In [1]:
```
pip install pandas
```

```
Requirement already satisfied: pandas in c:\users\aktha\appdata\local\programs\py
thon\python312\lib\site-packages (2.2.3)
Requirement already satisfied: numpy>=1.26.0 in c:\users\aktha\appdata\local\prog
rams\python\python312\lib\site-packages (from pandas) (2.2.6)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\aktha\appdata\l
ocal\programs\python\python312\lib\site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\aktha\appdata\local\progr
ams\python\python312\lib\site-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in c:\users\aktha\appdata\local\pro
grams\python\python312\lib\site-packages (from pandas) (2025.2)
Requirement already satisfied: six>=1.5 in c:\users\aktha\appdata\local\programs
\python\python312\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.17.
0)
Note: you may need to restart the kernel to use updated packages.
[notice] A new release of pip is available: 24.2 -> 25.1.1
[notice] To update, run: C:\Users\aktha\AppData\Local\Programs\Python\Python312\p
ython.exe -m pip install --upgrade pip
```

In [2]:
```
pip install numpy
```

```
Requirement already satisfied: numpy in c:\users\aktha\appdata\local\programs\pyt
hon\python312\lib\site-packages (2.2.6)
Note: you may need to restart the kernel to use updated packages.
[notice] A new release of pip is available: 24.2 -> 25.1.1
[notice] To update, run: C:\Users\aktha\AppData\Local\Programs\Python\Python312\p
ython.exe -m pip install --upgrade pip
```

In [3]:
```
pip install matplotlib
```

```
Requirement already satisfied: matplotlib in c:\users\aktha\appdata\local\program
s\python\python312\lib\site-packages (3.10.3)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\aktha\appdata\local\p
rograms\python\python312\lib\site-packages (from matplotlib) (1.3.2)
Requirement already satisfied: cycler>=0.10 in c:\users\aktha\appdata\local\progr
ams\python\python312\lib\site-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\aktha\appdata\local
\programs\python\python312\lib\site-packages (from matplotlib) (4.58.0)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\aktha\appdata\local
\programs\python\python312\lib\site-packages (from matplotlib) (1.4.8)
Requirement already satisfied: numpy>=1.23 in c:\users\aktha\appdata\local\progra
ms\python\python312\lib\site-packages (from matplotlib) (2.2.6)
Requirement already satisfied: packaging>=20.0 in c:\users\aktha\appdata\local\pr
ograms\python\python312\lib\site-packages (from matplotlib) (25.0)
Requirement already satisfied: pillow>=8 in c:\users\aktha\appdata\local\programs
\python\python312\lib\site-packages (from matplotlib) (11.2.1)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\aktha\appdata\local\p
rograms\python\python312\lib\site-packages (from matplotlib) (3.2.3)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\aktha\appdata\loc
al\programs\python\python312\lib\site-packages (from matplotlib) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in c:\users\aktha\appdata\local\programs
\python\python312\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.1
7.0)
Note: you may need to restart the kernel to use updated packages.
[notice] A new release of pip is available: 24.2 -> 25.1.1
[notice] To update, run: C:\Users\aktha\AppData\Local\Programs\Python\Python312\p
ython.exe -m pip install --upgrade pip
```

In [4]:
```
pip install seaborn
```

```
Requirement already satisfied: seaborn in c:\users\aktha\appdata\local\programs\p
ython\python312\lib\site-packages (0.13.2)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in c:\users\aktha\appdata\loc
al\programs\python\python312\lib\site-packages (from seaborn) (2.2.6)
Requirement already satisfied: pandas>=1.2 in c:\users\aktha\appdata\local\progra
ms\python\python312\lib\site-packages (from seaborn) (2.2.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in c:\users\aktha\appdata
\local\programs\python\python312\lib\site-packages (from seaborn) (3.10.3)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\aktha\appdata\local\p
rograms\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seabor
n) (1.3.2)
Requirement already satisfied: cycler>=0.10 in c:\users\aktha\appdata\local\progr
ams\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn)
(0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\aktha\appdata\local
\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seabo
rn) (4.58.0)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\aktha\appdata\local
\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seabo
rn) (1.4.8)
Requirement already satisfied: packaging>=20.0 in c:\users\aktha\appdata\local\pr
ograms\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn)
(25.0)
Requirement already satisfied: pillow>=8 in c:\users\aktha\appdata\local\programs
\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (11.
2.1)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\aktha\appdata\local\p
rograms\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seabor
n) (3.2.3)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\aktha\appdata\loc
al\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->sea
born) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\aktha\appdata\local\progr
ams\python\python312\lib\site-packages (from pandas>=1.2->seaborn) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in c:\users\aktha\appdata\local\pro
grams\python\python312\lib\site-packages (from pandas>=1.2->seaborn) (2025.2)
Requirement already satisfied: six>=1.5 in c:\users\aktha\appdata\local\programs
\python\python312\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.
1,>=3.4->seaborn) (1.17.0)
Note: you may need to restart the kernel to use updated packages.
[notice] A new release of pip is available: 24.2 -> 25.1.1
[notice] To update, run: C:\Users\aktha\AppData\Local\Programs\Python\Python312\p
ython.exe -m pip install --upgrade pip
```

In [5]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

sns.set(style='whitegrid')
%matplotlib inline
```

In [6]:
```
df = pd.read_csv("train.csv")
df.head()
```

Out[6]:

	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	7.2
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0

In [11]:
```
# Shape of the dataset
print("Shape of dataset:", df.shape)

# Info about columns, types, nulls
df.info()

# Statistical summary of numerical columns
df.describe()

# Count unique values in 'Survived' column
df['Survived'].value_counts()
```

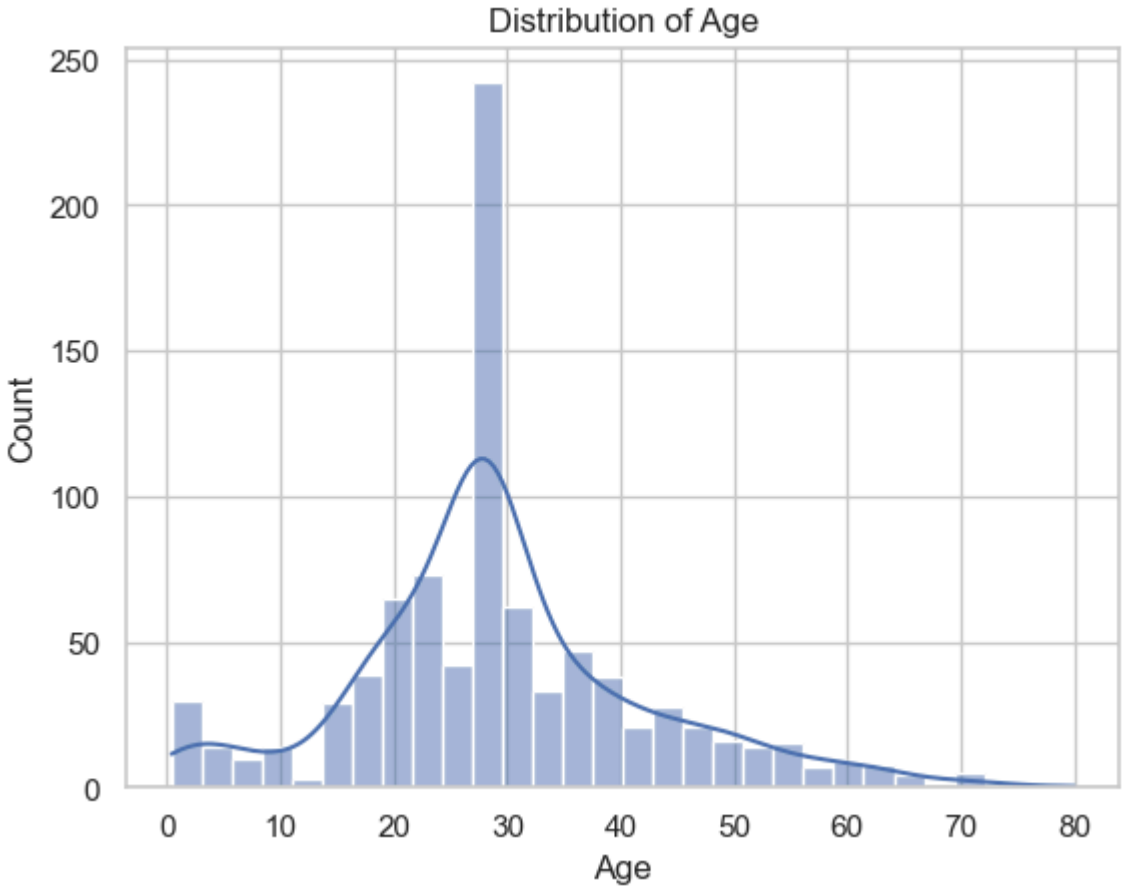
```
Shape of dataset: (891, 11)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 11 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          891 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  Embarked     891 non-null    object 
dtypes: float64(2), int64(5), object(4)
memory usage: 76.7+ KB
```

Out[11]:
```
Survived
0    549
1    342
Name: count, dtype: int64
```

HISTOGRAM (AGE)

In [15]:
```
# Histogram of Age
sns.histplot(df['Age'].dropna(), bins=30, kde=True)
plt.title('Distribution of Age')
plt.xlabel('Age')
```

Out[15]:
```
Text(0.5, 0, 'Age')
```

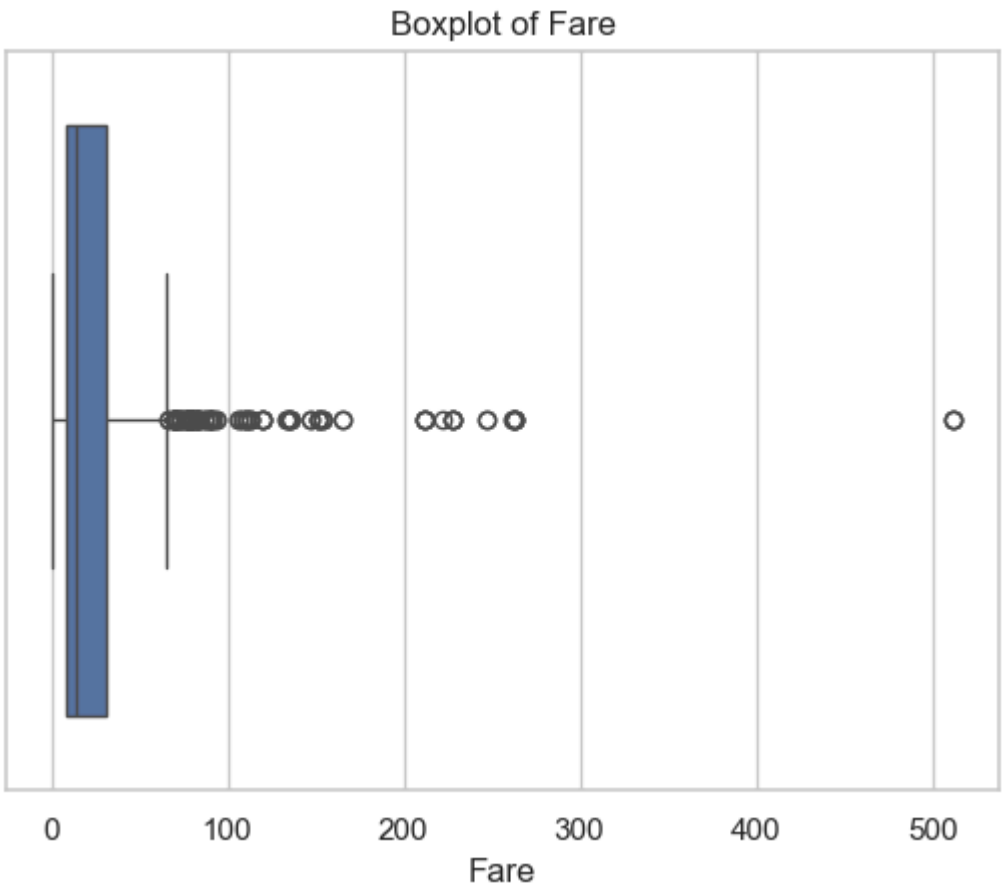


From the histogram of Age, we observe that most passengers are between 20 and 40 years old. Very few are above 60, and there are children as young as 0–10 years. This shows the majority of passengers were young adults.

BOXLOT (FARE)

In [13]:
```
# Boxplot of Fare
sns.boxplot(x=df['Fare'])
plt.title('Boxplot of Fare')
```

Out[13]:
```
Text(0.5, 1.0, 'Boxplot of Fare')
```

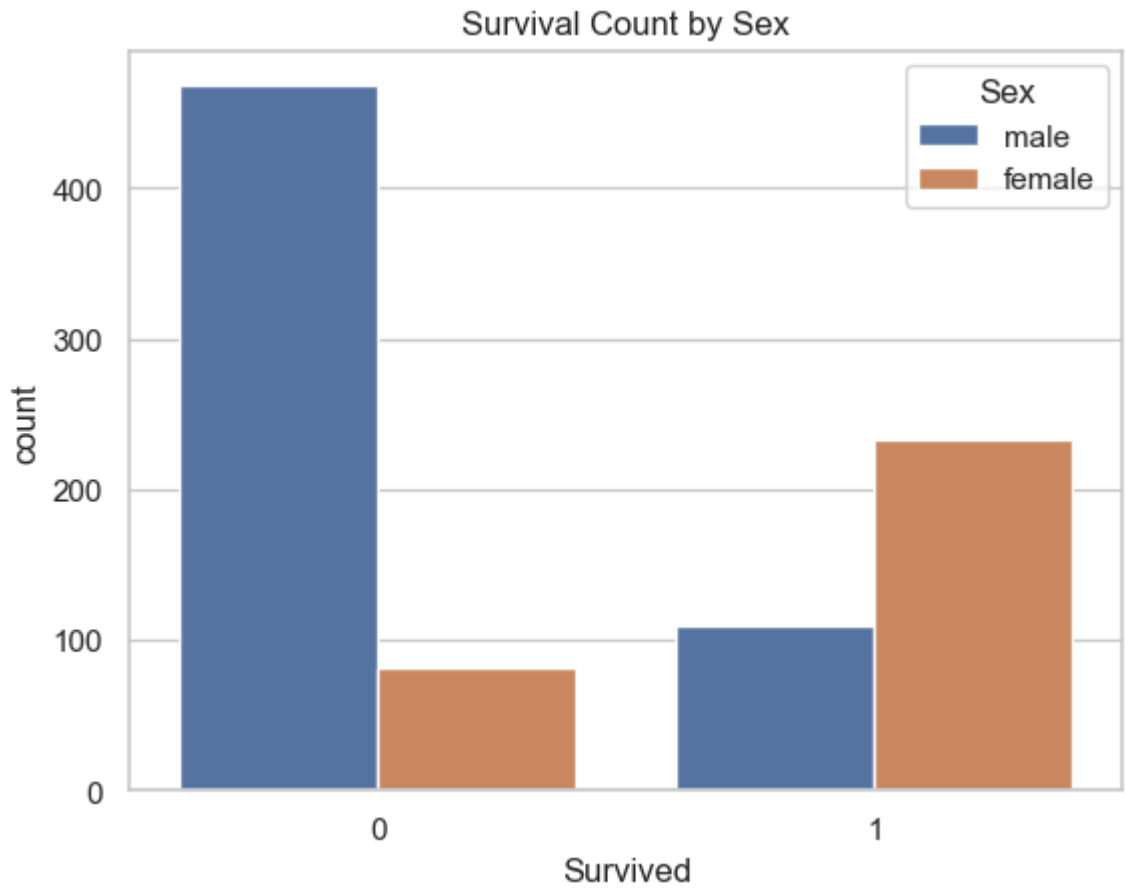


The boxplot of Fare shows that there are multiple outliers where passengers paid significantly higher fares. Most fares lie below $100, with a few going upto 500+$. This indicates some passengers were traveling in luxury classes.

COUNTPLOT(SURVIVED VS SEX)

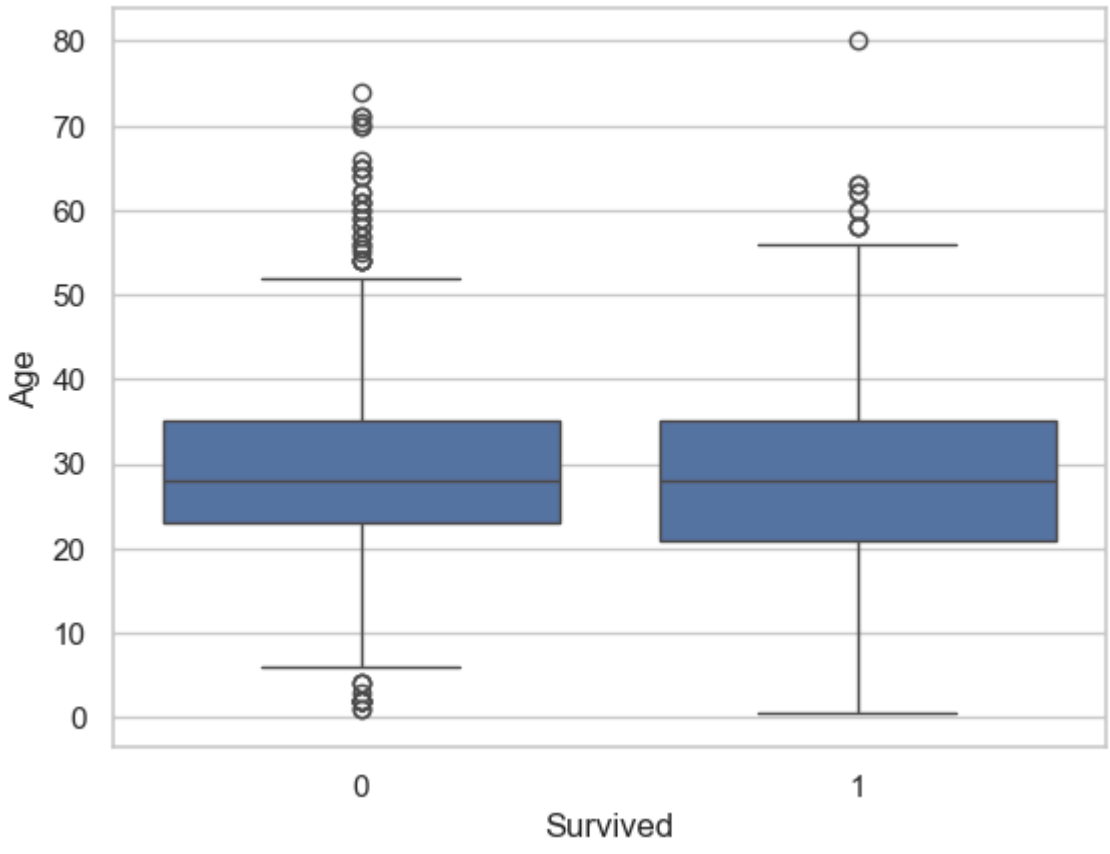
In [16]:
```
# Count plot of Survived vs Sex
sns.countplot(x='Survived', hue='Sex', data=df)
plt.title('Survival Count by Sex')
```

Out[16]:
```
Text(0.5, 1.0, 'Survival Count by Sex')
```



In [17]:
```
# Boxplot of Age vs Survived
sns.boxplot(x='Survived', y='Age', data=df)
```

Out[17]: <Axes: xlabel='Survived', ylabel='Age'>



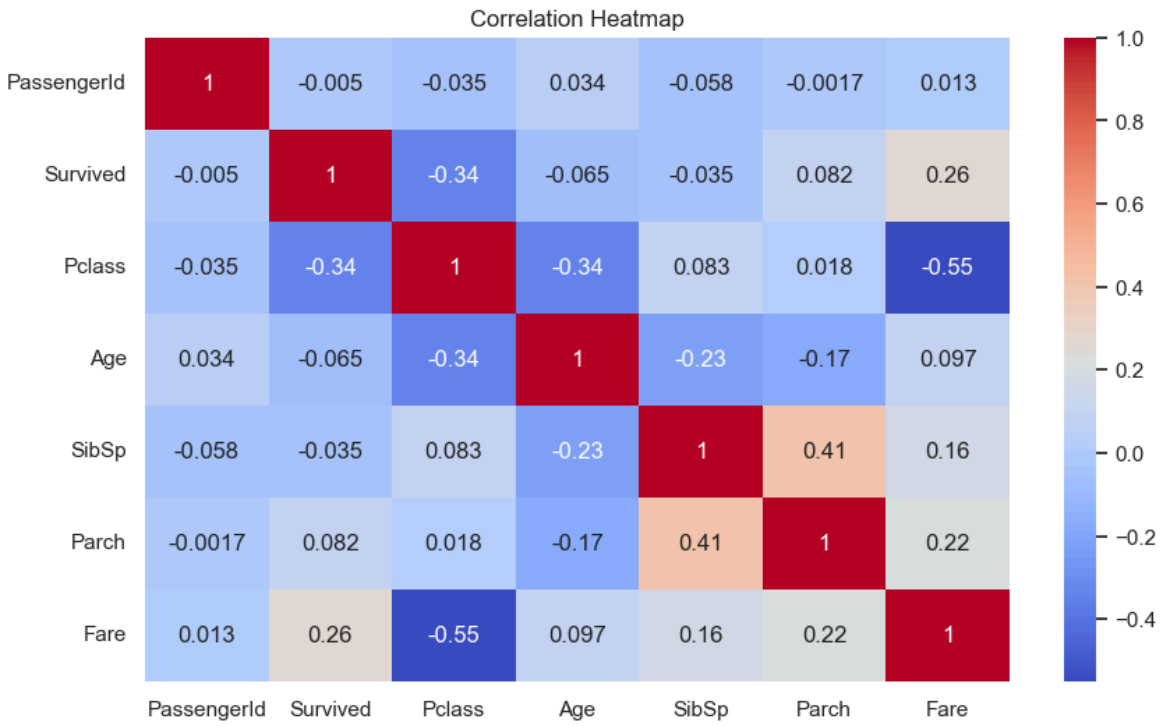
More women survived compared to men. This suggests that women may have been given priority during rescue, following the "women and children first" principle.

HEATMAP(CORRELATION)

In [18]:
```
# Correlation matrix
corr = df.corr(numeric_only=True)

# Heatmap
plt.figure(figsize=(10,6))
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
```

Out[18]:
```
Text(0.5, 1.0, 'Correlation Heatmap')
```

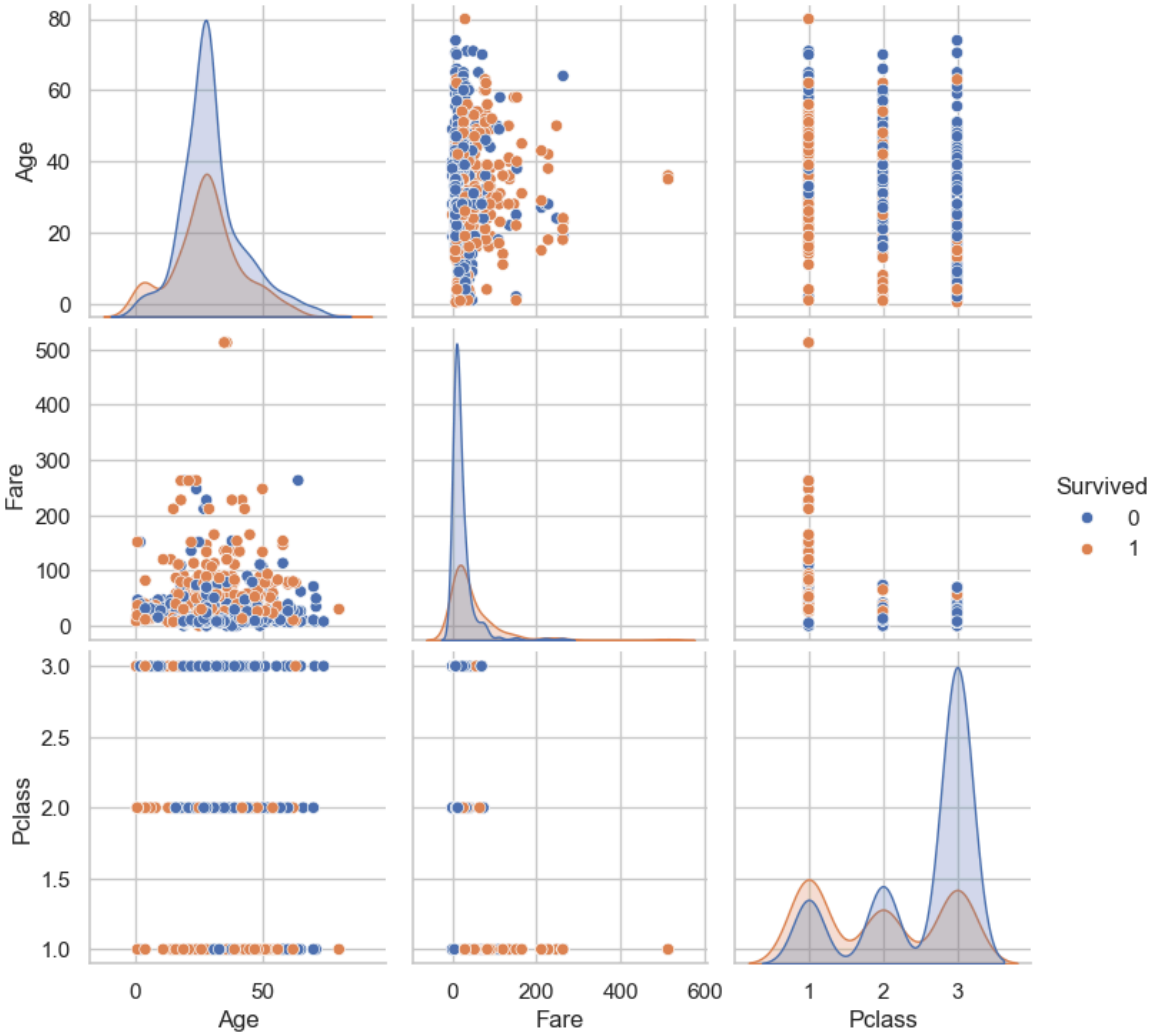


There is a negative correlation between Pclass and Fare, which means higher-class passengers (Pclass = 1) paid higher fares. Survived has a positive correlation with Fare and negative with Pclass.

PAIRPLOT

In [19]:
```
# Pairplot
sns.pairplot(df[['Survived', 'Age', 'Fare', 'Pclass']], hue='Survived')
```

Out[19]:
```
<seaborn.axisgrid.PairGrid at 0x16a00f4bce0>
```



The pairplot shows that survivors are mostly clustered in lower age groups with higher fares. This indicates younger passengers who paid more (likely 1st class) had a higher survival rate.

SUMMARY

- The dataset contains 891 rows and 12 columns with both numerical and categorical features.
- The 'Age' and 'Cabin' columns have missing values, with 'Cabin' having over 75% null values.
- Most passengers are between 20 to 40 years old.
- Females had a much higher survival rate than males, likely due to rescue priorities.
- Passengers in 1st class had a higher chance of survival than those in 2nd or 3rd class.
- Fare values show some high outliers, indicating a few passengers paid significantly more.
- There is a positive correlation between Fare and Survival, and a negative correlation between Pclass and Survival.
- No strong correlation between Age and Survival was observed.
- Overall, social class and gender played a major role in survival.