15. Pandas – DataFrame – Sorting

Contents

15. Pandas – DataFrame – Sorting

1. Sorting

- ✓ DataFrame contains group of rows, column, index and values.
- ✓ Based on requirement we can applying sorting on column name, index etc.

Creating a dataframe:

- ✓ If we load a csv file in pandas then it returns dataframe

Program Name: Loading csv file
Input file: demo1.py
sales4.csv

```
import pandas as pd

df = pd.read_csv("sales4.csv")
print(df)
```

Output

```
        Order_Id Customer_Name  Customer_Id               Product_Name  Product_Cost
0           1023         Venki           15            27in FHD Monitor         59000
1           1024     Chaithanya          14                  iPhone 11         69000
2           1025        Shahid           20  Bose SoundSport Headphones         65999
3           1026         Veeru            3       Apple iPad 10.2-inch         63999
4           1027          Venu           23               Google Phone         63999
...          ...           ...          ...                         ...           ...
299995    301018       Karteek            4       Apple iPad 10.2-inch         51999
299996    301019         Veeru            3         Macbook Pro Laptop         51999
299997    301020        Harsha            5         LG Washing Machine         65000
299998    301021     Nireekshan           1                  LG Mobile         60000
299999    301022       Pradhan           17     34in Ultrawide Monitor         55000

[300000 rows x 5 columns]
```

1.1. sort_values(by="column name")

- ✓ sort_values(p) is predefined method in DataFrame class.
- ✓ We should access this method by using DataFrame object.
- ✓ This method sort the values based on column which we specified.
  - Number default sorting is ascending order
  - String default sorting is alphabetical order

Program Name: Sorting dataframe by using sort_values(p) method
demo2.py
Input file: sales4.csv

```
import pandas as pd

df1 = pd.read_csv("sales4.csv")
df2 = df1.sort_values(by = "Product_Cost")

print(df2)
```

Output

```
        Order_Id Customer_Name  Customer_Id             Product_Name  Product_Cost
52994      54017       Sumanth           22   Samsung Galaxy S9 Plus         50000
122987    124010       Karteek            4         27in FHD Monitor         50000
16661      17684          Venu           23             Google Phone         50000
122986    124009        Harsha            5                iPhone 9         50000
122974    123997       Lavanya           16   34in Ultrawide Monitor         50000
...          ...           ...          ...                      ...           ...
273569    274592  Jaya Chandra           21       Macbook Pro Laptop         75999
273565    274588        Partha            8  Apple Airpods Headphones         75999
171470    172493         Veeru            3         27in FHD Monitor         75999
76102      77125       Lavanya           16       Samsung Galaxy S20         75999
76472      77495     Nireekshan            1       Samsung Galaxy S20         75999

[300000 rows x 5 columns]
```

Program Name: Sorting dataframe by using sort_values(p) method
Input file: demo3.py
sales4.csv

```
import pandas as pd

df1 = pd.read_csv("sales4.csv")
df2 = df1.sort_values(by = "Customer_Id")

print(df2)
```

Output

```
        Order_Id Customer_Name  Customer_Id           Product_Name  Product_Cost
160540    161563    Nireekshan            1       27in FHD Monitor         75000
158658    159681    Nireekshan            1     Macbook Pro Laptop         63999
266895    267918    Nireekshan            1     LG Washing Machine         59000
266908    267931    Nireekshan            1           Google Phone         55000
183395    184418    Nireekshan            1  LG ThinQ Refrigerator         69000
...           ...           ...          ...                    ...           ...
104589    105612         Shafi           25  LG ThinQ Refrigerator         69999
194864    195887         Shafi           25              iPhone 11         63999
223127    224150         Shafi           25     Samsung Galaxy S20         75999
249626    250649         Shafi           25     Macbook Pro Laptop         61000
109653    110676         Shafi           25  Samsung Galaxy S9 Plus         65999

[300000 rows x 5 columns]
```

Program Name: Sorting dataframe by using sort_values(p) method
Input file: demo4.py
sales4.csv

```
import pandas as pd

df1 = pd.read_csv("sales4.csv")
df2 = df1.sort_values(by = "Customer_Id", ascending = False)

print(df2)
```

Output

```
        Order_Id Customer_Name  Customer_Id               Product_Name  Product_Cost
225837    226860         Shafi           25       Apple iPad 10.2-inch         69999
58299      59322         Shafi           25            ThinkPad Laptop         63999
122755    123778         Shafi           25          Samsung Galaxy S20         61000
220596    221619         Shafi           25  Bose SoundSport Headphones         69999
9252       10275         Shafi           25              Flatscreen TV         60000
...          ...           ...          ...                        ...           ...
191789    192812    Nireekshan            1              Flatscreen TV         69999
146543    147566    Nireekshan            1                  iPhone 7s         60000
283770    284793    Nireekshan            1   Apple Airpods Headphones         69000
79317      80340    Nireekshan            1               20in Monitor         69000
78470      79493    Nireekshan            1                  iPhone 7s         61000

[300000 rows x 5 columns]
```

Program Name: Sorting dataframe by using sort_values(p) method
Input file: demo5.py
sales4.csv

```
import pandas as pd

df1 = pd.read_csv("sales4.csv")
df2 = df1.sort_values(by = "Customer_Id", ascending = 0)

print(df2)
```

Output

```
        Order_Id Customer_Name  Customer_Id               Product_Name  Product_Cost
225837    226860         Shafi           25       Apple iPad 10.2-inch         69999
58299      59322         Shafi           25           ThinkPad Laptop         63999
122755    123778         Shafi           25         Samsung Galaxy S20         61000
220596    221619         Shafi           25  Bose SoundSport Headphones         69999
9252       10275         Shafi           25              Flatscreen TV         60000
...          ...           ...          ...                        ...           ...
191789    192812    Nireekshan            1              Flatscreen TV         69999
146543    147566    Nireekshan            1                  iPhone 7s         60000
283770    284793    Nireekshan            1   Apple Airpods Headphones         69000
79317      80340    Nireekshan            1               20in Monitor         69000
78470      79493    Nireekshan            1                  iPhone 7s         61000

[300000 rows x 5 columns]
```

Program Name: Sorting dataframe by using sort_values() method
Input file: demo6.py
sales4.csv

```
import pandas as pd

df1 = pd.read_csv("sales4.csv")
df2 = df1.sort_values(by = "Customer_Name")

print(df2)
```

Output

```
        Order_Id Customer_Name  Customer_Id            Product_Name  Product_Cost
136196    137219        Balaji           12              LG Mobile         60000
128754    129777        Balaji           12     Samsung Galaxy S20         61000
292113    293136        Balaji           12              LG Mobile         65999
128753    129776        Balaji           12  LG ThinQ Refrigerator         65999
73777      74800        Balaji           12   Apple iPad 10.2-inch         51999
...          ...           ...          ...                     ...           ...
280203    281226         Vinay           10               iPhone 8         65999
155180    156203         Vinay           10  Samsung Galaxy S9 Plus         65000
82486      83509         Vinay           10  LG ThinQ Refrigerator         59000
26986      28009         Vinay           10              iPhone 11         59000
186628    187651         Vinay           10  LG ThinQ Refrigerator         51999

[300000 rows x 5 columns]
```

1.2. sort_index()

- ✓ sort_index() is predefined method in DataFrame class.
- ✓ We should access this method by using DataFrame object.
- ✓ This method sort the indexes in DataFrame

Program Name: Sorting dataframe by using sort_index()
demo7.py

```
import pandas as pd

d = {
      'Order id': [11, 21, 31],
      'Customer name': ['Kedar', 'Nireekshan', 'Daniel'],
      'Product': ['iPhone 11','hTC', 'macbook']
}

i = [555, 444, 333]

df1 = pd.DataFrame(d, index = i)
df2 = df1.sort_index()

print(df1)
print()
print(df2)
```

Output

```
     Order id Customer name    Product
555        11         Kedar  iPhone 11
444        21    Nireekshan        hTC
333        31        Daniel    macbook

     Order id Customer name    Product
333        31        Daniel    macbook
444        21    Nireekshan        hTC
555        11         Kedar  iPhone 11
```