

Task 2: Exploratory Data Analysis (EDA) using Titanic Dataset

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

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

# 1. Summary Statistics
print("Summary Statistics:\n")
print(df.describe(include='all'))

# 2. Histograms for Numeric Features
numeric_cols = df.select_dtypes(include=['float64', 'int64']).columns

df[numeric_cols].hist(bins=20, figsize=(12, 10), edgecolor='black')
plt.suptitle('Histograms of Numeric Features')
plt.show()

# 3. Boxplots to detect outliers
for col in numeric_cols:
    plt.figure(figsize=(6, 3))
    sns.boxplot(x=df[col])
    plt.title(f'Boxplot of {col}')
    plt.show()
```

4. Pairplot to see relationships

```
sns.pairplot(df.dropna(), vars=['Age', 'Fare', 'Pclass'], hue='Survived')  
plt.suptitle('Pairplot of Selected Features by Survival', y=1.02)  
plt.show()
```

5. Correlation Matrix

```
corr_matrix = df[numeric_cols].corr()  
plt.figure(figsize=(10, 8))  
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', fmt=".2f")  
plt.title('Correlation Matrix')  
plt.show()
```

6. Inferences

```
print("\nBasic Inferences:")  
print("- Age and Fare show some outliers (see boxplots).")  
print("- Strong correlation between SibSp and Parch may indicate family relationships.")  
print("- Fare has moderate positive correlation with Pclass (as expected).")  
print("- Survival rates vary across gender and class — deeper analysis needed.")
```