

TASK – 2

1. Library Install:

```
Anaconda Prompt
(base) C:\Users\Lenovo>pip install pandas
Requirement already satisfied: pandas in c:\users\lenovo\anaconda3\lib\site-packages (0.23.4)
Requirement already satisfied: python-dateutil>=2.5.0 in c:\users\lenovo\anaconda3\lib\site-packages (from pandas) (2.7.5)
Requirement already satisfied: pytz>=2011k in c:\users\lenovo\anaconda3\lib\site-packages (from pandas) (2018.7)
Requirement already satisfied: numpy>=1.9.0 in c:\users\lenovo\anaconda3\lib\site-packages (from pandas) (1.15.4)
Requirement already satisfied: six>=1.5 in c:\users\lenovo\anaconda3\lib\site-packages (from python-dateutil>=2.5.0->pandas) (1.12.0)

(base) C:\Users\Lenovo>pip install numpy
Requirement already satisfied: numpy in c:\users\lenovo\anaconda3\lib\site-packages (1.15.4)
```

2. Object Creation:

import pandas as pd

3. Import File:

csv_path="Desktop/PS EXP 5 Sheet.csv"

4. Read CSV File:

df=pd.read_csv(csv_path)

5. Understand the csv file:

df.info()

df.describe()

In [4]: df.info()
df.describe()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 8 columns):
Student_ID      10 non-null int64
Student Name    10 non-null object
Maths Marks     10 non-null int64
Science Marks   10 non-null int64
English Marks   10 non-null int64
History Marks   10 non-null int64
Hindi Marks     10 non-null int64
Result         10 non-null object
dtypes: int64(6), object(2)
memory usage: 720.0+ bytes
```

Out[4]:

	Student_ID	Maths Marks	Science Marks	English Marks	History Marks	Hindi Marks
count	10.000000	10.000000	10.000000	10.000000	10.000000	10.000000
mean	1037.500000	18.500000	17.300000	16.500000	18.400000	17.100000
std	3.02765	1.581139	2.496664	3.27448	1.837873	1.66333
min	1033.000000	15.000000	13.000000	10.000000	15.000000	15.000000
25%	1035.250000	18.000000	15.250000	14.500000	17.250000	16.000000
50%	1037.500000	19.000000	17.500000	17.500000	19.000000	17.000000
75%	1039.750000	19.750000	19.750000	18.750000	20.000000	18.000000
max	1042.000000	20.000000	20.000000	20.000000	20.000000	20.000000

6.To read first five of the csv file: df.head()

```
In [5]: df.head()
```

```
Out[5]:
```

	Student_ID	Student Name	Maths Marks	Science Marks	English Marks	History Marks	Hindi Marks	Result
0	1033	Harsh	19	20	18	20	15	PASS
1	1034	Durvesh	20	15	14	19	16	PASS
2	1035	Shivansu	18	16	13	15	17	PASS
3	1036	Rudra	17	18	10	16	18	PASS
4	1037	Rishi	20	19	20	17	19	PASS

7.To read last five lines of the csv file: df.tail()

```
In [6]: df.tail()
```

```
Out[6]:
```

	Student_ID	Student Name	Maths Marks	Science Marks	English Marks	History Marks	Hindi Marks	Result
5	1038	Saurabh	19	20	20	18	20	PASS
6	1039	Vedangi	20	13	19	19	15	PASS
7	1040	Nischay	19	20	18	20	16	PASS
8	1041	Shivendra	18	15	17	20	17	PASS
9	1042	Jay	15	17	16	20	18	PASS

8.To Identify duplicate data: df.duplicated().sum()

```
In [7]: df.duplicated().sum()
```

```
Out[7]: 0
```

9.To find null vale: df.isnull().sum()

```
In [8]: df.isnull().sum()
```

```
Out[8]: Student_ID      0
Student Name      0
Maths Marks      0
Science Marks      0
English Marks      0
History Marks      0
Hindi Marks      0
Result      0
dtype: int64
```

To find null value: df.isnull().sum()

```
In [8]: df.isnull().sum()
```

```
Out[8]: Student_ID      0
Student Name      0
Maths Marks      0
Science Marks      0
English Marks      0
History Marks      0
Hindi Marks      0
Result      0
dtype: int64
```

ISO 9001 : 2015

```
import numpy as np
```

```
data = np.array([14,3,24,30,32,2,7,6,98,34,29,33])
```

```
1 #Calculating the quantile at 0.5(median)
2 median=np.quantile(data,0.5)
3 print("Median(Quantile at 0.5):",median)
```

Median(Quantile at 0.5): 26.5

```
1 #Calculating the 75th percentile
2 percentile_75=np.percentile(data, 75)
3 print("75th Percentile:",percentile_75)
```

75th Percentile: 32.25

Addition: df['Student_ID']=df['Student_ID']+2

```
In [9]: df['Student_ID']=df['Student_ID']+2
```

```
In [10]: print(df)
```

	Student_ID	Student Name	Maths Marks	Science Marks	English Marks	\
0	1035	Harsh	19	20	18	
1	1036	Durvesh	20	15	14	
2	1037	Shivansu	18	16	13	
3	1038	Rudra	17	18	10	
4	1039	Rishi	20	19	20	
5	1040	Saurabh	19	20	20	
6	1041	Vedangi	20	13	19	
7	1042	Nischay	19	20	18	
8	1043	Shivendra	18	15	17	
9	1044	Jay	15	17	16	
	History Marks	Hindi Marks	Result			
0	20	15	PASS			
1	19	16	PASS			
2	15	17	PASS			
3	16	18	PASS			
4	17	19	PASS			
5	18	20	PASS			
6	19	15	PASS			
7	20	16	PASS			
8	20	17	PASS			
9	20	18	PASS			

Subtraction: df['Student_ID']=df['Student_ID']-3

```
In [11]: df['Student_ID']=df['Student_ID']-3
```

```
In [12]: print(df)
```

	Student_ID	Student Name	Maths Marks	Science Marks	English Marks	\
0	1032	Harsh	19	20	18	
1	1033	Durvesh	20	15	14	
2	1034	Shivansu	18	16	13	
3	1035	Rudra	17	18	10	
4	1036	Rishi	20	19	20	
5	1037	Saurabh	19	20	20	
6	1038	Vedangi	20	13	19	
7	1039	Nischay	19	20	18	
8	1040	Shivendra	18	15	17	
9	1041	Jay	15	17	16	

	History Marks	Hindi Marks	Result
0	20	15	PASS
1	19	16	PASS
2	15	17	PASS
3	16	18	PASS
4	17	19	PASS
5	18	20	PASS
6	19	15	PASS
7	20	16	PASS
8	20	17	PASS
9	20	18	PASS