TASK 01

TASK: create a bar chart or histogram to visualise the distribution of categorical or continuous variable such as distribution of ages or gender in a population.

INSTALLING REQUIRED LIBRARIES

In [18]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Load the Titanic dataset from seaborn
titanic_data = sns.load_dataset('titanic')

# Display the first few rows of the dataset
print(titanic_data.head())
```

```
   survived  pclass     sex   age  sibsp  parch     fare embarked  class
\
0         0       3    male  22.0      1      0   7.2500        S  Third
1         1       1  female  38.0      1      0  71.2833        C  First
2         1       3  female  26.0      0      0   7.9250        S  Third
3         1       1  female  35.0      1      0  53.1000        S  First
4         0       3    male  35.0      0      0   8.0500        S  Third

     who  adult_male deck  embark_town alive  alone
0    man        True  NaN  Southampton    no  False
1  woman       False    C    Cherbourg   yes  False
2  woman       False  NaN  Southampton   yes   True
3  woman       False    C  Southampton   yes  False
4    man        True  NaN  Southampton    no   True
```

INSPECTING THE DATA

In [19]:
```
# Check the shape of the dataset
print("Shape of the dataset:", titanic_data.shape)

# Check for missing values
print(titanic_data.isnull().sum())

# Display column names
print("Column names:", titanic_data.columns.tolist())
```

```
Shape of the dataset: (891, 15)
survived         0
pclass           0
sex              0
age            177
sibsp            0
parch            0
fare             0
embarked         2
class            0
who              0
adult_male       0
deck           688
embark_town      2
alive            0
alone            0
dtype: int64
Column names: ['survived', 'pclass', 'sex', 'age', 'sibsp', 'parch', 'far
e', 'embarked', 'class', 'who', 'adult_male', 'deck', 'embark_town', 'aliv
e', 'alone']
```

PREPARING THE DATA

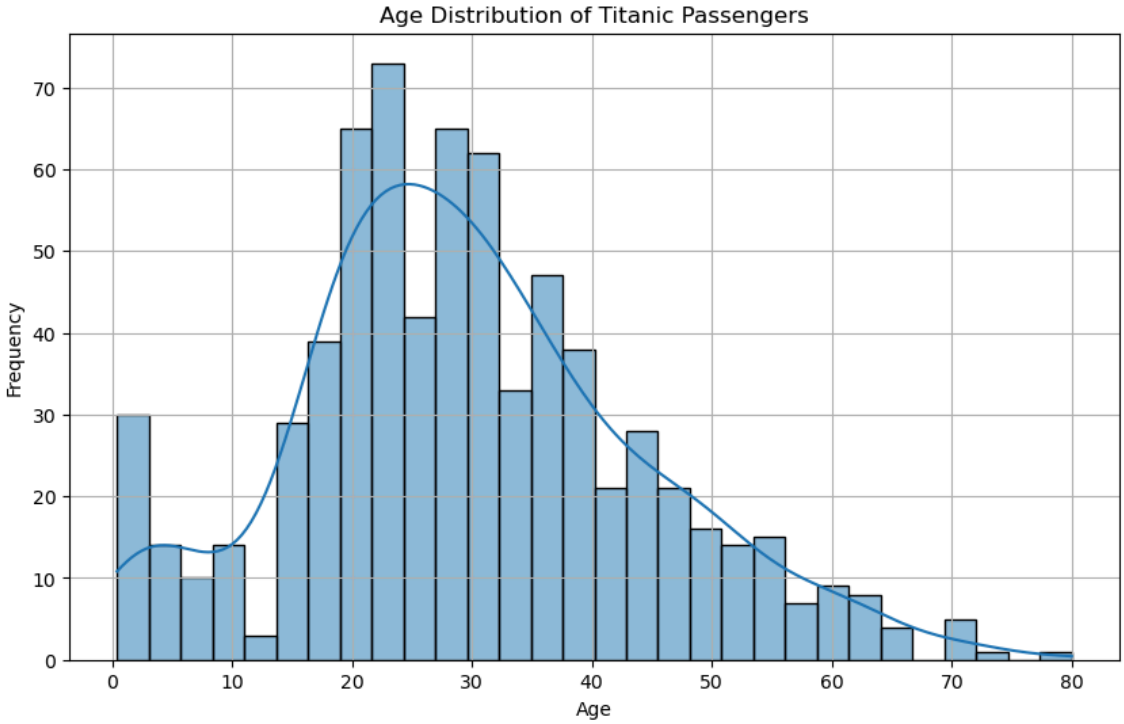
In [20]:
```
# Drop rows with missing values in 'age' and 'sex' columns
titanic_data_cleaned = titanic_data.dropna(subset=['age', 'sex'])

# Check the shape after cleaning
print("Shape after cleaning:", titanic_data_cleaned.shape)
```

```
Shape after cleaning: (714, 15)
```

VISUALIZING AGE DISTRIBUTION

In [21]:
```
plt.figure(figsize=(10, 6))
sns.histplot(titanic_data_cleaned['age'], bins=30, kde=True)
plt.title('Age Distribution of Titanic Passengers')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.grid()
plt.show()
```



The histogram shows how the ages of Titanic passengers are distributed. We can analyze which age groups were more prevalent on the Titanic.

Age Distribution Histogram

Key Insights:

Age Range: Passengers' ages range from 0 to 80 years.

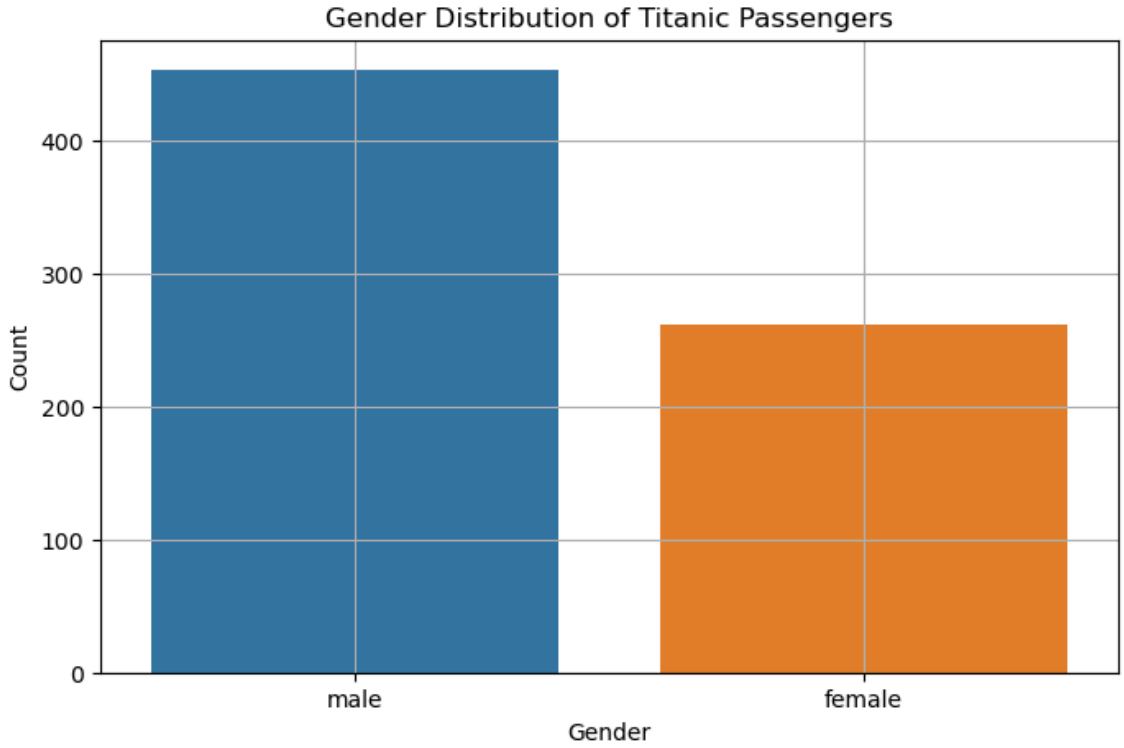
Main Age Group: The largest group of passengers is between 20-30 years old.

Younger Passengers: There are many infants and young children (ages 0-5).

Older Passengers: Few passengers are over 60 years old.

VISUALIZATION OF GENDER DISTRIBUTION

In [22]:
```
plt.figure(figsize=(8, 5))
sns.countplot(data=titanic_data_cleaned, x='sex', order=titanic_data_cleane
plt.title('Gender Distribution of Titanic Passengers')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.grid()
plt.show()
```



The bar chart reveals the number of male and female passengers on the Titanic, allowing us to understand the gender ratio among the passengers.

Key Insights:

Total Passengers: Approximately 891 passengers.

Males vs. Females: Males: About 577 (64.7%) Females: About 314 (35.3%)

Conclusion:

Most passengers were young adults, with a significant number of children.

There were notably more males than females aboard the Titanic.

This summary provides a clear and concise understanding of the demographics of Titanic passengers based on the visualizations.

In [ ]: