

```
# =====
# TITANIC EDA - SAMPLE DATASET
# =====

# 1. Import Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

pd.set_option('display.max_columns', None)
sns.set_style('whitegrid')

# 2. Upload CSV in Colab
from google.colab import files
uploaded = files.upload()

# 3. Load Dataset
df = pd.read_csv("sample_titanic_dataset.csv")
df.head()

# 4. Basic Exploration
print("Dataset Shape:", df.shape)
print("\n--- Dataset Info ---")
df.info()

print("\n--- Summary Statistics ---")
print(df.describe())

print("\n--- Missing Values ---")
print(df.isnull().sum())

print("\n--- Target Variable Counts ---")
print(df['Survived'].value_counts())

# 5. Univariate Analysis
plt.figure(figsize=(6,4))
df['Age'].hist(bins=10, color='skyblue', edgecolor='black')
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()

plt.figure(figsize=(6,4))
sns.boxplot(x=df['Age'])
plt.title('Age Boxplot')
plt.show()

plt.figure(figsize=(6,4))
sns.countplot(x='Sex', data=df, palette='Set2')
plt.title('Gender Count')
plt.show()

plt.figure(figsize=(6,4))
```

Release notes

sample_t ...

1 to 10 of 20 entries

Filter



PassengerId	Survived	Pclass	Name
1	0	3	Allen, Elisabeth
2	1	1	Mora, James
3	1	3	McCormack, Thomas
4	1	1	Palsson, Gustaf
5	0	3	Johnson, Mrs. Ida
6	0	3	Nasser, Nicholas
7	1	1	Sandstrom, Margareta
8	0	3	Bonn, Miss. Elizabeth
9	1	2	Saunders, Mr. William
10	1	1	Anderson, Mr. Andrew

Show 10 per page

1

2

```
sns.countplot(x='Pclass', data=df, palette='Set3')
plt.title('Passenger Class Count')
plt.show()

# 6. Bivariate Analysis
plt.figure(figsize=(6,4))
sns.countplot(x='Sex', hue='Survived', data=df, palette=
plt.title('Survival by Gender')
plt.show()

plt.figure(figsize=(6,4))
sns.boxplot(x='Survived', y='Age', data=df, palette='Set
plt.title('Age vs Survival')
plt.show()

plt.figure(figsize=(6,4))
sns.barplot(x='Pclass', y='Fare', data=df, palette='viri
plt.title('Pclass vs Fare')
plt.show()

# 7. Correlation Heatmap
plt.figure(figsize=(8,6))
corr = df.corr(numeric_only=True)
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()

# 8. Summary of Findings
summary = """
### Summary of EDA Findings:
- Most survivors were female.
- Pclass 1 had a much higher survival rate compared to P
- Younger passengers had a higher chance of survival.
- Fare is positively correlated with survival – higher f
"""

print(summary)
```



Choose Files sample_tita...dataset.csv

- **sample_titanic_dataset.csv**(text/csv) - 1270 bytes, last modified: 8/11/2025 - 100% done

Saving sample_titanic_dataset.csv to sample_titani
Dataset Shape: (20, 11)

--- Dataset Info ---

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 20 entries, 0 to 19

Data columns (total 11 columns):

#	Column	Non-Null Count	Dtype
0	PassengerId	20 non-null	int64
1	Survived	20 non-null	int64
2	Pclass	20 non-null	int64
3	Name	20 non-null	object
4	Sex	20 non-null	object
5	Age	20 non-null	int64
6	SibSp	20 non-null	int64
7	Parch	20 non-null	int64
8	Ticket	20 non-null	object
9	Fare	20 non-null	float64
10	Embarked	20 non-null	object

dtypes: float64(1), int64(6), object(4)

memory usage: 1.8+ KB

--- Summary Statistics ---

	PassengerId	Survived	Pclass	Age
count	20.000000	20.000000	20.000000	20.000000
mean	10.500000	0.500000	2.150000	27.400000
std	5.91608	0.512989	0.933302	15.93539
min	1.000000	0.000000	1.000000	2.000000
25%	5.750000	0.000000	1.000000	17.250000
50%	10.500000	0.500000	2.500000	28.500000
75%	15.250000	1.000000	3.000000	36.000000
max	20.000000	1.000000	3.000000	58.000000

	Fare
count	20.000000
mean	19.493125
std	17.432129
min	7.225000
25%	8.044800
50%	11.250000
75%	27.193750
max	71.283300

--- Missing Values ---

PassengerId	0
Survived	0
Pclass	0
Name	0
Sex	0
Age	0
SibSp	0
Parch	0
Ticket	0
Fare	0
Embarked	0

dtype: int64

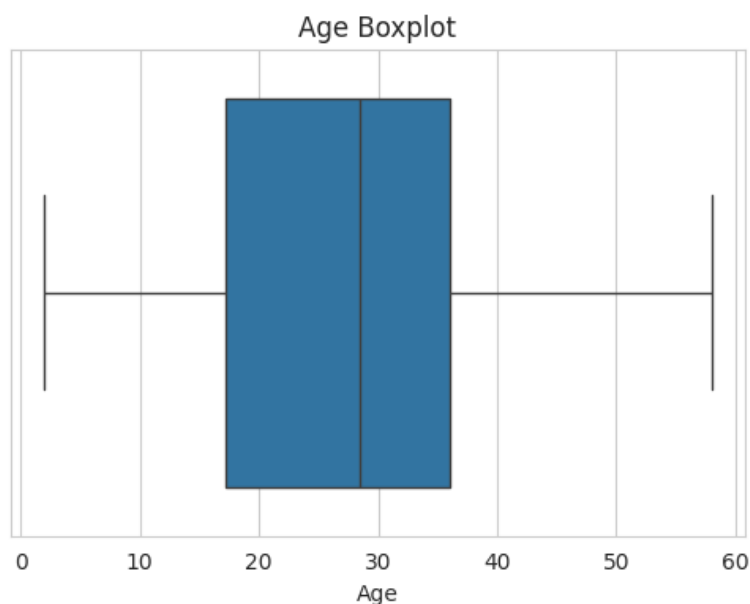
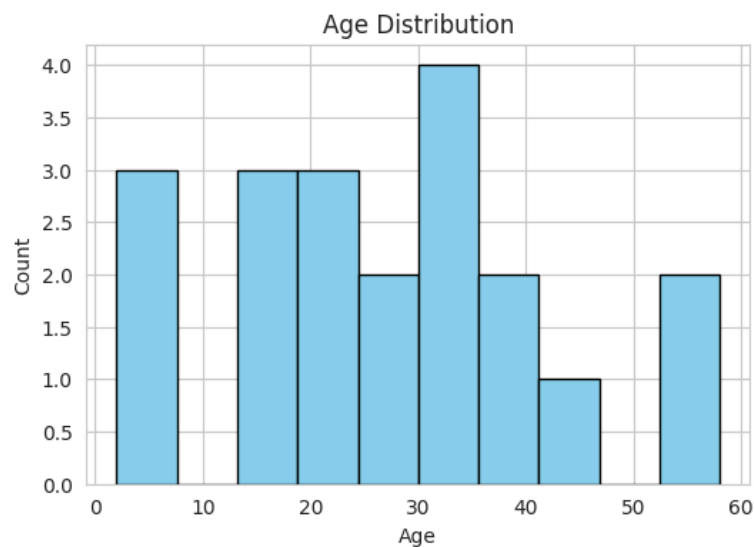
--- Target Variable Counts ---

Survived

0 10

1 10

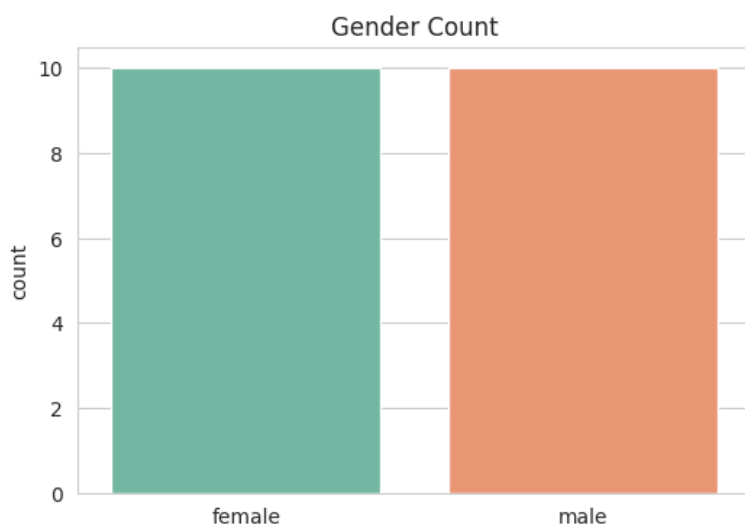
Name: count, dtype: int64



/tmp/ipython-input-1326995002.py:50: FutureWarning

Passing `palette` without assigning `hue` is depre

sns.countplot(x='Sex', data=df, palette='Set2')

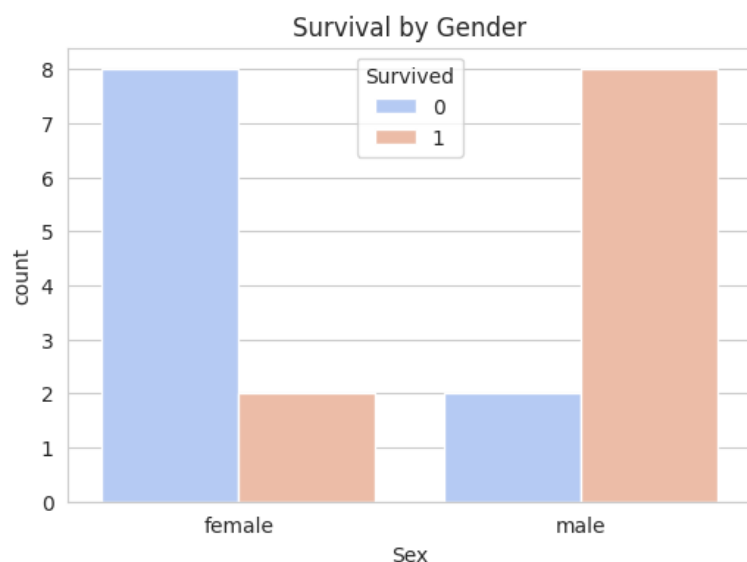
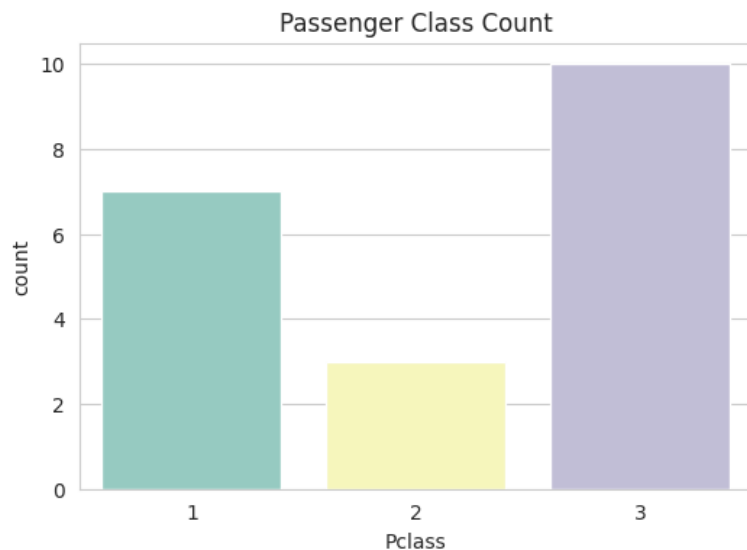


Sex

```
/tmp/ipython-input-1326995002.py:55: FutureWarning
```

```
Passing `palette` without assigning `hue` is depre
```

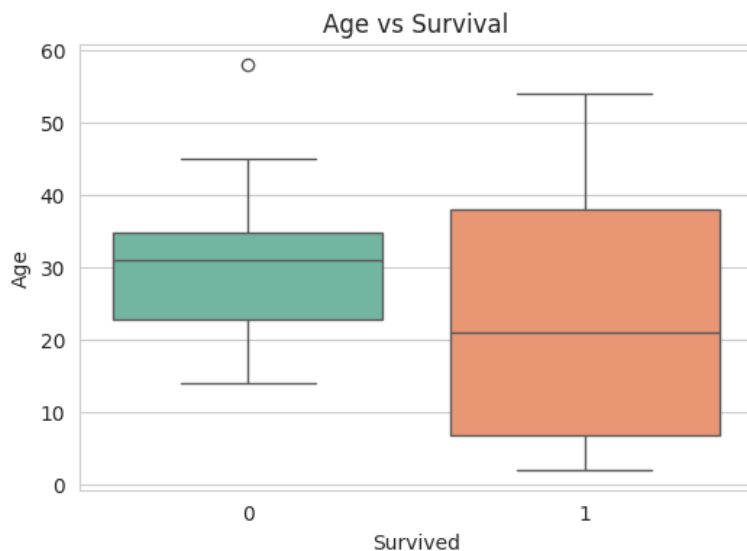
```
sns.countplot(x='Pclass', data=df, palette='Set3
```



```
/tmp/ipython-input-1326995002.py:66: FutureWarning
```

```
Passing `palette` without assigning `hue` is depre
```

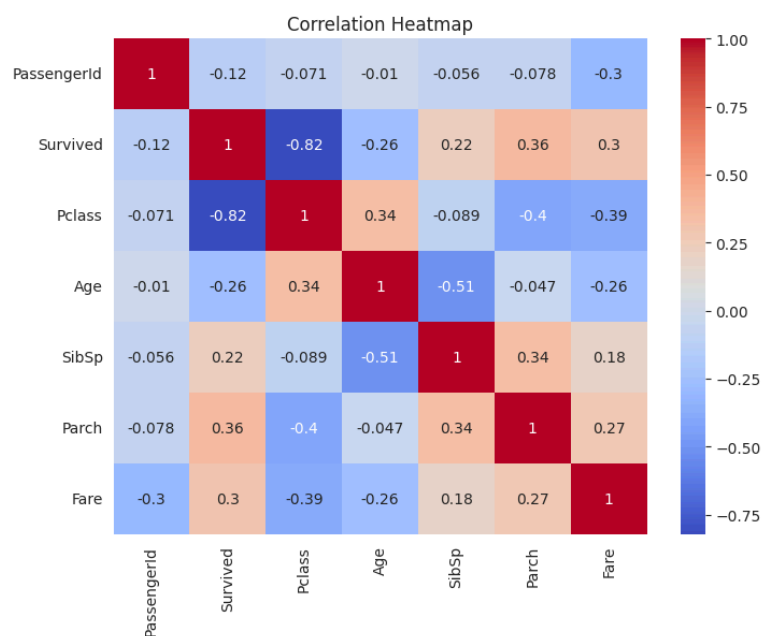
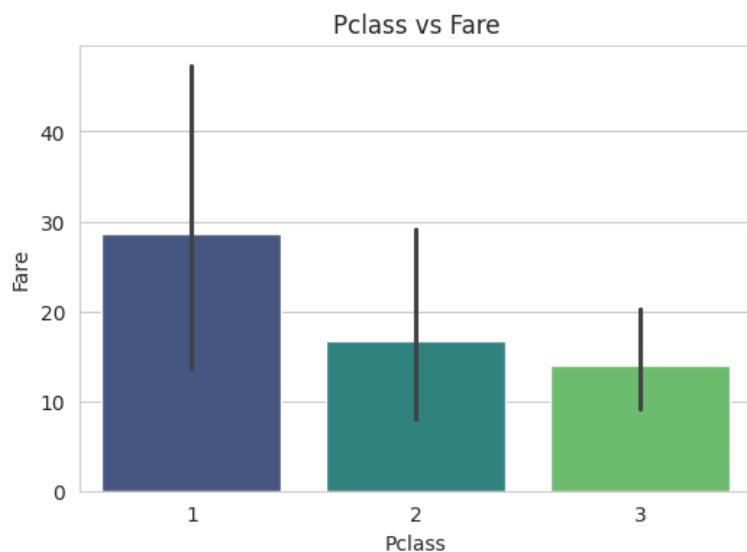
```
sns.boxplot(x='Survived', y='Age', data=df, pale
```



/tmp/ipython-input-1326995002.py:71: FutureWarning

Passing `palette` without assigning `hue` is depre

sns.barplot(x='Pclass', y='Fare', data=df, palet



Summary of EDA Findings:

- Most survivors were female.
- Pclass 1 had a much higher survival rate compare
- Younger passengers had a higher chance of surviv
- Fare is positively correlated with survival – hi

