In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
df=pd.DataFrame(np.arange(0,20).reshape(5,4),index=['row1','row2','row3','row4','row
```

In [3]:
```
df.head()
```

Out[3]:

	column1	column2	column3	column4
row1	0	1	2	3
row2	4	5	6	7
row3	8	9	10	11
row4	12	13	14	15
row5	16	17	18	19

In [5]:
```
df.to_csv('C:/Users/DGVC/Downloads/Catfish_Trout (1).csv')
```

In [8]:
```
df.loc['row1']
```

Out[8]:
```
column1    0
column2    1
column3    2
column4    3
Name: row1, dtype: int32
```

In [9]:
```
type(df.loc['row1'])
```

Out[9]:
```
pandas.core.series.Series
```

In [10]:
```
df.loc[:,:]
```

Out[10]:

	column1	column2	column3	column4
row1	0	1	2	3
row2	4	5	6	7
row3	8	9	10	11
row4	12	13	14	15
row5	16	17	18	19

In [12]:
```
df.loc[['row1','row2']]
```

Out[12]:

	column1	column2	column3	column4
row1	0	1	2	3
row2	4	5	6	7

In [13]:
```
df.isnull().sum()
```

Out[13]:
```
column1    0
column2    0
column3    0
```

```
column4    0
dtype: int64
```

In [14]:
```
df['column1'].value_counts()
```

Out[14]:
```
12    1
4     1
16    1
8     1
0     1
Name: column1, dtype: int64
```

In [15]:
```
df['column1'].unique()
```

Out[15]:
```
array([ 0,  4,  8, 12, 16])
```

In [16]:
```
df.iloc[:,1:].values
```

Out[16]:
```
array([[ 1,  2,  3],
       [ 5,  6,  7],
       [ 9, 10, 11],
       [13, 14, 15],
       [17, 18, 19]])
```

In [17]:
```
df.iloc[:,1:].values.shape
```

Out[17]:
```
(5, 3)
```

In [30]:
```
df=pd.read_csv('C:/Users/DGVC/Downloads/genome_scores.csv')
```

In [31]:
```
df
```

Out[31]:

	movieId	tagId	relevance
0	1	1	0.02500
1	1	2	0.02500
2	1	3	0.05775
3	1	4	0.09675
4	1	5	0.14675
...	...	...	...
11709763	131170	1124	0.58775
11709764	131170	1125	0.01075
11709765	131170	1126	0.01575
11709766	131170	1127	0.11450
11709767	131170	1128	0.02175

11709768 rows × 3 columns

In [32]:
```
df.head()
```

Out[32]:

	movieId	tagId	relevance
0	1	1	0.02500
1	1	2	0.02500

	movieId	tagId	relevance
2	1	3	0.05775
3	1	4	0.09675
4	1	5	0.14675

In [33]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11709768 entries, 0 to 11709767
Data columns (total 3 columns):
 #   Column     Dtype  
---  ------     -----  
 0   movieId    int64  
 1   tagId      int64  
 2   relevance  float64
dtypes: float64(1), int64(2)
memory usage: 268.0 MB
```

In [34]:
```
df.describe()
```

Out[34]:

	movieId	tagId	relevance
count	1.170977e+07	1.170977e+07	1.170977e+07
mean	2.584297e+04	5.645000e+02	1.164833e-01
std	3.467615e+04	3.256254e+02	1.542463e-01
min	1.000000e+00	1.000000e+00	2.500000e-04
25%	2.926000e+03	2.827500e+02	2.425000e-02
50%	6.017000e+03	5.645000e+02	5.650000e-02
75%	4.606200e+04	8.462500e+02	1.415000e-01
max	1.311700e+05	1.128000e+03	1.000000e+00

In [35]:
```
df.corr()
```

Out[35]:

	movieId	tagId	relevance
movieId	1.000000e+00	-4.271317e-19	0.006900
tagId	-4.271317e-19	1.000000e+00	0.012325
relevance	6.900077e-03	1.232533e-02	1.000000

In [36]:
```
df.isnull()
```

Out[36]:

	movieId	tagId	relevance
0	False	False	False
1	False	False	False
2	False	False	False
3	False	False	False
4	False	False	False
...	...	...	...

	movieId	tagId	relevance
11709763	False	False	False
11709764	False	False	False
11709765	False	False	False
11709766	False	False	False
11709767	False	False	False

11709768 rows × 3 columns

In [37]:
```
df.notnull()
```

Out[37]:

	movieId	tagId	relevance
0	True	True	True
1	True	True	True
2	True	True	True
3	True	True	True
4	True	True	True
...	...	...	...
11709763	True	True	True
11709764	True	True	True
11709765	True	True	True
11709766	True	True	True
11709767	True	True	True

11709768 rows × 3 columns

In [38]:
```
df.dropna()
```

Out[38]:

	movieId	tagId	relevance
0	1	1	0.02500
1	1	2	0.02500
2	1	3	0.05775
3	1	4	0.09675
4	1	5	0.14675
...	...	...	...
11709763	131170	1124	0.58775
11709764	131170	1125	0.01075
11709765	131170	1126	0.01575
11709766	131170	1127	0.11450
11709767	131170	1128	0.02175

11709768 rows × 3 columns

In [44]:
```
df.std()
```

Out[44]:
```
movieId      34676.151996
tagId          325.625438
relevance        0.154246
dtype: float64
```

In [45]:
```
df['tagId'].value_counts()
```

Out[45]:
```
1128    10381
379     10381
373     10381
374     10381
375     10381
        ...
746     10381
745     10381
744     10381
743     10381
1       10381
Name: tagId, Length: 1128, dtype: int64
```

In [49]:
```
df[df['movieId']>100]
```

Out[49]:

	movieId	tagId	relevance
110544	101	1	0.02075
110545	101	2	0.02600
110546	101	3	0.03100
110547	101	4	0.04750
110548	101	5	0.09150
...	...	...	...
11709763	131170	1124	0.58775
11709764	131170	1125	0.01075
11709765	131170	1126	0.01575
11709766	131170	1127	0.11450
11709767	131170	1128	0.02175

11599224 rows × 3 columns

In [50]:
```
df.head()
```

Out[50]:

	movieId	tagId	relevance
0	1	1	0.02500
1	1	2	0.02500
2	1	3	0.05775
3	1	4	0.09675
4	1	5	0.14675

In [52]:
```
import numpy as np
```

In [53]:
```
lst_data=[[1,2,3],[3,4,np.nan],[5,6,np.nan],[np.nan,np.nan,np.nan]]
```

In [55]:
```
type(lst_data)
```

Out[55]:
```
list
```

In [56]:
```
df=pd.DataFrame(lst_data)
```

In [57]:
```
df.head()
```

Out[57]:

	0	1	2
0	1.0	2.0	3.0
1	3.0	4.0	NaN
2	5.0	6.0	NaN
3	NaN	NaN	NaN

In [58]:
```
df.dropna(axis=0)
```

Out[58]:

	0	1	2
0	1.0	2.0	3.0

In [59]:
```
df.dropna(axis=1)
```

Out[59]:

0
1
2
3

In [60]:
```
df = pd.DataFrame(np.random.randn(5, 3), index=['a', 'c', 'e', 'f', 'h'],
                  columns=['one', 'two', 'three'])
```

In [61]:
```
df.head()
```

Out[61]:

	one	two	three
a	1.474252	-1.250456	0.135534
c	0.761218	1.036719	0.162914
e	-0.050396	-1.360936	0.488408
f	-0.359059	-1.865125	-1.109839
h	-0.323184	1.229046	0.528253

In [62]:
```
df2=df.reindex(['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h'])
```

In [63]:
```
df2
```

Out[63]:

	one	two	three
a	1.474252	-1.250456	0.135534
b	NaN	NaN	NaN
c	0.761218	1.036719	0.162914
d	NaN	NaN	NaN
e	-0.050396	-1.360936	0.488408
f	-0.359059	-1.865125	-1.109839
g	NaN	NaN	NaN
h	-0.323184	1.229046	0.528253

In [64]:
```
df.tail()
```

Out[64]:

	one	two	three
a	1.474252	-1.250456	0.135534
c	0.761218	1.036719	0.162914
e	-0.050396	-1.360936	0.488408
f	-0.359059	-1.865125	-1.109839
h	-0.323184	1.229046	0.528253

In [66]:
```
df.min()
```

Out[66]:
```
one     -0.359059
two     -1.865125
three   -1.109839
dtype: float64
```

In [67]:
```
df.max()
```

Out[67]:
```
one      1.474252
two      1.229046
three    0.528253
dtype: float64
```

In [69]:
```
df.mean()
```

Out[69]:
```
one      0.300566
two     -0.442151
three    0.041054
dtype: float64
```

In [70]:
```
df.median()
```

Out[70]:
```
one     -0.050396
two     -1.250456
three    0.162914
dtype: float64
```

In [71]:
```
df.count()
```

Out[71]:
```
one      5
two      5
three    5
dtype: int64
```

In [83]:
```
df2.dropna(axis=0)
```

Out[83]:

	one	two	three
a	1.474252	-1.250456	0.135534
c	0.761218	1.036719	0.162914
e	-0.050396	-1.360936	0.488408
f	-0.359059	-1.865125	-1.109839
h	-0.323184	1.229046	0.528253

In [84]:
```
pd.isna(df2['one'])
```

Out[84]:
```
a    False
b     True
c    False
d     True
e    False
f    False
g     True
h    False
Name: one, dtype: bool
```

In [85]:
```
df2['one'].notna()
```

Out[85]:
```
a     True
b    False
c     True
d    False
e     True
f     True
g    False
h     True
Name: one, dtype: bool
```

In [86]:
```
df2.fillna('Missing')
```

Out[86]:

	one	two	three
a	1.47425	-1.25046	0.135534
b	Missing	Missing	Missing
c	0.761218	1.03672	0.162914
d	Missing	Missing	Missing
e	-0.0503963	-1.36094	0.488408
f	-0.359059	-1.86513	-1.10984
g	Missing	Missing	Missing
h	-0.323184	1.22905	0.528253

In [87]:
```
df2['one'].values
```

Out[87]:
```
array([ 1.47425192,         nan,  0.76121784,         nan, -0.05039626,
       -0.35905947,         nan, -0.32318361])
```

In [92]:
```
df2['one'].unique
```

Out[92]:
```
<bound method Series.unique of a    1.474252
b         NaN
c    0.761218
d         NaN
e   -0.050396
f   -0.359059
g         NaN
h   -0.323184
Name: one, dtype: float64>
```

In [93]:
```
df2['one'].shape
```

Out[93]:
```
(8,)
```

In [99]:
```
df.count()
```

Out[99]:
```
one      5
two      5
three    5
dtype: int64
```

In [ ]:

In [ ]: