**TASK 5**

**Step 1:** import pandas as pd

**Step 2**: data =pd.read\_csv('tested.csv')

**Step 3:** data.info()

**Step 4:** data.describe()

**Step 5:** print(data['Survived'].value\_counts())

print(data['Pclass'].value\_counts())

print(data['Sex'].value\_counts())

**step 6:** import seaborn as sns

import matplotlib.pyplot as plt

sns.pairplot(data, hue='Survived')

plt.show()

**step 7:** numeric\_cols = data.select\_dtypes(include=['int64', 'float64']).columns

correlation\_matrix = data[numeric\_cols].corr()

plt.figure(figsize=(10,7))

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', fmt='.2f')

plt.title('Correlation Heatmap')

plt.show()

**step 8:** plt.figure(figsize=(10, 6))

sns.histplot(data=data, x='Age', hue='Survived', multiple='stack', bins=30)

plt.title('Age Distribution of Survivors vs Non-Survivors')

plt.xlabel('Age')

plt.ylabel('Count')

plt.show()

**step 9**: plt.figure(figsize=(10, 6))

sns.histplot(data['Age'].dropna(), bins=30, kde=True)

plt.title('Age Distribution of Passengers')

plt.xlabel('Age')

plt.ylabel('Frequency')

plt.show()

**step 10:** plt.figure(figsize=(10, 6))

sns.boxplot(x='Survived', y='Age', data=data)

plt.title('Boxplot of Age by Survival')

plt.xlabel('Survived (0 = No, 1 = Yes)')

plt.ylabel('Age')

plt.show()

**step 11:** plt.figure(figsize=(10, 6))

sns.scatterplot(x='Age', y='Fare', hue='Survived', data=data, alpha=0.6)

plt.title('Scatterplot of Age vs. Fare')

plt.xlabel('Age')

plt.ylabel('Fare')

plt.show()

**Summary:**

**1. Age Distribution**

* The age distribution of passengers is right-skewed, indicating a higher frequency of younger individuals.
* Most passengers are concentrated in the younger age groups, particularly between 20-30 years old.
* There are significantly fewer older passengers, suggesting that the demographic on board was predominantly young.

**2. Survival Rates**

* **By Passenger Class (Pclass)**:
  + Higher survival rates are observed in 1st class passengers compared to those in 2nd and 3rd class. This indicates that socio-economic status played a crucial role in survival chances.
* **By Sex**:
  + Female passengers have a notably higher survival rate than male passengers. This aligns with historical accounts of women and children being prioritized during evacuation.

**3. Age and Survival**

* The boxplot analysis shows that the median age of survivors is slightly higher than that of non-survivors, suggesting that older children and adults had better survival rates.
* The presence of outliers indicates that some older passengers survived, while many younger passengers did not.

**4. Fare and Survival**

* The scatterplot analysis indicates no direct correlation between age and fare; however, higher fares are associated with higher survival rates.
* Wealthier passengers (who typically paid higher fares) had better chances of survival, likely due to their access to lifeboats and priority during evacuation.

**5. General Trends**

* The analysis highlights the impact of socio-economic factors (class and fare) and demographic factors (age and sex) on survival outcomes during the Titanic disaster.
* The combination of these factors suggests that those in higher social classes, particularly women and children, had significantly better chances of survival.