

|  |  |
|--|--|
| <pre>print(df.isna().sum().sum())</pre> <p>مجموعه داده ای titanic.csv که شامل Null نیز است.</p> <pre>titanic_data = pd.read_csv('titanic.csv') print(titanic_data.columns)</pre> <pre>titanic_data.head(5) titanic_data['Age'].mean()</pre> <pre>titanic_data['Sex'].unique() titanic_data['Sex'].unique().size</pre> <p>#?</p> <pre>titanic_data.isnull().any(axis=1).sum()</pre> <p>مجموعه فوق در مجموع چند داده null دارد؟<br/>چه تعداد از ردیف های داده فوق حداقل شامل یک داده (cell) null است؟<br/>خط اول برنامه فوق ((titanic_data = pd.read_csv('titanic.csv'))،<br/>فایل را به یکباره از دیسک خوانده و در رم لوود می کند، اما اگر<br/>فایل بزرگ باشد تکلیف چیست؟</p> <pre>titanic_data_p = pd.read_csv('titanic.csv', skiprows=10, nrows=10)</pre> | <pre>import pandas as pd</pre> <pre>df = pd.DataFrame(columns=['fname', 'lname', 'std_id', 'age', 'sex']) df.loc[0] = ['Ali', 'Mahmoodi', 555, 22, 'M'] df.loc[1] = ['Samira', 'Sadeghi', 333, 23, 'F'] df.loc[2] = ['Parisa', 'Alivand', 666, None, 'F'] df.loc[3] = ['Armin', 'Vahedi', 111, 24, 'M'] df.loc[4] = ['Reza', 'Yasi', 222, 27, 'M'] df.loc[5] = ['Hassan', 'Bahadori', 444, None, None]</pre> <pre>print(df['age'].mean()) print(df['fname'].count())</pre> <pre>x = df['age'] &gt; 23 print(x.count()) print(x.sum()) print(df['age'].isna().sum())</pre> <p>بررسی خروجی:</p> <pre>print(df.isna())</pre> <pre>fname lname std_id age sex 0 False False False False False 1 False False False False False 2 False False False True False 3 False False False False False 4 False False False False False 5 False False False True True</pre> <pre>print(df.isna().sum())</pre> <pre>fname    0 lname    0 std_id    0 age       2 sex       1 dtype: int64</pre> |
|  |  |