**Data Aggregation:**

Data aggregation is any process whereby data is gathered and expressed in a summary form.

**Data Frame**

import pandas as pd

data={'corporation':['YAHOO','YAHOO','MSFT','MSFT','GOOGLE','GOOGLE'],

'person':['Sanjay','Chetan','Smiti','Anjali','Shaliendra','Jagrati'],

'sales\_in\_USD':[100,140,540,670,240,551]}

df=pd.DataFrame(data)

print(df)

**output**

corporation person sales\_in\_USD

0 YAHOO Sanjay 100

1 YAHOO Chetan 140

2 MSFT Smiti 540

3 MSFT Anjali 670

4 GOOGLE Shaliendra 240

5 GOOGLE Jagrati 551

print(df.groupby('corporation'))

**output**

<pandas.core.groupby.generic.DataFrameGroupBy object at 0x000001E9324FC9A0>

print(type(df.groupby('corporation')))

**Output**

<class 'pandas.core.groupby.generic.DataFrameGroupBy'>

group\_data=df.groupby('corporation')

**Aggregation function:**

1. **Sum() :**

print(group\_data.sum())

**output**

sales\_in\_USD

corporation

GOOGLE 791

MSFT 1210

YAHOO 240

1. **mean():**

print(group\_data.mean())

**output**

corporation

GOOGLE 395.5

MSFT 605.0

YAHOO 120.0

1. **std():**

print(group\_data.std())

**output**

sales\_in\_USD

corporation

GOOGLE 219.910209

MSFT 91.923882

YAHOO 28.284271

1. **min():**

print(group\_data.min())

**output**

person sales\_in\_USD

corporation

GOOGLE Jagrati 240

MSFT Anjali 540

YAHOO Chetan 100

1. **max():**

print(group\_data.max())

**output**

person sales\_in\_USD

corporation

GOOGLE Shaliendra 551

MSFT Smiti 670

YAHOO Sanjay 140

1. **count():**

print(group\_data.count())

**output**

person sales\_in\_USD

corporation

GOOGLE 2 2

MSFT 2 2

YAHOO 2 2

1. **describe() :**

print(group\_data.describe())

**output**

sales\_in\_USD ...

count mean std ... 50% 75% max

corporation ...

GOOGLE 2.0 395.5 219.910209 ... 395.5 473.25 551.0

MSFT 2.0 605.0 91.923882 ... 605.0 637.50 670.0

YAHOO 2.0 120.0 28.284271 ... 120.0 130.00 140.0

print(group\_data.describe().transpose())

**output**

corporation GOOGLE MSFT YAHOO

sales\_in\_USD count 2.000000 2.000000 2.000000

mean 395.500000 605.000000 120.000000

std 219.910209 91.923882 28.284271

min 240.000000 540.000000 100.000000

25% 317.750000 572.500000 110.000000

50% 395.500000 605.000000 120.000000

75% 473.250000 637.500000 130.000000

max 551.000000 670.000000 140.000000

print(group\_data.describe().transpose()['GOOGLE'])

**output**

sales\_in\_USD count 2.000000

mean 395.500000

std 219.910209

min 240.000000

25% 317.750000

50% 395.500000

75% 473.250000

max 551.000000