[ ]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[3]:
```
dat=pd.read_csv("/content/Titanic-Dataset.csv")
```

[5]:
```
dat.head()
```

[5]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th…	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

[7]:
```
dat.describe()
```

[7]:

	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
```

[8]:
```
dat.info()
```

```
<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
```

[9]:
```
corr=dat.corr
```

[10]:
```
corr
```

[10]:
```
<bound method DataFrame.corr of      PassengerId  Survived  Pclass  \
0              1         0       3
1              2         1       1
2              3         1       3
3              4         1       1
4              5         0       3
```

```
..           …       …      …
886          887         0       2
887          888         1       1
888          889         0       3
889          890         1       1
890          891         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
..                                                 …     …   …    …
886                             Montvila, Rev. Juozas    male  27.0      0
887                      Graham, Miss. Margaret Edith  female  19.0      0
888          Johnston, Miss. Catherine Helen "Carrie"  female   NaN      1
889                             Behr, Mr. Karl Howell    male  26.0      0
890                               Dooley, Mr. Patrick    male  32.0      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
..     …               …       …   …      …
886      0            211536  13.0000   NaN        S
887      0            112053  30.0000   B42        S
888      2        W./C. 6607  23.4500   NaN        S
889      0            111369  30.0000  C148        C
890      0            370376   7.7500   NaN        Q

[891 rows x 12 columns]>
```

[11]:
```
dat.isnull().sum()
```

[11]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
```

```
Cabin          687
Embarked         2
dtype: int64
```

[12]:
```
dat.isnull().any()
```

[12]:
```
PassengerId    False
Survived       False
Pclass         False
Name           False
Sex            False
Age             True
SibSp          False
Parch          False
Ticket         False
Fare           False
Cabin           True
Embarked        True
dtype: bool
```

[13]:
```
dat.median()
```

```
<ipython-input-13-899baa360a33>:1: FutureWarning: The default value of
numeric_only in DataFrame.median is deprecated. In a future version, it will
default to False. In addition, specifying 'numeric_only=None' is deprecated.
Select only valid columns or specify the value of numeric_only to silence this
warning.
  dat.median()
```

[13]:
```
PassengerId    446.0000
Survived         0.0000
Pclass           3.0000
Age             28.0000
SibSp            0.0000
Parch            0.0000
Fare            14.4542
dtype: float64
```

[14]:
```
m_dat=dat.Age.median()
```

[15]:
```
dat["Age"]=dat["Age"].fillna(m_dat)
```

[16]:
```
dat.drop("Cabin",axis=1,inplace=True)
```

[17]:
```
dat.head()
```

[17]:
```
   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 Embarked
0      0         A/5 21171   7.2500        S
1      0          PC 17599  71.2833        C
2      0  STON/O2. 3101282   7.9250        S
3      0            113803  53.1000        S
4      0            373450   8.0500        S
```

[18]:
```
dat.isnull().any()
```

[18]:
```
PassengerId    False
Survived       False
Pclass         False
Name           False
Sex            False
Age            False
SibSp          False
Parch          False
Ticket         False
Fare           False
Embarked        True
dtype: bool
```

[19]:
```
mode=dat["Embarked"].mode()
```

[20]:
```
mode
```

[20]:
```
0    S
Name: Embarked, dtype: object
```

[21]:
```
dat["Embarked"]=dat["Embarked"].fillna(m_dat)
```

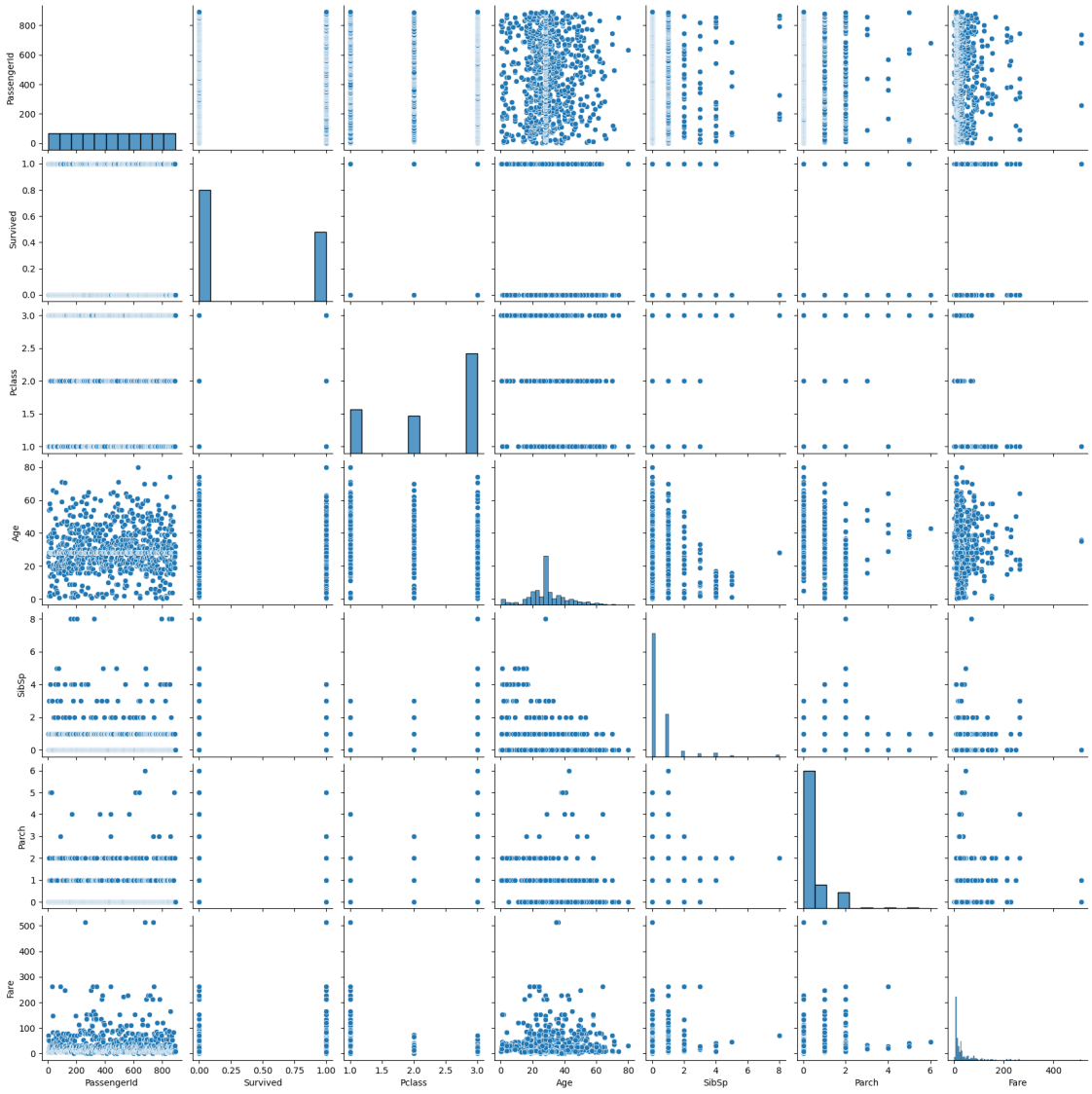
[22]:
```
dat.isnull().any()
```

[22]:
```
PassengerId    False
Survived       False
Pclass         False
Name           False
Sex            False
Age            False
SibSp          False
Parch          False
Ticket         False
Fare           False
Embarked       False
dtype: bool
```

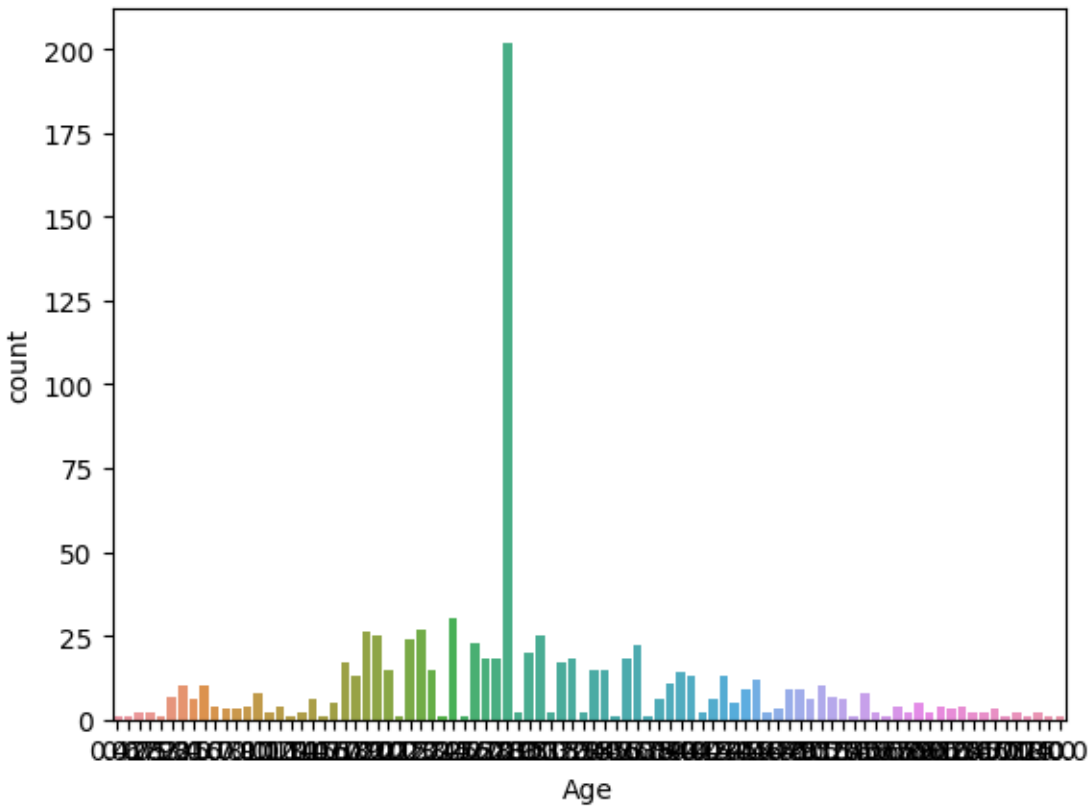
[23]:
```
dat.isnull().sum()
```

[23]:
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       0
dtype: int64
```

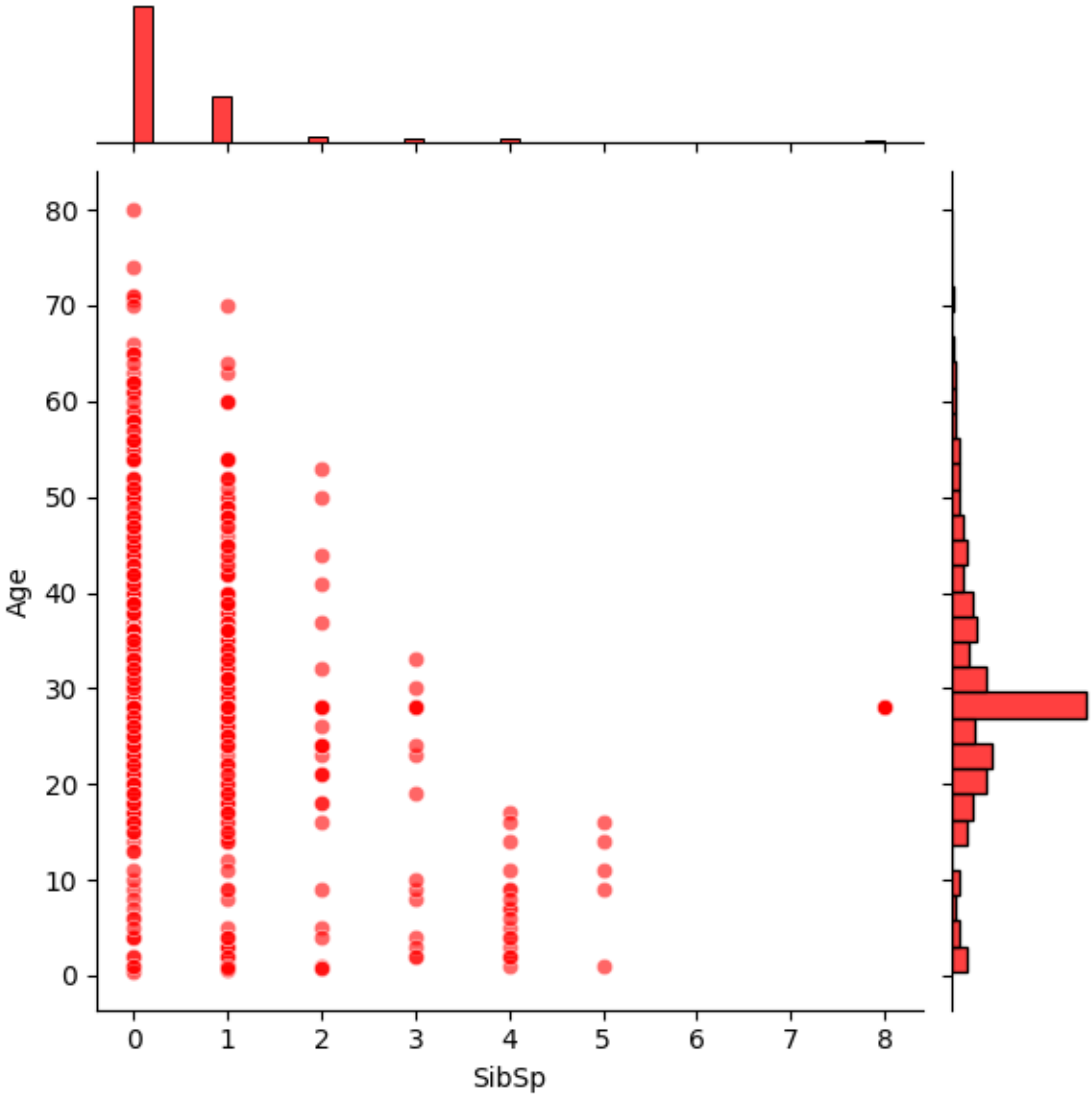
[25]:
```
import seaborn as sns
import matplotlib.pyplot as plt

sns.pairplot(dat)
plt.show()
```



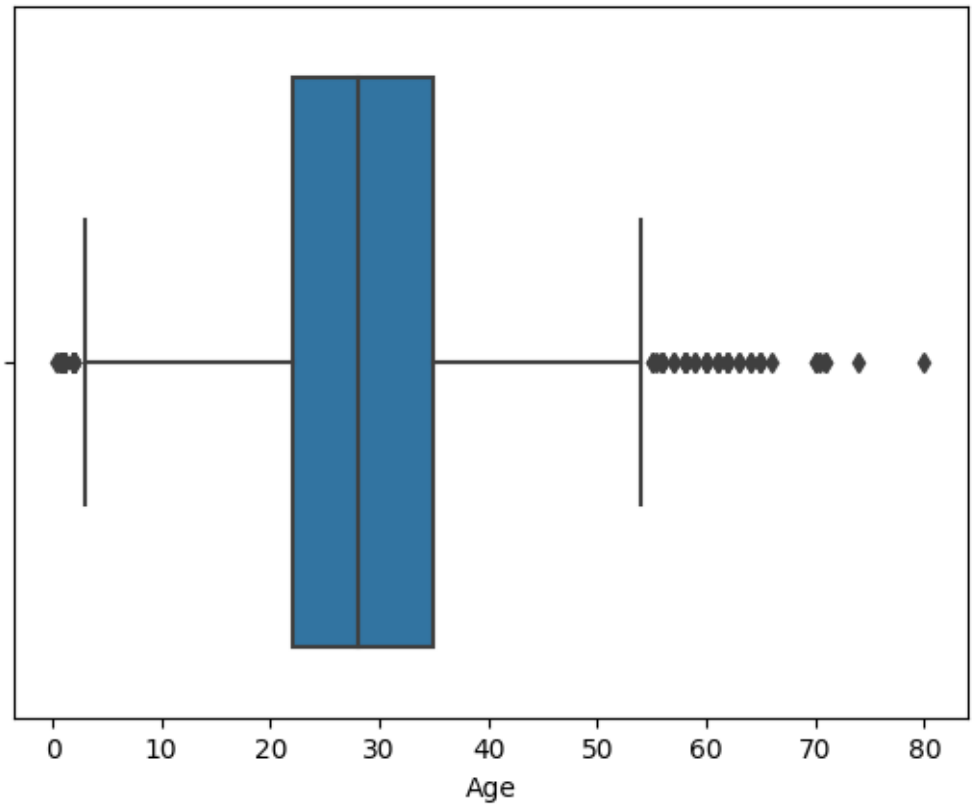
[26]:
```
import seaborn as sns
import matplotlib.pyplot as plt

sns.countplot(x="Age", data=dat)
plt.show()
```



[27]:
```
import seaborn as sns
import matplotlib.pyplot as plt

sns.jointplot(data=dat, x="SibSp", y="Age", kind="scatter", color="red",␣
 ↪alpha=0.6)
plt.show()
```



[28]:
```
import seaborn as sns
import matplotlib.pyplot as plt

sns.boxplot(x=dat["Age"])
plt.show()
```



[29]:
```
q1=dat.Age.quantile(0.25)
```

[30]:
```
q3=dat.Age.quantile(0.75)
```

[31]:
```
q3
```

[31]: 35.0

[32]:
```
q1
```

[32]: 22.0

[33]:
```
IQR=q3-q1
```

[34]:
```
IQR
```

[34]: 13.0

[35]:
```
upp_lim=q3+1.5*IQR
```

[36]:
```
upp_lim
```

[36]:
```
54.5
```

[37]:
```
low_lim=q1-1.5*IQR
```

[38]:
```
low_lim
```

[38]:
```
2.5
```

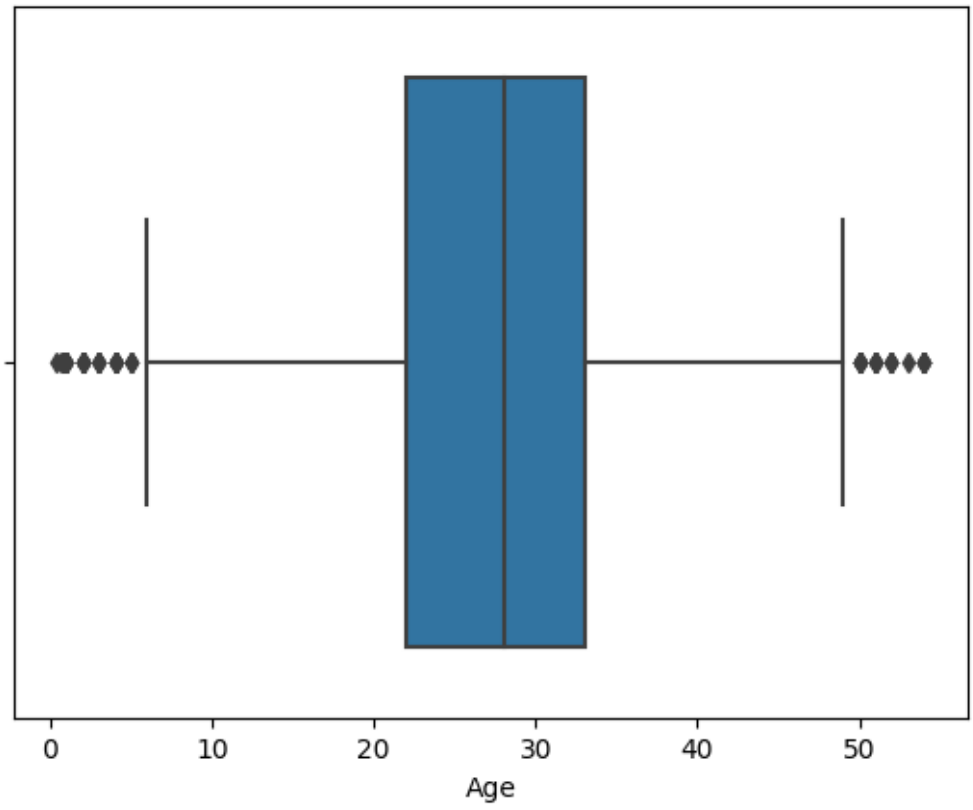
[39]:
```
d_Age = dat["Age"].median()
```

[42]:
```
import numpy as np

dat["Age"] = np.where(dat["Age"] > upp_lim, d_Age, dat["Age"])
```

[43]:
```
dat.describe()
```

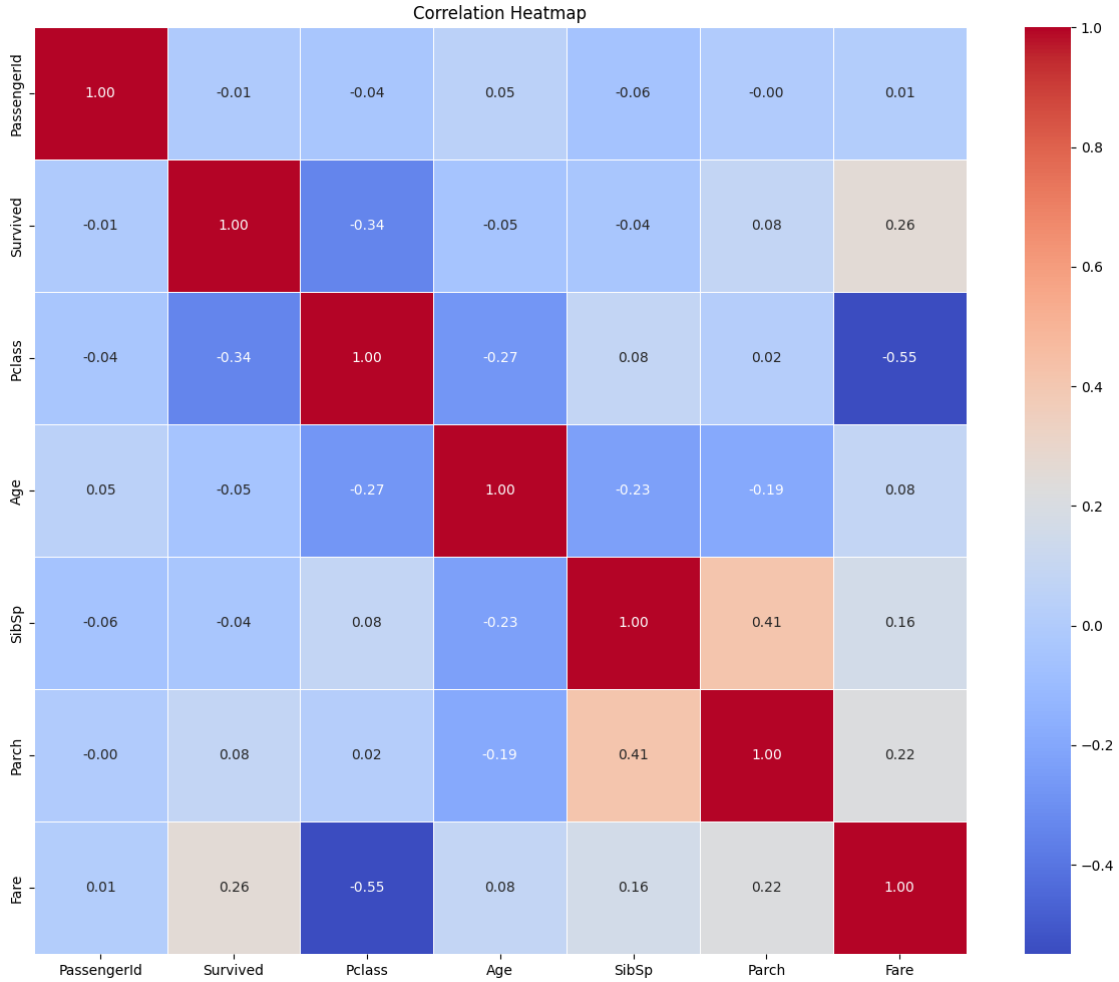
[43]:
```
       PassengerId    Survived      Pclass         Age       SibSp  \
count   891.000000  891.000000  891.000000  891.000000  891.000000
mean    446.000000    0.383838    2.308642   27.758889    0.523008
std     257.353842    0.486592    0.836071   10.735830    1.102743
min       1.000000    0.000000    1.000000    0.420000    0.000000
25%     223.500000    0.000000    2.000000   22.000000    0.000000
50%     446.000000    0.000000    3.000000   28.000000    0.000000
75%     668.500000    1.000000    3.000000   33.000000    1.000000
max     891.000000    1.000000    3.000000   54.000000    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
```

[44]:
```
import seaborn as sns
import matplotlib.pyplot as plt

sns.boxplot(x=dat["Age"])
plt.show()
```



[46]:
```
import seaborn as sns
import matplotlib.pyplot as plt

correlation_matrix = dat.corr()
plt.figure(figsize=(15, 12))
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", fmt=".2f",␣
 ↪linewidths=0.6)
plt.title("Correlation Heatmap")
plt.show()
```

```
<ipython-input-46-239ffa6745ff>:4: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  correlation_matrix = dat.corr()
```



[93]:
```
x = dat.iloc[:, 3:15]
```

[94]:
```
x
```

[94]:

	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	
..	…	…	…	…	
886	Montvila, Rev. Juozas	male	27.0	0	
887	Graham, Miss. Margaret Edith	female	19.0	0	
888	Johnston, Miss. Catherine Helen "Carrie"	female	28.0	1	
889	Behr, Mr. Karl Howell	male	26.0	0	
890	Dooley, Mr. Patrick	male	32.0	0	

```
     Parch            Ticket     Fare Embarked
0        0         A/5 21171   7.2500        S
1        0          PC 17599  71.2833        C
2        0  STON/O2. 3101282   7.9250        S
3        0            113803  53.1000        S
4        0            373450   8.0500        S
..     …               …      …      …
886      0            211536  13.0000        S
887      0            112053  30.0000        S
888      2        W./C. 6607  23.4500        S
889      0            111369  30.0000        C
890      0            370376   7.7500        Q

[891 rows x 8 columns]
```

[91]:
```
y=dat["Survived"]
```

[92]:
```
y
```

[92]:
```
0      0
1      1
2      1
3      1
4      0
      ..
886    0
887    1
888    0
889    1
890    0
Name: Survived, Length: 891, dtype: int64
```

[55]:
```
from sklearn.preprocessing import LabelEncoder

# Create an instance of the LabelEncoder
le = LabelEncoder()
```

[60]:
```
x["Age"] = le.fit_transform(x["Age"])
```

[61]:
```
x["Age"]
```

[61]:
```
0      28
1      51
2      34
3      47
4      47
```

```
        ..
886     35
887     24
888     36
889     34
890     42
Name: Age, Length: 891, dtype: int64
```

[58]:
```
x["Name"]=le.fit_transform(x["Name"])
```

[59]:
```
x["Name"]
```

[59]:
```
0      108
1      190
2      353
3      272
4       15
      ...
886    548
887    303
888    413
889     81
890    220
Name: Name, Length: 891, dtype: int64
```

[62]:
```
x["Ticket"]=le.fit_transform(x["Ticket"])
```

[63]:
```
x["Ticket"]
```

[63]:
```
0      523
1      596
2      669
3       49
4      472
      ...
886    101
887     14
888    675
889      8
890    466
Name: Ticket, Length: 891, dtype: int64
```

[64]:
```
x_Embarked = pd.get_dummies(x["Embarked"], drop_first=True)
```

[65]:
```
x_Embarked
```

[65]:
```
     C  Q  S
0    0  0  1
1    1  0  0
2    0  0  1
3    0  0  1
4    0  0  1
..  .. .. ..
886  0  0  1
887  0  0  1
888  0  0  1
889  1  0  0
890  0  1  0

[891 rows x 3 columns]
```

[66]:
```
x = pd.concat([x, x_Embarked], axis=1)
```

[67]:
```
x.drop("Embarked",axis=1,inplace=True)
```

[68]:
```
x.head
```

[68]:
```
<bound method NDFrame.head of      Name  Sex  Age  SibSp  Parch  Ticket     Fare
C  Q  S
0     108    1   28      1      0     523   7.2500  0  0  1
1     190    0   51      1      0     596  71.2833  1  0  0
2     353    0   34      0      0     669   7.9250  0  0  1
3     272    0   47      1      0      49  53.1000  0  0  1
4      15    1   47      0      0     472   8.0500  0  0  1
..    ...  ...  ...    ...    ...     ...      ... .. .. ..
886   548    1   35      0      0     101  13.0000  0  0  1
887   303    0   24      0      0      14  30.0000  0  0  1
888   413    0   36      1      2     675  23.4500  0  0  1
889    81    1   34      0      0       8  30.0000  1  0  0
890   220    1   42      0      0     466   7.7500  0  1  0

[891 rows x 10 columns]>
```

[76]:
```
from sklearn.model_selection import train_test_split
```

[95]:
```
x_train,x_test,y_train,y_test = train_test_split(x, y, test_size=0.2,␣
↪random_state=0)
```

[96]:
```
x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

[96]:
```
((712, 8), (179, 8), (712,), (179,))
```