Aim :

To perform Preprocessing , Normalization and Standardization on the given dataset using 1. By creating functions. 2. By using sklearn library. And validating the results

Observation and Result-

1. Importing Data.

In [15]:
```
import pandas as pd

path = r"E:\DYPIU\SEM 6\KDD\KDD LABS\LAB_03\kc_house_data.csv"
missing_val_format = ["n.a","?","NA","n/a","na","--","NaN"]
df = pd.read_csv(path,na_values=missing_val_format)
df.head()
```

Out[15]:

	id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	...	grade	sqft_above	sqft_basement	yr_built
0	7129300520	20141013T000000	221900.0	3	1.00	1180	5650	1.0	0	0	...	7	1180.0	0	1955
1	6414100192	20141209T000000	538000.0	3	2.25	2570	7242	2.0	0	0	...	7	2170.0	400	1951
2	5631500400	20150225T000000	180000.0	2	1.00	770	10000	1.0	0	0	...	6	770.0	0	1933
3	2487200875	20141209T000000	604000.0	4	3.00	1960	5000	1.0	0	0	...	7	1050.0	910	1965
4	1954400510	20150218T000000	510000.0	3	2.00	1680	8080	1.0	0	0	...	8	1680.0	0	1987

5 rows × 21 columns

2. Normalization

Normalization

In [19]:
```
from numpy import asarray
from sklearn.preprocessing import MinMaxScaler

#data as array.
data  = df[["price","bedrooms","bathrooms","sqft_living","floors"]].to_numpy()
data
#define scaler

scaler = MinMaxScaler()
scaler = scaler.fit(data)

scaler.data_min_
scaler.data_max_

scaled = scaler.transform(data)
df2 = pd.DataFrame(scaled,columns=["price","bedrooms","bathrooms","sqft_living","floors"])
df2.head()
```

Out[19]:

	price	bedrooms	bathrooms	sqft_living	floors
0	0.019266	0.090909	0.12500	0.067170	0.0
1	0.060721	0.090909	0.28125	0.172075	0.4
2	0.013770	0.060606	0.12500	0.036226	0.0
3	0.069377	0.121212	0.37500	0.126038	0.0
4	0.057049	0.090909	0.25000	0.104906	0.0

3. Standardization

Standardization

In [21]:
```
from numpy import asarray
from sklearn.preprocessing import StandardScaler

#data as array.
data  = df[["price","bedrooms","bathrooms","sqft_living","floors"]].to_numpy()
data
#define scaler

scaler = StandardScaler()
scaler = scaler.fit(data)

scaler.mean_
scaler.scale_

scaled = scaler.transform(data)
df2 = pd.DataFrame(scaled,columns=["price","bedrooms","bathrooms","sqft_living","floors"])
df2.head()
```

Out[21]:

	price	bedrooms	bathrooms	sqft_living	floors
0	-0.866717	-0.398737	-1.447464	-0.979835	-0.915427
1	-0.005688	-0.398737	0.175607	0.533634	0.936506
2	-0.980849	-1.473959	-1.447464	-1.426254	-0.915427
3	0.174090	0.676485	1.149449	-0.130550	-0.915427
4	-0.081958	-0.398737	-0.149007	-0.435422	-0.915427

Conclusion –

Implemented Normalization and Standardization using various libraries in Python.