In [69]:
```
#Loading & Previewing the Data
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [71]:
```
df=pd.read_csv("D:/Purvava/Datasets/train.csv")
df.head()
```

Out[71]:

	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

In [73]:
```
# Data Summary & Initial Exploration
df.info()
df.describe()
df.isnull().sum()
df['Survived'].value_counts()
```

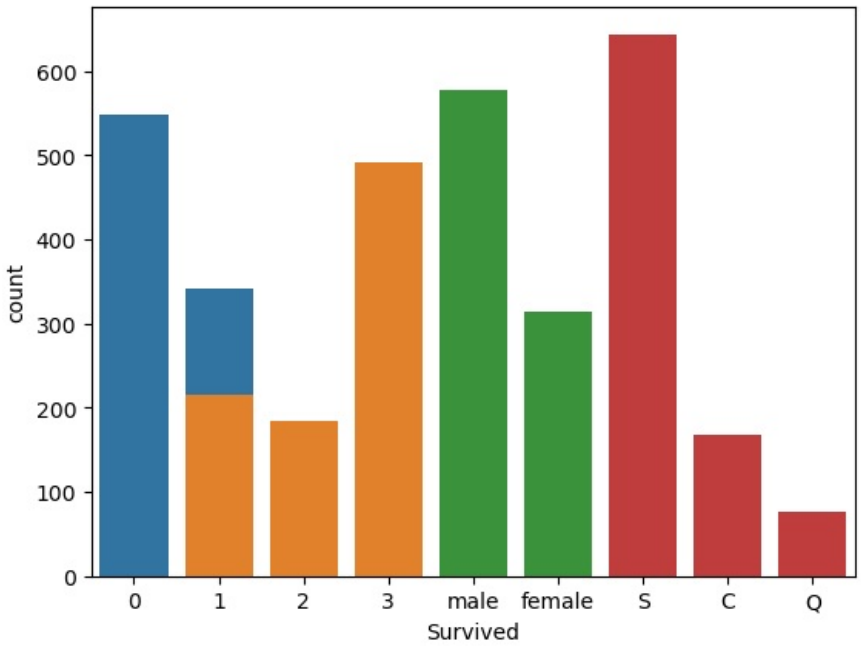
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    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
```

Out[73]:
```
Survived
0    549
1    342
Name: count, dtype: int64
```

In [75]:
```
# Missing values in Age, Cabin, Embarked
# Survived: 0 (not survived), 1 (survived)
```

In [77]:
```
# Univariate Analysis (Distribution of Individual Variables)

sns.countplot(x='Survived', data=df)
sns.countplot(x='Pclass', data=df)
sns.countplot(x='Sex', data=df)
sns.countplot(x='Embarked', data=df)
```

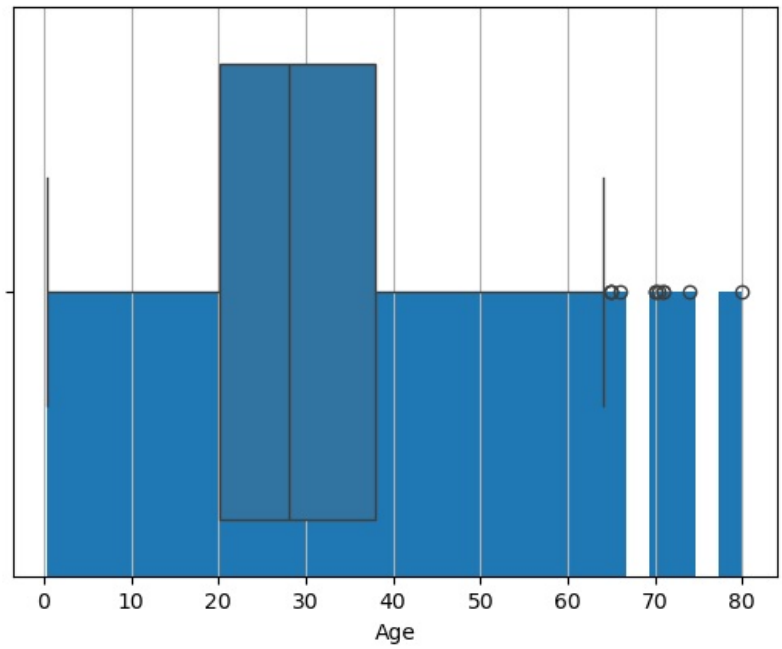
Out[77]:
```
<Axes: xlabel='Survived', ylabel='count'>
```



In [79]:
```
# Class 3 had more passengers
# Most survivors were in Class 1
# Males had higher mortality
```

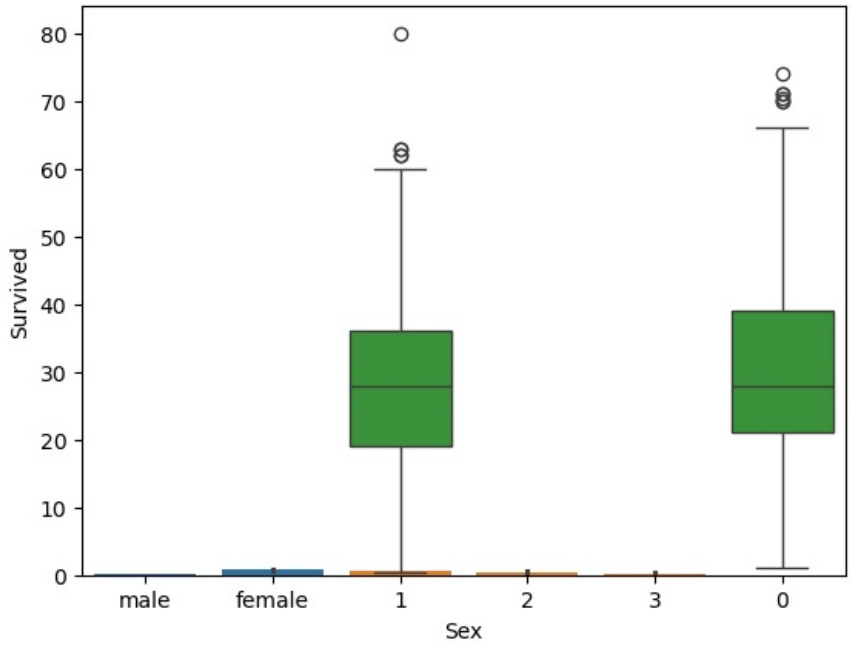
In [81]:
```
df['Age'].hist(bins=30)
sns.boxplot(x='Age', data=df)
```

Out[81]:
```
<Axes: xlabel='Age'>
```



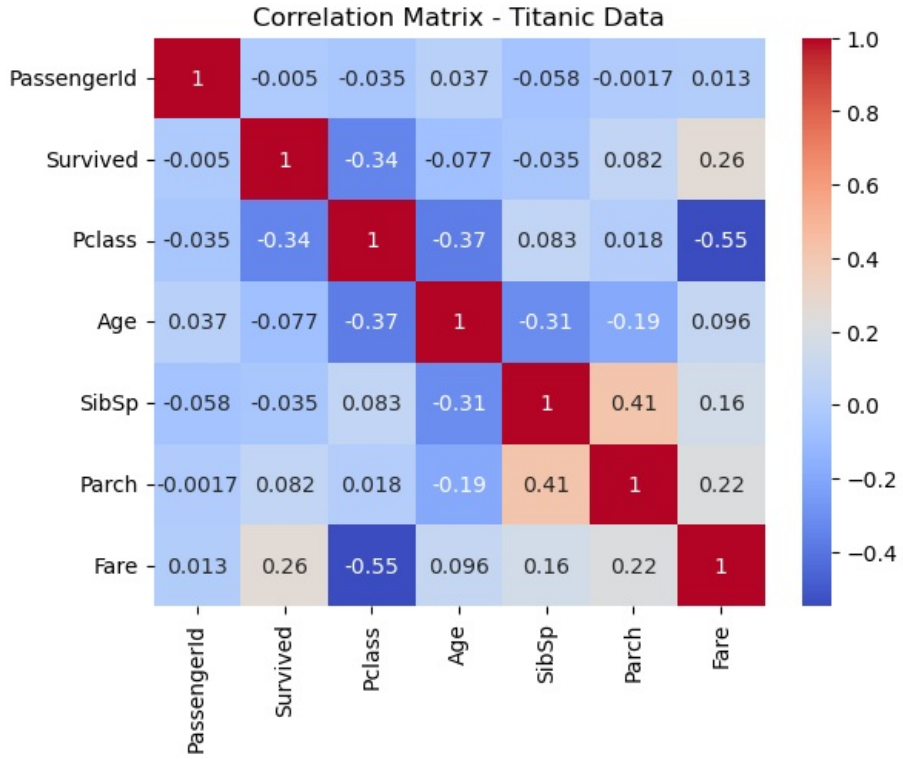
In [83]:
```
# Bivariate Analysis (Relations Between Variables)
sns.barplot(x='Sex', y='Survived', data=df)
sns.barplot(x='Pclass', y='Survived', data=df)
sns.boxplot(x='Survived', y='Age', data=df)
```

Out[83]: <Axes: xlabel='Sex', ylabel='Survived'>



In [85]:
```
# Females had higher survival rate
# Younger passengers were more likely to survive
```

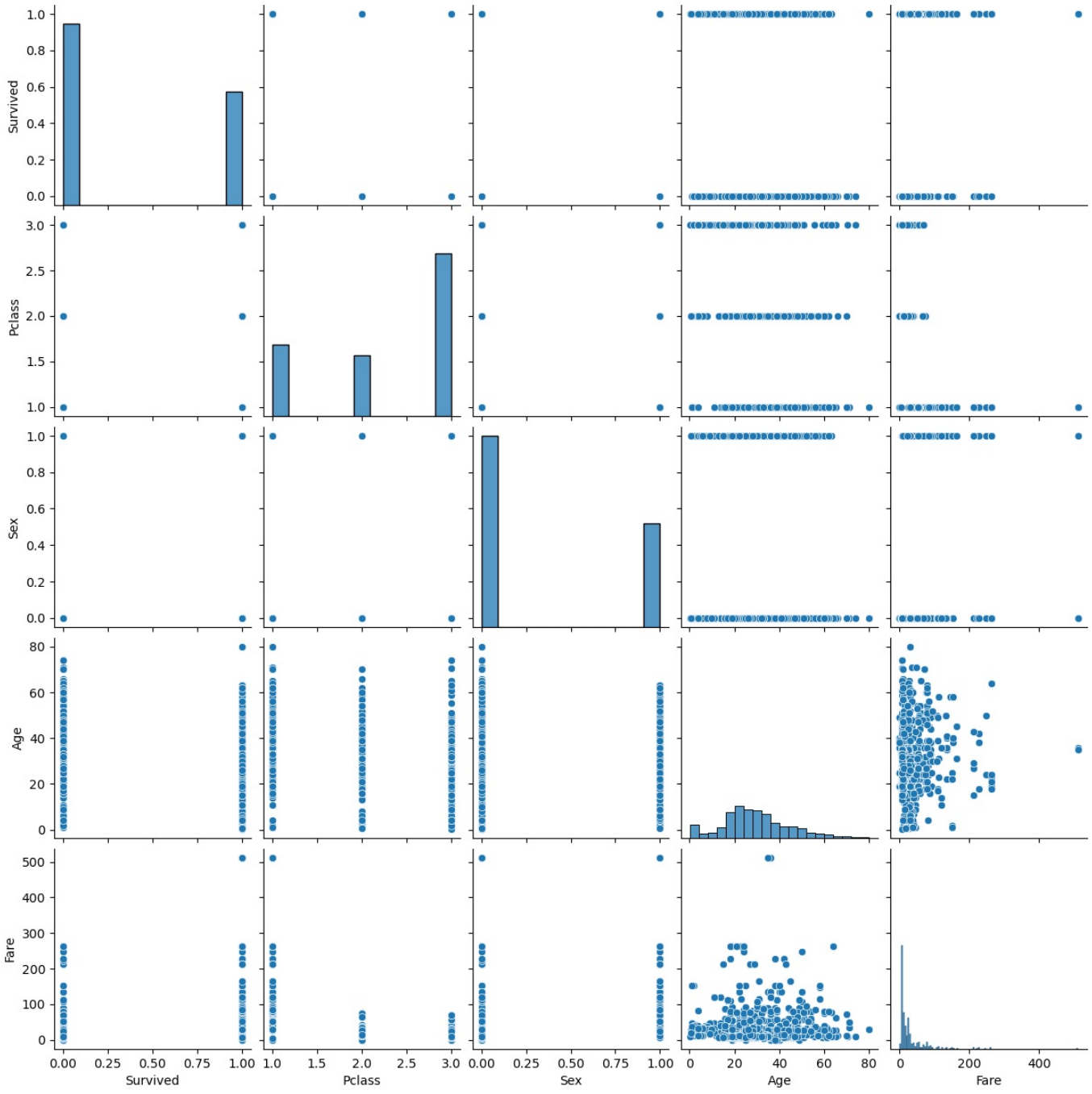
In [87]:
```
# Heatmap of Correlation Matrix
numeric_df = df.select_dtypes(include='number')

sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Matrix - Titanic Data')
plt.show()
```



In [89]:
```
# Fare has weak positive correlation with survival
# Age and survival have slightly negative correlation
```

In [91]:
```
# Pairplot for Relationships
df['Sex'] = df['Sex'].map({'male': 0, 'female': 1})
sns.pairplot(df[['Survived', 'Pclass','Sex', 'Age', 'Fare']])
```

Out[91]: <seaborn.axisgrid.PairGrid at 0x1d01164b470>



In [92]:
```
# Handling Missing Values
df.fillna({'Age':df['Age'].median()}, inplace=True)
print(df.columns)
df.drop(columns=['Cabin'], inplace=True,errors='ignore')
df.dropna(subset=['Embarked'], inplace=True)
```

```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

In [93]:
```
df.head()
```

Out[93]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	0	22.0	1	0	A/5 21171	7.2500	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	1	38.0	1	0	PC 17599	71.2833	C
2	3	1	3	Heikkinen, Miss. Laina	1	26.0	0	0	STON/O2. 3101282	7.9250	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	1	35.0	1	0	113803	53.1000	S
4	5	0	3	Allen, Mr. William Henry	0	35.0	0	0	373450	8.0500	S

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js