Course Name: AIML Lab **Course Code: 21CSH-316**

Experiment:3.3

Student Name: UTKARSH JOSHI **UID**: 21BCS9158
Branch: BE-CSE **Section/Group**:802-A
Semester: 05 **Date of Performance**: 15/11/23
Subject Name: AI&ML with Lab **Subject Code**: 21CSH-316

1. **Aim:** Implement exploratory Data Analysis on any data set.
2. **Software used:** Google colab, PYTHON V2023.14.0 .
3. **Pseudo code/Algorithms/Flowchart/Steps:**

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
# Load the Titanic dataset
data = pd.read_csv('/content/titanic.csv')
# Display basic information about the dataset
print(data.info())
# Summary statistics of numerical features
print(data.describe())
# Distribution of categorical variables
print(data['Sex'].value_counts())
print(data['Embarked'].value_counts())
# Missing data
print(data.isnull().sum())
# Visualizing data
# Plot age distribution
plt.figure(figsize=(8, 6))
sns.histplot(data['Age'], bins=20, kde=True)
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Count')
```

```
plt.show()
# Plot survival rate by sex
plt.figure(figsize=(8, 6))
sns.countplot(data=data, x='Sex', hue='Survived')
plt.title('Survival by Sex')
plt.xlabel('Sex')
plt.ylabel('Count')
plt.show()
# Plot correlation heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(data.corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

4. Output:

```
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
```

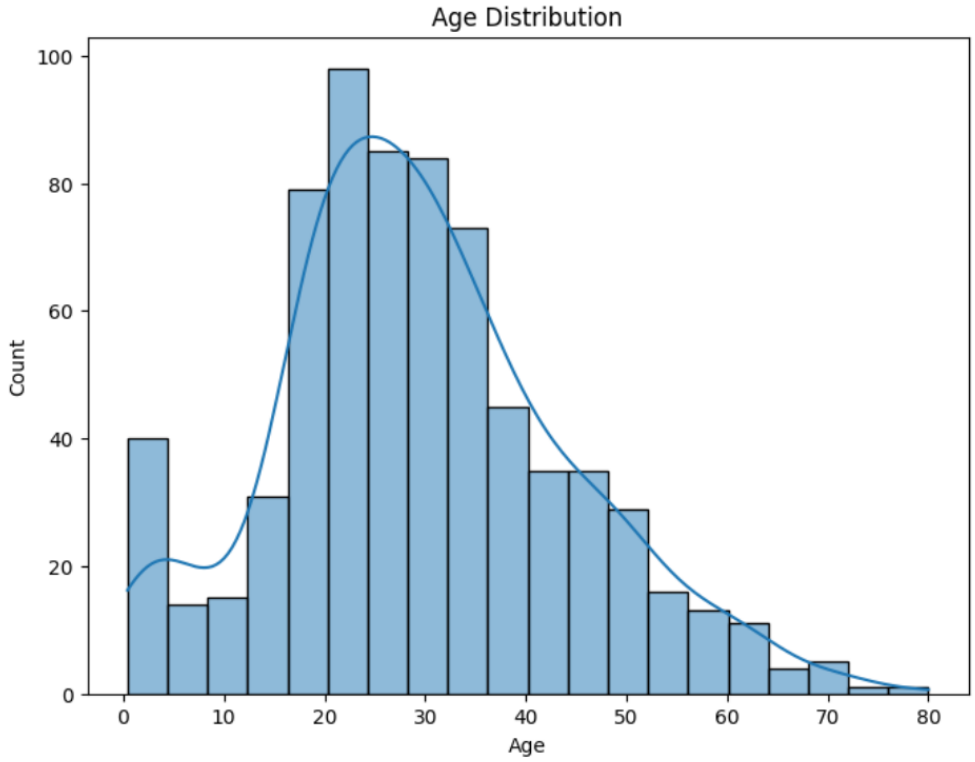
Course Name: AIML Lab **Course Code: 21CSH-316**

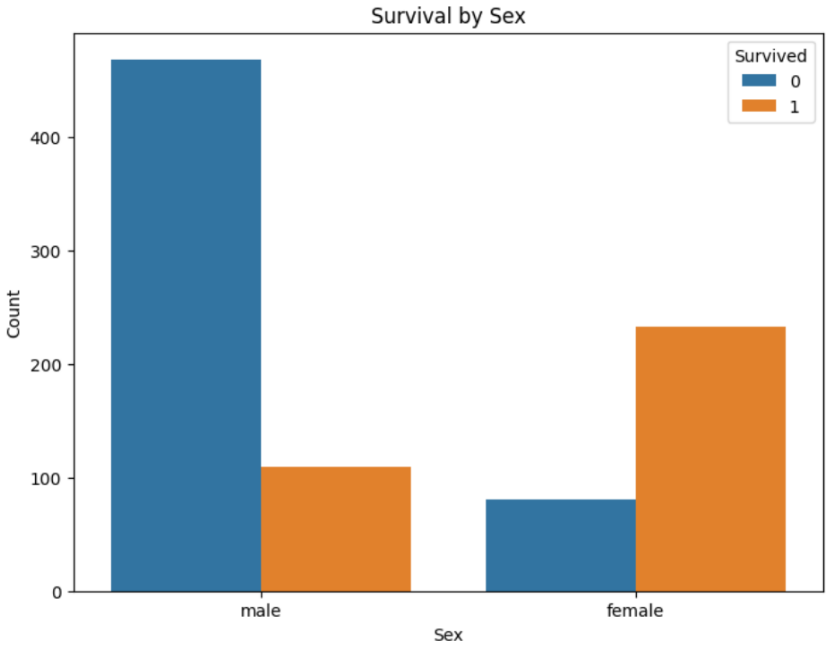
```
       PassengerId    Survived      Pclass         Age       SibSp  \
count   891.000000  891.000000  891.000000  714.000000  891.000000
mean    446.000000    0.383838    2.308642   29.699118    0.523008
std     257.353842    0.486592    0.836071   14.526497    1.102743
min       1.000000    0.000000    1.000000    0.420000    0.000000
25%     223.500000    0.000000    2.000000   20.125000    0.000000
50%     446.000000    0.000000    3.000000   28.000000    0.000000
75%     668.500000    1.000000    3.000000   38.000000    1.000000
max     891.000000    1.000000    3.000000   80.000000    8.000000
```

```
            Parch        Fare
count  891.000000  891.000000
mean     0.381594   32.204208
std      0.806057   49.693429
min      0.000000    0.000000
25%      0.000000    7.910400
50%      0.000000   14.454200
75%      0.000000   31.000000
max      6.000000  512.329200
male      577
female    314
Name: Sex, dtype: int64
S    644
C    168
Q     77
Name: Embarked, dtype: int64
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

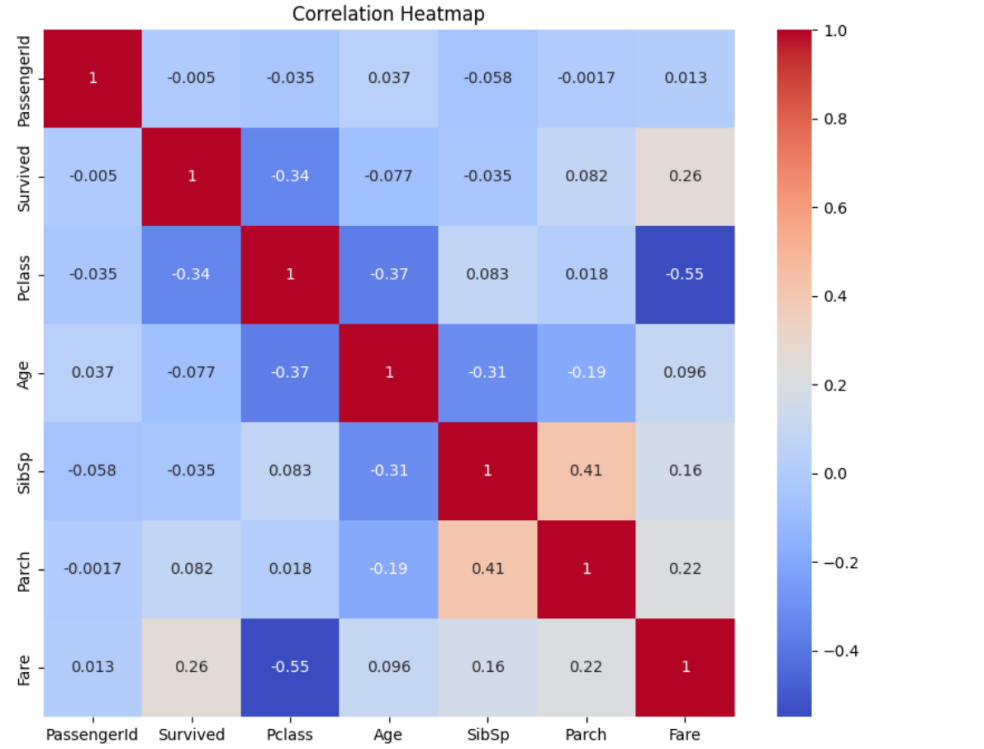
Course Name: AIML Lab **Course Code: 21CSH-316**

```
dtype: int64
```





Course Name: AIML Lab **Course Code: 21CSH-316**



5. Learning and Outcomes:

- Understand the concept of plotting.
- Learned the fundamentals of Python.
- Usage of libraries to preprocess useful data.