In [1]:

**import** pandas **as** pd

**import** numpy **as** np

pd.read\_csv("tested.csv")

Out[1]:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PassengerId** | **Survived** | **Pclass** | **Name** | **Sex** | **Age** | **SibSp** | **Parch** | **Ticket** | **F** |

**0** 892 0 3 Kelly, Mr. James

Wilkes,

Mrs. James (Ellen Needs)

|  |  |  |  |
| --- | --- | --- | --- |
| **1** | 893 | 1 | 3 |
| **2** | 894 | 0 | 2 |
| **3** | 895 | 0 | 3 |

Myles, Mr. Thomas Francis

Wirz, Mr.

Albert

Hirvonen,

Mrs. Alexander (Helga E Lindqvist)

|  |  |  |  |
| --- | --- | --- | --- |
| **4** | 896 | 1 | 3 |
| **...** | ... | ... | ... |
| **413** | 1305 | 0 | 3 |

male 34.5 0 0 330911 7.8

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| female | 47.0 | 1 | 0 | 363272 | 7.0 |
| male | 62.0 | 0 | 0 | 240276 | 9.6 |
| male | 27.0 | 0 | 0 | 315154 | 8.6 |

female 22.0 1 1 3101298 12.2

... ... ... ... ... ...

**414** 1306 1 1

Spector, Mr. Woolf

Oliva y Ocana, Dona. Fermina

Saether, Mr. Simon Sivertsen

|  |  |  |  |
| --- | --- | --- | --- |
| **415** | 1307 | 0 | 3 |
| **416** | 1308 | 0 | 3 |
| **417** | 1309 | 0 | 3 |

Ware, Mr. Frederick

Peter, Master. Michael J

male NaN 0 0 A.5. 3236 8.0

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| female | 39.0 | 0 | 0 | PC 17758 | 108.9 |
| male | 38.5 | 0 | 0 | SOTON/O.Q.  3101262 | 7.2 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| male | NaN | 0 | 0 | 359309 | 8.0 |
| male | NaN | 1 | 1 | 2668 | 22.3 |

418 rows × 12 columns

In [2]:

data**=**pd.read\_csv("tested.csv")

In [ ]:

data.head()

Out[4]:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PassengerId** | **Survived** | **Pclass** | **Name** | **Sex** | **Age** | **SibSp** | **Parch** | **Ticket** | **Fare** | **Ca** |

**0** 892 0 3 Kelly, Mr. James

Wilkes,

Mrs. James (Ellen Needs)

|  |  |  |  |
| --- | --- | --- | --- |
| **1** | 893 | 1 | 3 |
| **2** | 894 | 0 | 2 |
| **3** | 895 | 0 | 3 |

Myles, Mr. Thomas Francis

Wirz, Mr.

Albert

male 34.5 0 0 330911 7.8292 N

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| female | 47.0 | 1 | 0 | 363272 | 7.0000 | N |
| male | 62.0 | 0 | 0 | 240276 | 9.6875 | N |
| male | 27.0 | 0 | 0 | 315154 | 8.6625 | N |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | Hirvonen, |  | | | | | | |
| Mrs. |
| **4** | 896 | 1 | 3 | Alexander | female | 22.0 | 1 | 1 | 3101298 | 12.2875 | N |
|  |  |  |  | (Helga E |  |  |  |  |  |  |  |
|  |  |  |  | Lindqvist) |  |  |  |  |  |  |  |

In [5]:

data.dtypes

Out[5]:

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 [6]:

data.columns

Out[6]:

Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp', 'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],

dtype='object')

data.tail()

Out[7]:

**PassengerId Survived Pclass Name Sex Age SibSp Parch Ticket F**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **413** | 1305 | 0 | 3 | Spector, Mr. Woolf | male | NaN | 0 | 0 | A.5. 3236 | 8.05 |
| **414** | 1306 | 1 | 1 | Oliva y Ocana, Dona. | female | 39.0 | 0 | 0 | PC 17758 | 108.90 |
|  |  |  |  | Fermina |  |  |  |  |  |  |
| **415** | 1307 | 0 | 3 | Saether,  Mr. Simon | male | 38.5 | 0 | 0 | SOTON/O.Q.  3101262 | 7.25 |
|  |  |  |  | Sivertsen |  |  |  |  |  |  |
| **416** | 1308 | 0 | 3 | Ware,  Mr. | male | NaN | 0 | 0 | 359309 | 8.05 |
|  |  |  |  | Frederick |  |  |  |  |  |  |
| **417** | 1309 | 0 | 3 | Peter, Master. | male | NaN | 1 | 1 | 2668 | 22.35 |
|  |  |  |  | Michael J |  |  |  |  |  |  |

In [8]:

data.info()

<class 'pandas.core.frame.DataFrame'> RangeIndex: 418 entries, 0 to 417

Data columns (total 12 columns):

# Column Non-Null Count Dtype

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 0 |  | PassengerId | 418 | non-null |  | int64 |
| 1 |  | Survived | 418 | non-null |  | int64 |
| 2 |  | Pclass | 418 | non-null |  | int64 |
| 3 |  | Name | 418 | non-null |  | object |
| 4 |  | Sex | 418 | non-null |  | object |
| 5 |  | Age | 332 | non-null |  | float64 |
| 6 |  | SibSp | 418 | non-null |  | int64 |
| 7 |  | Parch | 418 | non-null |  | int64 |
| 8 |  | Ticket | 418 | non-null |  | object |
| 9 |  | Fare | 417 | non-null |  | float64 |
| 10 | Cabin | | 91 non-null | | object | |
| 11 | Embarked | | 418 non-null | | object | |

dtypes: float64(2), int64(5), object(5) memory usage: 39.3+ KB

In [9]:

data.shape

Out[9]:

(418, 12)

data.describe()

Out[10]:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **PassengerId** | **Survived** | **Pclass** | **Age** | **SibSp** | **Parch** | **Fare** |
| **count** | 418.000000 | 418.000000 | 418.000000 | 332.000000 | 418.000000 | 418.000000 | 417.000000 |
| **mean** | 1100.500000 | 0.363636 | 2.265550 | 30.272590 | 0.447368 | 0.392344 | 35.627188 |
| **std** | 120.810458 | 0.481622 | 0.841838 | 14.181209 | 0.896760 | 0.981429 | 55.907576 |
| **min** | 892.000000 | 0.000000 | 1.000000 | 0.170000 | 0.000000 | 0.000000 | 0.000000 |
| **25%** | 996.250000 | 0.000000 | 1.000000 | 21.000000 | 0.000000 | 0.000000 | 7.895800 |
| **50%** | 1100.500000 | 0.000000 | 3.000000 | 27.000000 | 0.000000 | 0.000000 | 14.454200 |
| **75%** | 1204.750000 | 1.000000 | 3.000000 | 39.000000 | 1.000000 | 0.000000 | 31.500000 |
| **max** | 1309.000000 | 1.000000 | 3.000000 | 76.000000 | 8.000000 | 9.000000 | 512.329200 |

In [ ]:

In [12]:

data.isnull().sum()

|  |  |
| --- | --- |
| Out[12]: |  |
| PassengerId | 0 |
| Survived | 0 |
| Pclass | 0 |
| Name | 0 |
| Sex | 0 |
| Age | 86 |
| SibSp | 0 |
| Parch | 0 |
| Ticket | 0 |
| Fare | 1 |
| Cabin | 327 |
| Embarked  dtype: int64 | 0 |
| In [ ]: |  |

In [ ]: