Pandas

- DataFrame
  - it is a table with multiple rows and columns
  - it can read different types of files
  - example : csv,excel,html,json

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

In [2]:
```
pd.Series([1,2,3])
```

Out[2]:
```
0    1
1    2
2    3
dtype: int64
```

In [3]:
```
# creating a dataframe

d1 = pd.DataFrame([1,2,3,4])
d1
```

Out[3]:

	0
0	1
1	2
2	3
3	4

In [4]:
```
d2 = pd.DataFrame([["apssdc",22,2021],["workshop",5,2021]])
d2
```

Out[4]:

	0	1	2
0	apssdc	22	2021
1	workshop	5	2021

In [5]:
```
d2 = pd.DataFrame([["apssdc",22,2021],["workshop",5,2021]],columns = ("name","da
                  index = ["a","b"])
```

In [6]:
```
d2
```

Out[6]:

	name	date	year
a	apssdc	22	2021
b	workshop	5	2021

In [7]:
```
dict1 = {"Name":["apssdc","workshop"],"date":[22,5],"year":[2021,2020]}
dict1
```

Out[7]:
```
{'Name': ['apssdc', 'workshop'], 'date': [22, 5], 'year': [2021, 2020]}
```

In [10]:
```
d3 = pd.DataFrame(dict1,index = ["a","b"])
d3
```

Out[10]:

	Name	date	year
a	apssdc	22	2021
b	workshop	5	2020

In [11]:
```
d3.index
```

Out[11]:
```
Index(['a', 'b'], dtype='object')
```

In [12]:
```
d3.columns
```

Out[12]:
```
Index(['Name', 'date', 'year'], dtype='object')
```

In [13]:
```
d3.shape
```

Out[13]:
```
(2, 3)
```

In [14]:
```
# Accessing the elements
```

In [16]:
```
d3["Name"]# getting columns
```

Out[16]:
```
a      apssdc
b    workshop
Name: Name, dtype: object
```

In [19]:
```
d3["Name"]["b"]
```

Out[19]:
```
'workshop'
```

In [20]:
```
d3["Name"][1]
```

Out[20]:
```
'workshop'
```

In [21]:
```
d3[-1:]
```

Out[21]:

	Name	date	year
b	workshop	5	2020

In [24]:
```
d3[0:1]
```

Out[24]:

	Name	date	year
a	apssdc	22	2021

In [25]:
```
# Load dataset
```

In [26]:
```
data = pd.read_csv("market.csv")
```

In [27]:
```
data
```

Out[27]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Sh
0	Ord_5446	Prod_16	SHP_7609	Cust_1818	136.8100	0.01	23	-30.51	
1	Ord_5406	Prod_13	SHP_7549	Cust_1818	42.2700	0.01	13	4.56	
2	Ord_5446	Prod_4	SHP_7610	Cust_1818	4701.6900	0.00	26	1148.90	
3	Ord_5456	Prod_6	SHP_7625	Cust_1818	2337.8900	0.09	43	729.34	
4	Ord_5485	Prod_17	SHP_7664	Cust_1818	4233.1500	0.08	35	1219.87	
5	Ord_5446	Prod_6	SHP_7608	Cust_1818	164.0200	0.03	23	-47.64	
6	Ord_31	Prod_12	SHP_41	Cust_26	14.7600	0.01	5	1.32	
7	Ord_4725	Prod_4	SHP_6593	Cust_1641	3410.1575	0.10	48	1137.91	
8	Ord_4725	Prod_13	SHP_6593	Cust_1641	162.0000	0.01	33	45.84	
9	Ord_4725	Prod_6	SHP_6593	Cust_1641	57.2200	0.07	8	-27.72	
10	Ord_4743	Prod_2	SHP_6615	Cust_1641	4072.0100	0.01	43	1675.98	
11	Ord_1925	Prod_6	SHP_2637	Cust_708	465.9000	0.05	38	79.34	
12	Ord_2978	Prod_16	SHP_4112	Cust_1088	305.0500	0.04	27	23.12	
13	Ord_2207	Prod_11	SHP_3093	Cust_839	3364.2480	0.10	15	-693.23	
14	Ord_2207	Prod_10	SHP_3006	Cust_839	1410.9300	0.08	10	-317.48	
15	Ord_2280	Prod_5	SHP_3114	Cust_839	460.6900	0.06	48	-103.48	
16	Ord_2282	Prod_9	SHP_3122	Cust_839	443.4600	0.06	30	193.12	
17	Ord_4471	Prod_15	SHP_6228	Cust_1521	13255.9300	0.02	25	4089.27	
18	Ord_4427	Prod_6	SHP_6171	Cust_1521	283.1300	0.08	45	-141.26	
19	Ord_996	Prod_13	SHP_1378	Cust_371	41.9700	0.05	12	-37.03	
20	Ord_996	Prod_13	SHP_1378	Cust_371	57.1700	0.08	18	-24.03	
21	Ord_996	Prod_6	SHP_1378	Cust_371	81.2500	0.01	11	-44.54	
22	Ord_996	Prod_5	SHP_1377	Cust_371	3202.2500	0.09	44	991.26	
23	Ord_996	Prod_7	SHP_1378	Cust_371	35.6400	0.05	10	-0.71	
24	Ord_2573	Prod_3	SHP_3525	Cust_931	197.6100	0.08	13	3.46	
25	Ord_2335	Prod_13	SHP_3204	Cust_931	38.2600	0.03	22	-2.34	
26	Ord_2456	Prod_5	SHP_3367	Cust_931	109.5800	0.00	13	31.32	
27	Ord_2405	Prod_9	SHP_3300	Cust_931	1062.6900	0.01	28	401.80	
28	Ord_2573	Prod_4	SHP_3527	Cust_931	3594.7435	0.05	38	1016.97	
29	Ord_2478	Prod_12	SHP_3395	Cust_931	139.9800	0.07	33	-140.54	
...	...	...	...	...	...	...	...	...	
8369	Ord_3633	Prod_3	SHP_5031	Cust_1274	1169.2600	0.02	41	515.62	
8370	Ord_2696	Prod_13	SHP_3690	Cust_1006	62.7800	0.04	20	-17.75	
8371	Ord_2624	Prod_4	SHP_3591	Cust_1006	4924.1350	0.07	28	1049.54	

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Sh
8372	Ord_2772	Prod_9	SHP_3806	Cust_1006	56.9000	0.03	7	12.64	
8373	Ord_2600	Prod_16	SHP_3560	Cust_1006	106.6400	0.10	30	-31.95	
8374	Ord_2658	Prod_5	SHP_3637	Cust_1006	1082.6600	0.08	14	-256.93	
8375	Ord_2772	Prod_3	SHP_3806	Cust_1006	1413.8200	0.10	47	226.53	
8376	Ord_2624	Prod_8	SHP_3590	Cust_1006	1211.0000	0.00	36	-27.99	
8377	Ord_2722	Prod_12	SHP_3729	Cust_1006	34.0100	0.00	12	10.58	
8378	Ord_2706	Prod_2	SHP_3705	Cust_1006	1361.9100	0.05	20	312.52	
8379	Ord_2722	Prod_5	SHP_3730	Cust_1006	1008.9500	0.04	41	69.31	
8380	Ord_2772	Prod_6	SHP_3807	Cust_1006	308.9200	0.04	45	-143.58	
8381	Ord_2696	Prod_4	SHP_3691	Cust_1006	2836.0505	0.01	25	561.13	
8382	Ord_2658	Prod_3	SHP_3636	Cust_1006	120.9800	0.00	28	-92.85	
8383	Ord_2722	Prod_1	SHP_3731	Cust_1006	3508.3300	0.04	21	-546.98	
8384	Ord_4620	Prod_3	SHP_6435	Cust_1577	59.6200	0.04	10	-56.30	
8385	Ord_1833	Prod_3	SHP_2527	Cust_637	611.1600	0.04	46	100.22	
8386	Ord_2324	Prod_7	SHP_3189	Cust_851	121.8700	0.07	39	11.32	
8387	Ord_2220	Prod_3	SHP_3019	Cust_851	41.0600	0.04	4	-16.39	
8388	Ord_4424	Prod_1	SHP_6165	Cust_1519	994.0400	0.03	10	-335.06	
8389	Ord_4444	Prod_13	SHP_6192	Cust_1519	159.4100	0.00	44	34.68	
8390	Ord_5435	Prod_16	SHP_7594	Cust_1798	316.9900	0.04	47	-276.54	
8391	Ord_5435	Prod_4	SHP_7594	Cust_1798	1991.8985	0.07	20	88.36	
8392	Ord_5384	Prod_9	SHP_7519	Cust_1798	181.5000	0.08	43	-6.24	
8393	Ord_5348	Prod_8	SHP_7470	Cust_1798	356.7200	0.07	9	12.61	
8394	Ord_5353	Prod_4	SHP_7479	Cust_1798	2841.4395	0.08	28	374.63	
8395	Ord_5411	Prod_6	SHP_7555	Cust_1798	127.1600	0.10	20	-74.03	
8396	Ord_5388	Prod_6	SHP_7524	Cust_1798	243.0500	0.02	39	-70.85	
8397	Ord_5348	Prod_15	SHP_7469	Cust_1798	3872.8700	0.03	23	565.34	
8398	Ord_5459	Prod_6	SHP_7628	Cust_1798	603.6900	0.00	47	131.39	

8399 rows × 10 columns

In [28]:
```
data.shape
```

Out[28]: (8399, 10)

In [29]:
```
data.head()
```

Out[29]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Shipping
0	Ord_5446	Prod_16	SHP_7609	Cust_1818	136.81	0.01	23	-30.51	
1	Ord_5406	Prod_13	SHP_7549	Cust_1818	42.27	0.01	13	4.56	
2	Ord_5446	Prod_4	SHP_7610	Cust_1818	4701.69	0.00	26	1148.90	
3	Ord_5456	Prod_6	SHP_7625	Cust_1818	2337.89	0.09	43	729.34	
4	Ord_5485	Prod_17	SHP_7664	Cust_1818	4233.15	0.08	35	1219.87	

In [31]:
```
data.tail(2)
```

Out[31]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Shippi
8397	Ord_5348	Prod_15	SHP_7469	Cust_1798	3872.87	0.03	23	565.34	
8398	Ord_5459	Prod_6	SHP_7628	Cust_1798	603.69	0.00	47	131.39	

In [32]:
```
data.sample(3)
```

Out[32]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Shippi
1102	Ord_4663	Prod_9	SHP_6498	Cust_1612	2026.42	0.03	50	655.36	
337	Ord_4062	Prod_5	SHP_5665	Cust_1385	2504.41	0.01	30	-547.89	
1293	Ord_5	Prod_3	SHP_6	Cust_5	93.54	0.03	12	-54.04	

In [33]:
```
data.columns
```

Out[33]:
```
Index(['Ord_id', 'Prod_id', 'Ship_id', 'Cust_id', 'Sales', 'Discount',
       'Order_Quantity', 'Profit', 'Shipping_Cost', 'Product_Base_Margin'],
      dtype='object')
```

In [34]:
```
data.index
```

Out[34]:
```
RangeIndex(start=0, stop=8399, step=1)
```

In [35]:
```
# statisical Analysis
```

In [36]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8399 entries, 0 to 8398
Data columns (total 10 columns):
Ord_id                  8399 non-null object
Prod_id                 8399 non-null object
Ship_id                 8399 non-null object
Cust_id                 8399 non-null object
Sales                   8399 non-null float64
Discount                8399 non-null float64
Order_Quantity          8399 non-null int64
Profit                  8399 non-null float64
Shipping_Cost           8399 non-null float64
Product_Base_Margin     8336 non-null float64
dtypes: float64(5), int64(1), object(4)
memory usage: 656.2+ KB
```

In [37]:
```
data.describe()
```

Out[37]:

	Sales	Discount	Order_Quantity	Profit	Shipping_Cost	Product_Base_Mar
count	8399.000000	8399.000000	8399.000000	8399.000000	8399.000000	8336.000
mean	1775.878179	0.049671	25.571735	181.184424	12.838557	0.512
std	3585.050525	0.031823	14.481071	1196.653371	17.264052	0.135
min	2.240000	0.000000	1.000000	-14140.700000	0.490000	0.350
25%	143.195000	0.020000	13.000000	-83.315000	3.300000	0.380
50%	449.420000	0.050000	26.000000	-1.500000	6.070000	0.520
75%	1709.320000	0.080000	38.000000	162.750000	13.990000	0.590
max	89061.050000	0.250000	50.000000	27220.690000	164.730000	0.850

In [38]:
```
# selecting rows
```

In [39]:
```
data[10:20]
```

Out[39]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Shipp
10	Ord_4743	Prod_2	SHP_6615	Cust_1641	4072.010	0.01	43	1675.98	
11	Ord_1925	Prod_6	SHP_2637	Cust_708	465.900	0.05	38	79.34	
12	Ord_2978	Prod_16	SHP_4112	Cust_1088	305.050	0.04	27	23.12	
13	Ord_2207	Prod_11	SHP_3093	Cust_839	3364.248	0.10	15	-693.23	
14	Ord_2207	Prod_10	SHP_3006	Cust_839	1410.930	0.08	10	-317.48	
15	Ord_2280	Prod_5	SHP_3114	Cust_839	460.690	0.06	48	-103.48	
16	Ord_2282	Prod_9	SHP_3122	Cust_839	443.460	0.06	30	193.12	
17	Ord_4471	Prod_15	SHP_6228	Cust_1521	13255.930	0.02	25	4089.27	
18	Ord_4427	Prod_6	SHP_6171	Cust_1521	283.130	0.08	45	-141.26	
19	Ord_996	Prod_13	SHP_1378	Cust_371	41.970	0.05	12	-37.03	

In [40]:
```
data[10:20:2]
```

Out[40]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Shippin
10	Ord_4743	Prod_2	SHP_6615	Cust_1641	4072.01	0.01	43	1675.98	
12	Ord_2978	Prod_16	SHP_4112	Cust_1088	305.05	0.04	27	23.12	
14	Ord_2207	Prod_10	SHP_3006	Cust_839	1410.93	0.08	10	-317.48	
16	Ord_2282	Prod_9	SHP_3122	Cust_839	443.46	0.06	30	193.12	
18	Ord_4427	Prod_6	SHP_6171	Cust_1521	283.13	0.08	45	-141.26	

In [41]:
```
# selecting columns
```

In [43]:
```
data["Sales"].head()
```

Out[43]:
```
0     136.81
1      42.27
2    4701.69
3    2337.89
4    4233.15
Name: Sales, dtype: float64
```

In [48]:
```
data[["Sales","Profit","Discount"]].head()
```

Out[48]:

	Sales	Profit	Discount
0	False	False	False
1	False	False	False
2	False	False	False
3	False	False	False
4	False	False	False

In [52]:
```
data.isnull().sum()
```

Out[52]:
```
Ord_id                 0
Prod_id                0
Ship_id                0
Cust_id                0
Sales                  0
Discount               0
Order_Quantity         0
Profit                 0
Shipping_Cost          0
Product_Base_Margin    63
dtype: int64
```

In [ ]:
```
# iloc and loc
# iloc-->Position based index
#Loc -->Label based index
```

In [53]:
```
data.iloc[1000]
```

Out[53]:
```
Ord_id                Ord_1895
Prod_id                 Prod_4
Ship_id               SHP_2601
Cust_id               Cust_686
Sales                 2433.55
Discount                 0.05
Order_Quantity             23
Profit                355.93
Shipping_Cost            8.08
Product_Base_Margin      0.57
Name: 1000, dtype: object
```

In [54]:
```
data.iloc[1000:2000]
```

Out[54]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Sh
1000	Ord_1895	Prod_4	SHP_2601	Cust_686	2433.5500	0.05	23	355.93	
1001	Ord_1910	Prod_8	SHP_3095	Cust_891	587.2000	0.08	29	134.25	
1002	Ord_1910	Prod_12	SHP_3096	Cust_891	110.7900	0.05	39	42.68	
1003	Ord_2025	Prod_6	SHP_2769	Cust_897	162.3900	0.10	33	-101.68	
1004	Ord_97	Prod_5	SHP_132	Cust_70	426.3400	0.00	31	73.53	
1005	Ord_187	Prod_12	SHP_252	Cust_70	35.1700	0.03	11	5.58	
1006	Ord_319	Prod_1	SHP_426	Cust_70	178.4500	0.09	31	-104.96	
1007	Ord_319	Prod_5	SHP_425	Cust_70	413.8600	0.00	50	39.29	
1008	Ord_28	Prod_3	SHP_38	Cust_23	42.6600	0.06	5	-11.83	
1009	Ord_1173	Prod_6	SHP_1619	Cust_449	355.9700	0.03	7	-4.46	
1010	Ord_1182	Prod_13	SHP_1631	Cust_449	79.1400	0.09	30	-4.73	
1011	Ord_1172	Prod_6	SHP_1617	Cust_449	240.1800	0.06	40	-160.21	
1012	Ord_1173	Prod_3	SHP_1619	Cust_449	887.4500	0.10	47	365.79	
1013	Ord_1182	Prod_10	SHP_1630	Cust_449	4514.8600	0.09	23	-641.09	
1014	Ord_1172	Prod_1	SHP_1618	Cust_449	529.2800	0.10	45	-30.27	
1015	Ord_1003	Prod_9	SHP_1387	Cust_376	49.0400	0.02	8	9.29	
1016	Ord_1025	Prod_9	SHP_1416	Cust_376	364.8000	0.01	42	149.45	
1017	Ord_1003	Prod_11	SHP_1388	Cust_376	2320.3500	0.00	6	-237.62	
1018	Ord_907	Prod_15	SHP_1249	Cust_322	159.5100	0.04	1	-103.63	
1019	Ord_907	Prod_4	SHP_1250	Cust_322	5015.0510	0.05	30	1090.43	
1020	Ord_918	Prod_6	SHP_1266	Cust_322	676.2600	0.06	34	181.98	
1021	Ord_918	Prod_8	SHP_1265	Cust_322	397.5500	0.08	5	-152.47	
1022	Ord_4794	Prod_5	SHP_6683	Cust_1657	633.9100	0.00	49	141.17	
1023	Ord_4628	Prod_4	SHP_6445	Cust_1586	849.3370	0.07	8	-258.85	
1024	Ord_4735	Prod_2	SHP_6604	Cust_1586	1756.2700	0.09	38	649.05	
1025	Ord_4835	Prod_6	SHP_6741	Cust_1586	514.8600	0.08	45	-10.84	
1026	Ord_4835	Prod_8	SHP_6740	Cust_1586	437.7700	0.01	26	-125.80	
1027	Ord_4835	Prod_6	SHP_6740	Cust_1586	356.2700	0.10	12	84.53	
1028	Ord_4735	Prod_5	SHP_6605	Cust_1586	345.5800	0.05	50	-38.04	
1029	Ord_4738	Prod_1	SHP_6610	Cust_1586	5290.5700	0.08	30	1389.18	
...	...	...	...	...	...	...	...	...	
1970	Ord_1624	Prod_17	SHP_2242	Cust_525	5744.2400	0.09	12	1118.17	
1971	Ord_1516	Prod_13	SHP_2094	Cust_525	764.3200	0.02	35	-34.59	
1972	Ord_1516	Prod_1	SHP_2093	Cust_525	2536.1800	0.10	19	350.62	

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Sh
1973	Ord_1484	Prod_15	SHP_2049	Cust_525	7834.7700	0.04	37	297.79	
1974	Ord_1398	Prod_5	SHP_1929	Cust_525	1737.0600	0.02	42	339.17	
1975	Ord_428	Prod_6	SHP_572	Cust_148	119.0300	0.06	13	29.62	
1976	Ord_394	Prod_5	SHP_528	Cust_148	742.2100	0.03	50	122.21	
1977	Ord_370	Prod_6	SHP_496	Cust_148	188.0500	0.04	26	-21.41	
1978	Ord_450	Prod_6	SHP_602	Cust_148	945.0300	0.00	47	267.64	
1979	Ord_428	Prod_5	SHP_573	Cust_148	1445.6000	0.08	35	126.03	
1980	Ord_428	Prod_12	SHP_572	Cust_148	28.8100	0.02	10	7.15	
1981	Ord_964	Prod_6	SHP_1333	Cust_351	201.0900	0.08	30	-56.22	
1982	Ord_993	Prod_1	SHP_1374	Cust_368	2645.8000	0.00	31	-684.78	
1983	Ord_993	Prod_5	SHP_1373	Cust_368	1140.2600	0.04	39	387.20	
1984	Ord_992	Prod_6	SHP_1371	Cust_368	282.9800	0.06	45	-237.47	
1985	Ord_992	Prod_15	SHP_1370	Cust_368	5041.4600	0.06	43	-2136.66	
1986	Ord_1014	Prod_5	SHP_1404	Cust_368	131.0900	0.01	10	10.94	
1987	Ord_992	Prod_11	SHP_1372	Cust_368	8817.7100	0.09	27	-2946.05	
1988	Ord_1014	Prod_6	SHP_1404	Cust_368	1817.9000	0.06	33	700.31	
1989	Ord_925	Prod_4	SHP_1280	Cust_331	381.1995	0.08	7	-233.56	
1990	Ord_5075	Prod_3	SHP_7088	Cust_1723	2753.0700	0.07	42	551.34	
1991	Ord_5082	Prod_5	SHP_7095	Cust_1723	93.7300	0.03	6	51.14	
1992	Ord_5069	Prod_3	SHP_7078	Cust_1723	390.4600	0.02	50	-171.17	
1993	Ord_5071	Prod_3	SHP_7080	Cust_1723	211.8600	0.01	34	-239.32	
1994	Ord_5069	Prod_7	SHP_7078	Cust_1723	113.4300	0.02	43	24.27	
1995	Ord_727	Prod_10	SHP_7074	Cust_1723	3645.4500	0.04	30	-765.50	
1996	Ord_5071	Prod_6	SHP_7081	Cust_1723	355.0300	0.00	50	-33.31	
1997	Ord_727	Prod_8	SHP_992	Cust_252	302.0500	0.00	9	-58.81	
1998	Ord_5265	Prod_4	SHP_7350	Cust_1775	994.2280	0.09	8	-446.58	
1999	Ord_5265	Prod_2	SHP_7349	Cust_1775	174.6400	0.04	40	-143.23	

1000 rows × 10 columns

In [55]:
```
data.iloc[[2,5,7,1000]]
```

Out[55]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Shi
2	Ord_5446	Prod_4	SHP_7610	Cust_1818	4701.6900	0.00	26	1148.90	
5	Ord_5446	Prod_6	SHP_7608	Cust_1818	164.0200	0.03	23	-47.64	
7	Ord_4725	Prod_4	SHP_6593	Cust_1641	3410.1575	0.10	48	1137.91	
1000	Ord_1895	Prod_4	SHP_2601	Cust_686	2433.5500	0.05	23	355.93	

In [57]:
```
data.iloc[3:10,4:]
```

Out[57]:

	Sales	Discount	Order_Quantity	Profit	Shipping_Cost	Product_Base_Margin
3	2337.8900	0.09	43	729.34	14.30	0.37
4	4233.1500	0.08	35	1219.87	26.30	0.38
5	164.0200	0.03	23	-47.64	6.15	0.37
6	14.7600	0.01	5	1.32	0.50	0.36
7	3410.1575	0.10	48	1137.91	0.99	0.55
8	162.0000	0.01	33	45.84	0.71	0.52
9	57.2200	0.07	8	-27.72	6.60	0.37

In [58]:
```
data.loc[4,"Sales"]
```

Out[58]: 4233.15

In [59]:
```
data.head()
```

Out[59]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Shipping
0	Ord_5446	Prod_16	SHP_7609	Cust_1818	136.81	0.01	23	-30.51	
1	Ord_5406	Prod_13	SHP_7549	Cust_1818	42.27	0.01	13	4.56	
2	Ord_5446	Prod_4	SHP_7610	Cust_1818	4701.69	0.00	26	1148.90	
3	Ord_5456	Prod_6	SHP_7625	Cust_1818	2337.89	0.09	43	729.34	
4	Ord_5485	Prod_17	SHP_7664	Cust_1818	4233.15	0.08	35	1219.87	

In [62]:
```
data.loc[8,"Sales"]
```

Out[62]: 162.0

In [63]:
```
# Filtering
```

In [70]:
```
data
```

Out[70]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Sh
0	Ord_5446	Prod_16	SHP_7609	Cust_1818	136.8100	0.01	23	-30.51	
1	Ord_5406	Prod_13	SHP_7549	Cust_1818	42.2700	0.01	13	4.56	
2	Ord_5446	Prod_4	SHP_7610	Cust_1818	4701.6900	0.00	26	1148.90	
3	Ord_5456	Prod_6	SHP_7625	Cust_1818	2337.8900	0.09	43	729.34	
4	Ord_5485	Prod_17	SHP_7664	Cust_1818	4233.1500	0.08	35	1219.87	
5	Ord_5446	Prod_6	SHP_7608	Cust_1818	164.0200	0.03	23	-47.64	
6	Ord_31	Prod_12	SHP_41	Cust_26	14.7600	0.01	5	1.32	
7	Ord_4725	Prod_4	SHP_6593	Cust_1641	3410.1575	0.10	48	1137.91	
8	Ord_4725	Prod_13	SHP_6593	Cust_1641	162.0000	0.01	33	45.84	
9	Ord_4725	Prod_6	SHP_6593	Cust_1641	57.2200	0.07	8	-27.72	
10	Ord_4743	Prod_2	SHP_6615	Cust_1641	4072.0100	0.01	43	1675.98	
11	Ord_1925	Prod_6	SHP_2637	Cust_708	465.9000	0.05	38	79.34	
12	Ord_2978	Prod_16	SHP_4112	Cust_1088	305.0500	0.04	27	23.12	
13	Ord_2207	Prod_11	SHP_3093	Cust_839	3364.2480	0.10	15	-693.23	
14	Ord_2207	Prod_10	SHP_3006	Cust_839	1410.9300	0.08	10	-317.48	
15	Ord_2280	Prod_5	SHP_3114	Cust_839	460.6900	0.06	48	-103.48	
16	Ord_2282	Prod_9	SHP_3122	Cust_839	443.4600	0.06	30	193.12	
17	Ord_4471	Prod_15	SHP_6228	Cust_1521	13255.9300	0.02	25	4089.27	
18	Ord_4427	Prod_6	SHP_6171	Cust_1521	283.1300	0.08	45	-141.26	
19	Ord_996	Prod_13	SHP_1378	Cust_371	41.9700	0.05	12	-37.03	
20	Ord_996	Prod_13	SHP_1378	Cust_371	57.1700	0.08	18	-24.03	
21	Ord_996	Prod_6	SHP_1378	Cust_371	81.2500	0.01	11	-44.54	
22	Ord_996	Prod_5	SHP_1377	Cust_371	3202.2500	0.09	44	991.26	
23	Ord_996	Prod_7	SHP_1378	Cust_371	35.6400	0.05	10	-0.71	
24	Ord_2573	Prod_3	SHP_3525	Cust_931	197.6100	0.08	13	3.46	
25	Ord_2335	Prod_13	SHP_3204	Cust_931	38.2600	0.03	22	-2.34	
26	Ord_2456	Prod_5	SHP_3367	Cust_931	109.5800	0.00	13	31.32	
27	Ord_2405	Prod_9	SHP_3300	Cust_931	1062.6900	0.01	28	401.80	
28	Ord_2573	Prod_4	SHP_3527	Cust_931	3594.7435	0.05	38	1016.97	
29	Ord_2478	Prod_12	SHP_3395	Cust_931	139.9800	0.07	33	-140.54	
...	...	...	...	...	...	...	...	...	
8369	Ord_3633	Prod_3	SHP_5031	Cust_1274	1169.2600	0.02	41	515.62	
8370	Ord_2696	Prod_13	SHP_3690	Cust_1006	62.7800	0.04	20	-17.75	
8371	Ord_2624	Prod_4	SHP_3591	Cust_1006	4924.1350	0.07	28	1049.54	

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Sh
8372	Ord_2772	Prod_9	SHP_3806	Cust_1006	56.9000	0.03	7	12.64	
8373	Ord_2600	Prod_16	SHP_3560	Cust_1006	106.6400	0.10	30	-31.95	
8374	Ord_2658	Prod_5	SHP_3637	Cust_1006	1082.6600	0.08	14	-256.93	
8375	Ord_2772	Prod_3	SHP_3806	Cust_1006	1413.8200	0.10	47	226.53	
8376	Ord_2624	Prod_8	SHP_3590	Cust_1006	1211.0000	0.00	36	-27.99	
8377	Ord_2722	Prod_12	SHP_3729	Cust_1006	34.0100	0.00	12	10.58	
8378	Ord_2706	Prod_2	SHP_3705	Cust_1006	1361.9100	0.05	20	312.52	
8379	Ord_2722	Prod_5	SHP_3730	Cust_1006	1008.9500	0.04	41	69.31	
8380	Ord_2772	Prod_6	SHP_3807	Cust_1006	308.9200	0.04	45	-143.58	
8381	Ord_2696	Prod_4	SHP_3691	Cust_1006	2836.0505	0.01	25	561.13	
8382	Ord_2658	Prod_3	SHP_3636	Cust_1006	120.9800	0.00	28	-92.85	
8383	Ord_2722	Prod_1	SHP_3731	Cust_1006	3508.3300	0.04	21	-546.98	
8384	Ord_4620	Prod_3	SHP_6435	Cust_1577	59.6200	0.04	10	-56.30	
8385	Ord_1833	Prod_3	SHP_2527	Cust_637	611.1600	0.04	46	100.22	
8386	Ord_2324	Prod_7	SHP_3189	Cust_851	121.8700	0.07	39	11.32	
8387	Ord_2220	Prod_3	SHP_3019	Cust_851	41.0600	0.04	4	-16.39	
8388	Ord_4424	Prod_1	SHP_6165	Cust_1519	994.0400	0.03	10	-335.06	
8389	Ord_4444	Prod_13	SHP_6192	Cust_1519	159.4100	0.00	44	34.68	
8390	Ord_5435	Prod_16	SHP_7594	Cust_1798	316.9900	0.04	47	-276.54	
8391	Ord_5435	Prod_4	SHP_7594	Cust_1798	1991.8985	0.07	20	88.36	
8392	Ord_5384	Prod_9	SHP_7519	Cust_1798	181.5000	0.08	43	-6.24	
8393	Ord_5348	Prod_8	SHP_7470	Cust_1798	356.7200	0.07	9	12.61	
8394	Ord_5353	Prod_4	SHP_7479	Cust_1798	2841.4395	0.08	28	374.63	
8395	Ord_5411	Prod_6	SHP_7555	Cust_1798	127.1600	0.10	20	-74.03	
8396	Ord_5388	Prod_6	SHP_7524	Cust_1798	243.0500	0.02	39	-70.85	
8397	Ord_5348	Prod_15	SHP_7469	Cust_1798	3872.8700	0.03	23	565.34	
8398	Ord_5459	Prod_6	SHP_7628	Cust_1798	603.6900	0.00	47	131.39	

8399 rows × 10 columns

In [66]:
```
data[data["Sales"]>20000]
```

Out[66]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	S
84	Ord_5232	Prod_17	SHP_7306	Cust_1758	21366.510	0.00	3	-11984.40	
197	Ord_2753	Prod_3	SHP_3778	Cust_997	23106.460	0.08	28	9527.47	
385	Ord_3707	Prod_17	SHP_5136	Cust_1307	28359.400	0.05	49	14440.39	
734	Ord_4705	Prod_17	SHP_6566	Cust_1592	21320.580	0.09	27	5381.02	
798	Ord_2477	Prod_11	SHP_3394	Cust_939	20596.580	0.09	39	-1331.55	
1184	Ord_5181	Prod_17	SHP_7238	Cust_1750	23775.560	0.03	44	7080.99	
1265	Ord_5250	Prod_17	SHP_7331	Cust_1769	24105.870	0.07	14	4073.25	
1363	Ord_82	Prod_14	SHP_108	Cust_62	21506.770	0.06	44	1260.51	
1448	Ord_2673	Prod_1	SHP_3661	Cust_1027	21337.270	0.02	49	7606.00	
1740	Ord_3667	Prod_17	SHP_5082	Cust_1291	21532.260	0.09	44	8323.39	
1771	Ord_5347	Prod_17	SHP_7468	Cust_1797	21134.710	0.06	6	-4266.09	
1787	Ord_2872	Prod_14	SHP_7219	Cust_1748	21555.600	0.02	47	8965.83	
1835	Ord_3875	Prod_17	SHP_5370	Cust_1351	41343.210	0.09	8	3852.19	
1848	Ord_5315	Prod_14	SHP_7423	Cust_1780	21141.070	0.06	49	6225.36	
1942	Ord_4732	Prod_11	SHP_6600	Cust_1642	20701.928	0.08	49	3918.54	
2059	Ord_4543	Prod_14	SHP_6324	Cust_1538	23949.510	0.00	46	-1312.84	
2216	Ord_4216	Prod_14	SHP_5881	Cust_1432	26126.920	0.04	42	9498.60	
2253	Ord_3143	Prod_14	SHP_4362	Cust_1170	28664.520	0.09	50	13340.26	
2259	Ord_1978	Prod_17	SHP_2703	Cust_725	25312.000	0.01	48	8788.81	
2349	Ord_2373	Prod_14	SHP_3259	Cust_942	33367.850	0.01	9	3992.52	
2680	Ord_1963	Prod_11	SHP_2687	Cust_732	26622.550	0.08	49	3146.22	
2738	Ord_3084	Prod_17	SHP_4279	Cust_1151	89061.050	0.00	13	27220.69	
2800	Ord_417	Prod_11	SHP_561	Cust_156	20333.816	0.02	45	-1430.45	
2862	Ord_3067	Prod_3	SHP_4259	Cust_1152	20175.480	0.00	50	9373.96	
3241	Ord_118	Prod_3	SHP_161	Cust_63	23792.930	0.00	18	10951.31	
3329	Ord_2130	Prod_15	SHP_2906	Cust_796	24701.120	0.04	48	8022.94	
3620	Ord_432	Prod_14	SHP_578	Cust_138	23255.610	0.00	38	-734.33	
3637	Ord_882	Prod_14	SHP_1212	Cust_305	23281.050	0.09	41	9097.65	
3725	Ord_1920	Prod_17	SHP_3108	Cust_852	20329.800	0.00	19	9224.10	
3784	Ord_2338	Prod_17	SHP_3207	Cust_932	45923.760	0.07	7	102.61	
...	...	...	...	...	...	...	...	...	
4389	Ord_2412	Prod_15	SHP_3309	Cust_962	21390.440	0.09	46	5365.43	
4399	Ord_4614	Prod_14	SHP_6423	Cust_1571	29884.600	0.05	49	12748.86	
4963	Ord_4161	Prod_17	SHP_5798	Cust_1421	25313.340	0.05	35	8612.11	

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	S
5042	Ord_2425	Prod_14	SHP_3329	Cust_934	27820.340	0.08	48	11630.15	
5199	Ord_4842	Prod_14	SHP_6754	Cust_1672	22319.580	0.02	41	-734.31	
5395	Ord_5097	Prod_14	SHP_7119	Cust_1731	21046.740	0.04	29	5217.27	
5436	Ord_2934	Prod_15	SHP_4053	Cust_1045	21425.910	0.01	40	7360.43	
5598	Ord_3421	Prod_17	SHP_4742	Cust_1233	23300.120	0.06	9	-217.47	
5800	Ord_5114	Prod_15	SHP_7144	Cust_1729	24639.800	0.01	46	8614.79	
5865	Ord_4166	Prod_14	SHP_5804	Cust_1401	21956.030	0.10	38	5903.09	
5971	Ord_3059	Prod_17	SHP_4249	Cust_1143	21921.280	0.04	43	10521.33	
6037	Ord_3727	Prod_17	SHP_5171	Cust_1310	29186.490	0.05	38	11562.08	
6116	Ord_1791	Prod_14	SHP_2711	Cust_685	21717.360	0.06	32	8249.86	
6245	Ord_997	Prod_14	SHP_1379	Cust_365	28761.520	0.04	8	285.11	
6384	Ord_4473	Prod_17	SHP_6232	Cust_1535	20265.220	0.08	47	6168.64	
6484	Ord_3627	Prod_17	SHP_5022	Cust_1281	24559.910	0.01	47	7358.66	
6653	Ord_5391	Prod_14	SHP_7527	Cust_1799	24233.540	0.07	43	6492.67	
6660	Ord_5425	Prod_14	SHP_7580	Cust_1799	27720.980	0.07	46	11984.40	
6765	Ord_5186	Prod_17	SHP_7247	Cust_1763	26095.130	0.03	35	12606.81	
6797	Ord_1265	Prod_14	SHP_1749	Cust_466	24051.490	0.07	41	9791.04	
6926	Ord_2171	Prod_15	SHP_2958	Cust_800	23239.960	0.06	47	6888.36	
6972	Ord_669	Prod_14	SHP_916	Cust_224	20872.160	0.03	29	-4437.91	
6979	Ord_3236	Prod_10	SHP_4488	Cust_1194	21062.910	0.01	23	5713.53	
7006	Ord_5361	Prod_3	SHP_7489	Cust_1793	21752.010	0.03	25	9296.35	
7091	Ord_911	Prod_10	SHP_1255	Cust_302	28180.080	0.02	32	7513.88	
7318	Ord_546	Prod_14	SHP_858	Cust_198	27875.540	0.00	46	-635.69	
7547	Ord_3170	Prod_10	SHP_4400	Cust_1162	29345.270	0.03	34	7497.55	
7706	Ord_79	Prod_17	SHP_105	Cust_42	22079.470	0.06	12	5322.14	
8046	Ord_825	Prod_14	SHP_1132	Cust_247	27663.920	0.05	8	-391.92	
8217	Ord_3359	Prod_10	SHP_7245	Cust_1762	28389.140	0.07	33	7132.18	

66 rows × 10 columns

In [69]:
```
data[(data["Sales"]>20000) & (data["Profit"]>10000)]
```

Out[69]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Shi
385	Ord_3707	Prod_17	SHP_5136	Cust_1307	28359.40	0.05	49	14440.39	
2253	Ord_3143	Prod_14	SHP_4362	Cust_1170	28664.52	0.09	50	13340.26	
2738	Ord_3084	Prod_17	SHP_4279	Cust_1151	89061.05	0.00	13	27220.69	
3241	Ord_118	Prod_3	SHP_161	Cust_63	23792.93	0.00	18	10951.31	
4124	Ord_2345	Prod_3	SHP_3218	Cust_937	25409.63	0.02	20	11535.28	
4399	Ord_4614	Prod_14	SHP_6423	Cust_1571	29884.60	0.05	49	12748.86	
5042	Ord_2425	Prod_14	SHP_3329	Cust_934	27820.34	0.08	48	11630.15	
5971	Ord_3059	Prod_17	SHP_4249	Cust_1143	21921.28	0.04	43	10521.33	
6037	Ord_3727	Prod_17	SHP_5171	Cust_1310	29186.49	0.05	38	11562.08	
6660	Ord_5425	Prod_14	SHP_7580	Cust_1799	27720.98	0.07	46	11984.40	
6765	Ord_5186	Prod_17	SHP_7247	Cust_1763	26095.13	0.03	35	12606.81	

In [71]:
```
# sorting
# Ascending order
# Descending order
```

In [79]:
```
data.sort_values(by = ["Prod_id"],inplace=True)
```

In [80]:
```
data
```

Out[80]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Shi
1593	Ord_1067	Prod_1	SHP_1474	Cust_397	7333.45	0.07	27	-207.28	
5429	Ord_2414	Prod_1	SHP_3312	Cust_940	825.82	0.00	36	-1049.26	
3986	Ord_168	Prod_1	SHP_227	Cust_61	18235.47	0.03	48	1166.93	
8102	Ord_3954	Prod_1	SHP_5495	Cust_1369	1556.61	0.06	17	-485.55	
4794	Ord_4923	Prod_1	SHP_6867	Cust_1681	1438.33	0.08	43	163.38	
2595	Ord_4464	Prod_1	SHP_6219	Cust_1445	279.77	0.07	8	-14.02	
7772	Ord_3210	Prod_1	SHP_4452	Cust_1183	494.34	0.05	15	-30.19	
5570	Ord_909	Prod_1	SHP_1253	Cust_307	189.04	0.08	16	-74.77	
2596	Ord_4249	Prod_1	SHP_5933	Cust_1445	299.01	0.09	38	-142.86	
2131	Ord_4792	Prod_1	SHP_6679	Cust_1568	292.38	0.05	8	6.11	
2597	Ord_4250	Prod_1	SHP_5936	Cust_1445	6152.80	0.04	27	554.55	
1014	Ord_1172	Prod_1	SHP_1618	Cust_449	529.28	0.10	45	-30.27	
2992	Ord_1382	Prod_1	SHP_1906	Cust_532	2367.99	0.04	14	320.10	
3996	Ord_479	Prod_1	SHP_645	Cust_175	1132.54	0.00	22	-628.38	
4773	Ord_5048	Prod_1	SHP_7045	Cust_1714	447.25	0.08	35	-15.92	
3001	Ord_5249	Prod_1	SHP_7330	Cust_1771	17129.97	0.07	43	5616.08	
1006	Ord_319	Prod_1	SHP_426	Cust_70	178.45	0.09	31	-104.96	
3979	Ord_1662	Prod_1	SHP_2295	Cust_530	479.84	0.02	34	5.71	
403	Ord_1579	Prod_1	SHP_2179	Cust_566	1221.62	0.01	18	-468.64	
4798	Ord_4903	Prod_1	SHP_6839	Cust_1681	389.98	0.07	32	-14.98	
5914	Ord_1243	Prod_1	SHP_1719	Cust_472	2227.34	0.09	20	-609.09	
3936	Ord_4164	Prod_1	SHP_5802	Cust_1400	6448.69	0.00	45	1663.35	
3939	Ord_4199	Prod_1	SHP_5854	Cust_1400	3194.05	0.08	23	608.52	
1828	Ord_4485	Prod_1	SHP_6245	Cust_1537	1116.03	0.04	32	203.44	
5580	Ord_3003	Prod_1	SHP_4156	Cust_1111	621.44	0.08	43	-130.96	
1035	Ord_2294	Prod_1	SHP_3140	Cust_850	572.40	0.07	36	-39.50	
396	Ord_4144	Prod_1	SHP_5773	Cust_1418	1144.35	0.05	14	-448.60	
3966	Ord_2625	Prod_1	SHP_3592	Cust_994	4890.58	0.05	35	310.51	
6584	Ord_3354	Prod_1	SHP_4650	Cust_1236	133.85	0.03	9	-5.28	
4821	Ord_1628	Prod_1	SHP_2248	Cust_570	1390.17	0.07	13	-469.84	
...	...	...	...	...	...	...	...	...	
4989	Ord_693	Prod_9	SHP_948	Cust_236	253.89	0.01	45	-83.66	
8106	Ord_3428	Prod_9	SHP_4751	Cust_1222	35.68	0.01	3	-11.43	
1583	Ord_1571	Prod_9	SHP_2168	Cust_514	32.49	0.08	3	-5.05	

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Shi
7310	Ord_950	Prod_9	SHP_1315	Cust_334	4906.85	0.09	32	1907.94	
843	Ord_4057	Prod_9	SHP_5656	Cust_1391	305.68	0.10	19	128.17	
3295	Ord_409	Prod_9	SHP_551	Cust_151	239.35	0.05	42	78.96	
4526	Ord_2604	Prod_9	SHP_3565	Cust_993	8.34	0.09	1	-3.44	
3307	Ord_3895	Prod_9	SHP_5401	Cust_1353	405.83	0.06	41	178.86	
773	Ord_5241	Prod_9	SHP_7320	Cust_1746	55.66	0.09	10	8.82	
740	Ord_51	Prod_9	SHP_67	Cust_34	751.77	0.05	50	353.20	
7368	Ord_5017	Prod_9	SHP_7000	Cust_1706	1323.31	0.05	38	146.64	
4886	Ord_3480	Prod_9	SHP_4817	Cust_1184	155.89	0.10	17	50.19	
4835	Ord_2894	Prod_9	SHP_3993	Cust_572	531.48	0.08	50	232.58	
7822	Ord_4578	Prod_9	SHP_6369	Cust_1492	729.75	0.02	44	375.25	
7823	Ord_4370	Prod_9	SHP_6092	Cust_1492	2325.42	0.02	25	739.91	
4775	Ord_5108	Prod_9	SHP_7133	Cust_1733	15.00	0.10	1	-13.74	
7337	Ord_780	Prod_9	SHP_1071	Cust_240	181.13	0.00	47	-11.39	
4766	Ord_2383	Prod_9	SHP_3271	Cust_955	34.66	0.06	7	-20.87	
6784	Ord_1234	Prod_9	SHP_1707	Cust_466	217.35	0.00	23	-67.19	
716	Ord_2165	Prod_9	SHP_2951	Cust_783	561.06	0.07	49	-2.23	
7831	Ord_3313	Prod_9	SHP_4597	Cust_1226	782.93	0.03	48	401.07	
4697	Ord_702	Prod_9	SHP_960	Cust_241	108.26	0.06	7	25.51	
4693	Ord_2970	Prod_9	SHP_5273	Cust_1326	141.90	0.02	12	1.46	
7343	Ord_701	Prod_9	SHP_959	Cust_240	355.69	0.09	42	-149.46	
8134	Ord_2619	Prod_9	SHP_3584	Cust_990	775.77	0.07	20	179.37	
1669	Ord_5073	Prod_9	SHP_7083	Cust_1720	741.57	0.10	49	352.41	
6942	Ord_4033	Prod_9	SHP_5615	Cust_1375	197.45	0.06	34	65.44	
6870	Ord_2705	Prod_9	SHP_3704	Cust_1010	60.19	0.01	7	6.44	
3264	Ord_4644	Prod_9	SHP_6469	Cust_1601	473.46	0.03	13	128.11	
8177	Ord_5166	Prod_9	SHP_7220	Cust_1755	347.84	0.00	37	-99.25	

8399 rows × 10 columns

In [75]:
```
data.sort_values(by = ["Prod_id"],inplace=True,ascending=False)
```

In [81]:
```
data.head(10)
```

Out[81]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Shi
1593	Ord_1067	Prod_1	SHP_1474	Cust_397	7333.45	0.07	27	-207.28	
5429	Ord_2414	Prod_1	SHP_3312	Cust_940	825.82	0.00	36	-1049.26	
3986	Ord_168	Prod_1	SHP_227	Cust_61	18235.47	0.03	48	1166.93	
8102	Ord_3954	Prod_1	SHP_5495	Cust_1369	1556.61	0.06	17	-485.55	
4794	Ord_4923	Prod_1	SHP_6867	Cust_1681	1438.33	0.08	43	163.38	
2595	Ord_4464	Prod_1	SHP_6219	Cust_1445	279.77	0.07	8	-14.02	
7772	Ord_3210	Prod_1	SHP_4452	Cust_1183	494.34	0.05	15	-30.19	
5570	Ord_909	Prod_1	SHP_1253	Cust_307	189.04	0.08	16	-74.77	
2596	Ord_4249	Prod_1	SHP_5933	Cust_1445	299.01	0.09	38	-142.86	
2131	Ord_4792	Prod_1	SHP_6679	Cust_1568	292.38	0.05	8	6.11	

Data Preprocessing with scikit -Learn

- problems :
  - insufficient of data
  - too much of data
  - outliers
  - Missing data
  - Duplicate data
- Scaling Techniques
  - Standard Scalar
  - Robust Scalar
  - Normalization
  - Range Scalar(Min Max Scalar)

- Standard Scalar
- [xi-mean(x)]/std(x)

In [82]:
```
from sklearn import preprocessing
```

In [86]:
```
sc1 = preprocessing.StandardScaler()
```

In [84]:
```
print(dir(preprocessing))
```

```
['Binarizer', 'CategoricalEncoder', 'FunctionTransformer', 'Imputer', 'KBinsDis
cretizer', 'KernelCenterer', 'LabelBinarizer', 'LabelEncoder', 'MaxAbsScaler',
'MinMaxScaler', 'MultiLabelBinarizer', 'Normalizer', 'OneHotEncoder', 'OrdinalE
ncoder', 'PolynomialFeatures', 'PowerTransformer', 'QuantileTransformer', 'Robu
stScaler', 'StandardScaler', '__all__', '__builtins__', '__cached__', '__doc_
_', '__file__', '__loader__', '__name__', '__package__', '__path__', '__spec_
_', '_discretization', '_encoders', '_function_transformer', 'add_dummy_featur
e', 'base', 'binarize', 'data', 'imputation', 'label', 'label_binarize', 'maxab
s_scale', 'minmax_scale', 'normalize', 'power_transform', 'quantile_transform',
'robust_scale', 'scale']
```

In [93]:
```
data.head()
```

Out[93]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Shi
1593	Ord_1067	Prod_1	SHP_1474	Cust_397	7333.45	0.07	27	-207.28	
5429	Ord_2414	Prod_1	SHP_3312	Cust_940	825.82	0.00	36	-1049.26	
3986	Ord_168	Prod_1	SHP_227	Cust_61	18235.47	0.03	48	1166.93	
8102	Ord_3954	Prod_1	SHP_5495	Cust_1369	1556.61	0.06	17	-485.55	
4794	Ord_4923	Prod_1	SHP_6867	Cust_1681	1438.33	0.08	43	163.38	

In [104]:
```
data1 = data[["Sales","Discount","Order_Quantity","Profit","Shipping_Cost"]]
```

In [105]:
```
data1.head()
```

Out[105]:

	Sales	Discount	Order_Quantity	Profit	Shipping_Cost
1593	7333.45	0.07	27	-207.28	35.00
5429	825.82	0.00	36	-1049.26	35.00
3986	18235.47	0.03	48	1166.93	99.00
8102	1556.61	0.06	17	-485.55	35.00
4794	1438.33	0.08	43	163.38	8.22

In [107]:
```
sc1_d = sc1.fit_transform(data1)
sc1_d
```

```
C:\Users\Alekhya\Anaconda3\lib\site-packages\sklearn\preprocessing\data.py:645:
DataConversionWarning: Data with input dtype int64, float64 were all converted
to float64 by StandardScaler.
  return self.partial_fit(X, y)
C:\Users\Alekhya\Anaconda3\lib\site-packages\sklearn\base.py:464: DataConversio
nWarning: Data with input dtype int64, float64 were all converted to float64 by
StandardScaler.
  return self.fit(X, **fit_params).transform(X)
```

Out[107]:
```
array([[ 1.55029968,  0.63884009,  0.09863568, -0.32464502,  1.28375207],
       [-0.2650213 , -1.56095642,  0.72017367, -1.02829918,  1.28375207],
       [ 4.59144764, -0.61818648,  1.54889099,  0.82380102,  4.99109786],
       ...,
       [-0.47859586, -1.24669977, -1.28255985, -0.1460363 , -0.59888446],
       [-0.36331307, -0.61818648, -0.86820119, -0.04435502, -0.35790699],
       [-0.39835511, -1.56095642,  0.78923345, -0.23436287, -0.26348552]])
```

In [110]:
```
data1.columns
```

Out[110]:
```
Index(['Sales', 'Discount', 'Order_Quantity', 'Profit', 'Shipping_Cost'], dtype
='object')
```

In [113]:
```
data.info()
```

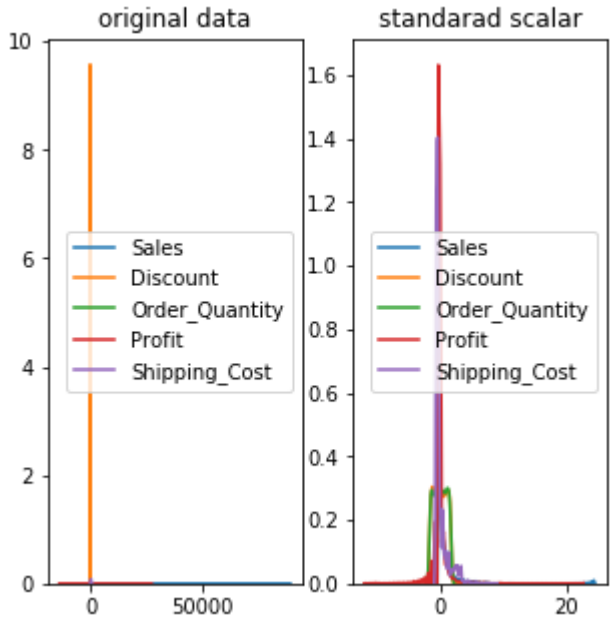
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8399 entries, 1593 to 8177
Data columns (total 10 columns):
Ord_id                 8399 non-null object
Prod_id                8399 non-null object
Ship_id                8399 non-null object
Cust_id                8399 non-null object
Sales                  8399 non-null float64
Discount               8399 non-null float64
Order_Quantity         8399 non-null int64
Profit                 8399 non-null float64
Shipping_Cost          8399 non-null float64
Product_Base_Margin    8336 non-null float64
dtypes: float64(5), int64(1), object(4)
memory usage: 1.0+ MB
```

In [111]:
```
sc1_data = pd.DataFrame(sc1_d,columns=data1.columns)
sc1_data.head()
```

Out[111]:

	Sales	Discount	Order_Quantity	Profit	Shipping_Cost
0	1.550300	0.638840	0.098636	-0.324645	1.283752
1	-0.265021	-1.560956	0.720174	-1.028299	1.283752
2	4.591448	-0.618186	1.548891	0.823801	4.991098
3	-0.061165	0.324583	-0.591962	-0.557199	1.283752
4	-0.094160	0.953097	1.203592	-0.014879	-0.267540

In [116]:
```
import matplotlib.pyplot as plt
import seaborn as sns
fig,(ax1,ax2)=plt.subplots(ncols=2,figsize=(5,5))
ax1.set_title("original data")
sns.kdeplot(data1["Sales"],ax=ax1)
sns.kdeplot(data1["Discount"],ax=ax1)
sns.kdeplot(data1["Order_Quantity"],ax=ax1)
sns.kdeplot(data1["Profit"],ax=ax1)
sns.kdeplot(data1["Shipping_Cost"],ax=ax1)
ax2.set_title("standarad scalar")
sns.kdeplot(sc1_data["Sales"],ax=ax2)
sns.kdeplot(sc1_data["Discount"],ax=ax2)
sns.kdeplot(sc1_data["Order_Quantity"],ax=ax2)
sns.kdeplot(sc1_data["Profit"],ax=ax2)
sns.kdeplot(sc1_data["Shipping_Cost"],ax=ax2)
plt.show()
```



In [117]:
```
# robust scalar
```

In [118]:
```
rs = preprocessing.RobustScaler()
```

In [119]:
```
rs_d = rs.fit_transform(data1)
rs_d
```

Out[119]:
```
array([[ 4.39558145,  0.33333333,  0.04      , -0.8362831 ,  2.70626754],
       [ 0.24033841, -0.83333333,  0.4       , -4.25806189,  2.70626754],
       [11.3567244 , -0.33333333,  0.88      ,  4.74846077,  8.69317119],
       ...,
       [-0.24853061, -0.66666667, -0.76      ,  0.0322679 , -0.33395697],
       [ 0.01534999, -0.33333333, -0.52      ,  0.52673074,  0.05519177],
       [-0.06486072, -0.83333333,  0.44      , -0.39725276,  0.20767072]])
```

In [122]:
```
da = pd.DataFrame(rs_d,columns=data1.columns)
da.head()
```

Out[122]:

	Sales	Discount	Order_Quantity	Profit	Shipping_Cost
0	4.395581	0.333333	0.04	-0.836283	2.706268
1	0.240338	-0.833333	0.40	-4.258062	2.706268
2	11.356724	-0.333333	0.88	4.748461	8.693171
3	0.706961	0.166667	-0.36	-1.967163	2.706268
4	0.631437	0.500000	0.68	0.670067	0.201123

In [ ]:

In [ ]: