Name:- Harsh sinha
Reg. No:- 21BLC1419
Assignment:- 3

```python
# 1. Import the necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, LabelEncoder

# 2. Import the dataset
titanic_df = pd.read_csv('/Titanic-Dataset.csv')

# 3. Checking for Null Values
# Let's check for missing values in the dataset
print(titanic_df.isnull().sum())

# 4. Data Visualization
# You can use various visualization techniques to understand the data better.
# For example, let's create a histogram for the 'Age' column
sns.histplot(data=titanic_df, x='Age', bins=20, kde=True)
plt.xlabel('Age')
plt.ylabel('Count')
plt.title('Age Distribution')
plt.show()

# 5. Outlier Detection
# You can use box plots or other visualization methods to identify outliers
sns.boxplot(data=titanic_df, x='Fare')
plt.xlabel('Fare')
plt.title('Fare Distribution')
plt.show()

# 6. Splitting Dependent and Independent Variables
# Assuming 'Survived' is the target variable
X = titanic_df.drop('Survived', axis=1)
y = titanic_df['Survived']
```

```python
# 7. Encoding Categorical Variables
# You can use Label Encoding or One-Hot Encoding for categorical variables
# For example, let's label encode 'Sex' and 'Embarked' columns
label_encoder = LabelEncoder()
X['Sex'] = label_encoder.fit_transform(X['Sex'])
X['Embarked'] = label_encoder.fit_transform(X['Embarked'])

# 8. Feature Scaling
# Standardize numerical features (e.g., 'Age' and 'Fare') to have mean=0 and std=1
scaler = StandardScaler()
X[['Age', 'Fare']] = scaler.fit_transform(X[['Age', 'Fare']])

# 9. Splitting Data into Train and Test Sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Now you have preprocessed data ready for machine learning modeling.
# X_train, X_test are your independent variables, and y_train, y_test are your target variables.
```

```
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
```