```
#step 1 import required Libaries
%pip install seaborn
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

#step 2 Load dataset
df = sns.load_dataset('titanic')

# Step 3: Show head of your data
display(df.head())

# Step 4: Show info of your data
df.info()

# Step 5: Missing values
missing_values = df.isnull().sum()
print(missing_values)

# Step 6: Visualize age distribution
plt.figure(figsize=(5, 5))
sns.histplot(df['age'].dropna(), kde=True, bins=30, color='green')
plt.title("Age Distribution of Titanic dataset")
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()

# Step 7: Survivor count by gender
plt.figure(figsize=(5, 5))
sns.countplot(x='sex', hue='survived', data=df)
plt.title("Survivor count by gender")
plt.xlabel('Gender')
plt.ylabel('Count')
plt.legend(title='Survived', labels=['No', 'Yes'])
plt.show()

# Step 8: Survivor count by class
plt.figure(figsize=(5, 5))
sns.countplot(x='pclass', hue='survived', data=df)
plt.title("Survival count by passenger class")
plt.xlabel('Passenger Class')
plt.ylabel('Count')
plt.legend(title='Survived', labels=['No', 'Yes'])
plt.show()

# Step 9: Correlation Matrix
plt.figure(figsize=(8, 6))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm')
```

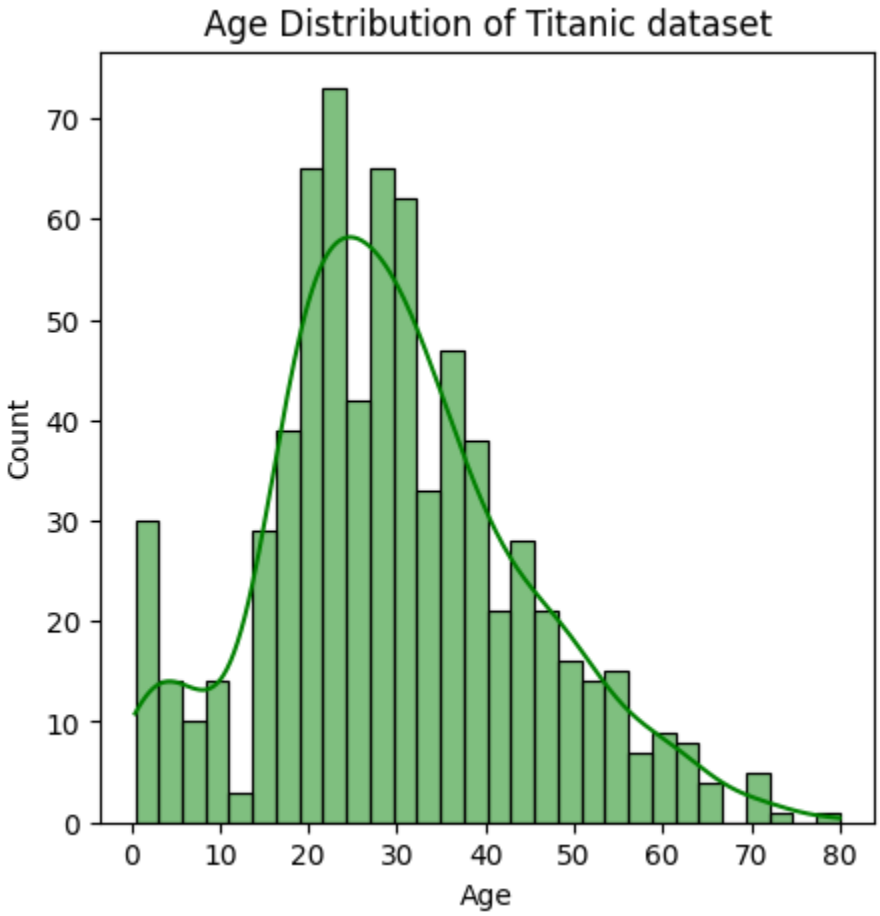
```
plt.title('Correlation Matrix')
plt.show()

# Step 10: Summary
print("Summary of Titanic dataset")
print(df.describe())
```

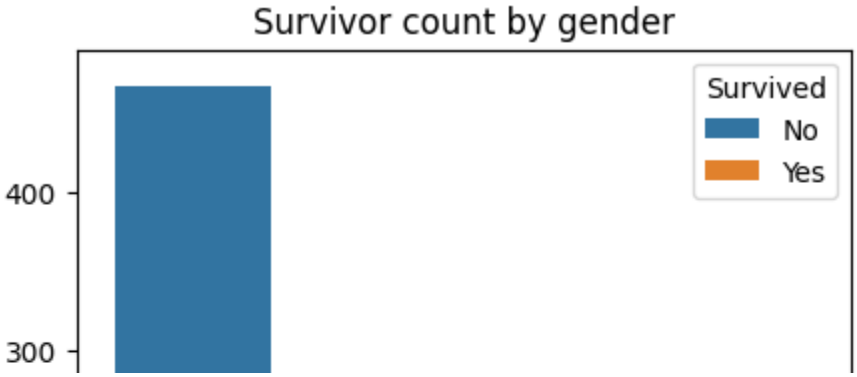
```
Requirement already satisfied: seaborn in /usr/local/lib/python3.11/dist-packages (0.
Requirement already satisfied: numpy!=1.24.0,>=1.20 in /usr/local/lib/python3.11/dist
Requirement already satisfied: pandas>=1.2 in /usr/local/lib/python3.11/dist-packages
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in /usr/local/lib/python3.11/d
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-pac
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.11/dist-package
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-pa
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-pa
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-pack
Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.11/dist-packages (
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-pac
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.11/dist
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-package
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packa
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (f
```

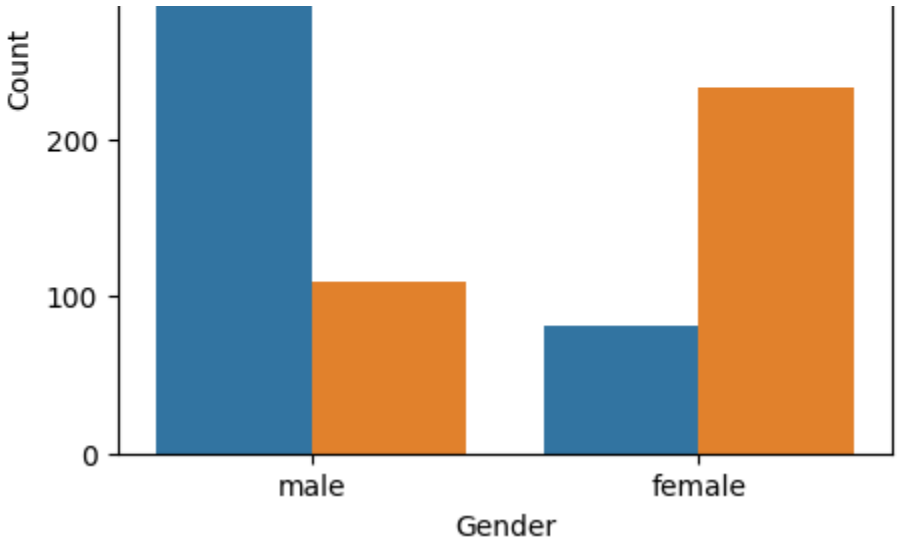
	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult
0	0	3	male	22.0	1	0	7.2500	S	Third	man	
1	1	1	female	38.0	1	0	71.2833	C	First	woman	
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	
3	1	1	female	35.0	1	0	53.1000	S	First	woman	
4	0	3	male	35.0	0	0	8.0500	S	Third	man	

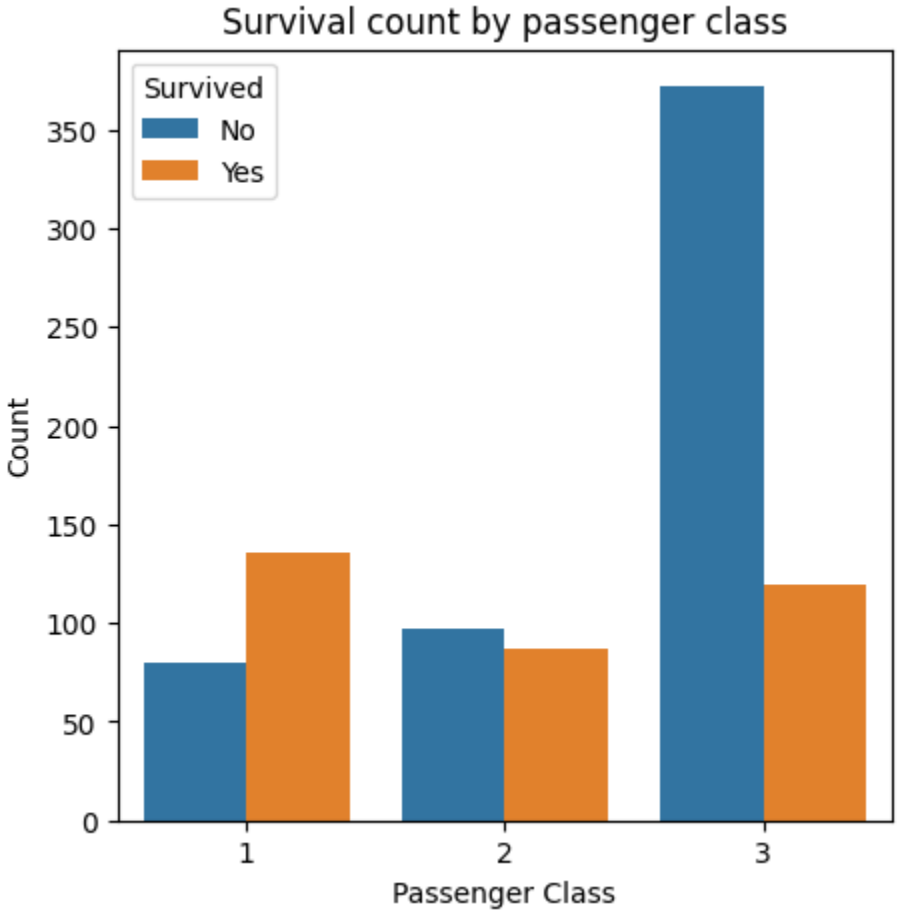
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 15 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   survived     891 non-null    int64
 1   pclass       891 non-null    int64
 2   sex          891 non-null    object
 3   age          714 non-null    float64
 4   sibsp        891 non-null    int64
 5   parch        891 non-null    int64
 6   fare         891 non-null    float64
 7   embarked     889 non-null    object
 8   class        891 non-null    category
 9   who          891 non-null    object
 10  adult_male   891 non-null    bool
 11  deck         203 non-null    category
 12  embark_town  889 non-null    object
 13  alive        891 non-null    object
 14  alone        891 non-null    bool
dtypes: bool(2), category(2), float64(2), int64(4), object(5)
```

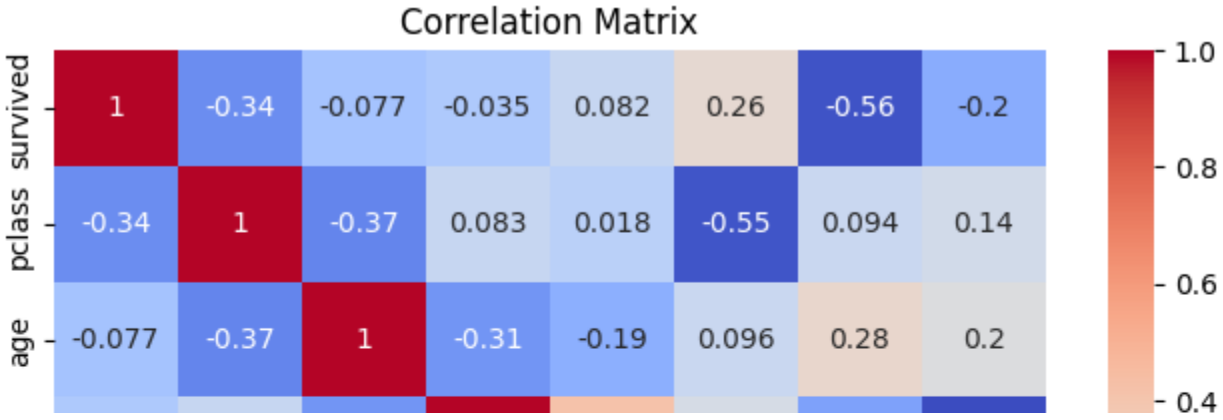
```
memory usage: 80.7+ KB
survived         0
pclass           0
sex              0
age            177
sibsp            0
parch            0
fare             0
embarked         2
class            0
who              0
adult_male       0
deck           688
embark_town      2
alive            0
alone            0
dtype: int64
```

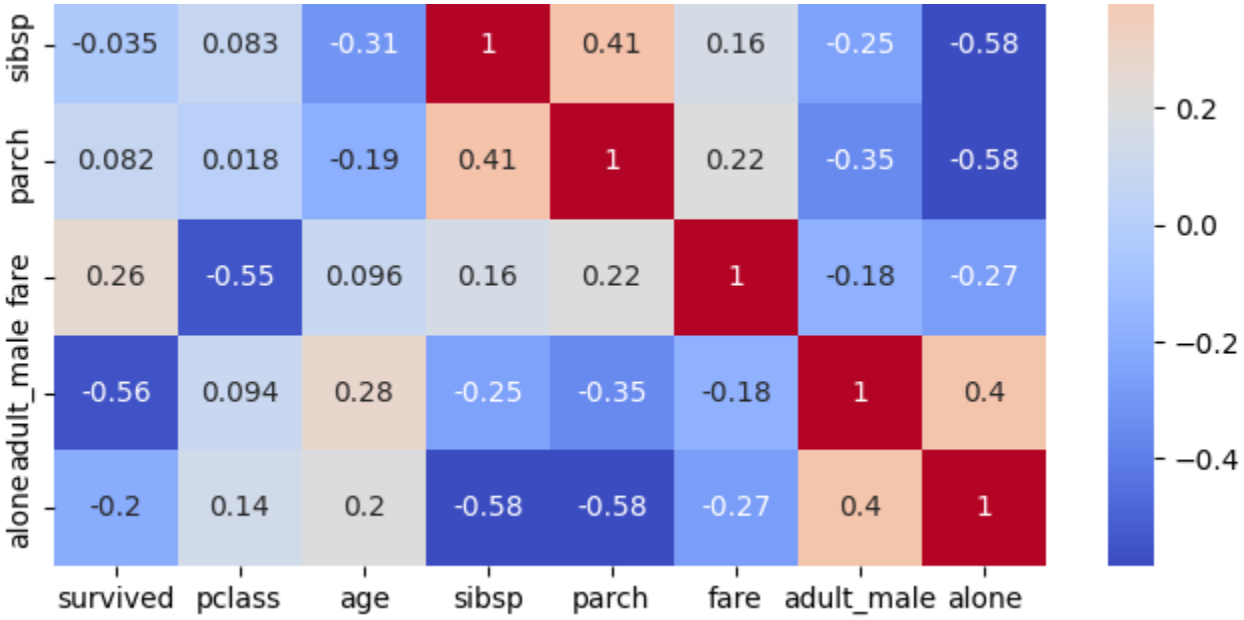





Count
200
100
0
male
female
Gender


Survival count by passenger class
Survived
No
Yes
350
300
250
200
150
100
50
0
Count
1
2
3
Passenger Class


Correlation Matrix
survived
pclass
age
1
-0.34
-0.077
-0.035
0.082
0.26
-0.56
-0.2
-0.34
1
-0.37
0.083
0.018
-0.55
0.094
0.14
-0.077
-0.37
1
-0.31
-0.19
0.096
0.28
0.2
1.0
0.8
0.6
0.4



Summary of Titanic dataset

	survived	pclass	age	sibsp	parch	fare
count	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000