titanic_eda_visualizations

October 3, 2024

1 Titanic Dataset - Exploratory Data Analysis (EDA) - Visualizations

[28]:
```
# Required Libraries
import warnings
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# # Suppress FutureWarnings
warnings.simplefilter(action='ignore', category=FutureWarning)

# Load the dataset (assuming it's stored locally as 'titanic.csv')
df = pd.read_csv('data/titanic.csv',index_col='PassengerId')

# Display first few rows to verify data
df.head()
```

[28]:

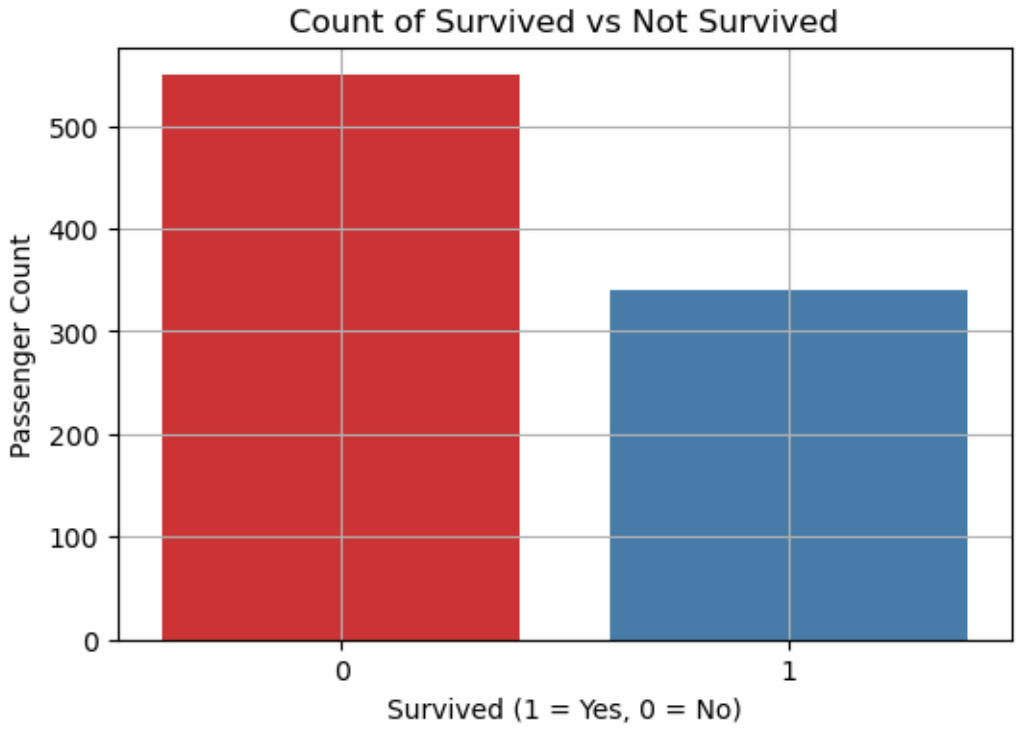
PassengerId	Survived	Pclass	Name	Sex	Age	Sibling_Spouse	Parents_Child	Ticket	Fare_Npr
1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	1290.5000
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th…	female	38.0				
3	1	3	Heikkinen, Miss. Laina	female	26.0				
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0				
5	0	3	Allen, Mr. William Henry	male	35.0				

```
2                         1             0          PC 17599  12688.4274
3                         0             0  STON/O2. 3101282   1410.6500
4                         1             0            113803   9451.8000
5                         0             0            373450   1432.9000

                    Port Age_Category
PassengerId
1            Southampton        Adult
2              Cherbourg        Adult
3            Southampton        Adult
4            Southampton        Adult
5            Southampton        Adult
```

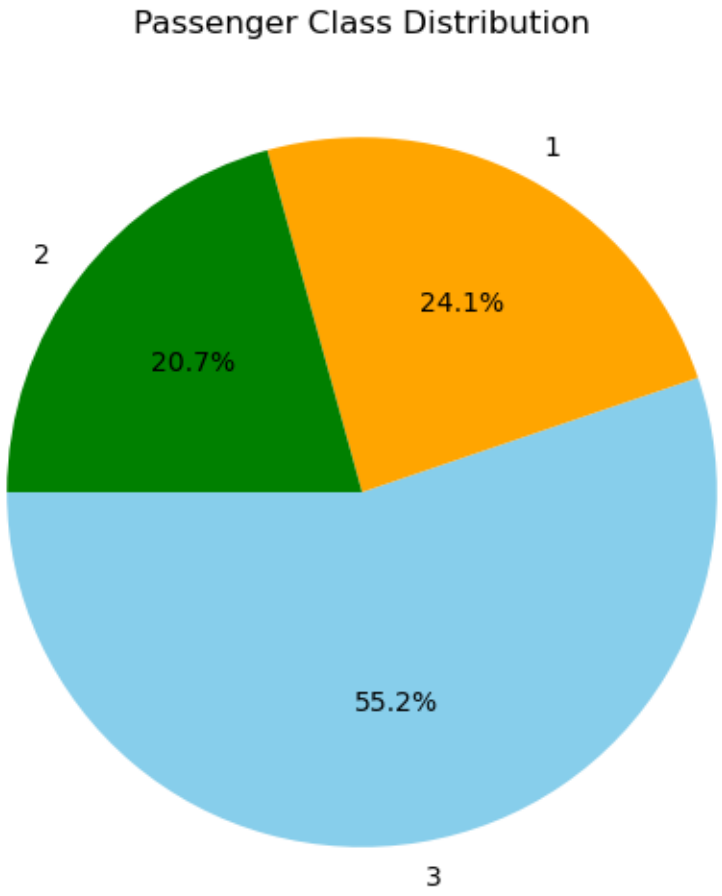
1.1 1. Bar Plot for Survival Count

[45]:
```
# Bar plot for Survived column
plt.figure(figsize=(6,4))
sns.countplot(x='Survived', data=df, hue='Survived', palette='Set1',␣
 ↪legend=False)
plt.title('Count of Survived vs Not Survived')
plt.xlabel('Survived (1 = Yes, 0 = No)')
plt.ylabel('Passenger Count')
plt.grid(True)
plt.show()
```



1.2 2. Pie Chart for Passenger Class Distribution

[51]:
```
# Pie chart for Pclass
pclass_counts = df['Pclass'].value_counts()
plt.figure(figsize=(6,6))
plt.pie(pclass_counts, labels=pclass_counts.index, autopct='%1.1f%%',␣
 ↪colors=['skyblue', 'orange', 'green'], startangle=180)
plt.title('Passenger Class Distribution')
plt.show()
```



[46]:
```
df['Pclass'].value_counts()
```

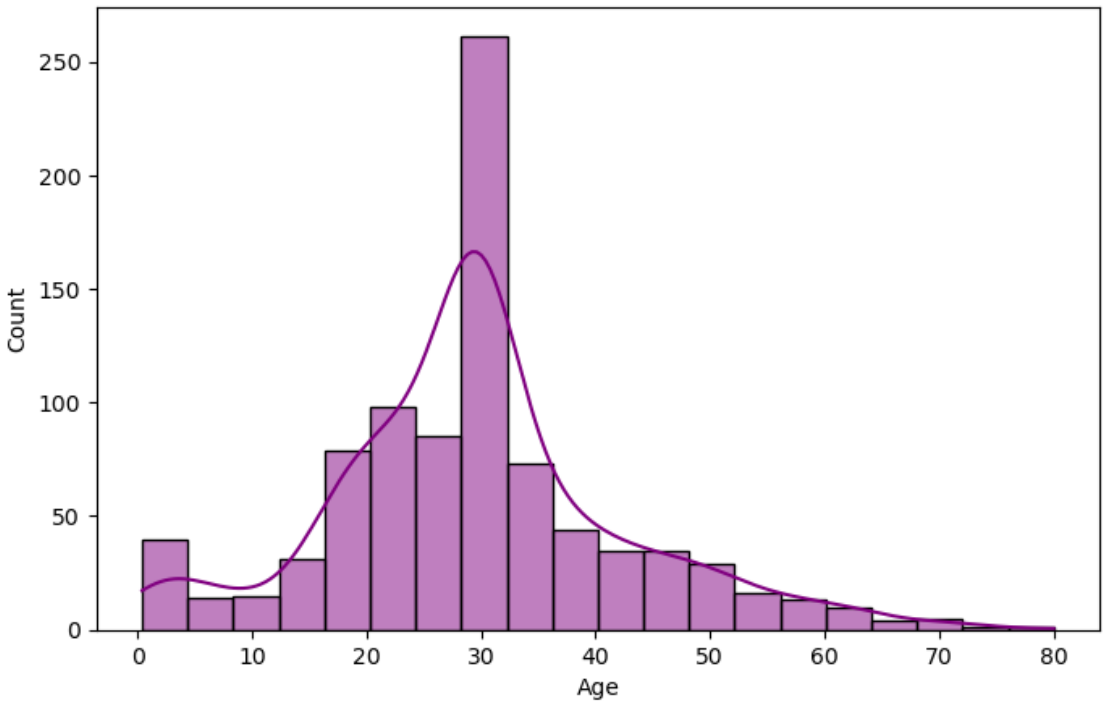
[46]:
```
Pclass
3    491
1    214
```

```
2    184
Name: count, dtype: int64
```

1.3 3. Histogram for Age Distribution

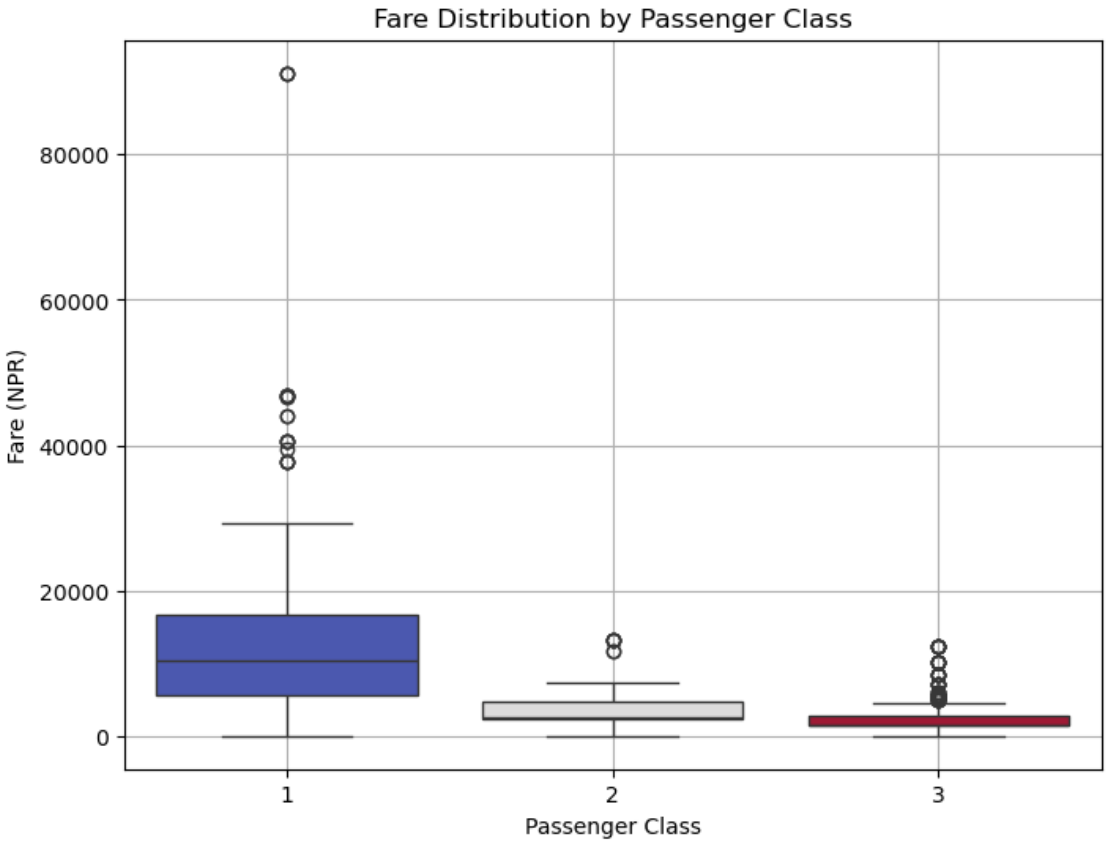
[31]:
```
# Histogram for Age
plt.figure(figsize=(8,5))
sns.histplot(df['Age'].dropna(), bins=20, kde=True, color='purple')
```

[31]:
```
<Axes: xlabel='Age', ylabel='Count'>
```



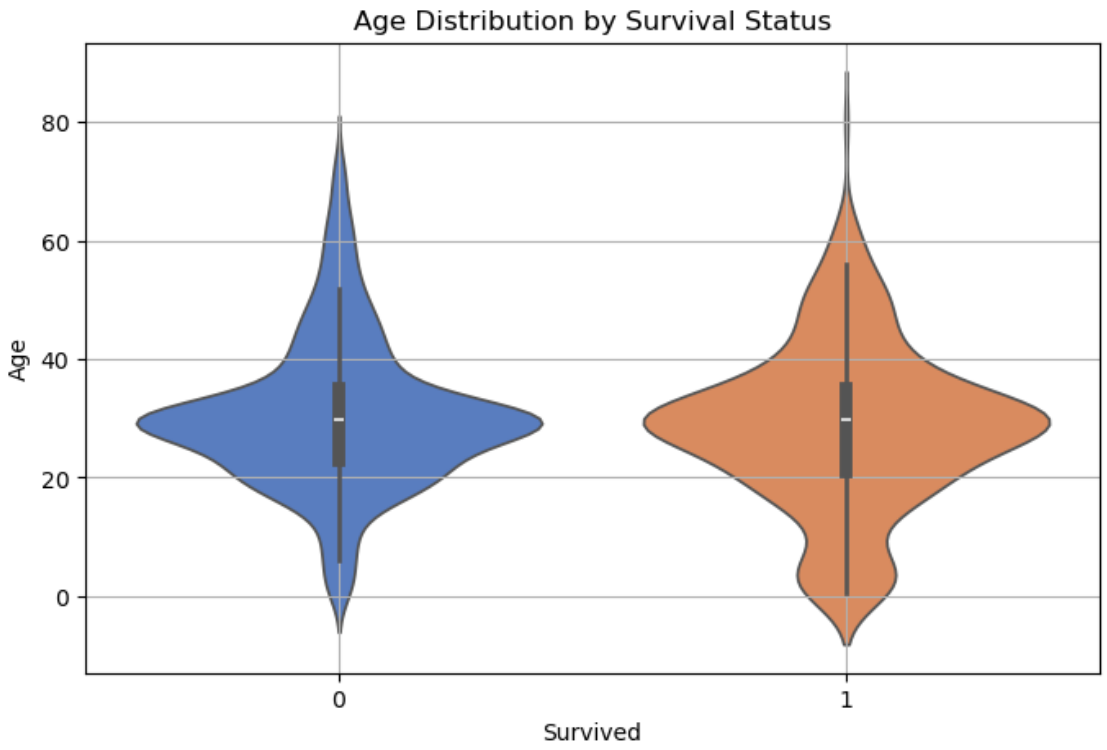
1.4 4. Box Plot for Fare Distribution by Passenger Class

[32]:
```
# Box plot for Fare by Pclass
plt.figure(figsize=(8,6))
sns.boxplot(x='Pclass', y='Fare_Npr', data=df, hue='Pclass',␣
 ↪palette='coolwarm', legend=False)
plt.title('Fare Distribution by Passenger Class')
plt.xlabel('Passenger Class')
plt.ylabel('Fare (NPR)')
plt.grid(True)
plt.show()
```



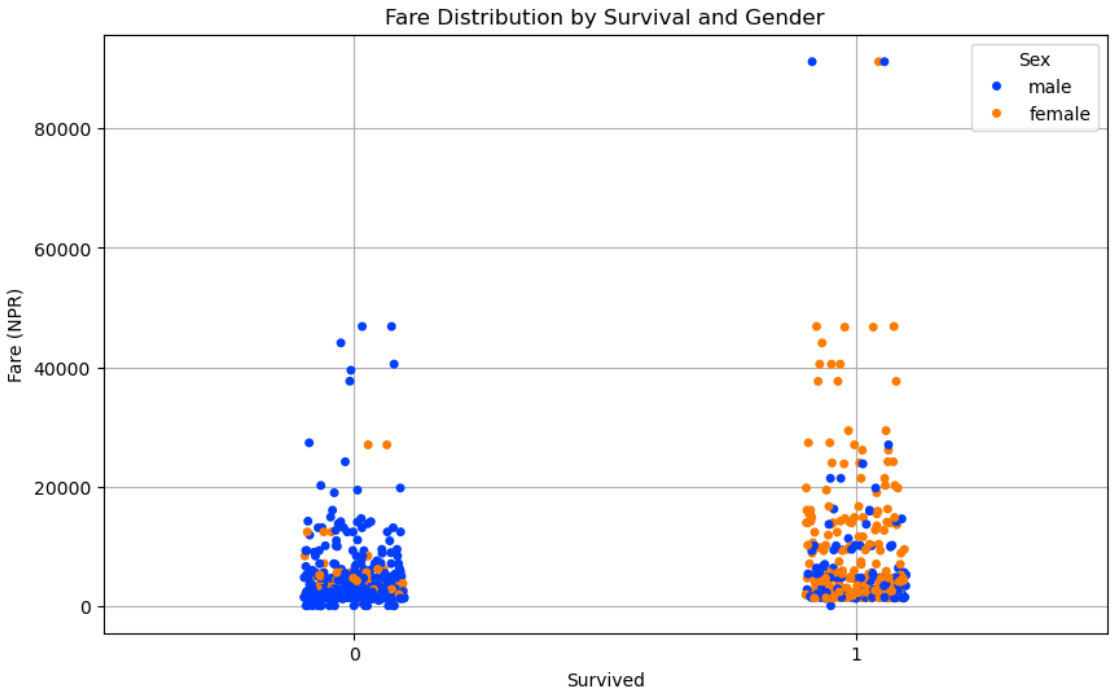
1.5 5. Violin Plot for Age Distribution by Survival Status

[33]:
```
# Violin plot for Age distribution by Survival
plt.figure(figsize=(8,5))
sns.violinplot(x='Survived', y='Age', data=df,hue='Survived',␣
↪palette='muted',legend=False)
plt.title('Age Distribution by Survival Status')
plt.xlabel('Survived')
plt.ylabel('Age')
plt.grid(True)
plt.show()
```



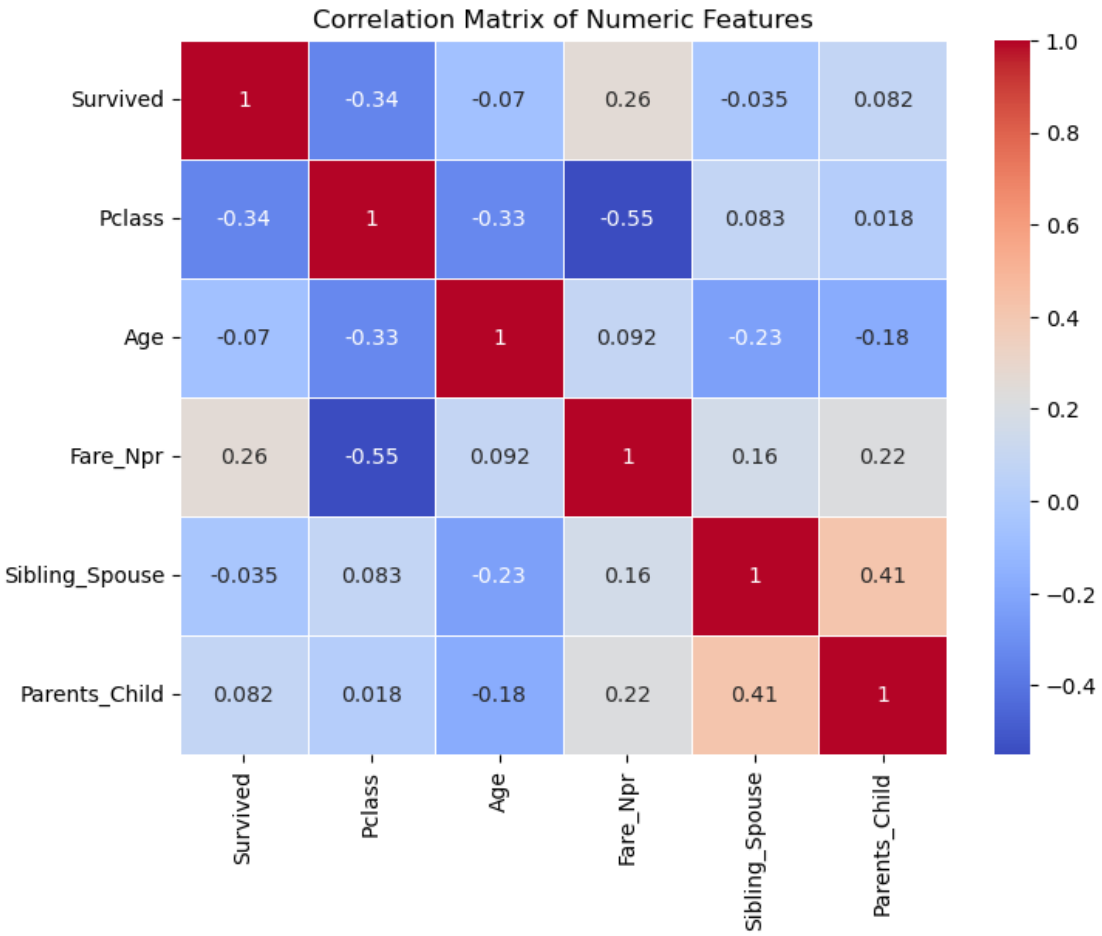
1.6 6. Swarm Plot for Fare by Survival Status and Gender

[58]:
```
# Swarm plot for Fare by Survived and Sex
plt.figure(figsize=(10,6))
sns.stripplot(x='Survived', y='Fare_Npr', hue='Sex', data=df, palette='bright')
plt.title('Fare Distribution by Survival and Gender')
plt.xlabel('Survived')
plt.ylabel('Fare (NPR)')
plt.legend(title='Sex')
plt.grid(True)
plt.show()
```



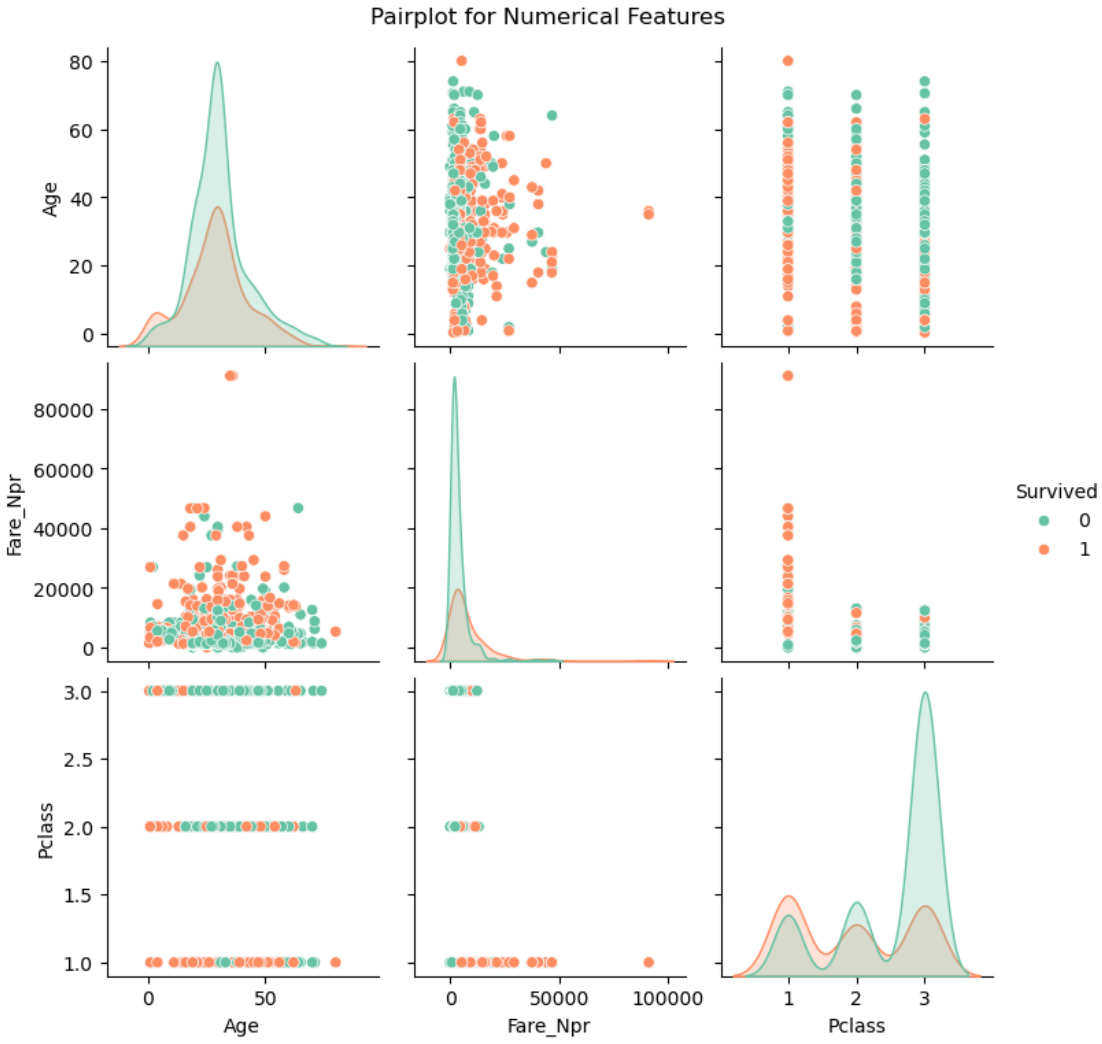
1.7 7. Heatmap for Correlation Matrix

[19]:
```
# Correlation heatmap
plt.figure(figsize=(8,6))
corr_matrix = df[['Survived', 'Pclass', 'Age', 'Fare_Npr', 'Sibling_Spouse',␣
 ↪'Parents_Child']].corr()
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', linewidths=0.5)
plt.title('Correlation Matrix of Numeric Features')
plt.show()
```



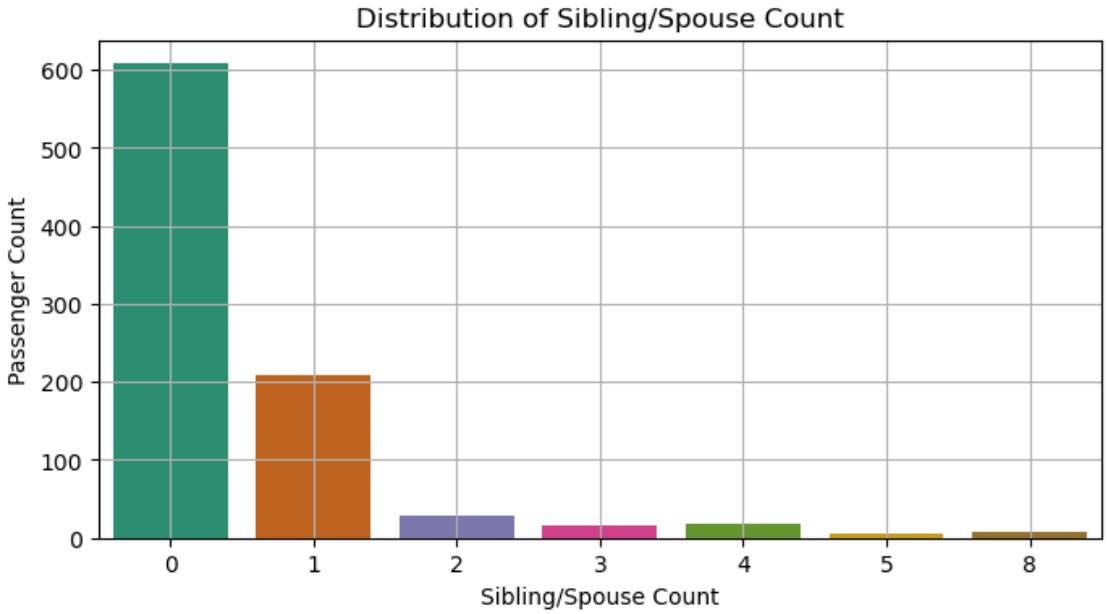
1.8 8. Pairplot for Visualizing Relationships

[23]:
```
# Pairplot for numerical columns
sns.pairplot(df[['Survived', 'Age', 'Fare_Npr', 'Pclass']], hue='Survived',␣
↪palette='Set2')
plt.suptitle('Pairplot for Numerical Features', y=1.02)
plt.show()
```



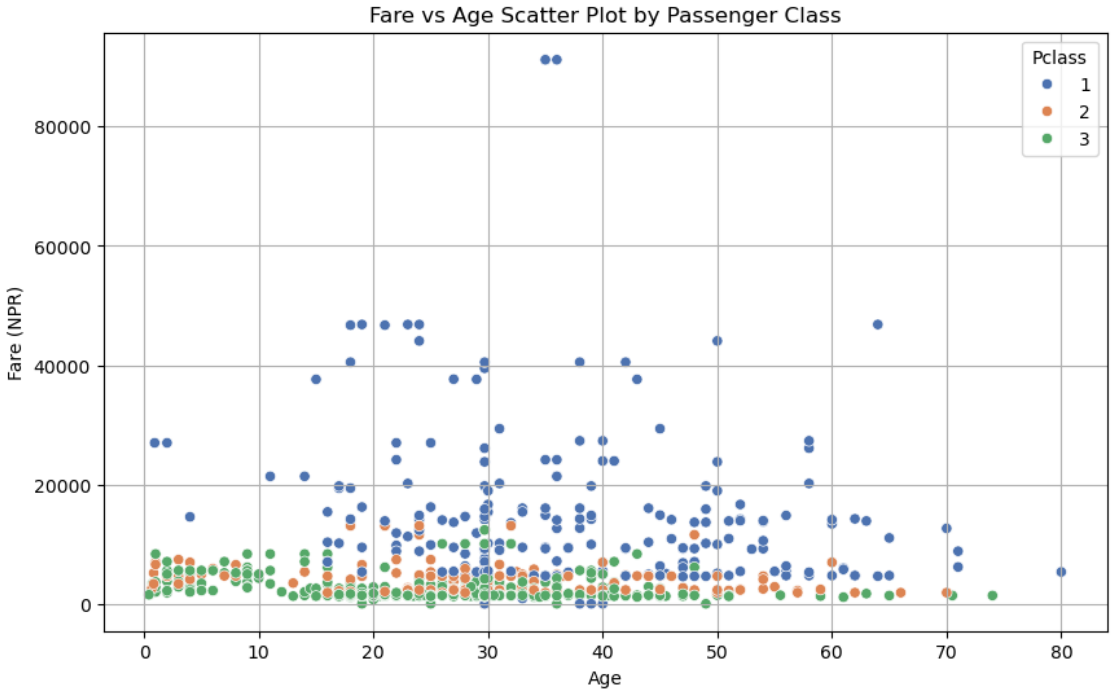
1.9 9. Bar Plot for Sibling/Spouse Count Distribution

[24]:
```
# Bar plot for Sibling_Spouse
plt.figure(figsize=(8,4))
sns.countplot(x='Sibling_Spouse', data=df, palette='Dark2')
plt.title('Distribution of Sibling/Spouse Count')
plt.xlabel('Sibling/Spouse Count')
plt.ylabel('Passenger Count')
plt.grid(True)
plt.show()
```



1.10 10. Scatter Plot for Fare vs Age Colored by Class

[25]:
```
# Scatter plot for Fare vs Age
plt.figure(figsize=(10,6))
sns.scatterplot(x='Age', y='Fare_Npr', hue='Pclass', data=df, palette='deep')
plt.title('Fare vs Age Scatter Plot by Passenger Class')
plt.xlabel('Age')
plt.ylabel('Fare (NPR)')
plt.grid(True)
plt.show()
```


Fare vs Age Scatter Plot by Passenger Class
Pclass
1
2
3
Fare (NPR)
80000
60000
40000
20000
0
0
10
20
30
40
50
60
70
80
Age