In [1]:

```
# ✅ Titanic Dataset - Exploratory Data Analysis
# Step 1: Import Librariesimport pandas as pdimport matplotlib.pyplot as pltimport seaborn as sns
sns.set(style="whitegrid")%matplotlib inline
# Step 2: Load Datasetdf = pd.read_csv('train.csv')  # Ensure train.csv is in the same directoryprint("✅ Dataset Loaded")display(df.head())
# Step 3: Understand the Dataprint("\n📊 Data Info:")df.info()print("\n📈 Statistical Summary:")display(df.describe())print("\n🔍 Missing Values:")print(df.isnull().sum())
# Step 4: Handle Missing Valuesdf['Age'].fillna(df['Age'].median(), inplace=True)df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)df.drop(columns=['Cabin', 'Ticket'], inplace=True)
# Step 5: Univariate Analysisplt.figure(figsize=(12, 4))sns.histplot(df['Age'], kde=True)plt.title('Age Distribution')plt.show()
sns.histplot(df['Fare'], kde=True)plt.title('Fare Distribution')plt.show()
sns.countplot(x='Pclass', data=df)plt.title('Passenger Class Count')plt.show()
# Step 6: Bivariate Analysissns.countplot(x='Sex', hue='Survived', data=df)plt.title('Survival by Gender')plt.show()
sns.countplot(x='Pclass', hue='Survived', data=df)plt.title('Survival by Passenger Class')plt.show()
sns.boxplot(x='Survived', y='Age', data=df)plt.title('Age vs Survival')plt.show()
# Step 7: Correlation Heatmapplt.figure(figsize=(10,6))sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm')plt.title('Correlation Heatmap')plt.show()
# Step 8: Pairplotsns.pairplot(df[['Survived', 'Age', 'Fare', 'Pclass']], hue='Survived')plt.suptitle('Pairplot of Key Features', y=1.02)plt.show()
# Step 9: Summaryprint("""✅ Summary of Findings:- Women had a much higher survival rate than men.- First-class passengers were more likely to survive.- Age and Fare have a mild correlation with survival.- Strong correlation found between Pclass and Fare.""")
```

✅ Dataset Loaded

	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
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

📊 Data 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
```

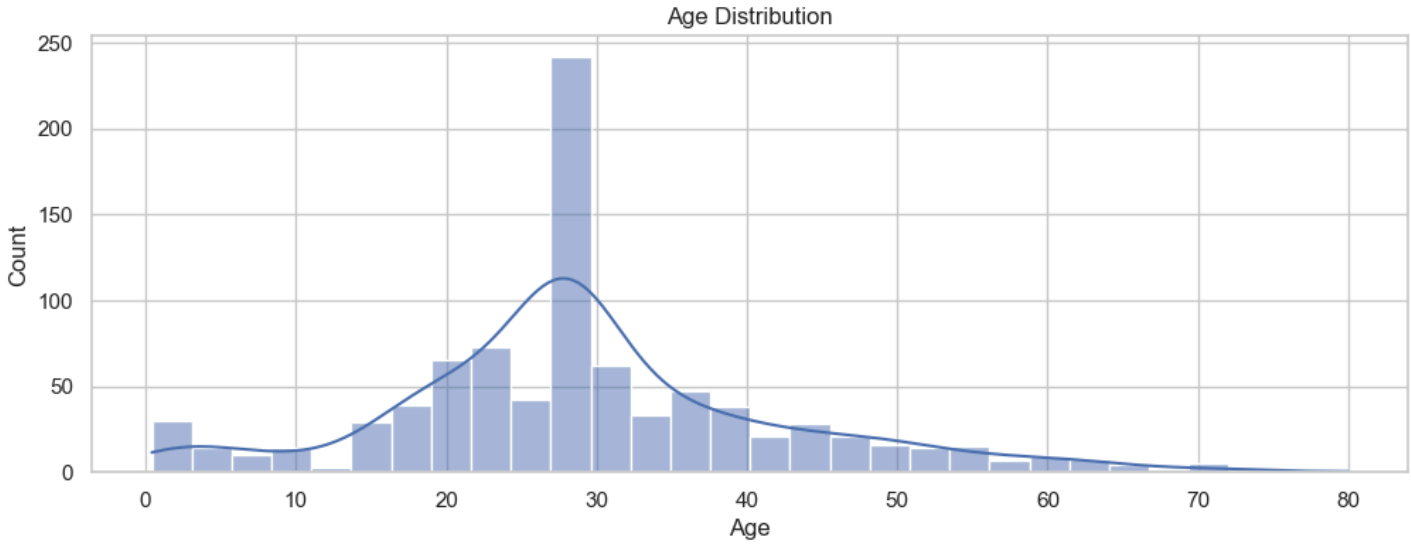
📈 Statistical Summary:

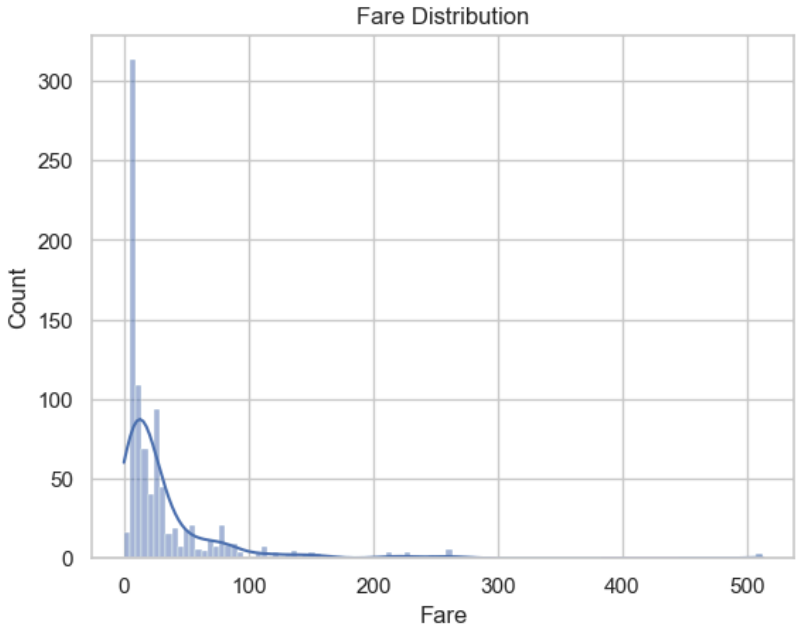
	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

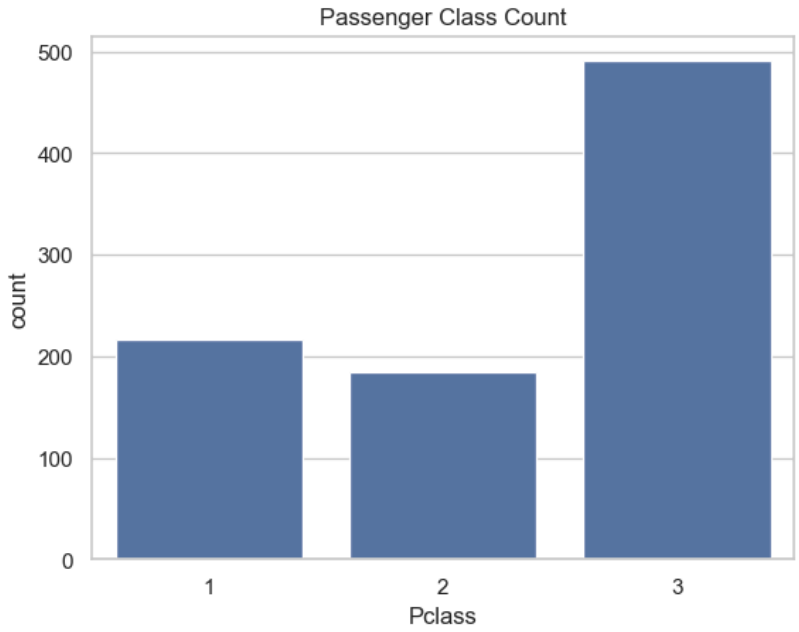
🔍 Missing Values:

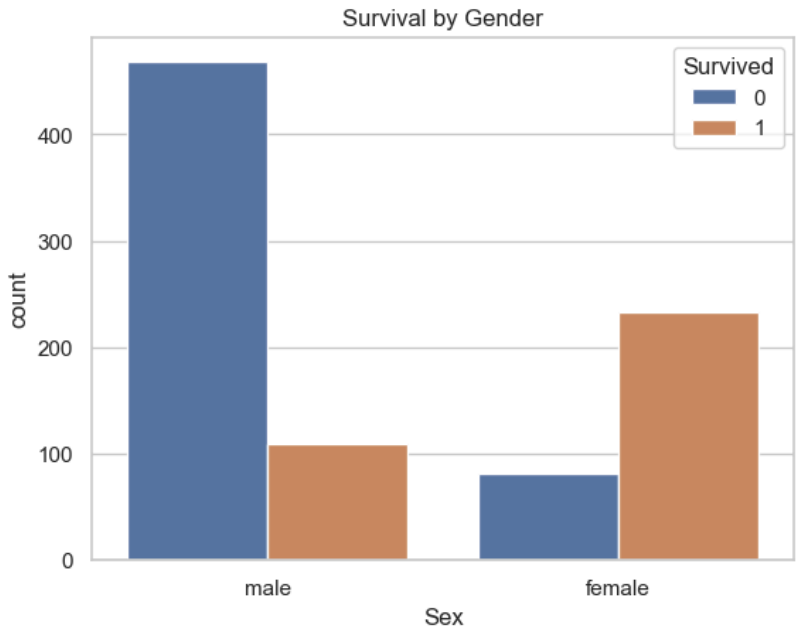
```
PassengerId          0
```

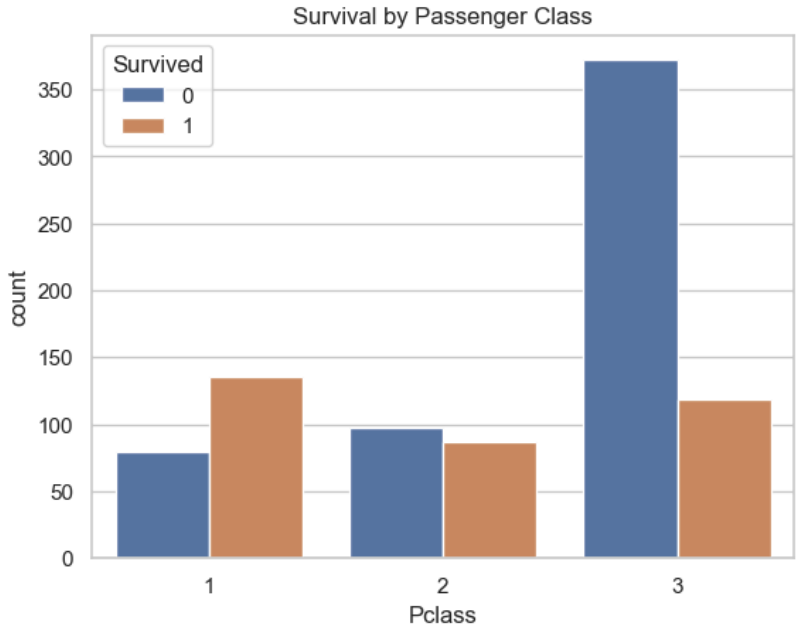
```
Survived      0
Pclass        0
Name          0
Sex           0
Age         177
SibSp         0
Parch         0
Ticket        0
Fare          0
Cabin       687
Embarked      2
dtype: int64
C:\Users\admin\AppData\Local\Temp\ipykernel_17372\1076483640.py:25: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.


  df['Age'].fillna(df['Age'].median(), inplace=True)
C:\Users\admin\AppData\Local\Temp\ipykernel_17372\1076483640.py:26: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.


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

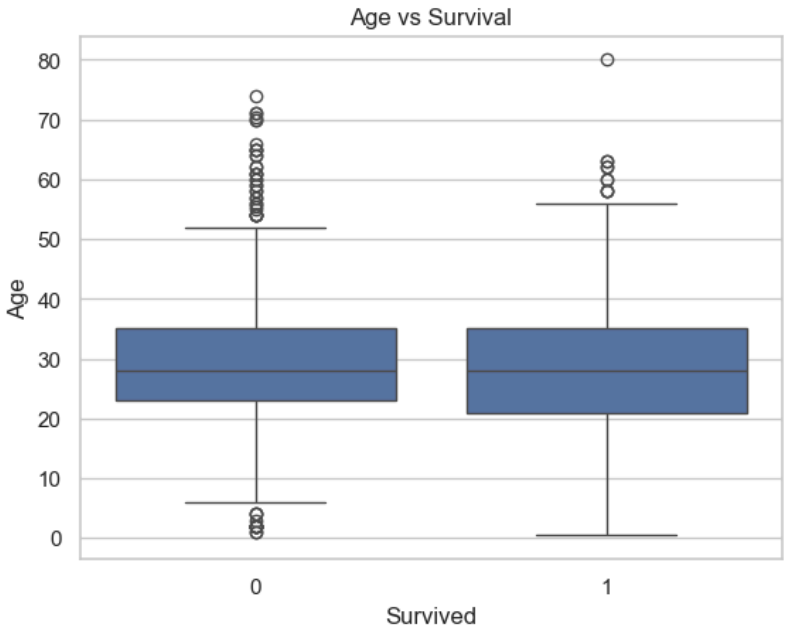


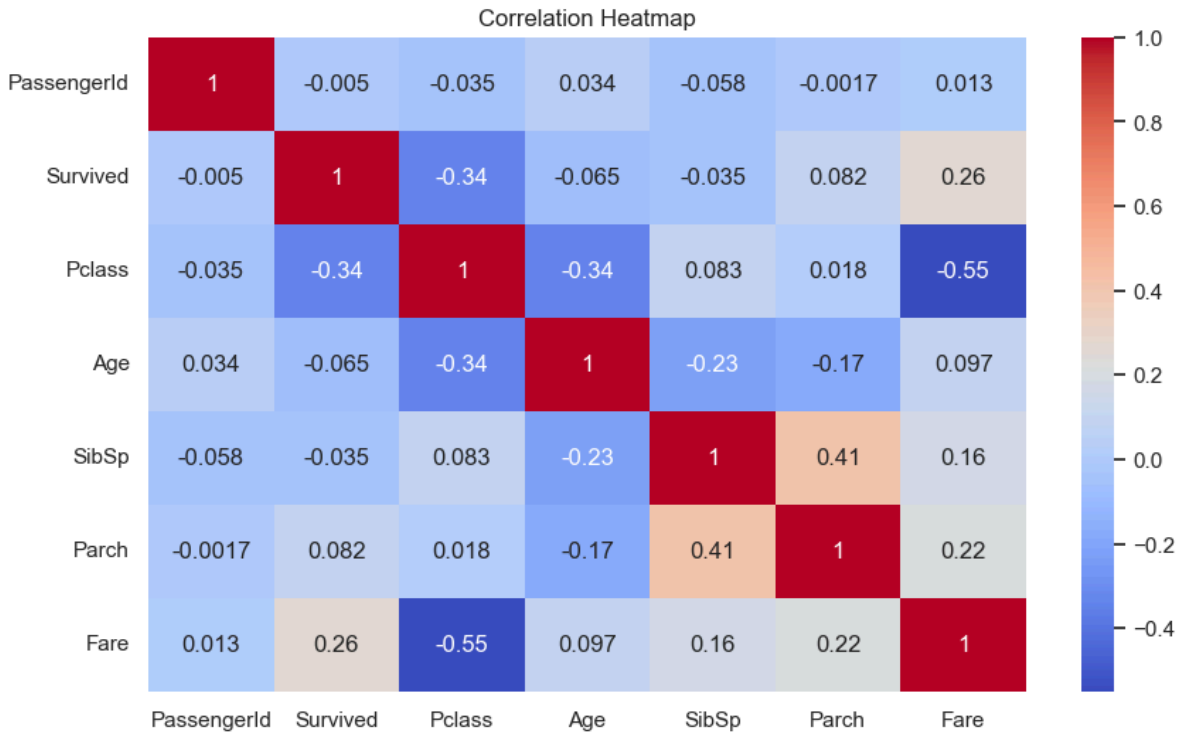

Fare Distribution
Count
Fare
0
50
100
150
200
250
300
0
100
200
300
400
500

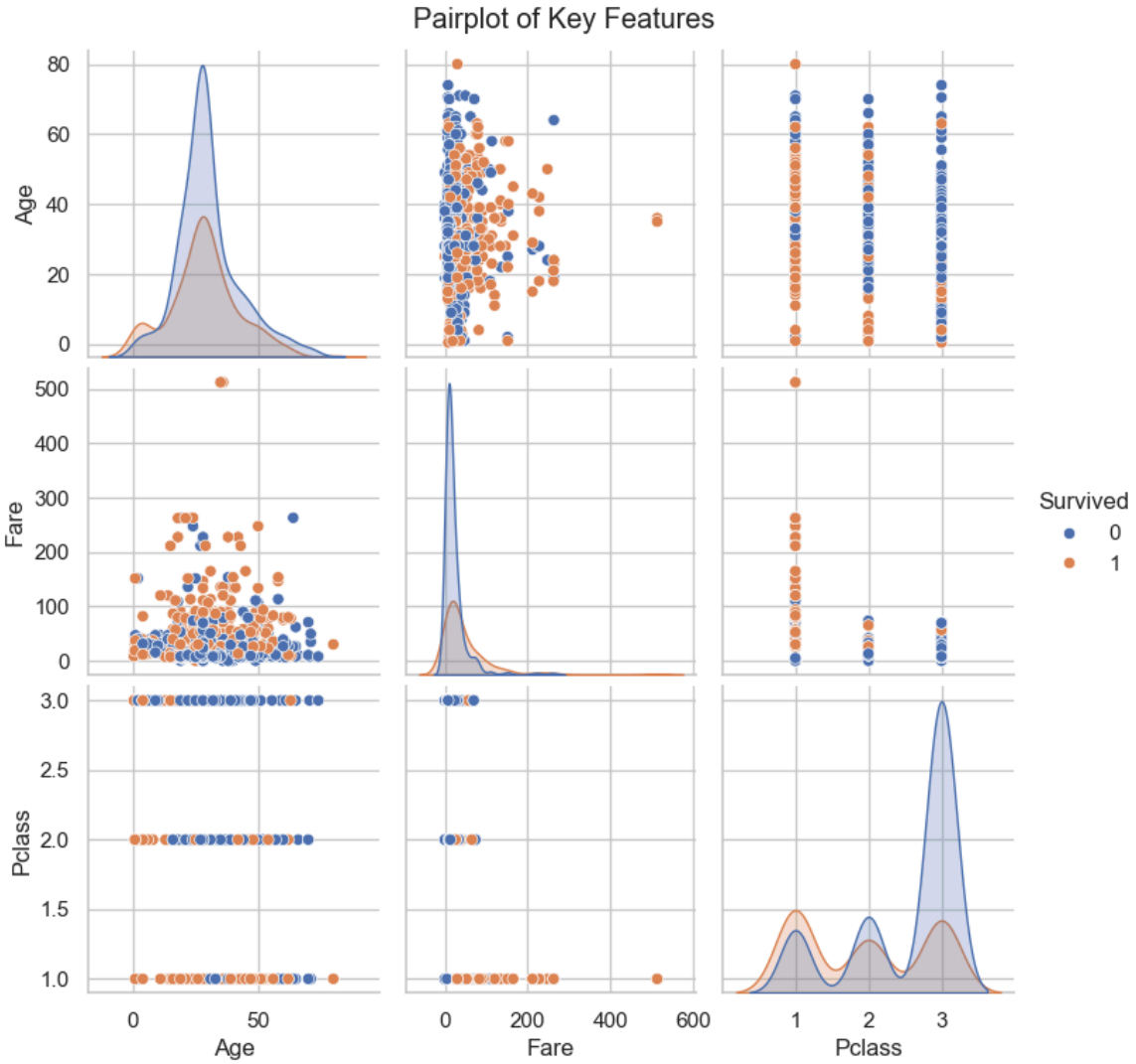

Passenger Class Count
count
Pclass
0
100
200
300
400
500
1
2
3


Survival by Gender
Survived
0
1
count
Sex
0
100
200
300
400
male
female


Survival by Passenger Class
Survived
0
1
count
Pclass
1
2
3


Age vs Survival
Age
Survived
0
1


Correlation Heatmap
PassengerId
Survived
Pclass
Age
SibSp
Parch
Fare



✅ Summary of Findings:
- Women had a much higher survival rate than men.
- First-class passengers were more likely to survive.
- Age and Fare have a mild correlation with survival.
- Strong correlation found between Pclass and Fare.