

Nayan Ranjan Das | A2305220148 | 5CSE2

Lab 2

August 19, 2022

In [1]:

```
1 import pandas as pd
2 import numpy as np
```

In [4]:

```
1 market      = pd.read_csv("global_sales_data/market_fact.csv")
2 cust         = pd.read_csv("global_sales_data/cust_dimen.csv")
3 orders       = pd.read_csv("global_sales_data/orders_dimen.csv")
4 prod         = pd.read_csv("global_sales_data/prod_dimen.csv")
5 ship         = pd.read_csv("global_sales_data/shipping_dimen.csv")
```

Columns in each dataframe

In [3]:

```
1 market.columns
```

Out[3]:

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

In [5]:

```
1 cust.columns
```

Out[5]:

```
Index(['Customer_Name', 'Province', 'Region', 'Customer_Segment', 'Cust_i
d'], dtype='object')
```

In [6]:

```
1 orders.columns
```

Out[6]:

```
Index(['Order_ID', 'Order_Date', 'Order_Priority', 'Ord_id'], dtype='objec
t')
```



In [7]:

```
1 prod.columns
```

Out[7]:

```
Index(['Product_Category', 'Product_Sub_Category', 'Prod_id'], dtype='object')
```

In [8]:

```
1 ship.columns
```

Out[8]:

```
Index(['Order_ID', 'Ship_Mode', 'Ship_Date', 'Ship_id'], dtype='object')
```

select all orders and select the sales of customers segment corporate

In [11]:

```
1 cust.Customer_Segment.unique()
```

Out[11]:

```
array(['SMALL BUSINESS', 'CONSUMER', 'CORPORATE', 'HOME OFFICE'],  
      dtype=object)
```

In [42]:

```
1 new = pd.merge(market, cust, on="Cust_id")  
2 new.shape
```

Out[42]:

```
(8399, 14)
```

In [43]:

```
1 new.loc[new.Customer_Segment == "CORPORATE", ['Ord_id', 'Sales']]
```

Out[43]:

	Ord_id	Sales
0	Ord_5446	136.81
1	Ord_5406	42.27
2	Ord_5446	4701.69
3	Ord_5456	2337.89
4	Ord_5485	4233.15
...
8385	Ord_1833	611.16
8386	Ord_2324	121.87
8387	Ord_2220	41.06
8388	Ord_4424	994.04
8389	Ord_4444	159.41

3076 rows × 2 columns

select from above dataframe where customer segment is home office and profit is greater than 500

In [44]:

```
1 new.loc[(new.Customer_Segment == "HOME OFFICE") & (new.Profit > 500)]
```

Out[44]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit
7	Ord_4725	Prod_4	SHP_6593	Cust_1641	3410.1575	0.10	48	1137.91
10	Ord_4743	Prod_2	SHP_6615	Cust_1641	4072.0100	0.01	43	1675.98
17	Ord_4471	Prod_15	SHP_6228	Cust_1521	13255.9300	0.02	25	4089.27
237	Ord_597	Prod_4	SHP_816	Cust_209	4283.2350	0.10	44	676.13
239	Ord_597	Prod_4	SHP_816	Cust_209	4374.6865	0.05	43	973.16
...
8338	Ord_2107	Prod_2	SHP_2882	Cust_785	2409.9600	0.07	32	575.10
8343	Ord_2178	Prod_4	SHP_2968	Cust_785	6030.5800	0.09	39	1197.86
8366	Ord_3593	Prod_3	SHP_4974	Cust_1274	12073.0600	0.03	39	5081.87
8367	Ord_3593	Prod_15	SHP_4975	Cust_1274	6685.0500	0.09	25	1653.60
8369	Ord_3633	Prod_3	SHP_5031	Cust_1274	1169.2600	0.02	41	515.62

305 rows × 14 columns

Show rows where customer name is Alex

In [45]:

```
1 new.loc[(new.Customer_Name == "ALEX AVILA") & (new.Profit > 500) & (new.Customer_Segment == "HOME OFFICE")]
```

Out[45]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	Shipper
	237	Ord_597	Prod_4	SHP_816	Cust_209	4283.2350	0.10	44	676.13
	239	Ord_597	Prod_4	SHP_816	Cust_209	4374.6865	0.05	43	973.16

Sales greater than or equal to 300 and profit greater than 1000

In [46]:

```
1 new.loc[(new.Sales >= 300) & (new.Profit > 1000)]
```

Out[46]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit
2	Ord_5446	Prod_4	SHP_7610	Cust_1818	4701.6900	0.00	26	1148.90
4	Ord_5485	Prod_17	SHP_7664	Cust_1818	4233.1500	0.08	35	1219.87
7	Ord_4725	Prod_4	SHP_6593	Cust_1641	3410.1575	0.10	48	1137.91
10	Ord_4743	Prod_2	SHP_6615	Cust_1641	4072.0100	0.01	43	1675.98
17	Ord_4471	Prod_15	SHP_6228	Cust_1521	13255.9300	0.02	25	4089.27
...
8317	Ord_2770	Prod_3	SHP_3802	Cust_1016	7535.9600	0.08	46	2745.87
8343	Ord_2178	Prod_4	SHP_2968	Cust_785	6030.5800	0.09	39	1197.86
8366	Ord_3593	Prod_3	SHP_4974	Cust_1274	12073.0600	0.03	39	5081.87
8367	Ord_3593	Prod_15	SHP_4975	Cust_1274	6685.0500	0.09	25	1653.60
8371	Ord_2624	Prod_4	SHP_3591	Cust_1006	4924.1350	0.07	28	1049.54

678 rows × 14 columns

select all orders from product category office supplies and from corporate sector

In [47]:

```
1 new2 = pd.merge(new, prod, on='Prod_id')
```

In [48]:

```
1 new2.loc[(new2.Product_Category == 'OFFICE SUPPLIES') & (new2.Customer_Segment == 'CO
```

Out[48]:

	Ord_id	Prod_id	Ship_id	Cust_id	Sales	Discount	Order_Quantity	Profit	St
0	Ord_5446	Prod_16	SHP_7609	Cust_1818	136.81	0.01	23	-30.51	
3	Ord_3730	Prod_16	SHP_5175	Cust_1314	459.08	0.04	34	61.57	
7	Ord_4506	Prod_16	SHP_6273	Cust_1544	92.02	0.07	9	-24.88	
9	Ord_1551	Prod_16	SHP_2145	Cust_531	184.77	0.00	29	-71.96	
11	Ord_1429	Prod_16	SHP_1976	Cust_510	539.06	0.05	42	-123.07	
...	
7545	Ord_4629	Prod_1	SHP_6447	Cust_1587	848.19	0.06	25	120.02	
7546	Ord_4604	Prod_1	SHP_6403	Cust_1522	234.24	0.09	24	-151.80	
7551	Ord_3543	Prod_1	SHP_4905	Cust_1266	1184.11	0.07	6	-145.07	
7552	Ord_2722	Prod_1	SHP_3731	Cust_1006	3508.33	0.04	21	-546.98	
7553	Ord_4424	Prod_1	SHP_6165	Cust_1519	994.04	0.03	10	-335.06	

1680 rows × 16 columns

merge shipping dataframe to new2

In [50]:

```
1 new3 = pd.merge(new2, ship, on='Ship_id')
```

merge order dataframe to new3

In [51]:

```
1 new4 = pd.merge(new3, orders, on='Ord_id')
```

create a new dataframe

In [54]:

```
1 dic = [  
2     {"Name": "Aman", "Age": 34, "Gender": "M"},  
3     {"Name": "Joy", "Age": 31, "Gender": "M"},  
4     {"Name": "Roshni", "Age": 32, "Gender": "F"},  
5     {"Name": "Saif", "Age": 33, "Gender": "M"}  
6 ]  
7 dic
```

Out[54]:

```
[{'Name': 'Aman', 'Age': 34, 'Gender': 'M'},  
{ 'Name': 'Joy', 'Age': 31, 'Gender': 'M'},  
{ 'Name': 'Roshni', 'Age': 32, 'Gender': 'F'},  
{ 'Name': 'Saif', 'Age': 33, 'Gender': 'M'}]
```

In [61]:

```
1 temp = [  
2     {  
3         "Name": ["Aman", "Joy", "Roshni", "Saif"],  
4         "Age": [34, 31, 32, 33],  
5         "Gender": ["M", "M", "F", "M"]  
6     }  
7 ]
```

In [62]:

```
1 temp
```

Out[62]:

```
[{'Name': ['Aman', 'Joy', 'Roshni', 'Saif'],  
  'Age': [34, 31, 32, 33],  
  'Gender': ['M', 'M', 'F', 'M']}]
```

In [60]:

```
1 df = pd.DataFrame(dic)  
2 df
```

Out[60]:

	Name	Age	Gender
0	Aman	34	M
1	Joy	31	M
2	Roshni	32	F
3	Saif	33	M

create another dataframe and merge with the previously created dataframe

In [73]:

```
1 df2 = pd.DataFrame([
2     ["Akhil", 31, "M"],
3     ["Asha", 22, "F"],
4     ["Preeti", 21, "F"]],columns=["Name", "Age", "Gender"])
```

In [75]:

```
1 pd.concat([df,df2])
```

Out[75]:

	Name	Age	Gender
0	Aman	34	M
1	Joy	31	M
2	Roshni	32	F
3	Saif	33	M
0	Akhil	31	M
1	Asha	22	F
2	Preeti	21	F

another example

In [77]:

```
1 grad = pd.DataFrame({"Grade":["A", 'B', 'C', 'D'], "marks":[45,54,60,75]})
```

In [82]:

```
1 pd.concat([df,grad],axis=1)
```

Out[82]:

	Name	Age	Gender	Grade	marks
0	Aman	34	M	A	45
1	Joy	31	M	B	54
2	Roshni	32	F	C	60
3	Saif	33	M	D	75

In [92]:

```
1 match_21 = pd.DataFrame([
2     ["CSK", 50,20],
3     ["MI", 50,15],
4     ["DC", 50, 35],
5     ["RCB", 50, 40],
6     ["RR", 50, 15],
7     ["TT", 50, 22]
8 ], columns=["Team", "Played", "Won"])
9
10 match_21.set_index("Team", inplace=True)
11 match_21
```

Out[92]:

	Played	Won
Team		
CSK	50	20
MI	50	15
DC	50	35
RCB	50	40
RR	50	15
TT	50	22

In [93]:

```
1 match_22 = pd.DataFrame([
2     ["CSK", 50,34],
3     ["MI", 50,27],
4     ["DC", 50, 41],
5     ["RCB", 50, 28],
6     ["RR", 50, 7],
7     ["TT", 50, 11],
8     ["KKR", 50, 20]
9 ], columns=["Team", "Played", "Won"])
10
11 match_22.set_index("Team", inplace=True)
12 match_22
```

Out[93]:

	Played	Won
Team		
CSK	50	34
MI	50	27
DC	50	41
RCB	50	28
RR	50	7
TT	50	11
KKR	50	20

displaying total result from two dataframes

In [96]:

```
1 total = match_21+match_22
2 total
```

Out[96]:

	Played	Won
Team		
CSK	100.0	54.0
DC	100.0	76.0
KKR	NaN	NaN
MI	100.0	42.0
RCB	100.0	68.0
RR	100.0	22.0
TT	100.0	33.0

In [100]:

```
1 match_21.add(match_22, fill_value=0)
```

Out[100]:

	Played	Won
Team		
CSK	100.0	54.0
DC	100.0	76.0
KKR	50.0	20.0
MI	100.0	42.0
RCB	100.0	68.0
RR	100.0	22.0
TT	100.0	33.0