1. Import Libraries and Dataset

```
# Import necessary libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

# Load dataset
url = 'https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv'
data = pd.read_csv(url)

# Display first 5 rows
print(data.head())
```

2. Explore Basic Information (nulls, data types)

```
# Check dataset shape
print("Dataset shape:", data.shape)

# Check data types and null values
print(data.info())

# Check number of missing values
print(data.isnull().sum())
```

3. Handle Missing Values

```
# Fill missing 'Age' values with median
data['Age'].fillna(data['Age'].median(), inplace=True)

# Fill missing 'Embarked' values with mode
data['Embarked'].fillna(data['Embarked'].mode()[0], inplace=True)

# Drop 'Cabin' column as it has too many missing values
data.drop(columns='Cabin', inplace=True)
```

4. Convert Categorical Features to Numerical (Encoding)

```
# Convert 'Sex' and 'Embarked' to numerical using Label Encoding
data['Sex'] = data['Sex'].map({'male': 0, 'female': 1})

# One-hot encoding for 'Embarked'
data = pd.get_dummies(data, columns=['Embarked'], drop_first=True)
```

5. Normalize/Standardize Numerical Features

```
from sklearn.preprocessing import StandardScaler

# Select numerical columns to scale
num_cols = ['Age', 'Fare']

# Apply StandardScaler
scaler = StandardScaler()
data[num_cols] = scaler.fit_transform(data[num_cols])

# Check the result
print(data.head())
```

6. Visualize Outliers using Boxplots

```
# Boxplot for 'Age'
sns.boxplot(x=data['Age'])
plt.title('Boxplot for Age')
plt.show()

# Boxplot for 'Fare'
sns.boxplot(x=data['Fare'])
plt.title('Boxplot for Fare')
plt.show()
```

7. Final Cleaned Data

```
# Check final cleaned dataset info
print(data.info())
```