```python
#Data Cleaning & Preprocessing

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler
from scipy.stats import zscore

# Loading dataset
df = pd.read_csv('/content/titanic-dataset')

# Step 2: Explore basic info
print("Initial Info:")
print(df.info())
print("\nMissing values:\n", df.isnull().sum())

#Handle missing values
df['Age'] = df['Age'].fillna(df['Age'].median())  # Safe assignment
df['Embarked'] = df['Embarked'].fillna(df['Embarked'].mode()[0])  # Safe
assignment
if 'Cabin' in df.columns:
    df = df.drop(columns='Cabin')

#  Encode categorical features
if 'Sex' in df.columns:
    df['Sex'] = df['Sex'].map({'male': 0, 'female': 1})
if 'Embarked' in df.columns:
    df = pd.get_dummies(df, columns=['Embarked'], drop_first=True)

# Normalize numerical features
scaler = StandardScaler()
for col in ['Age', 'Fare']:
    if col in df.columns:
        df[[col]] = scaler.fit_transform(df[[col]])

# Remove outliers using z-score
if 'Fare' in df.columns:
    z_scores = zscore(df[['Fare']])
    df = df[(np.abs(z_scores) < 3).all(axis=1)]

# Visualize Fare outliers (after cleaning)
if 'Fare' in df.columns:
    plt.figure(figsize=(8, 4))
    sns.boxplot(x=df['Fare'])
    plt.title('Boxplot of Fare (Cleaned)')
    plt.show()

# Final dataset info
```

```
print("\nCleaned dataset preview:")
print(df.head())
print(f"\nShape of cleaned dataset: {df.shape}")
```