In [1]:
```python
import pandas as pd
import numpy as np
df = pd.read_csv("https://raw.githubusercontent.com/datasciencedojo/datasets/
print(df)
```

```
     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]
```

In [2]:
```python
print(df.head())
```

```
   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
```

In [3]:
```python
print(df.head(10))
```

```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3
5            6         0       3
6            7         0       1
7            8         0       3
8            9         1       3
9           10         1       2

                                                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
5                                   Moran, Mr. James    male   NaN      0
6                            McCarthy, Mr. Timothy J    male  54.0      0
7                     Palsson, Master. Gosta Leonard    male   2.0      3
8  Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)  female  27.0      0
9                     Nasser, Mrs. Nicholas (Adele Achem)  female  14.0      1

   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
5      0           330877    8.4583   NaN        Q
6      0            17463   51.8625   E46        S
7      1           349909   21.0750   NaN        S
8      2           347742   11.1333   NaN        S
9      0           237736   30.0708   NaN        C
```

In [4]:
```python
print(df.tail())
```

```
        PassengerId  Survived  Pclass                                      Name
\
886             887         0       2                     Montvila, Rev. Juozas
887             888         1       1                 Graham, Miss. Margaret Edith
888             889         0       3   Johnston, Miss. Catherine Helen "Carrie"
889             890         1       1                      Behr, Mr. Karl Howell
890             891         0       3                         Dooley, Mr. Patrick

        Sex   Age  SibSp  Parch      Ticket    Fare Cabin Embarked
886    male  27.0      0      0      211536   13.00   NaN        S
887  female  19.0      0      0      112053   30.00   B42        S
888  female   NaN      1      2   W./C. 6607  23.45   NaN        S
889    male  26.0      0      0      111369   30.00  C148        C
890    male  32.0      0      0      370376    7.75   NaN        Q
```

In [5]:
```python
print(df.tail(8))
```

```
        PassengerId  Survived  Pclass                                      Name
\
883             884         0       2                 Banfield, Mr. Frederick James
884             885         0       3                       Sutehall, Mr. Henry Jr
885             886         0       3             Rice, Mrs. William (Margaret Norton)
886             887         0       2                     Montvila, Rev. Juozas
887             888         1       1                 Graham, Miss. Margaret Edith
888             889         0       3   Johnston, Miss. Catherine Helen "Carrie"
889             890         1       1                      Behr, Mr. Karl Howell
890             891         0       3                         Dooley, Mr. Patrick

        Sex   Age  SibSp  Parch             Ticket     Fare Cabin Embarked
883    male  28.0      0      0   C.A./SOTON 34068   10.500   NaN        S
884    male  25.0      0      0    SOTON/OQ 392076    7.050   NaN        S
885  female  39.0      0      5             382652   29.125   NaN        Q
886    male  27.0      0      0             211536   13.000   NaN        S
887  female  19.0      0      0             112053   30.000   B42        S
888  female   NaN      1      2        W./C. 6607   23.450   NaN        S
889    male  26.0      0      0             111369   30.000  C148        C
890    male  32.0      0      0             370376    7.750   NaN        Q
```

In [6]:
```python
print(df.tail(-5))
```

```
        PassengerId  Survived  Pclass  \
5                 6         0       3
6                 7         0       1
7                 8         0       3
8                 9         1       3
9                10         1       2
..              ...       ...     ...
886             887         0       2
887             888         1       1
888             889         0       3
889             890         1       1
890             891         0       3

                                                     Name     Sex   Age  SibSp
\
5                                        Moran, Mr. James    male   NaN      0
6                                    McCarthy, Mr. Timothy J    male  54.0      0
7                               Palsson, Master. Gosta Leonard    male   2.0      3
8    Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)  female  27.0      0
9                          Nasser, Mrs. Nicholas (Adele Achem)  female  14.0      1
..                                                    ...     ...   ...    ...
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
5        0    330877   8.4583   NaN        Q
6        0     17463  51.8625   E46        S
7        1    349909  21.0750   NaN        S
8        2    347742  11.1333   NaN        S
9        0    237736  30.0708   NaN        C
..     ...       ...      ...   ...      ...
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

[886 rows x 12 columns]
```

In [7]:
```python
print(df.columns)
```

```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

In [8]:
```python
print(df.index)
```

```
RangeIndex(start=0, stop=891, step=1)
```

In [9]:
```python
print(df.dtypes)
```

```
PassengerId      int64
Survived         int64
Pclass           int64
Name            object
Sex             object
Age            float64
SibSp            int64
Parch            int64
Ticket          object
Fare           float64
Cabin           object
Embarked        object
dtype: object
```

In [10]:
```python
print(df.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
None
```

In [11]:
```python
print(df.values)
```

```
[[1 0 3 ... 7.25 nan 'S']
 [2 1 1 ... 71.2833 'C85' 'C']
 [3 1 3 ... 7.925 nan 'S']
 ...
 [889 0 3 ... 23.45 nan 'S']
 [890 1 1 ... 30.0 'C148' 'C']
 [891 0 3 ... 7.75 nan 'Q']]
```

In [12]:
```python
print(df.ndim)
```

```
2
```

In [13]:
```python
print(df.size)
```

```
10692
```

In [14]:
```python
print(df.shape)
```

```
(891, 12)
```

In [15]:
```python
print(df.empty)
```

```
False
```

In [16]:
```python
print(df["Age"])
```

```
0      22.0
1      38.0
2      26.0
3      35.0
4      35.0
       ...
886    27.0
887    19.0
888     NaN
889    26.0
890    32.0
Name: Age, Length: 891, dtype: float64
```

In [17]:
```python
print(df.loc[5])
```

```
PassengerId                  6
Survived                     0
Pclass                       3
Name         Moran, Mr. James
Sex                       male
Age                        NaN
SibSp                        0
Parch                        0
Ticket                  330877
```

```
Fare                       8.4583
Cabin                         NaN
Embarked                        Q
Name: 5, dtype: object
```

In [18]:
```
print(df.loc[5, "Sex"])
```

```
male
```

In [19]:
```
print(df.loc[2:6])
```

```
   PassengerId  Survived  Pclass  \
2            3         1       3
3            4         1       1
4            5         0       3
5            6         0       3
6            7         0       1

                                          Name     Sex   Age  SibSp  Parch  \
2                          Heikkinen, Miss. Laina  female  26.0      0      0
3   Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1      0
4                        Allen, Mr. William Henry    male  35.0      0      0
5                                Moran, Mr. James    male   NaN      0      0
6                         McCarthy, Mr. Timothy J    male  54.0      0      0

            Ticket     Fare Cabin Embarked
2  STON/O2. 3101282   7.9250   NaN        S
3           113803  53.1000  C123        S
4           373450   8.0500   NaN        S
5           330877   8.4583   NaN        Q
6            17463  51.8625   E46        S
```

In [20]:
```
print(df.iloc[5])
```

```
PassengerId                    6
Survived                       0
Pclass                         3
Name           Moran, Mr. James
Sex                         male
Age                          NaN
SibSp                          0
Parch                          0
Ticket                    330877
Fare                      8.4583
Cabin                        NaN
Embarked                       Q
Name: 5, dtype: object
```

In [21]:
```
print(df["Age"].iloc[1])
```

```
38.0
```

In [22]:
```
print(df.iloc[5:10])
```

```
   PassengerId  Survived  Pclass  \
5            6         0       3
6            7         0       1
7            8         0       3
8            9         1       3
9           10         1       2
```

```
                                             Name     Sex   Age  SibSp  \
5                               Moran, Mr. James     male   NaN      0
6                           McCarthy, Mr. Timothy J     male  54.0      0
7                         Palsson, Master. Gosta Leonard     male   2.0      3
8   Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)   female  27.0      0
9                 Nasser, Mrs. Nicholas (Adele Achem)   female  14.0      1

   Parch  Ticket     Fare Cabin Embarked
5      0  330877   8.4583   NaN        Q
6      0   17463  51.8625   E46        S
7      1  349909  21.0750   NaN        S
8      2  347742  11.1333   NaN        S
9      0  237736  30.0708   NaN        C
```

In [23]:
```python
print(df[df.Sex == "female"])
```

```
     PassengerId  Survived  Pclass  \
1              2         1       1
2              3         1       3
3              4         1       1
8              9         1       3
9             10         1       2
..           ...       ...     ...
880          881         1       2
882          883         0       3
885          886         0       3
887          888         1       1
888          889         0       3

                                             Name     Sex   Age  SibSp
\
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
8   Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)   female  27.0      0
9                 Nasser, Mrs. Nicholas (Adele Achem)   female  14.0      1
..                                            ...      ...   ...    ...
880      Shelley, Mrs. William (Imanita Parrish Hall)   female  25.0      0
882               Dahlberg, Miss. Gerda Ulrika   female  22.0      0
885           Rice, Mrs. William (Margaret Norton)   female  39.0      0
887               Graham, Miss. Margaret Edith   female  19.0      0
888      Johnston, Miss. Catherine Helen "Carrie"   female   NaN      1

     Parch          Ticket     Fare Cabin Embarked
1        0        PC 17599  71.2833   C85        C
2        0  STON/O2. 3101282   7.9250   NaN        S
3        0          113803  53.1000  C123        S
8        2          347742  11.1333   NaN        S
9        0          237736  30.0708   NaN        C
..     ...             ...      ...   ...      ...
880      1          230433  26.0000   NaN        S
882      0            7552  10.5167   NaN        S
885      5          382652  29.1250   NaN        Q
887      0          112053  30.0000   B42        S
888      2       W./C. 6607  23.4500   NaN        S

[314 rows x 12 columns]
```

In [24]:
```python
print(df[(df.Sex == "male") & (df.Age > 60)])
```

```
     PassengerId  Survived  Pclass                             Name  \
33            34         0       2                Wheadon, Mr. Edward H
```

```
54           55         0        1            Ostby, Mr. Engelhart Cornelius
96           97         0        1              Goldschmidt, Mr. George B
116         117         0        3                     Connors, Mr. Patrick
170         171         0        1               Van der hoef, Mr. Wyckoff
252         253         0        1               Stead, Mr. William Thomas
280         281         0        3                        Duane, Mr. Frank
326         327         0        3               Nysveen, Mr. Johan Hansen
438         439         0        1                      Fortune, Mr. Mark
456         457         0        1                 Millet, Mr. Francis Davis
493         494         0        1                 Artagaveytia, Mr. Ramon
545         546         0        1             Nicholson, Mr. Arthur Ernest
555         556         0        1                     Wright, Mr. George
570         571         1        2                     Harris, Mr. George
625         626         0        1                  Sutton, Mr. Frederick
630         631         1        1      Barkworth, Mr. Algernon Henry Wilson
672         673         0        2              Mitchell, Mr. Henry Michael
745         746         0        1             Crosby, Capt. Edward Gifford
851         852         0        3                     Svensson, Mr. Johan

         Sex   Age  SibSp  Parch      Ticket      Fare       Cabin Embarked
33      male  66.0      0      0  C.A. 24579   10.5000         NaN        S
54      male  65.0      0      1      113509   61.9792         B30        C
96      male  71.0      0      0   PC 17754   34.6542          A5        C
116     male  70.5      0      0      370369    7.7500         NaN        Q
170     male  61.0      0      0      111240   33.5000         B19        S
252     male  62.0      0      0      113514   26.5500         C87        S
280     male  65.0      0      0      336439    7.7500         NaN        Q
326     male  61.0      0      0      345364    6.2375         NaN        S
438     male  64.0      1      4       19950  263.0000  C23 C25 C27       S
456     male  65.0      0      0       13509   26.5500         E38        S
493     male  71.0      0      0   PC 17609   49.5042          NaN        C
545     male  64.0      0      0         693   26.0000         NaN        S
555     male  62.0      0      0      113807   26.5500         NaN        S
570     male  62.0      0      0  S.W./PP 752  10.5000         NaN        S
625     male  61.0      0      0       36963   32.3208         D50        S
630     male  80.0      0      0       27042   30.0000         A23        S
672     male  70.0      0      0  C.A. 24580   10.5000         NaN        S
745     male  70.0      1      1   WE/P 5735   71.0000         B22        S
851     male  74.0      0      0      347060    7.7750         NaN        S
```

In [25]:
```python
print(df[df.Name.str.startswith("M")])
```

```
     PassengerId  Survived  Pclass                          Name     Sex   Age
\
5              6         0       3               Moran, Mr. James    male   NaN
6              7         0       1          McCarthy, Mr. Timothy J    male  54.0
19            20         1       3          Masselmani, Mrs. Fatima  female   NaN
22            23         1       3     McGowan, Miss. Anna "Annie"  female  15.0
34            35         0       1          Meyer, Mr. Edgar Joseph    male  28.0
..           ...       ...     ...                           ...     ...   ...
839          840         1       1          Marechal, Mr. Pierre    male   NaN
841          842         0       2       Mudd, Mr. Thomas Charles    male  16.0
847          848         0       3            Markoff, Mr. Marin    male  35.0
881          882         0       3           Markun, Mr. Johann    male  33.0
886          887         0       2         Montvila, Rev. Juozas    male  27.0

     SibSp  Parch     Ticket     Fare Cabin Embarked
5        0      0     330877   8.4583   NaN        Q
6        0      0      17463  51.8625   E46        S
19       0      0       2649   7.2250   NaN        C
22       0      0     330923   8.0292   NaN        Q
34       1      0   PC 17604  82.1708   NaN        C
..     ...    ...        ...      ...   ...      ...
```

```
839        0        0       11774   29.7000    C47        C
841        0        0   S.O./P.P. 3  10.5000   NaN        S
847        0        0      349213    7.8958    NaN        C
881        0        0      349257    7.8958    NaN        S
886        0        0      211536   13.0000    NaN        S

[74 rows x 12 columns]
```

In [27]:
```python
classes = [1, 2]
print(df[df.Pclass.isin(classes)])
```

```
     PassengerId  Survived  Pclass  \
1              2         1       1
3              4         1       1
6              7         0       1
9             10         1       2
11            12         1       1
..           ...       ...     ...
880          881         1       2
883          884         0       2
886          887         0       2
887          888         1       1
889          890         1       1

                                             Name     Sex   Age  SibSp  \
1       Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
3             Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
6                               McCarthy, Mr. Timothy J     male  54.0      0
9                     Nasser, Mrs. Nicholas (Adele Achem)  female  14.0      1
11                          Bonnell, Miss. Elizabeth  female  58.0      0
..                                                    ...     ...   ...    ...
880       Shelley, Mrs. William (Imanita Parrish Hall)  female  25.0      0
883                   Banfield, Mr. Frederick James     male  28.0      0
886                         Montvila, Rev. Juozas     male  27.0      0
887                   Graham, Miss. Margaret Edith  female  19.0      0
889                         Behr, Mr. Karl Howell     male  26.0      0

     Parch         Ticket     Fare Cabin Embarked
1        0      PC 17599  71.2833   C85        C
3        0        113803  53.1000  C123        S
6        0         17463  51.8625   E46        S
9        0        237736  30.0708   NaN        C
11       0        113783  26.5500  C103        S
..     ...           ...      ...   ...      ...
880      1        230433  26.0000   NaN        S
883      0  C.A./SOTON 34068  10.5000   NaN        S
886      0        211536  13.0000   NaN        S
887      0        112053  30.0000   B42        S
889      0        111369  30.0000  C148        C

[400 rows x 12 columns]
```

In [28]:
```python
print(df.query('Sex == "female" and Age > 50'))
```

```
     PassengerId  Survived  Pclass  \
11            12         1       1
15            16         1       2
195          196         1       1
268          269         1       1
275          276         1       1
366          367         1       1
483          484         1       3
```

```
496       497       1       1
513       514       1       1
571       572       1       1
591       592       1       1
765       766       1       1
772       773       0       2
774       775       1       2
820       821       1       1
829       830       1       1
879       880       1       1
```

|  | Name | Sex | Age | SibSp \ |
|---|---|---|---|---|
| 11 | Bonnell, Miss. Elizabeth | female | 58.0 | 0 |
| 15 | Hewlett, Mrs. (Mary D Kingcome) | female | 55.0 | 0 |
| 195 | Lurette, Miss. Elise | female | 58.0 | 0 |
| 268 | Graham, Mrs. William Thompson (Edith Junkins) | female | 58.0 | 0 |
| 275 | Andrews, Miss. Kornelia Theodosia | female | 63.0 | 1 |
| 366 | Warren, Mrs. Frank Manley (Anna Sophia Atkinson) | female | 60.0 | 1 |
| 483 | Turkula, Mrs. (Hedwig) | female | 63.0 | 0 |
| 496 | Eustis, Miss. Elizabeth Mussey | female | 54.0 | 1 |
| 513 | Rothschild, Mrs. Martin (Elizabeth L. Barrett) | female | 54.0 | 1 |
| 571 | Appleton, Mrs. Edward Dale (Charlotte Lamson) | female | 53.0 | 2 |
| 591 | Stephenson, Mrs. Walter Bertram (Martha Eustis) | female | 52.0 | 1 |
| 765 | Hogeboom, Mrs. John C (Anna Andrews) | female | 51.0 | 1 |
| 772 | Mack, Mrs. (Mary) | female | 57.0 | 0 |
| 774 | Hocking, Mrs. Elizabeth (Eliza Needs) | female | 54.0 | 1 |
| 820 | Hays, Mrs. Charles Melville (Clara Jennings Gr... | female | 52.0 | 1 |
| 829 | Stone, Mrs. George Nelson (Martha Evelyn) | female | 62.0 | 0 |
| 879 | Potter, Mrs. Thomas Jr (Lily Alexenia Wilson) | female | 56.0 | 0 |

|  | Parch | Ticket | Fare | Cabin | Embarked |
|---|---|---|---|---|---|
| 11 | 0 | 113783 | 26.5500 | C103 | S |
| 15 | 0 | 248706 | 16.0000 | NaN | S |
| 195 | 0 | PC 17569 | 146.5208 | B80 | C |
| 268 | 1 | PC 17582 | 153.4625 | C125 | S |
| 275 | 0 | 13502 | 77.9583 | D7 | S |
| 366 | 0 | 110813 | 75.2500 | D37 | C |
| 483 | 0 | 4134 | 9.5875 | NaN | S |
| 496 | 0 | 36947 | 78.2667 | D20 | C |
| 513 | 0 | PC 17603 | 59.4000 | NaN | C |
| 571 | 0 | 11769 | 51.4792 | C101 | S |
| 591 | 0 | 36947 | 78.2667 | D20 | C |
| 765 | 0 | 13502 | 77.9583 | D11 | S |
| 772 | 0 | S.O./P.P. 3 | 10.5000 | E77 | S |
| 774 | 3 | 29105 | 23.0000 | NaN | S |
| 820 | 1 | 12749 | 93.5000 | B69 | S |
| 829 | 0 | 113572 | 80.0000 | B28 | NaN |
| 879 | 1 | 11767 | 83.1583 | C50 | C |

In [29]:
```python
print(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
```

In [30]:
```python
print(df.describe(include="all"))
```

```
         PassengerId     Survived       Pclass                   Name   Sex  \
count     891.000000   891.000000   891.000000                    891   891
unique           NaN          NaN          NaN                    891     2
top              NaN          NaN          NaN    Dooley, Mr. Patrick  male
freq             NaN          NaN          NaN                      1   577
mean      446.000000     0.383838     2.308642                    NaN   NaN
std       257.353842     0.486592     0.836071                    NaN   NaN
min         1.000000     0.000000     1.000000                    NaN   NaN
25%       223.500000     0.000000     2.000000                    NaN   NaN
50%       446.000000     0.000000     3.000000                    NaN   NaN
75%       668.500000     1.000000     3.000000                    NaN   NaN
max       891.000000     1.000000     3.000000                    NaN   NaN

                Age         SibSp        Parch  Ticket         Fare Cabin Embarked
count    714.000000    891.000000   891.000000     891   891.000000   204      889
unique          NaN           NaN          NaN     681          NaN   147        3
top             NaN           NaN          NaN  347082          NaN    G6        S
freq            NaN           NaN          NaN       7          NaN     4      644
mean      29.699118      0.523008     0.381594     NaN    32.204208   NaN      NaN
std       14.526497      1.102743     0.806057     NaN    49.693429   NaN      NaN
min        0.420000      0.000000     0.000000     NaN     0.000000   NaN      NaN
25%       20.125000      0.000000     0.000000     NaN     7.910400   NaN      NaN
50%       28.000000      0.000000     0.000000     NaN    14.454200   NaN      NaN
75%       38.000000      1.000000     0.000000     NaN    31.000000   NaN      NaN
max       80.000000      8.000000     6.000000     NaN   512.329200   NaN      NaN
```

In [31]:
```python
print(df.describe(include=[np.number]))
```

```
         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
```

In [32]:
```python
print(df.describe(exclude=["number"]))
```

```
                    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
```

In [33]:
```python
print(df.isnull())
```

```
        PassengerId   Survived   Pclass    Name     Sex     Age   SibSp   Parch   Ticket
\
0             False      False    False   False   False   False   False   False    False
1             False      False    False   False   False   False   False   False    False
2             False      False    False   False   False   False   False   False    False
3             False      False    False   False   False   False   False   False    False
4             False      False    False   False   False   False   False   False    False
..              ...        ...      ...     ...     ...     ...     ...     ...      ...
886           False      False    False   False   False   False   False   False    False
887           False      False    False   False   False   False   False   False    False
888           False      False    False   False   False    True   False   False    False
889           False      False    False   False   False   False   False   False    False
890           False      False    False   False   False   False   False   False    False

        Fare  Cabin  Embarked
0      False   True     False
1      False  False     False
2      False   True     False
3      False  False     False
4      False   True     False
..       ...    ...       ...
886    False   True     False
887    False  False     False
888    False   True     False
889    False  False     False
890    False   True     False

[891 rows x 12 columns]
```

In [34]:
```python
print(df.isnull().values.any())
```

```
True
```

In [35]:
```python
print(df.isna().values.any())
```

```
True
```

In [36]:
```python
print(df.duplicated().values.any())
```

```
False
```

In [38]:
```python
df_no_missing = df.dropna()
print(df_no_missing.isnull().values.any())
```

```
False
```

In [39]:
```python
df_fill_mean = df.fillna(df.mean(numeric_only=True))
print(df_fill_mean.head())
```

```
    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
```

In [40]:
```python
df_fill_median = df.fillna(df.median(numeric_only=True))
print(df_fill_median.head())
```

```
   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
```

In [41]:
```python
df_drop = df.drop("Cabin", axis=1)
print(df_drop.head())
```

```
   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
```

In [42]:
```python
print(df.mean(numeric_only=True))
```

```
PassengerId    446.000000
Survived         0.383838
Pclass           2.308642
Age             29.699118
SibSp            0.523008
Parch            0.381594
Fare            32.204208
dtype: float64
```

In [43]:
```python
print(df.median(numeric_only=True))
```

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

In [44]:
```python
df["Age"] = df["Age"].astype("float")
print(df.dtypes)
```

```
PassengerId      int64
Survived         int64
Pclass           int64
Name            object
Sex             object
Age            float64
SibSp            int64
Parch            int64
Ticket          object
Fare           float64
Cabin           object
Embarked        object
dtype: object
```

In [ ]: