27/06/2025, 18:47 SCT DS 2 - Colab

```
# Import libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
# Load Titanic dataset (you can change the path or use seaborn's built-in one)
df= pd.read_csv('/content/drive/MyDrive/Titanic-Dataset.csv')
# Display the first few rows
print(df.head())
# -----
# Step 1: Data Cleaning
# ------
# Check for missing values
print("\nMissing values:\n", df.isnull().sum())
# Fill missing 'age' with median
df['Age'] = df['Age'].fillna(df['Age'].median())
# Drop rows with missing 'embarked'
df.dropna(subset=['Embarked'], inplace=True)
# Fill 'deck' with 'Unknown'
df['Cabin'] = df['Cabin'].fillna('Unknown')
# -----
# Step 2: EDA - Basic Statistics
# -----
print("\nSummary statistics:\n", df.describe(include='all'))
# -----
# Step 3: Visualizations
# -----
# Survival count plot
sns.countplot(x='Survived', data=df)
plt.title("Survival Count")
plt.xticks([0, 1], ['Not Survived', 'Survived'])
plt.show()
# Gender vs Survival
sns.countplot(x='Sex', hue='Survived', data=df)
nlt title("Survival hy Gender")
```

https://colab.research.google.com/drive/19mbcPFMpA7e-VDYhg 6W9kqySD-YlxqK?authuser=1#printMode=true

```
27/06/2025, 18:47
    plt.xticks(rotation=0)
   plt.legend(['Not Survived', 'Survived'])
   plt.show()
    # Class vs Survival
    sns.countplot(x='Pclass', hue='Survived', data=df)
    plt.title("Survival by Passenger Class")
   plt.legend(['Not Survived', 'Survived'])
    plt.show()
    # Age distribution
   sns.histplot(df['Age'], bins=30, kde=True)
   plt.title("Age Distribution of Passengers")
   plt.show()
   # Heatmap for correlation
   plt.figure(figsize=(10, 6))
   sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm')
```

plt.title("Correlation Heatmap")

plt.show()

27/06/2025, 18:47 SCT DS 2 - Colab

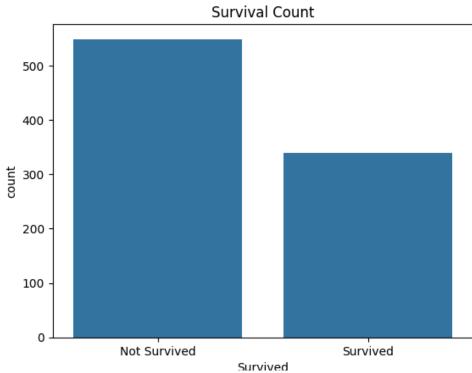
→

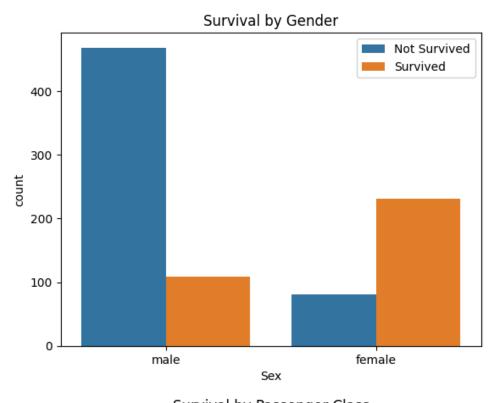
```
PassengerId Survived Pclass \
0
             1
                        0
                                3
             2
1
                        1
                                1
2
             3
                        1
                                3
3
             4
                        1
                                1
4
             5
                        0
                                3
                                                  Name
                                                           Sex
                                                                 Age SibSp \
0
                              Braund, Mr. Owen Harris
                                                          male
                                                                22.0
                                                        female
1
   Cumings, Mrs. John Bradley (Florence Briggs Th...
                                                                38.0
                                                                           1
2
                                                                           0
                               Heikkinen, Miss. Laina
                                                        female
                                                                26.0
3
        Futrelle, Mrs. Jacques Heath (Lily May Peel)
                                                               35.0
                                                        female
                                                                           1
4
                             Allen, Mr. William Henry
                                                          male 35.0
                                                                           0
   Parch
                     Ticket
                                Fare Cabin Embarked
0
       0
                 A/5 21171
                              7.2500
                                       NaN
1
                  PC 17599 71.2833
                                       C85
                                                   C
2
          STON/02. 3101282
                              7.9250
                                       NaN
                                                   S
3
                                                   S
                     113803
                             53.1000
                                      C123
                                                   S
4
       0
                     373450
                              8.0500
                                       NaN
Missing values:
PassengerId
                  0
Survived
                 0
Pclass
                 0
Name
                 0
Sex
                 0
               177
Age
SibSp
                 0
Parch
                 0
Ticket
                 0
Fare
                 0
Cabin
               687
                 2
Embarked
dtype: int64
Summary statistics:
         PassengerId
                         Survived
                                       Pclass
                                                                       Sex \
                                                               Name
         889.000000
                     889.000000
                                  889.000000
                                                               889
                                                                      889
count
unique
                NaN
                             NaN
                                          NaN
                                                                889
                                                                        2
top
                NaN
                             NaN
                                          NaN
                                               Dooley, Mr. Patrick
                                                                     male
freq
                NaN
                             NaN
                                          NaN
                                                                  1
                                                                      577
mean
         446.000000
                        0.382452
                                    2.311586
                                                               NaN
                                                                      NaN
std
         256.998173
                        0.486260
                                    0.834700
                                                                      NaN
                                                               NaN
min
           1.000000
                        0.000000
                                    1.000000
                                                                      NaN
                                                               NaN
25%
         224.000000
                        0.000000
                                    2.000000
                                                               NaN
                                                                      NaN
50%
         446.000000
                        0.000000
                                     3.000000
                                                               NaN
                                                                      NaN
75%
         668.000000
                        1.000000
                                     3.000000
                                                               NaN
                                                                      NaN
         891.000000
                        1.000000
                                     3.000000
                                                               NaN
                                                                      NaN
max
```

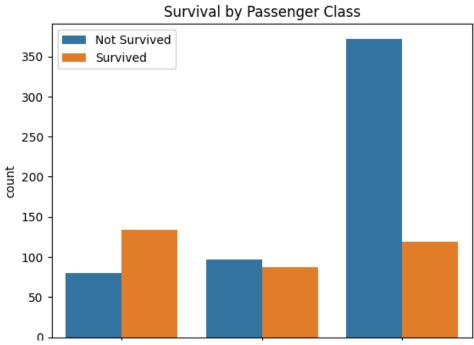
27/06/2025, 18:47 SCT_DS_2 - Colab

	Age	SibSp	Parch	Ticket	Fare	Cabin
count	889.000000	889.000000	889.000000	889	889.000000	889
unique	NaN	NaN	NaN	680	NaN	147
top	NaN	NaN	NaN	CA. 2343	NaN	Unknown
freq	NaN	NaN	NaN	7	NaN	687
mean	29.315152	0.524184	0.382452	NaN	32.096681	NaN
std	12.984932	1.103705	0.806761	NaN	49.697504	NaN
min	0.420000	0.000000	0.000000	NaN	0.000000	NaN
25%	22.000000	0.000000	0.000000	NaN	7.895800	NaN
50%	28.000000	0.000000	0.000000	NaN	14.454200	NaN
75%	35.000000	1.000000	0.000000	NaN	31.000000	NaN
max	80.000000	8.000000	6.000000	NaN	512.329200	NaN

	Embarked
count	889
unique	3
top	S
freq	644
mean	NaN
std	NaN
min	NaN
25%	NaN
50%	NaN
75%	NaN
max	NaN







27/06/2025, 18:47 SCT_DS_2 - Colab

1 2 Pclass

3

80

1.0

- 0.8

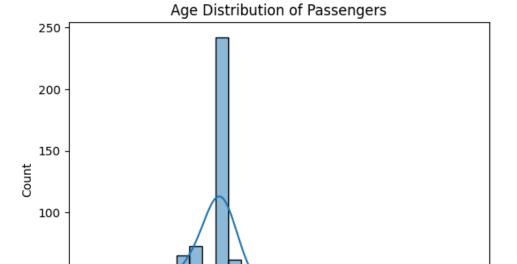
- 0.6

- 0.4

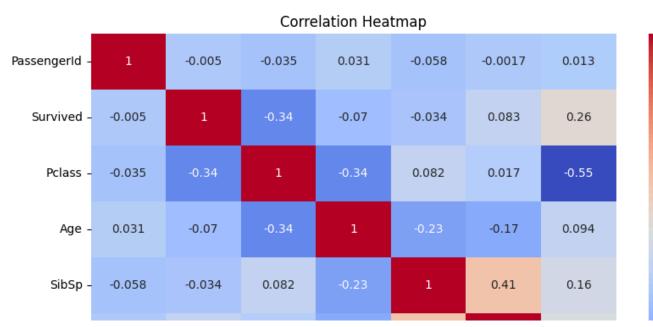
- 0.2

- 0.0

70



50



50

60

40

Age

20

10

30

27/06/2025, 18:47 SCT_DS_2 - Colab

