5))
sns.histplot(df['Fare'], bins=20, kde=True)
plt.xlabel('Fare')
plt.ylabel('Frequency')
plt.title('Distribution of Ticket Prices (Fares)')
plt.show()
```




```
sns.boxplot(x='Sex', y='Age', data=df, hue="Survived")
```

```
<Axes: xlabel='Sex', ylabel='Age'>
```




```
sns.violinplot(x='Sex', y='Age', data=df, hue="Survived", palette=['lightcoral', 'cyan'])
```

```
<Axes: xlabel='Sex', ylabel='Age'>
```