

Day 15

import pandas as pd

import numpy as np

a=np.random.randn(5,5)

**df=pd.DataFrame(a,columns=['col1','col2','col3',
 'col4','col5'])**

df1=pd.read_csv('Salary_Data.csv',names=['exp','salary'])

df2=pd.read_csv('sales_data.csv')

df2.index

df2.values

df2.shape

df2.size

df2.columns

df2.dtypes

df2.info()

df2.describe()

df3=df2.head(20)

df4=df2['MONTH_ID']

type(df4)

df5=df2[['QUANTITYORDERED','CITY']]

df2['vehicle-id']=np.arange(2823)

df2['QUANTITYORDERED']=df2['QUANTITYORDERED']+10

import pandas as pd

import numpy as np

a=[1,2,3,4,np.nan]

b=[1,2,3,np.nan,5]

c=[np.nan,1,2,3,4]

d=[1,2,3,4,5]

```
df=pd.DataFrame([a,b,c,d])
```

```
df.dropna(axis=0,how='any')
```

```
Out[3]:
```

```
   0  1  2   3   4
3  1.0  2  3  4.0  5.0
```

```
df.dropna(axis=0,how='all')
```

```
Out[4]:
```

```
   0  1  2   3   4
0  1.0  2  3  4.0 NaN
1  1.0  2  3  NaN  5.0
2  NaN  1  2  3.0  4.0
3  1.0  2  3  4.0  5.0
```

```
df.drop(axis=0,labels=[2])
```

```
Out[5]:
```

```
   0  1  2   3   4
0  1.0  2  3  4.0 NaN
1  1.0  2  3  NaN  5.0
3  1.0  2  3  4.0  5.0
```

```
df.drop(axis=0,labels=[2,3])
```

```
Out[6]:
```

```
   0  1  2   3   4
0  1.0  2  3  4.0 NaN
1  1.0  2  3  NaN  5.0
```

```
df
```

Out[7]:

```
   0 1 2 3 4
0 1.0 2 3 4.0 NaN
1 1.0 2 3 NaN 5.0
2 NaN 1 2 3.0 4.0
3 1.0 2 3 4.0 5.0
```

df.drop(axis=1,labels=[2])

Out[9]:

```
   0 1 3 4
0 1.0 2 4.0 NaN
1 1.0 2 NaN 5.0
2 NaN 1 3.0 4.0
3 1.0 2 4.0 5.0
```

df1=pd.read_csv('Data.csv')

df1.dropna(axis=0,how='any')

Out[11]:

```
   Country  Age  Salary  Purchased
0  France  44.0  72000.0      No
1  Spain   27.0  48000.0     Yes
2  Germany 30.0  54000.0      No
3  Spain   38.0  61000.0      No
5  France  35.0  58000.0     Yes
7  France  48.0  79000.0     Yes
8  Germany 50.0  83000.0      No
9  France  37.0  67000.0     Yes
```

df2=pd.read_csv('Social_Network.csv')

```
df2['Age']>25
```

```
Out[14]:
```

```
0    False
1     True
2     True
3     True
4    False
```

```
395    True
396    True
397    True
398    True
399    True
```

```
Name: Age, Length: 400, dtype: bool
```

```
con=df2['Age']>25
```

```
type(con)
```

```
Out[15]: pandas.core.series.Series
```

```
df2[con]
```

```
Out[16]:
```

| | Age | EstimatedSalary | Purchased |
|-----|-----|-----------------|-----------|
| 1 | 35 | 20000 | 0 |
| 2 | 26 | 43000 | 0 |
| 3 | 27 | 57000 | 0 |
| 5 | 27 | 58000 | 0 |
| 6 | 27 | 84000 | 0 |
| .. | ... | ... | ... |
| 395 | 46 | 41000 | 1 |

| | | | |
|-----|----|-------|---|
| 396 | 51 | 23000 | 1 |
| 397 | 50 | 20000 | 1 |
| 398 | 36 | 33000 | 0 |
| 399 | 49 | 36000 | 1 |

[351 rows x 3 columns]

```
df3=df2[con]
```

```
con1=df2['Age']>25
```

```
type(con1)
```

```
df3=df2[con1]
```

```
con2=df2['EstimatedSalary']<50000
```

```
df4=df[con1 & con2]
```

```
df5=df2[con1 | con2]
```

```
df6=df2[df2['Age'].isin([19,25,35])]
```

```
df7=df2[df2['Age'].between(20,30)]
```

```
df2.sort_values(by='Age')
```

Out[23]:

| | Age | EstimatedSalary | Purchased |
|-----|-----|-----------------|-----------|
| 76 | 18 | 52000 | 0 |
| 165 | 18 | 86000 | 0 |
| 51 | 18 | 44000 | 0 |
| 14 | 18 | 82000 | 0 |
| 141 | 18 | 68000 | 0 |

```
.. ...      ...      ...
393 60      42000      1
370 60      46000      1
223 60      102000     1
355 60      34000      1
371 60      83000      1
```

[400 rows x 3 columns]

df2.sort_values(by='Age')

Out[24]:

```
Age EstimatedSalary Purchased
76  18      52000      0
165 18      86000      0
51  18      44000      0
14  18      82000      0
141 18      68000      0
.. ...      ...      ...
393 60      42000      1
370 60      46000      1
223 60      102000     1
355 60      34000      1
371 60      83000      1
```

[400 rows x 3 columns]

df8=df2.sort_values(by='Age')

df8=df2.sort_values(by='EstimatedSalary')

```
df8=df2.sort_values(by='EstimatedSalary',ascending=False)
```

df2

Out[28]:

| | Age | EstimatedSalary | Purchased |
|-----|-----|-----------------|-----------|
| 0 | 19 | 19000 | 0 |
| 1 | 35 | 20000 | 0 |
| 2 | 26 | 43000 | 0 |
| 3 | 27 | 57000 | 0 |
| 4 | 19 | 76000 | 0 |
| .. | ... | ... | ... |
| 395 | 46 | 41000 | 1 |
| 396 | 51 | 23000 | 1 |
| 397 | 50 | 20000 | 1 |
| 398 | 36 | 33000 | 0 |
| 399 | 49 | 36000 | 1 |

[400 rows x 3 columns]

```
df2.sort_values(by='EstimatedSalary')
```

Out[29]:

| | Age | EstimatedSalary | Purchased |
|-----|-----|-----------------|-----------|
| 92 | 26 | 15000 | 0 |
| 43 | 30 | 15000 | 0 |
| 155 | 31 | 15000 | 0 |
| 150 | 26 | 15000 | 0 |
| 32 | 21 | 16000 | 0 |
| .. | ... | ... | ... |

| | | | |
|-----|----|--------|---|
| 168 | 29 | 148000 | 1 |
| 103 | 33 | 149000 | 1 |
| 240 | 42 | 149000 | 1 |
| 211 | 52 | 150000 | 1 |
| 7 | 32 | 150000 | 1 |

[400 rows x 3 columns]

```
df9=df2.sort_values(by=['Age','EstimatedSalary'])
```

```
dfr=pd.read_csv('drinks.csv')
```

```
dfr.set_index('country')
```

Out[35]:

| | beer_servings | ... | continent |
|-------------|---------------|-----|-----------|
| country | ... | | |
| Afghanistan | 0 | ... | AS |
| Albania | 89 | ... | EU |
| Algeria | 25 | ... | AF |
| Andorra | 245 | ... | EU |
| Angola | 217 | ... | AF |
| ... | ... | ... | |
| Venezuela | 333 | ... | SA |
| Vietnam | 111 | ... | AS |
| Yemen | 6 | ... | AS |
| Zambia | 32 | ... | AF |
| Zimbabwe | 64 | ... | AF |

[193 rows x 5 columns]


```
dfr1=dfr.set_index('country')
```

```
dfr1=dfr.set_index(['country','continent'])
```

```
dfr2=pd.read_csv('drinks.csv',index_col='country')
```

```
dfr2.loc['India']
```

Out[39]:

| | |
|------------------------------|-----|
| beer_servings | 9 |
| spirit_servings | 114 |
| wine_servings | 0 |
| total_litres_of_pure_alcohol | 2.2 |
| continent | AS |

Name: India, dtype: object

```
a=dfr2.loc['India']
```

```
b=dfr2.loc[['India','Australia','France']]
```

```
c=dfr2.iloc[5]
```

```
d=c=dfr2.iloc[5:20]
```

```
c=dfr2.iloc[5]
```

```
dfr1=dfr.set_index(['country','continent'])
```