Task 1: Data Visualization – Age and Gender Distribution

This task visualizes the distribution of gender and age using the Titanic dataset.
We use bar chart and histogram to understand how the passengers are distributed by gender and age.

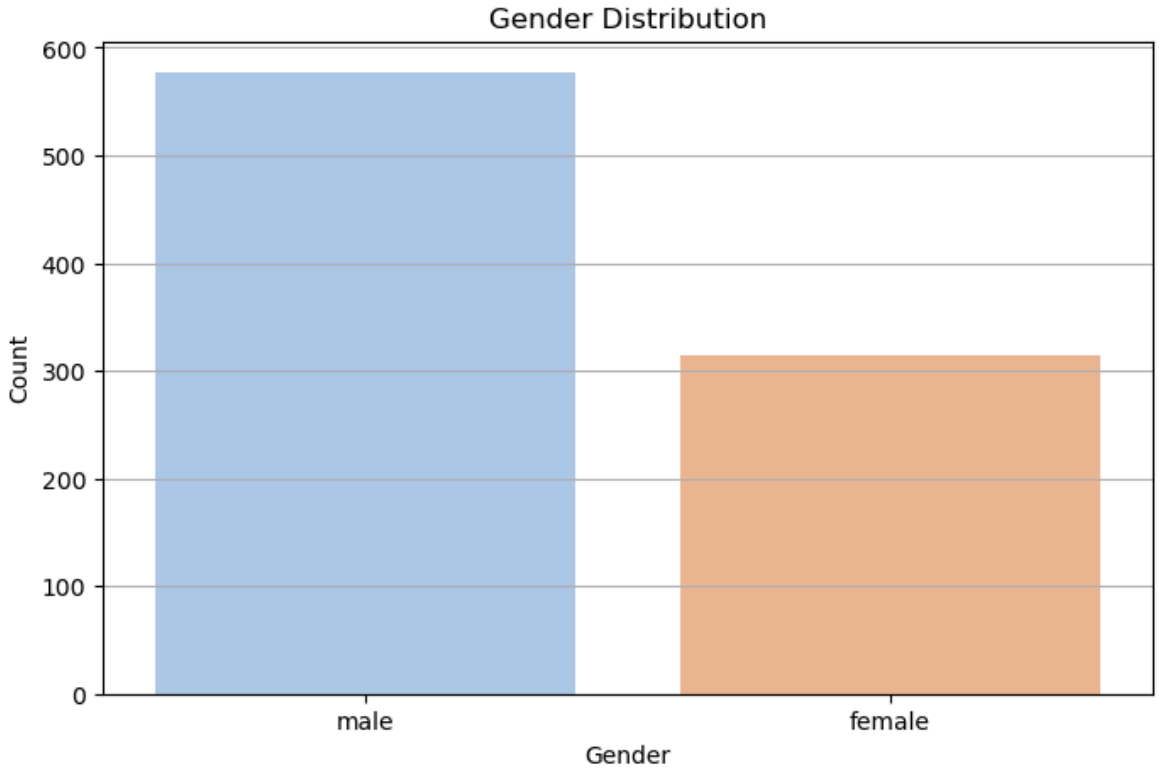
In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [11]:
```
url = 'https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic
df = pd.read_csv(url)
```

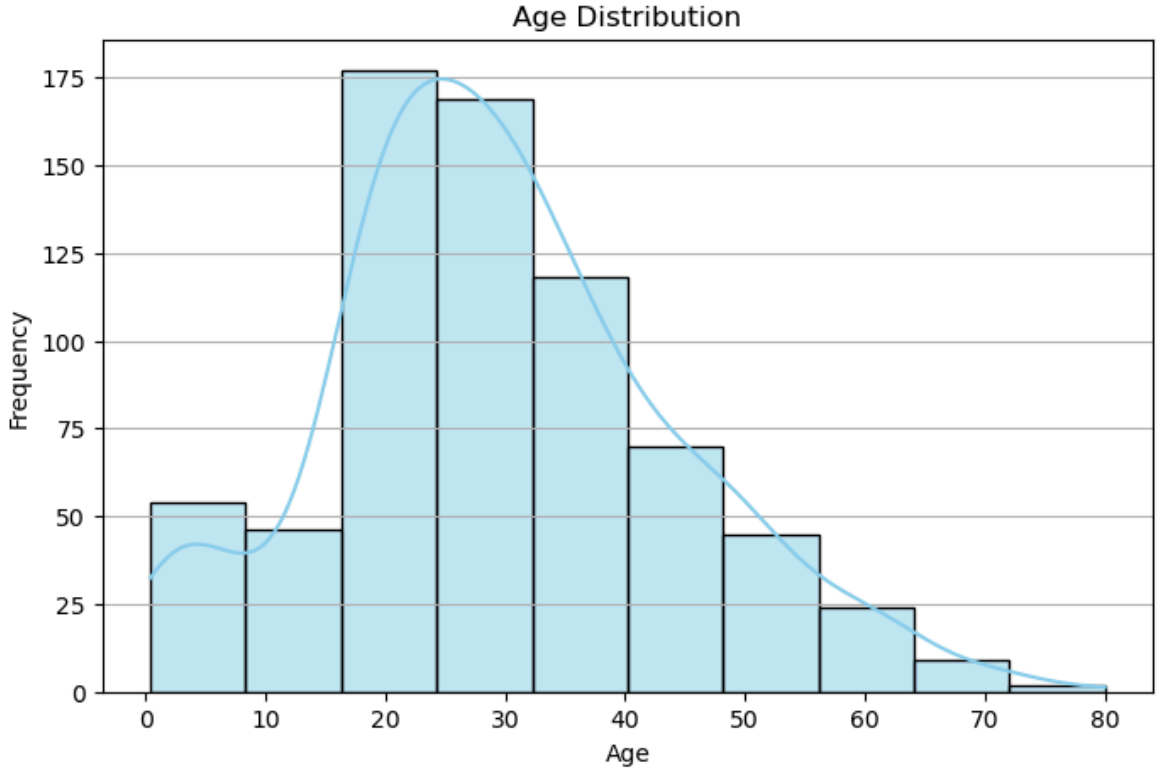
In [12]:
```
print("Dataset Preview:")
print(df.head())
```

```
Dataset Preview:
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3

                                                Name     Sex   Age  SibSp  \
0                            Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2                             Heikkinen, Miss. Laina  female  26.0      0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                           Allen, Mr. William Henry    male  35.0      0

   Parch            Ticket     Fare Cabin Embarked
0      0         A/5 21171   7.2500   NaN        S
1      0          PC 17599  71.2833   C85        C
2      0  STON/O2. 3101282   7.9250   NaN        S
3      0            113803  53.1000  C123        S
4      0            373450   8.0500   NaN        S
```

In [13]:
```
plt.figure(figsize=(8, 5))
sns.countplot(data=df, x='Sex', palette='pastel')
plt.title('Gender Distribution')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.grid(axis='y')
plt.show()
```



In [15]:
```
plt.figure(figsize=(8, 5))
sns.histplot(data=df, x='Age', bins=10, kde=True, color='skyblue')
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.grid(axis='y')
plt.show()
```



Conclusion

- Most passengers in the Titanic dataset were **male**.

- The age distribution shows that the majority of passengers were between **20 to 40 years**.
- These visualizations help us understand basic population characteristics in the dataset.

In [ ]: