ELEVATE LABS DATA ANALYST INTERN TASK ON 14th APRIL 2025

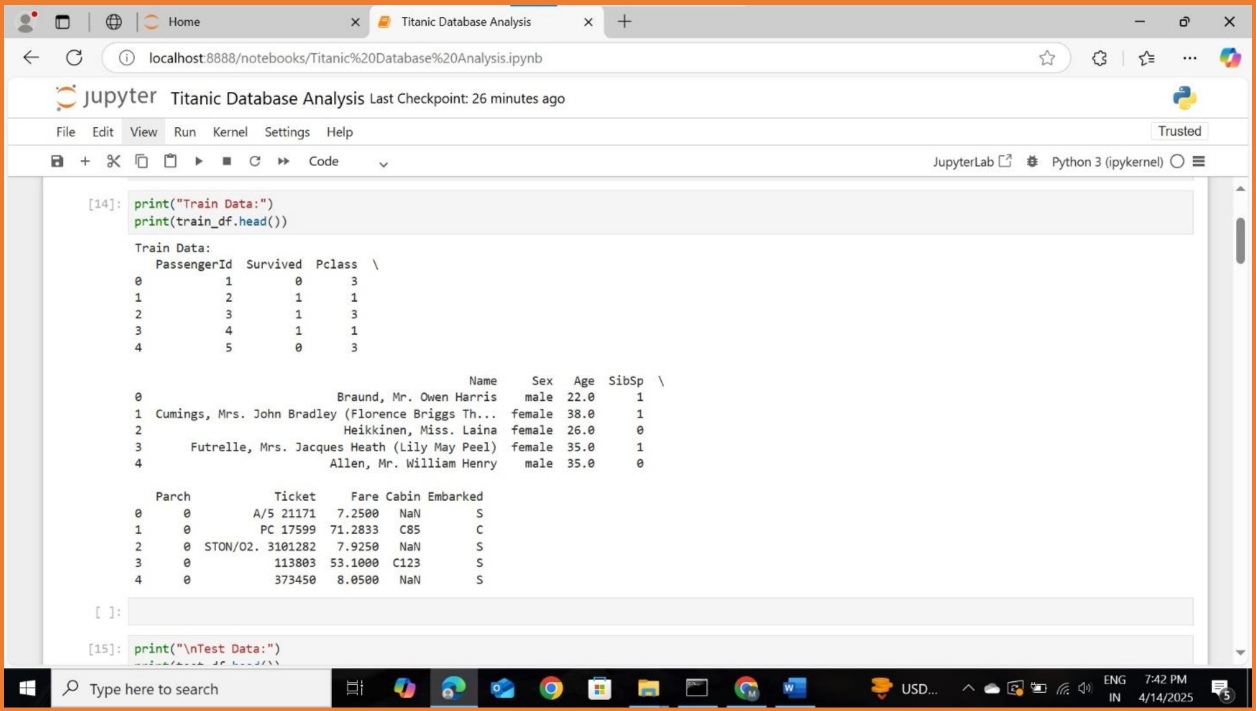
Got it! If your CSV files are located at:

C:\Users\LENOVO\Desktop\Elevate Labs Internship tasks\Day5

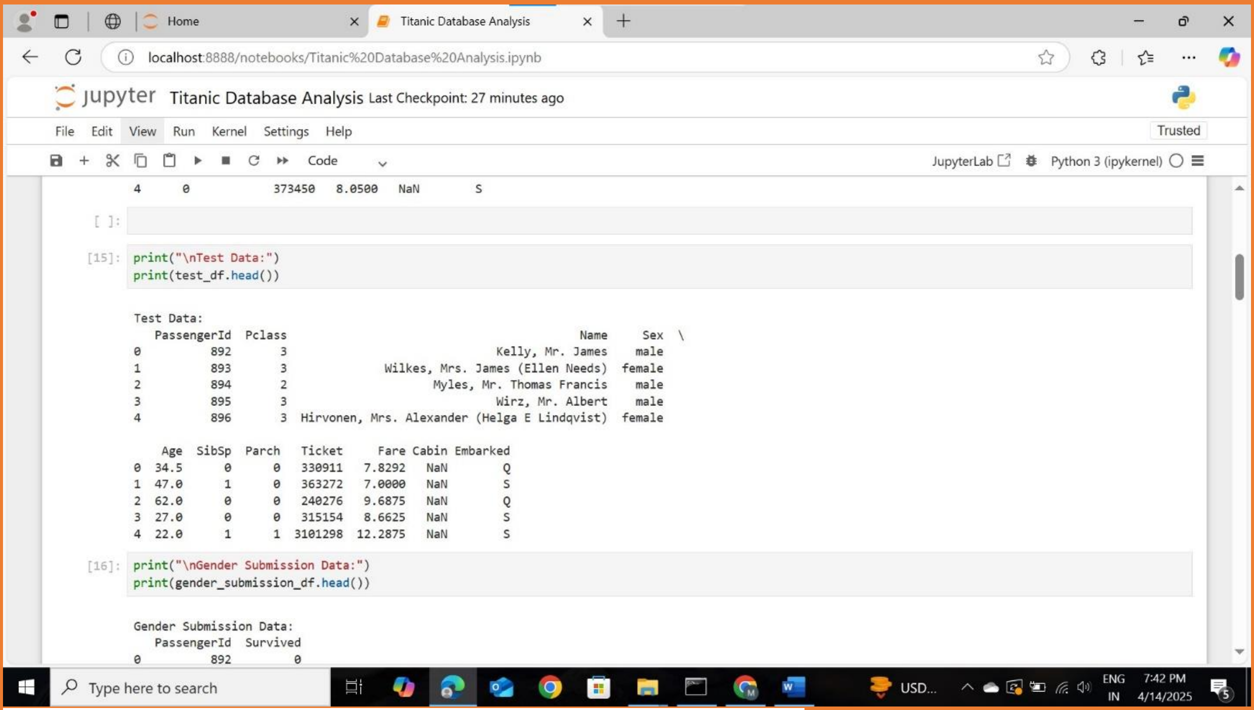
Here’s the Python code to load them in a Jupyter Notebook running on your local machine:

Python

```
import pandas as pd

# Base path to the CSV files
base_path = r'C:\Users\LENOVO\Desktop\Elevate Labs Internship tasks\Day5'

# Read the CSV files
train_df = pd.read_csv(f'{base_path}\\train.csv')
test_df = pd.read_csv(f'{base_path}\\test.csv')
gender_submission_df = pd.read_csv(f'{base_path}\\gender_submission.csv')
```


Home
Titanic Database Analysis
localhost:8888/notebooks/Titanic%20Database%20Analysis.ipynb
jupyter Titanic Database Analysis Last Checkpoint: 26 minutes ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
Type here to search
USD...
ENG IN
7:42 PM 4/14/2025

```
[14]: print("Train Data:")
      print(train_df.head())

Train Data:
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         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      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2                             Heikkinen, Miss. Laina  female  26.0      0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                           Allen, Mr. William Henry    male  35.0      0

   Parch            Ticket     Fare Cabin Embarked
0      0         A/5 21171   7.2500   NaN        S
1      0          PC 17599  71.2833   C85        C
2      0  STON/O2. 3101282   7.9250   NaN        S
3      0            113803  53.1000  C123        S
4      0            373450   8.0500   NaN        S

[ ]:

[15]: print("\nTest Data:")
```


Home
Titanic Database Analysis
localhost:8888/notebooks/Titanic%20Database%20Analysis.ipynb
Jupyter
Titanic Database Analysis Last Checkpoint: 27 minutes ago
File
Edit
View
Run
Kernel
Settings
Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
Type here to search
USD...
ENG
IN
7:42 PM
4/14/2025

```
4    0          373450   8.0500   NaN        S
```

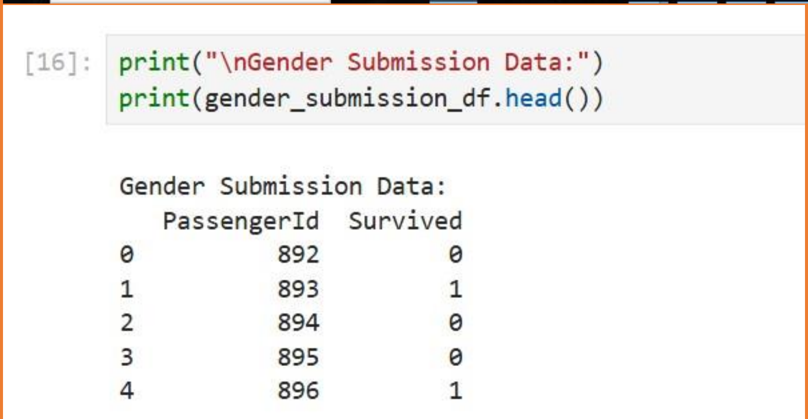
```
[ ]:
```

```
[15]: print("\nTest Data:")
      print(test_df.head())
```

```
Test Data:
   PassengerId  Pclass                                          Name     Sex  \
0          892       3                              Kelly, Mr. James    male   
1          893       3              Wilkes, Mrs. James (Ellen Needs)  female   
2          894       2                     Myles, Mr. Thomas Francis    male   
3          895       3                              Wirz, Mr. Albert    male   
4          896       3  Hirvonen, Mrs. Alexander (Helga E Lindqvist)  female   

    Age  SibSp  Parch   Ticket     Fare Cabin Embarked  
0  34.5      0      0   330911   7.8292   NaN        Q  
1  47.0      1      0   363272   7.0000   NaN        S  
2  62.0      0      0   240276   9.6875   NaN        Q  
3  27.0      0      0   315154   8.6625   NaN        S  
4  22.0      1      1  3101298  12.2875   NaN        S  
```

```
[16]: print("\nGender Submission Data:")
      print(gender_submission_df.head())
```

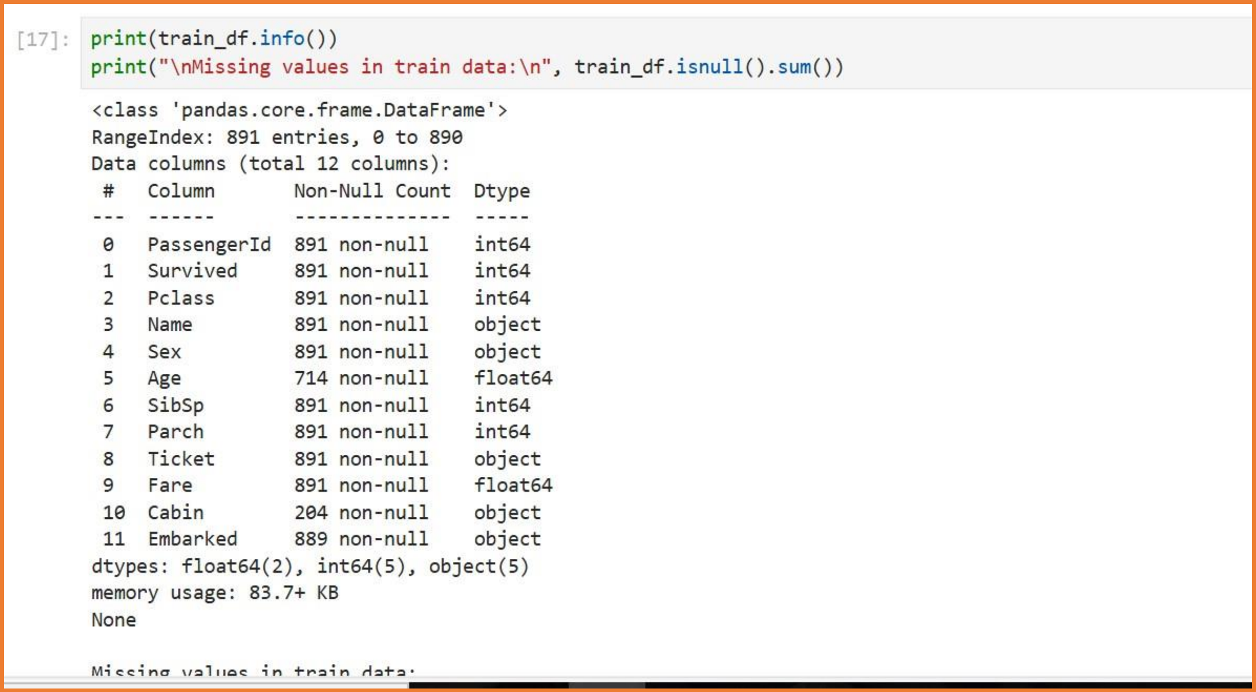
```
Gender Submission Data:
   PassengerId  Survived
0          892         0
```



```
[16]: print("\nGender Submission Data:")
      print(gender_submission_df.head())
```

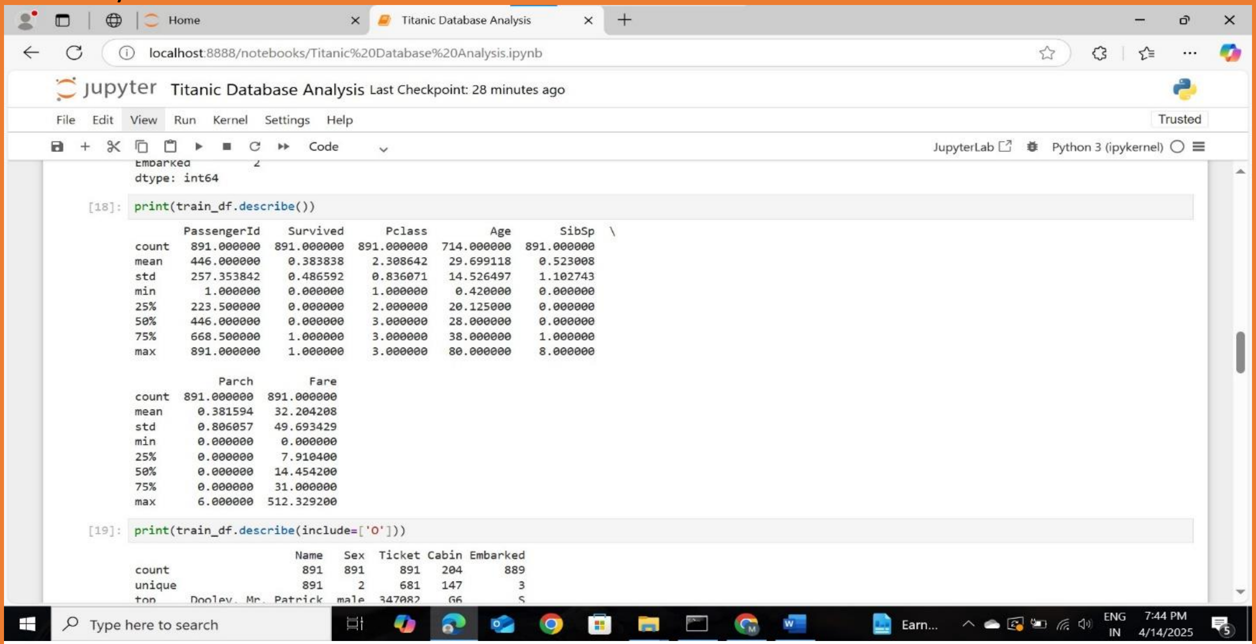
```
Gender Submission Data:
   PassengerId  Survived
0          892         0
1          893         1
2          894         0
3          895         0
4          896         1
```

Basic Info and Missing Values



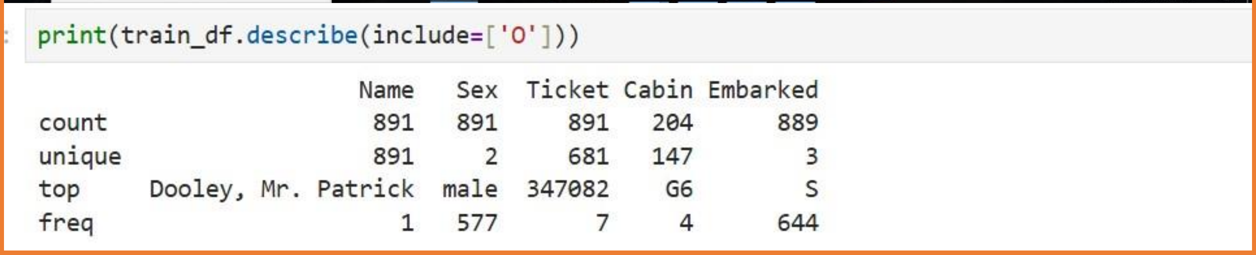
```
[17]: print(train_df.info())
      print("\nMissing values in train data:\n", train_df.isnull().sum())

      <class 'pandas.core.frame.DataFrame'>
      RangeIndex: 891 entries, 0 to 890
      Data columns (total 12 columns):
       #   Column       Non-Null Count  Dtype
      ---  ------       --------------  -----
       0   PassengerId  891 non-null    int64
       1   Survived     891 non-null    int64
       2   Pclass       891 non-null    int64
       3   Name         891 non-null    object
       4   Sex          891 non-null    object
       5   Age          714 non-null    float64
       6   SibSp        891 non-null    int64
       7   Parch        891 non-null    int64
       8   Ticket       891 non-null    object
       9   Fare         891 non-null    float64
       10  Cabin        204 non-null    object
       11  Embarked     889 non-null    object
      dtypes: float64(2), int64(5), object(5)
      memory usage: 83.7+ KB
      None

      Missing values in train data:
```

Summary Statistics



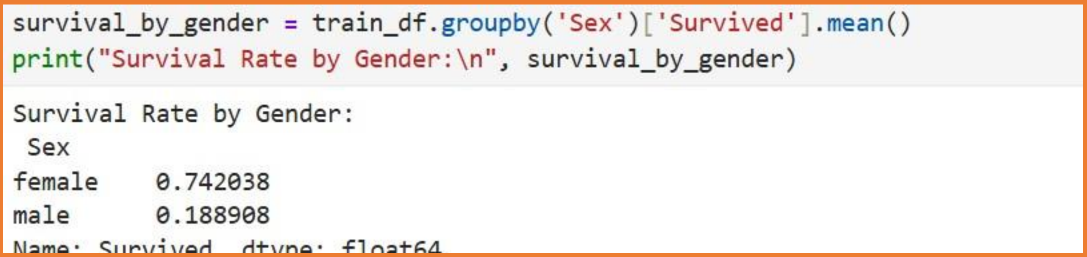
Home
Titanic Database Analysis
localhost:8888/notebooks/Titanic%20Database%20Analysis.ipynb
jupyter Titanic Database Analysis Last Checkpoint: 28 minutes ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
Type here to search
Earn...
ENG IN
7:44 PM
4/14/2025

```
Embarked         2
dtype: int64

[18]: print(train_df.describe())

       PassengerId    Survived      Pclass         Age       SibSp  \
count   891.000000  891.000000  891.000000  714.000000  891.000000
mean    446.000000    0.383838    2.308642   29.699118    0.523008
std     257.353842    0.486592    0.836071   14.526497    1.102743
min       1.000000    0.000000    1.000000    0.420000    0.000000
25%     223.500000    0.000000    2.000000   20.125000    0.000000
50%     446.000000    0.000000    3.000000   28.000000    0.000000
75%     668.500000    1.000000    3.000000   38.000000    1.000000
max     891.000000    1.000000    3.000000   80.000000    8.000000

            Parch        Fare
count  891.000000  891.000000
mean     0.381594   32.204208
std      0.806057   49.693429
min      0.000000    0.000000
25%      0.000000    7.910400
50%      0.000000   14.454200
75%      0.000000   31.000000
max      6.000000  512.329200

[19]: print(train_df.describe(include=['O']))

                           Name   Sex  Ticket Cabin Embarked
count                       891   891     891   204      889
unique                      891     2     681   147        3
top     Dooley, Mr. Patrick  male  347082    G6        S
```



```
print(train_df.describe(include=['O']))

                           Name   Sex  Ticket Cabin Embarked
count                       891   891     891   204      889
unique                      891     2     681   147        3
top         Dooley, Mr. Patrick  male  347082    G6        S
freq                          1   577       7     4      644
```

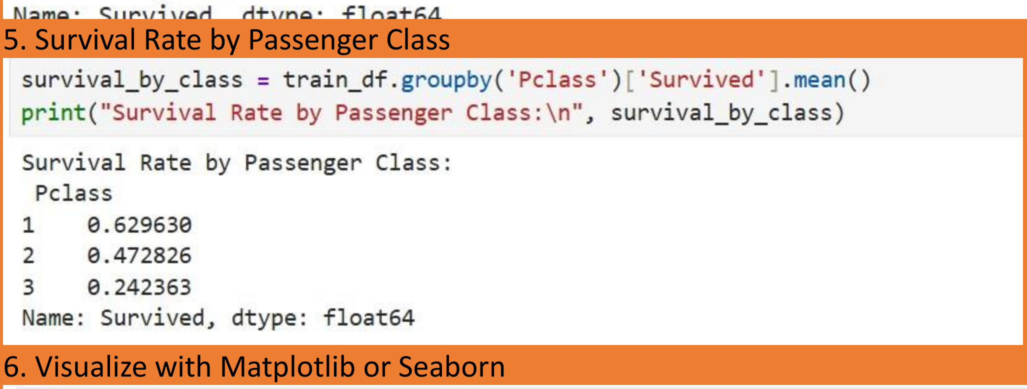
Survival Rate by Gender



```
survival_by_gender = train_df.groupby('Sex')['Survived'].mean()
print("Survival Rate by Gender:\n", survival_by_gender)
```

```
Survival Rate by Gender:
 Sex
female    0.742038
male      0.188908
Name: Survived, dtype: float64
```

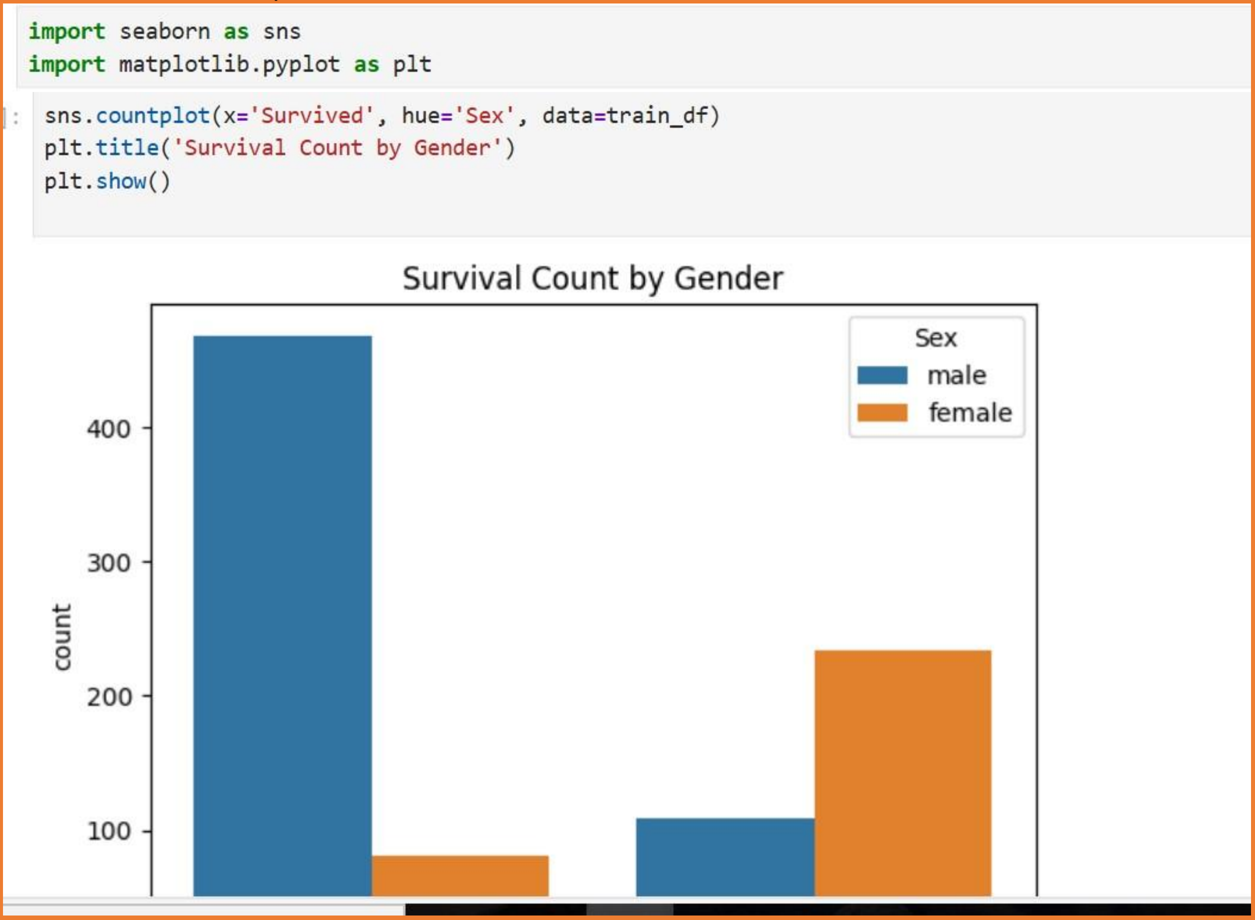
5. Survival Rate by Passenger Class



```
survival_by_class = train_df.groupby('Pclass')['Survived'].mean()
print("Survival Rate by Passenger Class:\n", survival_by_class)
```

```
Survival Rate by Passenger Class:
 Pclass
1    0.629630
2    0.472826
3    0.242363
Name: Survived, dtype: float64
```

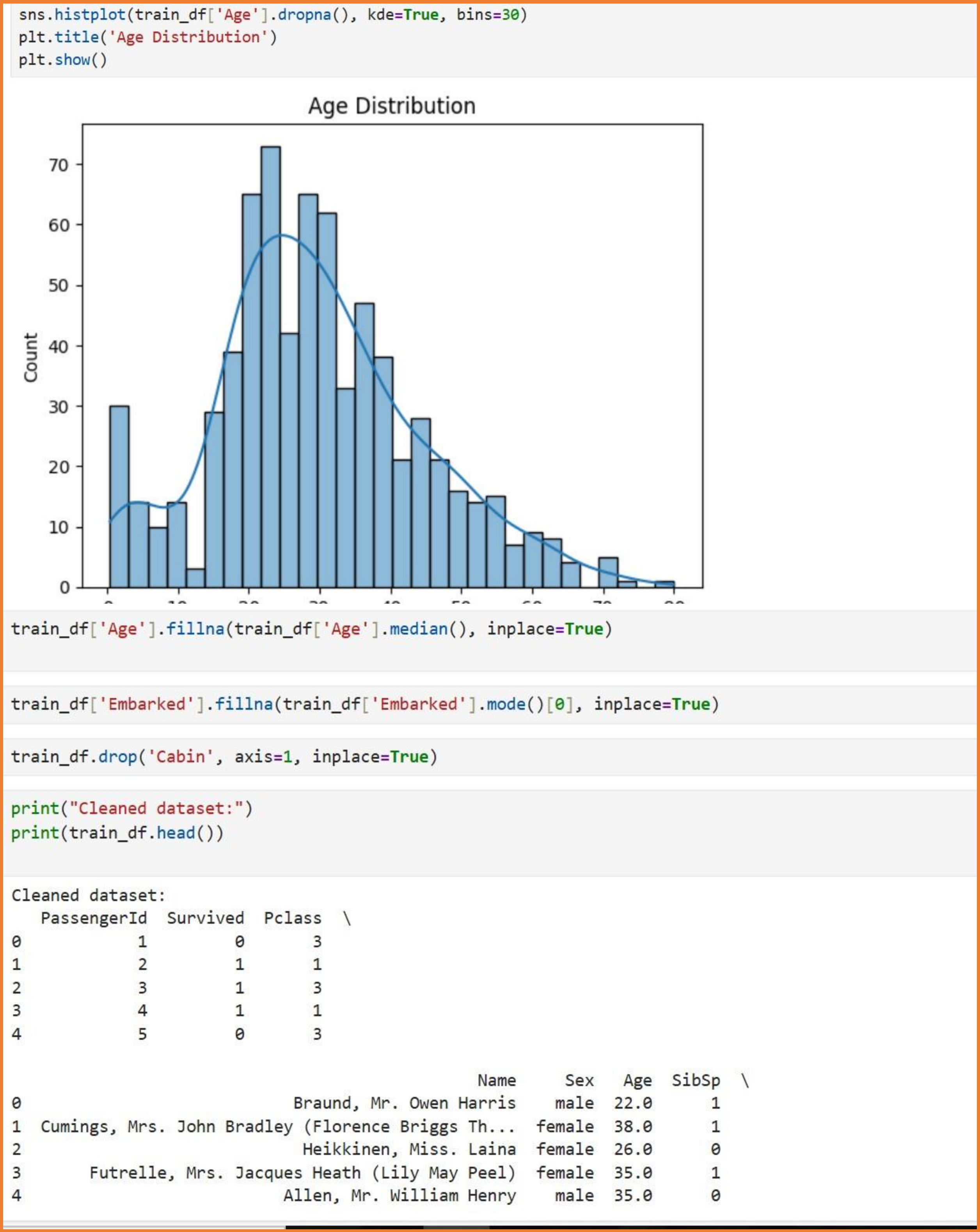
6. Visualize with Matplotlib or Seaborn



```
import seaborn as sns
import matplotlib.pyplot as plt
```

```
sns.countplot(x='Survived', hue='Sex', data=train_df)
plt.title('Survival Count by Gender')
plt.show()
```

7. Clean and Prepare for Modeling (Optional Preview)



```
sns.histplot(train_df['Age'].dropna(), kde=True, bins=30)
plt.title('Age Distribution')
plt.show()
```

```
train_df['Age'].fillna(train_df['Age'].median(), inplace=True)
```

```
train_df['Embarked'].fillna(train_df['Embarked'].mode()[0], inplace=True)
```

```
train_df.drop('Cabin', axis=1, inplace=True)
```

```
print("Cleaned dataset:")
print(train_df.head())
```

```
Cleaned dataset:
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         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      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2                             Heikkinen, Miss. Laina  female  26.0      0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                           Allen, Mr. William Henry    male  35.0      0
```