

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# 1. SETUP & LOADING
sns.set_theme(style="whitegrid")
df = sns.load_dataset('titanic')

print("--- DATA OVERVIEW ---")
print(f"Dataset Shape: {df.shape}")
print(df.head())

# 2. DATA CLEANING
print("\n--- MISSING VALUES ---")
missing_data = df.isnull().sum()
print(missing_data[missing_data > 0])

# Filling missing Age with Median, Embarked with Mode
df['age'] = df['age'].fillna(df['age'].median())
df['embarked'] = df['embarked'].fillna(df['embarked'].mode()[0])

# Dropping 'deck' because it has too many missing values
df.drop(columns=['deck'], inplace=True)

# 3. FEATURE ENGINEERING
# Creating a 'family_size' column to see if it impacts survival
df['family_size'] = df['sibsp'] + df['parch'] + 1

# 4. EXPLORATORY VISUALIZATION
fig, axes = plt.subplots(2, 2, figsize=(15, 12))

# Chart 1: Distribution of Age
sns.histplot(df['age'], bins=30, kde=True, ax=axes[0, 0], color='skyblue')
axes[0, 0].set_title('Age Distribution (Cleaned)')

# Chart 2: Categorical Survival Counts
sns.countplot(x='class', hue='survived', data=df, ax=axes[0, 1], palette='magma')
axes[0, 1].set_title('Survival Count by Class')

# Chart 3: Correlation Heatmap
# Selecting only numeric columns
numeric_df = df.select_dtypes(include=[np.number])
sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm', fmt='.2g', ax=axes[1, 0])
axes[1, 0].set_title('Correlation Matrix')

# Chart 4: Boxplot for Outlier Detection (Fare vs Pclass)
sns.boxplot(x='pclass', y='fare', data=df, ax=axes[1, 1])

```

```
axes[1, 1].set_title('Fare Outliers per Class')

plt.tight_layout()
plt.show()

# 5. FINAL SUMMARY STATS
print("\n--- GROUPED ANALYSIS ---")
# Calculating survival rate by Gender and Class
summary = df.groupby(['sex', 'class'])['survived'].mean().unstack()
print(summary)

print("\nEDA Project Code Execution Complete.")
```