**1 import pandas as pd**

**5/7/24, 6:43 PM Assignment 9 .ipynb - Colab**

**import numpy as np**

**import seaborn as sns**

**import matplotlib.pyplot as plt**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **data-pd.read\_csv(Vcontent/Titanic-Dataset.csv') df=pd.DataFrame(data)**  **df**  **Passengerld Survived Pclass Name**  **Braund,**  0 **1 0 3 Mr. Owen**  **Harris**  **Cumings,**  **Mrs. John**  **Bradley**  1 **2 1 1**  **(Florence**  **Briggs  Th...  Heikkinen,**  **2 3 1 3 Miss.**  **Laina**  **Futrelle,**  **Mrs.**  **Jacques**  **3 4 1 1**  **Heath**  **(Lily May**  **Peel)** | **Sex**  **male**  **female  female  female** | **Age**  **22.0**  **38.0  26.0  35.0** | **SibSp**  **1**  **1  0  1** | **Parch**  **0**  **0  0  0** | **Ticket N5 21171 PC 17599**  **STON/02.**  **3101282**  **113803** | **Fa**  **7.2E**  **71.2E 7.92 53.1C** |

**df.head()**

**Passengerld Survived Pclass Name Sex Age SibSp Parch**

**1 0 3 Braund, M r. Owen Harris male 22.0 1 0**

**Cumings, Mrs. John Bradley (Florence**

**2 1 1 female 38.0 1 0**

**Briggs Th...**

**3 1 3 Heikkinen, Miss. Laina female 26.0 0 0**

**Futrelle, Mrs. Jacques Heath (Lily May**

**4 1 1 female 35.0 1 0**

**Peel)**

0

1

**2**

**3**

|  |  |  |  |
| --- | --- | --- | --- |
| **Ticket** | **Fare** | **Cabin** | **Embarked** |
| *AA5* **21171** | **7.2500** | **NaN** | **S** |
| **PC 17599** | **71.2833** | **C85** | **C** |
| **STON/02.**  **3101282** | **7.9250** | **NaN** |  |
| **113803** | **53.1000** | **C123** |  |

**df.tail()**

**Passengerld Survived Pclass Name Sex Age SibSp**

**886 887 0 2 Montvila, Rev. Juozas male 27.0 0**

**887 888 1 1 Graham, Miss. Margaret Edith female 19.0 0**

**888 889 0 3 Johnston, Miss. Catherine Helen "Carrie" female NaN 1**

**889 890 1 1 Behr, Mr. Karl Howell male 26.0 0**

**890 891 0 3 Dooley, W. Patrick male 32.0 0**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Parch** | **Ticket** | **Fare** | **Cabin** | **Embarked** |
| **0** | **211536** | **13.00** | **NaN** | **S** |
| **0** | **112053** | **30.00** | **B42** | **S** |
| **2** | *W./C.* **6607** | **23.45** | **NaN** | **S** |
| **0** | **111369** | **30.00** | **C148** | **C** |
| **0** | **370376** | **7.75** | **NaN** | **Q** |

**df.describe()**

[**https://colab.research.google.com/drive/1FUv\_eM14-Ze7ikh-vdMVX-5asJTp\_h3L#printMode=true**](https://colab.research.google.com/drive/1FUv_eM14-Ze7ikh-vdMVX-5asJTp_h3L#printMode=true) **1/7**

[**https://colab.research.google.com/drive/1FUv\_eM14-Ze7ikh-vdMVX-5asJTp\_h3L#printMode=true**](https://colab.research.google.com/drive/1FUv_eM14-Ze7ikh-vdMVX-5asJTp_h3L#printMode=true) **2/7**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **5/7/24, 6:43 PM** | **Passengerld** | **Survived** | **Pclass** | **Age** | **Assignment 9 .ipynb - Colab**  **SibSp Parch Fare** | | |
| **count** | 891.000000 | 891.000000 | 891.000000 | 714.000000 | 891.000000 | 891.000000 | 891.000000 |
| **mean** | 446.000000 | 0.383838 | 2.308642 | 29.699118 | 0.523008 | 0.381594 | 32.204208 |
| **std** | 257.353842 | 0.486592 | 0.836071 | 14.526497 | 1.102743 | 0.806057 | 49.693429 |
| **min** | 1.000000 | 0.000000 | 1.000000 | 0.420000 | 0.000000 | 0.000000 | 0.000000 |
| **25%** | 223.500000 | 0.000000 | 2.000000 | 20.125000 | 0.000000 | 0.000000 | 7.910400 |
| **50%** | 446.000000 | 0.000000 | 3.000000 | 28.000000 | 0.000000 | 0.000000 | 14.454200 |
| **75%** | 668.500000 | 1.000000 | 3.000000 | 38.000000 | 1.000000 | 0.000000 | 31.000000 |
| **max** | 891.000000 | 1.000000 | 3.000000 | 80.000000 | 8.000000 | 6.000000 | 512.329200 |
| df.info() |  |  |  |  |  |  |  |

<class 'pandas.core.frame.DataFrame'> Rangelndex: 891 entries, 0 to 890 Data columns (total 12 columns):

# Column Non-Null Count Dtype

|  |  |  |  |
| --- | --- | --- | --- |
| 0 | Passengerld | 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) |

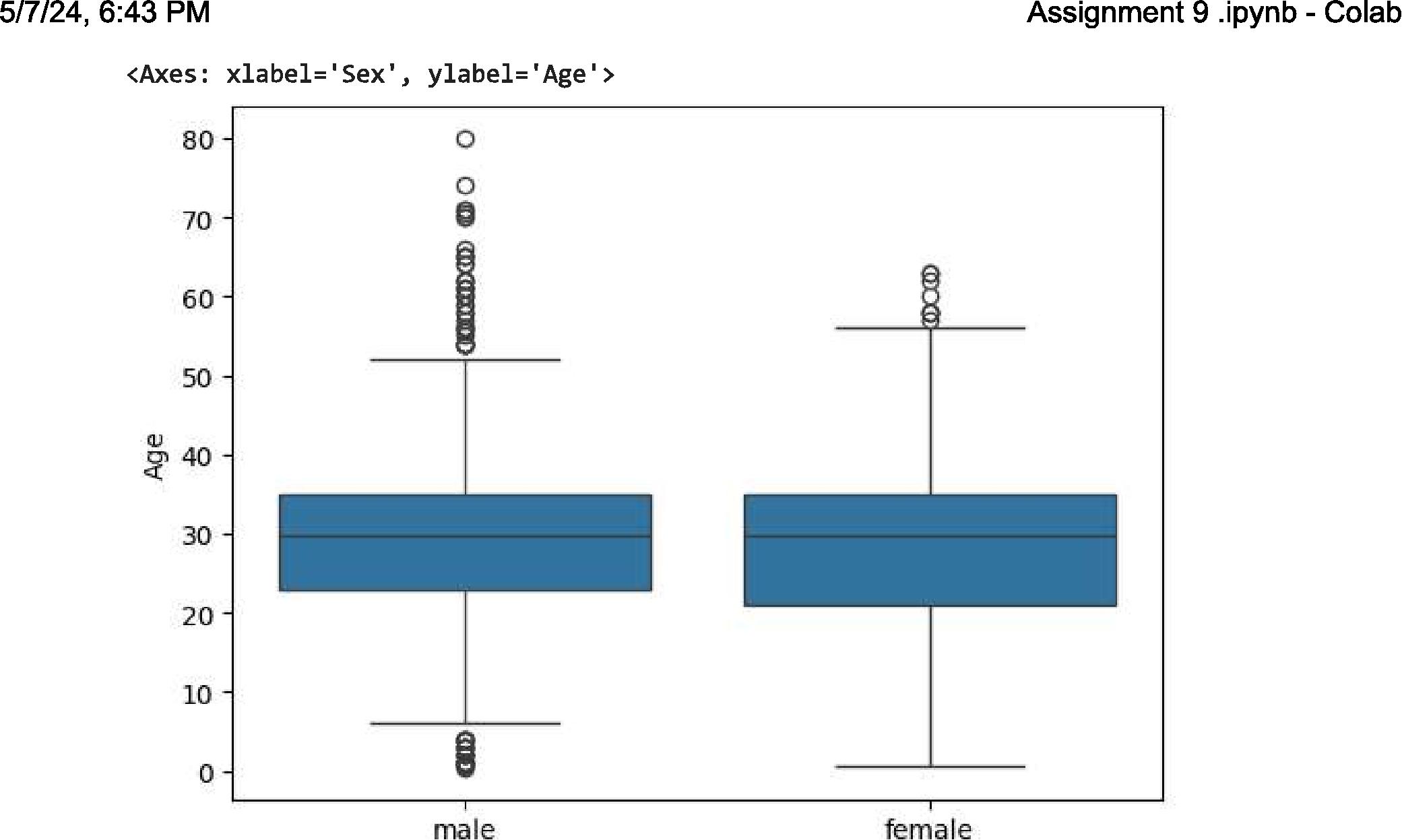
memory usage: 83.7+ KB

|  |  |
| --- | --- |
| df.isnull().sum() |  |
| Passengerld | **0** |
| Survived | **0** |
| Pclass | **0** |
| Name | **0** |
| Sex | **0** |
| Age | **0** |
| SibSp | **0** |
| Parch | **0** |
| Ticket | **0** |
| Fare | **0** |
| Cabin | 687 |
| Embarked dtype: int64 | 2 |

|  |  |
| --- | --- |
| dff'Cabin']=d-F['Cabin'].fillna(dfrCabinl.mode()[0])  df['Embarked']=df['Embarked'].fillna(df['Embarked'].mode()[0]) df.isnull().sum() | |
| Passengerld | 0 |
| Survived | **0** |
| Pclass | **0** |
| Name | **0** |
| Sex | **0** |
| Age | **0** |
| SibSp | **0** |
| Parch | **0** |
| Ticket | **0** |
| Fare | **0** |
| Cabin | **0** |
| Embarked dtype: int64 | **0** |

sns.boxplot(x=df['Sex'],y=df['Age'])

Sex



**5/7/24, 6:43 PM**

<Axes: xlabel='Sex', ylabel='Age'>

80-

70 - 60 - 50 - 40 - 30 -

20 -

10-

0-

male

**Assignment 9 .ipynb - Colab**

female

**male** female

sns.boxplot(x=df['Sex'],y=df['Age'],hue=df['Survived']) <Axes: xlabel='Sex', ylabel='Age'>

80-

70-

Survived
  
1111 0

**1===3 1**

60 -



8

•
  
•

50-

40

20 -

30-

10-

0 -

8

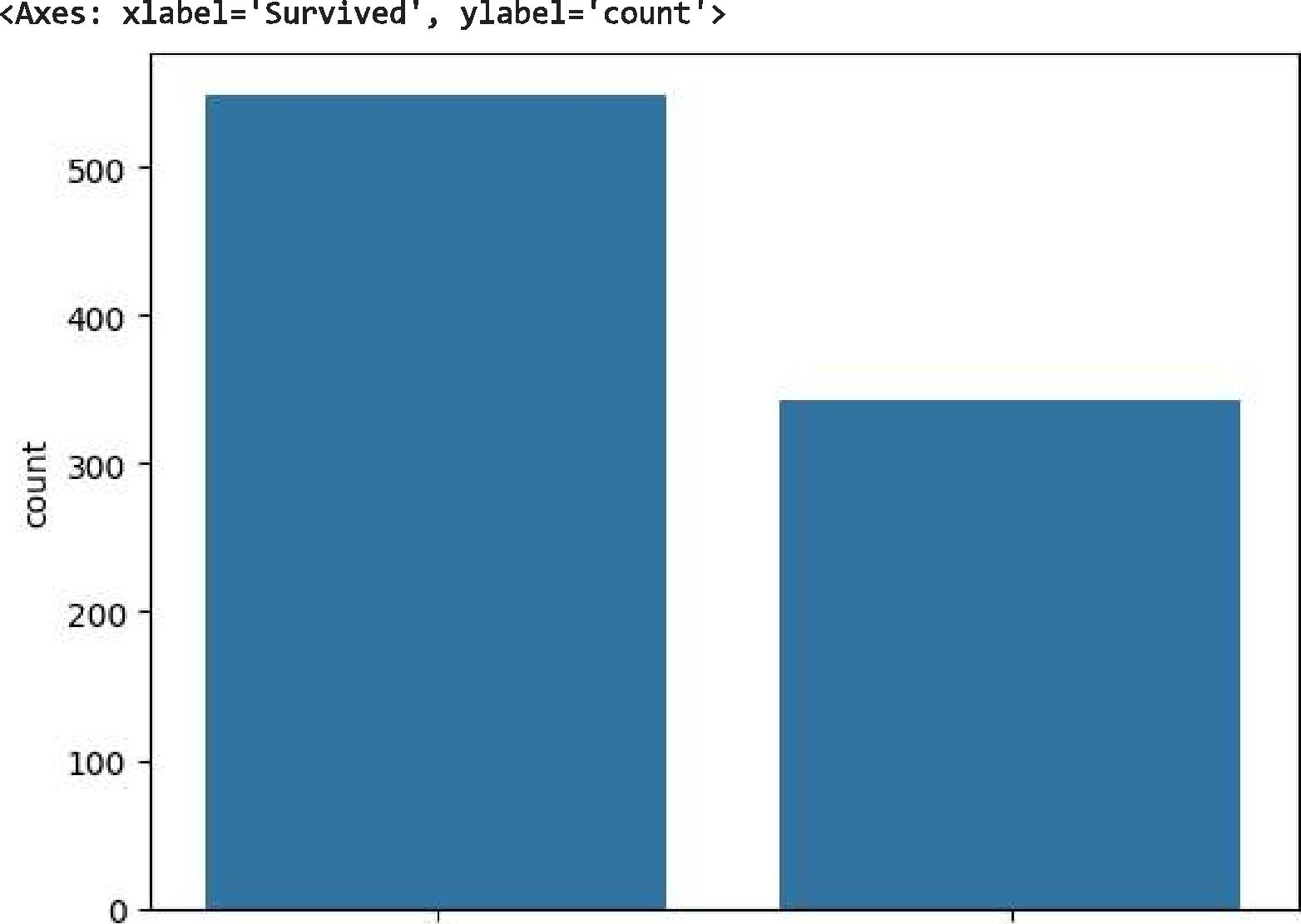
Sex

sns.countplot(x=data['Survived'])

[**https://colab.research.google.com/drive/1FUv\_eM14-Ze7ikh-vdMVX-5asJTp\_h3L#printMode=true**](https://colab.research.google.com/drive/1FUv_eM14-Ze7ikh-vdMVX-5asJTp_h3L#printMode=true) **3/7**

0 1

**5/7/24, 6:43 PM Assignment 9 .ipynb - Colab**



**<Axes: xlabel='Survived', ylabel='count'>**

500
  
400

c 300

200

100

0

[**https://colab.research.google.com/drive/1FUv\_eM14-Ze7ikh-vdMVX-5asJTp\_h3L#printMode=true**](https://colab.research.google.com/drive/1FUv_eM14-Ze7ikh-vdMVX-5asJTp_h3L#printMode=true) **4/7**

Survived

**data['Sex'].value\_counts().plot(kind="pie", autopct="%.2f") <Axes: ylabel='count'>**

male

**...W.I.:**

0

35.24

female

**plt.hist(df['Age'],bins=10)**

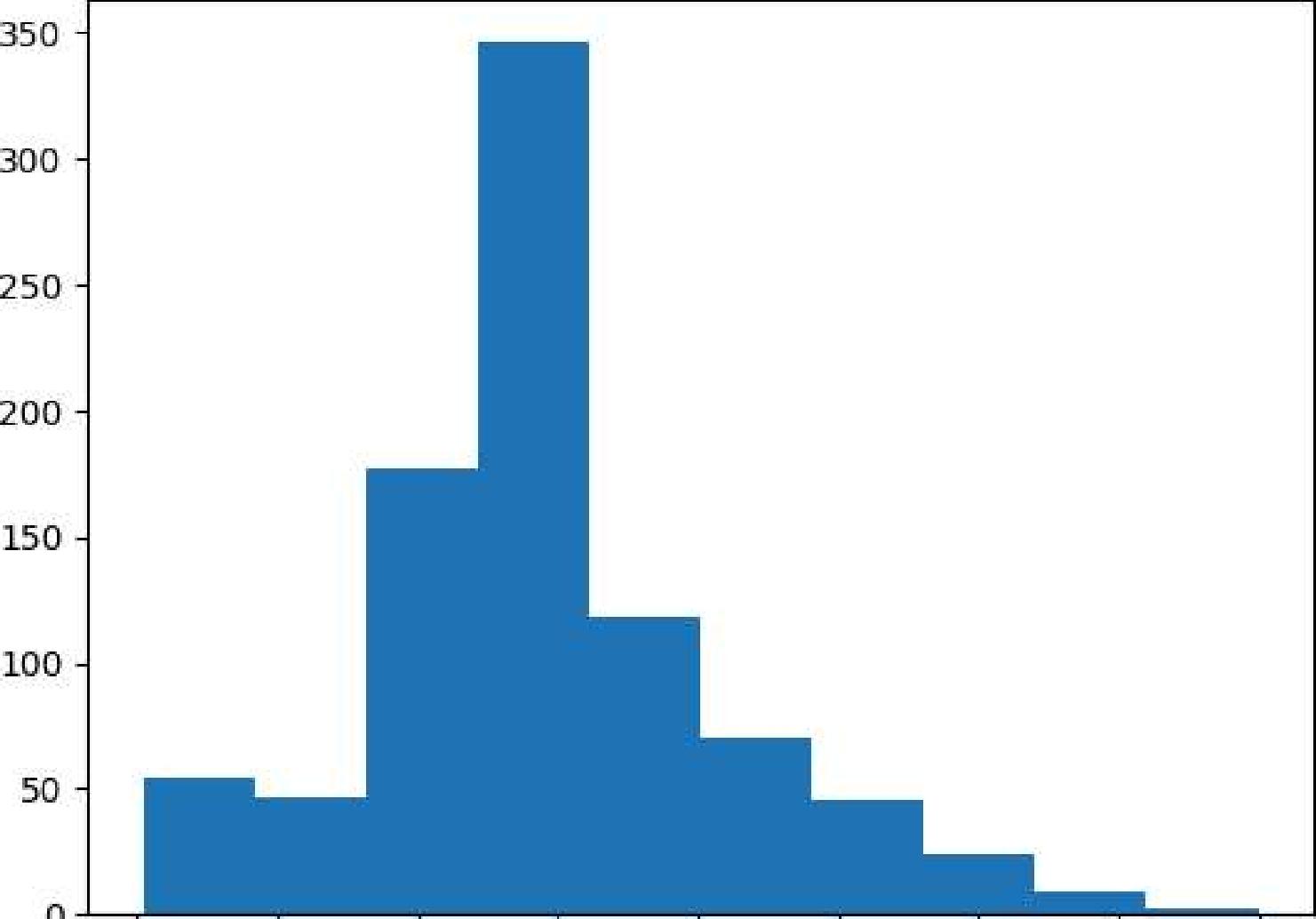
(array([ 54., 46., 177., 346., 118., 70., 45., 24., 9., 2.]),

**5/7/24, 6:43 PM Assignment 9 .ipynb - Colab**

array([ 0.42 , 8.378, 16.336, 24.294, 32.252, 40.21 , 48.168, 56.126, 64.084, 72.042, 80. ]),

<BarContainer object of 10 artists>)

10 20 30 40 50 60 10 80



350-

300 -

250-

200-

150-

100 -

5D -

sns.distplot(df['Age'])

<ipython-input-25-0fafe04ea3f6>:1: UserWarning:

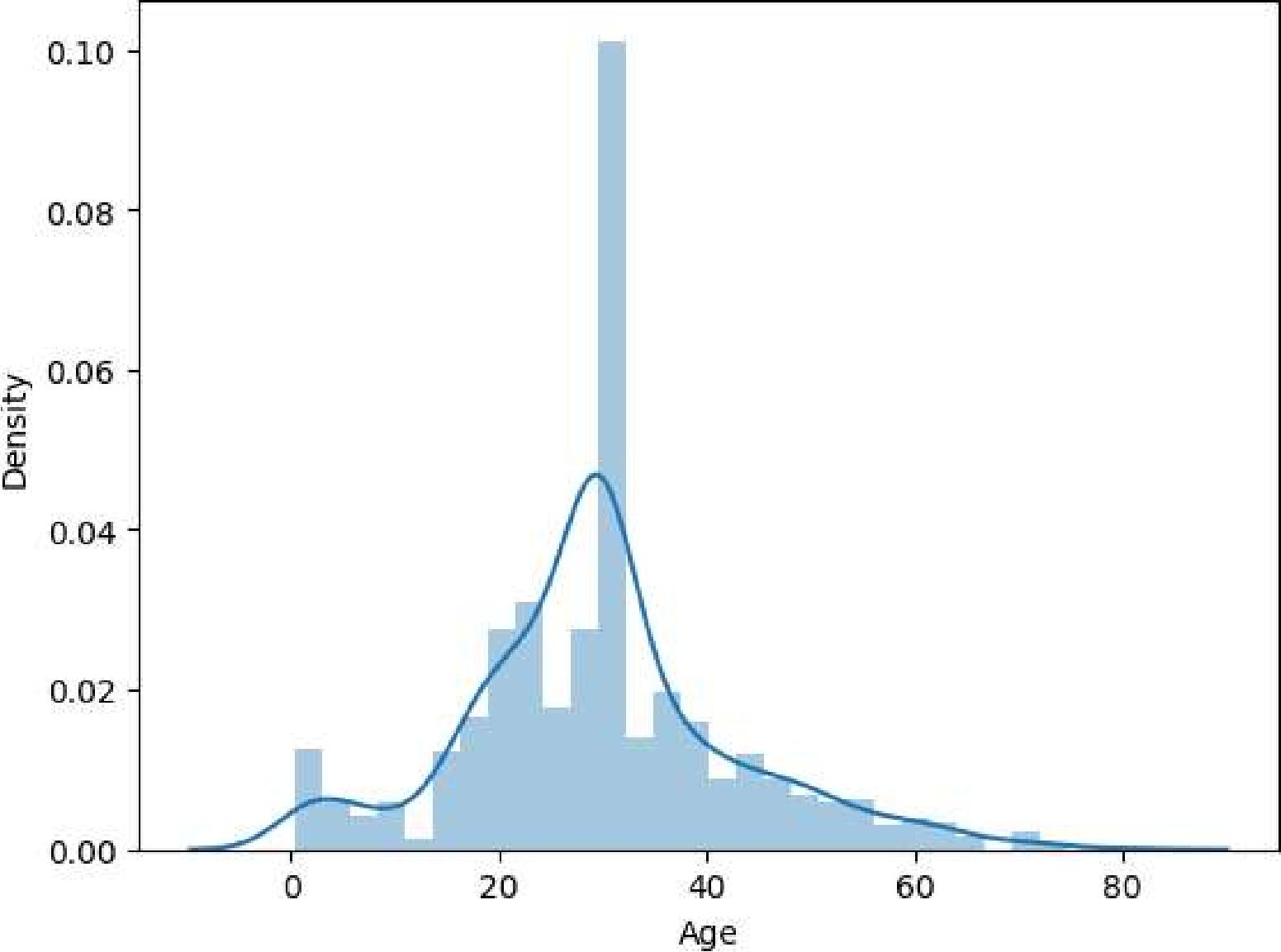
'distplot' is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either 'displot' (a figure-level function with similar flexibility) or 'histplot' (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see [httuILLgist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751](http://httuILLgist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751)

sns.distplot(df['Age'])

<Axes: xlabel='Age', ylabel='Density'>



sns.scatterplot(x=df['5ex'],y=dfPAge'bhue=df['Survived'])

[**https://colab.research.google.com/drive/1FUv\_eM14-Ze7ikh-vdMVX-5asJTp\_h3L#printMode=true**](https://colab.research.google.com/drive/1FUv_eM14-Ze7ikh-vdMVX-5asJTp_h3L#printMode=true) **5/7**