

PRACTICAL - 06

6. Consider any sales training/ weather forecasting dataset :

CODE :

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

```
In [2]: # Loading dataset into a DataFrame
data = pd.read_excel("Sales.xlsx")
data.head(10)
```

Out[2]:

	Order ID	Order Date	Ship yearMonth	State	Postal Code	Region	Product ID	Category	Sub-Category	
0	CA-2017-152156	2017-11-08	2017-11	Kentucky	42420	South	FUR-BO-10001798	Furniture	Bookcases	26
1	CA-2017-152156	2017-11-08	2017-11	Kentucky	42420	South	FUR-CH-10000454	Furniture	Chairs	73
2	CA-2017-138688	2017-06-12	2017-06	California	90036	West	OFF-LA-10000240	Office Supplies	Labels	1
3	US-2016-108966	NaT	2016-10	Florida	33311	South	FUR-TA-10000577	Furniture	Tables	95
4	US-2016-108966	2016-10-11	2016-10	Florida	33311	South	OFF-ST-10000760	Office Supplies	Storage	2
5	CA-2015-115812	2015-06-09	2015-06	California	90032	West	FUR-FU-10001487	Furniture	Furnishings	4
6	CA-2015-115812	2015-06-09	2015-06	California	90032	West	OFF-AR-10002833	Office Supplies	Art	
7	CA-2015-115812	2015-06-09	2015-06	California	90032	West	TEC-PH-10002275	Technology	Phones	90
8	CA-2015-115812	2015-06-09	2015-06	California	90032	West	OFF-BI-10003910	Office Supplies	Binders	1
9	CA-2015-115812	NaT	2015-06	California	90032	West	OFF-AP-10002892	Office Supplies	Appliances	11

- a. Compute mean of a series grouped by another series.

```
In [3]: grouped_mean = data.groupby('State')['Sales'].mean()

print("The Mean of a Series grouped by another :")
grouped_mean
```

The Mean of a Series grouped by another :

```
Out[3]: State
California      415.27600
Florida         489.97275
Kentucky        496.95000
North Carolina   15.55200
Texas           35.67700
Utah            55.50000
Washington      407.97600
Wisconsin       665.88000
Name: Sales, dtype: float64
```

- b. Fill an intermittent time series to replace all missing dates with values of previous non-missing date.

```
In [4]: #changing into date format
data['Order Date'] = pd.to_datetime(data['Order Date'])

# Forward-fill missing values
data['Order Date'] = data['Order Date'].ffill()
data.head(10)
```

SUBMITTED TO : DR. PRITI JAGWANI
NAME : GARIMA ROHILLA
ROLL NO : CSC/21/37
UNIVERSITY ROLL NO : 21059570010

Out[4]:

	Order ID	Order Date	Ship yearMonth	State	Postal Code	Region	Product ID	Category	Sub-Category	
0	CA-2017-152156	2017-11-08	2017-11	Kentucky	42420	South	FUR-BO-10001798	Furniture	Bookcases	26
1	CA-2017-152156	2017-11-08	2017-11	Kentucky	42420	South	FUR-CH-10000454	Furniture	Chairs	73
2	CA-2017-138688	2017-06-12	2017-06	California	90036	West	OFF-LA-10000240	Office Supplies	Labels	1
3	US-2016-108966	2017-06-12	2016-10	Florida	33311	South	FUR-TA-10000577	Furniture	Tables	95
4	US-2016-108966	2016-10-11	2016-10	Florida	33311	South	OFF-ST-10000760	Office Supplies	Storage	2
5	CA-2015-115812	2015-06-09	2015-06	California	90032	West	FUR-FU-10001487	Furniture	Furnishings	4
6	CA-2015-115812	2015-06-09	2015-06	California	90032	West	OFF-AR-10002833	Office Supplies	Art	
7	CA-2015-115812	2015-06-09	2015-06	California	90032	West	TEC-PH-10002275	Technology	Phones	90
8	CA-2015-115812	2015-06-09	2015-06	California	90032	West	OFF-BI-10003910	Office Supplies	Binders	1
9	CA-2015-115812	2015-06-09	2015-06	California	90032	West	OFF-AP-10002892	Office Supplies	Appliances	11

c. Perform appropriate year-month string to dates conversion.

```
In [5]: #conveting into date time

data['Ship yearMonth'] = pd.to_datetime(data['Ship yearMonth'])
data.head(10)
```

Out[5]:

	Order ID	Order Date	Ship yearMonth	State	Postal Code	Region	Product ID	Category	Sub-Category	
0	CA-2017-152156	2017-11-08	2017-11-01	Kentucky	42420	South	FUR-BO-10001798	Furniture	Bookcases	26
1	CA-2017-152156	2017-11-08	2017-11-01	Kentucky	42420	South	FUR-CH-10000454	Furniture	Chairs	73
2	CA-2017-138688	2017-06-12	2017-06-01	California	90036	West	OFF-LA-10000240	Office Supplies	Labels	1
3	US-2016-108966	2017-06-12	2016-10-01	Florida	33311	South	FUR-TA-10000577	Furniture	Tables	95
4	US-2016-108966	2016-10-11	2016-10-01	Florida	33311	South	OFF-ST-10000760	Office Supplies	Storage	2
5	CA-2015-115812	2015-06-09	2015-06-01	California	90032	West	FUR-FU-10001487	Furniture	Furnishings	4
6	CA-2015-115812	2015-06-09	2015-06-01	California	90032	West	OFF-AR-10002833	Office Supplies	Art	
7	CA-2015-115812	2015-06-09	2015-06-01	California	90032	West	TEC-PH-10002275	Technology	Phones	90
8	CA-2015-115812	2015-06-09	2015-06-01	California	90032	West	OFF-BI-10003910	Office Supplies	Binders	1
9	CA-2015-115812	2015-06-09	2015-06-01	California	90032	West	OFF-AP-10002892	Office Supplies	Appliances	11

- d. Split a dataset to group by two columns and then sort the aggregated results within the groups.

```
In [6]: #Lambda function for sorting
func = lambda x: x.sort_values(by='Sales')

sorted_data = data.groupby(['State', 'Region']).apply(func).reset_index(drop=True)
sorted_data.head(10)
```

Out[6]:

	Order ID	Order Date	Ship yearMonth	State	Postal Code	Region	Product ID	Category	Sub-Category	
0	CA-2015-115812	2015-06-09	2015-06-01	California	90032	West	OFF-AR-10002833	Office Supplies	Art	
1	CA-2015-143336	2015-08-27	2015-09-01	California	94109	West	OFF-AR-10003056	Office Supplies	Art	
2	CA-2017-138688	2017-06-12	2017-06-01	California	90036	West	OFF-LA-10000240	Office Supplies	Labels	
3	CA-2015-115812	2015-06-09	2015-06-01	California	90032	West	OFF-BI-10003910	Office Supplies	Binders	
4	CA-2015-115812	2015-06-09	2015-06-01	California	90032	West	FUR-FU-10001487	Furniture	Furnishings	
5	CA-2015-115812	2015-06-09	2015-06-01	California	90032	West	OFF-AP-10002892	Office Supplies	Appliances	1
6	CA-2015-115812	2015-06-09	2015-06-01	California	90032	West	TEC-PH-10002275	Technology	Phones	9
7	CA-2015-115812	2015-06-09	2015-06-01	California	90032	West	TEC-PH-10002033	Technology	Phones	9
8	CA-2015-115812	2015-06-09	2015-06-01	California	90032	West	FUR-TA-10001539	Furniture	Tables	17
9	US-2016-108966	2016-10-11	2016-10-01	Florida	33311	South	OFF-ST-10000760	Office Supplies	Storage	

e. Split a given dataframe into groups with bin counts.

```
In [9]: #creating bins
bin = pd.cut(data['Sales'], bins=3)

grouped_data = data.groupby(bin, observed=False)
for key, group in grouped_data:
    print("Key : ",key)
    print("GROUP \n",group)
    print("\n\n-----")
    print("-----\n\n")
```


SUBMITTED TO : DR. PRITI JAGWANI
NAME : GARIMA ROHILLA
ROLL NO : CSC/21/37
UNIVERSITY ROLL NO : 21059570010

Key : (0.84, 570.424]
GROUP

	Order ID	Order Date	Ship yearMonth	State	Postal Code
\					
0	CA-2017-152156	2017-11-08	2017-11-01	Kentucky	42420
2	CA-2017-138688	2017-06-12	2017-06-01	California	90036
4	US-2016-108966	2016-10-11	2016-10-01	Florida	33311
5	CA-2015-115812	2015-06-09	2015-06-01	California	90032
6	CA-2015-115812	2015-06-09	2015-06-01	California	90032
8	CA-2015-115812	2015-06-09	2015-06-01	California	90032
9	CA-2015-115812	2015-06-09	2015-06-01	California	90032
12	CA-2018-114412	2018-04-15	2018-04-01	North Carolina	28027
13	CA-2017-161389	2018-04-15	2017-12-01	Washington	98103
14	US-2016-118983	2016-11-22	2016-11-01	Texas	76106
15	US-2016-118983	2016-11-22	2016-11-01	Texas	76106
17	CA-2015-167164	2015-05-13	2015-05-01	Utah	84084
18	CA-2015-143336	2015-08-27	2015-09-01	California	94109

	Region	Product ID	Category	Sub-Category	Sales
0	South	FUR-BO-10001798	Furniture	Bookcases	261.960
2	West	OFF-LA-10000240	Office Supplies	Labels	14.620
4	South	OFF-ST-10000760	Office Supplies	Storage	22.368
5	West	FUR-FU-10001487	Furniture	Furnishings	48.860
6	West	OFF-AR-10002833	Office Supplies	Art	7.280
8	West	OFF-BI-10003910	Office Supplies	Binders	18.504
9	West	OFF-AP-10002892	Office Supplies	Appliances	114.900
12	South	OFF-PA-10002365	Office Supplies	Paper	15.552
13	West	OFF-BI-10003656	Office Supplies	Binders	407.976
14	Central	OFF-AP-10002311	Office Supplies	Appliances	68.810
15	Central	OFF-BI-10000756	Office Supplies	Binders	2.544
17	West	OFF-ST-10000107	Office Supplies	Storage	55.500
18	West	OFF-AR-10003056	Office Supplies	Art	8.560

Key : (570.424, 1138.304]
GROUP

	Order ID	Order Date	Ship yearMonth	State	Postal Code	\
1	CA-2017-152156	2017-11-08	2017-11-01	Kentucky	42420	
3	US-2016-108966	2017-06-12	2016-10-01	Florida	33311	
7	CA-2015-115812	2015-06-09	2015-06-01	California	90032	
11	CA-2015-115812	2015-06-09	2015-06-01	California	90032	
16	CA-2015-105893	2016-11-22	2015-11-01	Wisconsin	53711	

SUBMITTED TO : DR. PRITI JAGWANI
NAME : GARIMA ROHILLA
ROLL NO : CSC/21/37
UNIVERSITY ROLL NO : 21059570010

	Region	Product ID	Category	Sub-Category	Sales
1	South	FUR-CH-10000454	Furniture	Chairs	731.9400
3	South	FUR-TA-10000577	Furniture	Tables	957.5775
7	West	TEC-PH-10002275	Technology	Phones	907.1520
11	West	TEC-PH-10002033	Technology	Phones	911.4240
16	Central	OFF-ST-10004186	Office Supplies	Storage	665.8800

Key : (1138.304, 1706.184]
GROUP

	Order ID	Order Date	Ship yearMonth	State	Postal Code	Regi
on \						
10	CA-2015-115812	2015-06-09	2015-06-01	California	90032	Wes
t						

	Product ID	Category	Sub-Category	Sales
10	FUR-TA-10001539	Furniture	Tables	1706.184

