

```

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

# Load the Titanic dataset (adjust the file path as needed)
# Download from kaggle or use seaborn's built-in dataset
titanic = sns.load_dataset('titanic')

# Display the first few rows of the dataset
print("First 5 rows of the dataset:")
print(titanic.head())

# Data cleaning
print("\nSummary of missing values:")
print(titanic.isnull().sum())

# Filling missing 'age' with median
titanic['age'].fillna(titanic['age'].median(), inplace=True)

# Filling missing 'embarked' with the most frequent value
titanic['embarked'].fillna(titanic['embarked'].mode()[0], inplace=True)

# Dropping rows with missing 'deck' values due to a high percentage of missing data
titanic.drop(columns=['deck'], inplace=True)

# Check the updated missing values
print("\nUpdated summary of missing values:")
print(titanic.isnull().sum())

# Exploratory Data Analysis (EDA)
# Summary statistics
print("\nSummary statistics:")
print(titanic.describe())

# Distribution of numerical variables
plt.figure(figsize=(10, 5))
sns.histplot(titanic['age'], kde=True, bins=30, color='blue', alpha=0.7)
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()

# Bar chart of survival count
plt.figure(figsize=(6, 4))
sns.countplot(x='survived', data=titanic, palette='Set2')
plt.title('Survival Count')
plt.xticks([0, 1], ['Not Survived', 'Survived'])
plt.xlabel('Survived')
plt.ylabel('Count')
plt.show()

# Survival rate by gender
plt.figure(figsize=(6, 4))
sns.barplot(x='sex', y='survived', data=titanic, palette='Set3')
plt.title('Survival Rate by Gender')
plt.xlabel('Gender')
plt.ylabel('Survival Rate')
plt.show()

# Survival rate by passenger class
plt.figure(figsize=(6, 4))
sns.barplot(x='pclass', y='survived', data=titanic, palette='Set1')
plt.title('Survival Rate by Passenger Class')
plt.xlabel('Passenger Class')
plt.ylabel('Survival Rate')
plt.show()

# Correlation heatmap
plt.figure(figsize=(10, 8))
corr = titanic.corr()
sns.heatmap(corr, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Heatmap')
plt.show()

# Pair plot for numerical variables
sns.pairplot(titanic, hue='survived', diag_kind='kde', palette='husl')
plt.show()

# Insights
print("\nInsights from the analysis:")
print("- Younger passengers and women had higher survival rates.")
print("- First-class passengers had a significantly higher survival rate.")
print("- Survival rate shows correlation with age, gender, and class.")

# Save cleaned data (optional)
titanic.to_csv('cleaned_titanic.csv', index=False)
print("\nCleaned dataset saved as 'cleaned_titanic.csv'.")

```

First 5 rows of the dataset:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class \
0	0	3	male	22.0	1	0	7.2500	S	Third
1	1	1	female	38.0	1	0	71.2833	C	First
2	1	3	female	26.0	0	0	7.9250	S	Third
3	1	1	female	35.0	1	0	53.1000	S	First
4	0	3	male	35.0	0	0	8.0500	S	Third

	who	adult_male	deck	embark_town	alive	alone
0	man	True	NaN	Southampton	no	False
1	woman	False	C	Cherbourg	yes	False
2	woman	False	NaN	Southampton	yes	True
3	woman	False	C	Southampton	yes	False
4	man	True	NaN	Southampton	no	True

Summary of missing values:

survived	0
pclass	0
sex	0
age	177
sibsp	0
parch	0
fare	0
embarked	2
class	0
who	0
adult_male	0
deck	688
embark_town	2
alive	0
alone	0
dtype:	int64

Updated summary of missing values:

survived	0
pclass	0
sex	0
age	0
sibsp	0
parch	0
fare	0
embarked	0
class	0
who	0
adult_male	0
embark_town	2
alive	0
alone	0
dtype:	int64

Summary statistics:

	survived	pclass	age	sibsp	parch	fare
count	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000
mean	0.383838	2.308642	29.361582	0.523008	0.381594	32.204268
std	0.485592	0.836071	13.019697	1.102743	0.806057	49.693429
min	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	0.000000	2.000000	22.000000	0.000000	0.000000	7.910400
50%	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	1.000000	3.000000	35.000000	1.000000	0.000000	31.000000
max	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200