

# lec9 dealing with missing

In [1]:

```
import sys
import os
import pandas as pd
import numpy as np
```

In [2]:

```
cars_data=pd.read_csv('Toyota.csv',index_col=0,na_values=["??","????"])
```

In [15]:

```
carsdata=cars_data.copy(deep=True)
carsdata1=cars_data.copy()
```

In [4]:

```
carsdata.isnull().sum()
```

Out[4]:

```
Price      0
Age        100
KM          15
FuelType   100
HP          6
MetColor   150
Automatic   0
CC          0
Doors       0
Weight      0
dtype: int64
```

In [5]:

```
missing=carsdata[carsdata.isnull().any(axis=1)]
```

In [6]:

```
carsdata.describe()
```

Out[6]:

	Price	Age	KM	HP	MetColor
count	1436.000000	1336.000000	1421.000000	1430.000000	1286.000000
mean	10730.824513	55.672156	68647.239972	101.478322	0.674961
std	3626.964585	18.589804	37333.023589	14.768255	0.468572
min	4350.000000	1.000000	1.000000	69.000000	0.000000
25%	8450.000000	43.000000	43210.000000	90.000000	0.000000
50%	9900.000000	60.000000	63634.000000	110.000000	1.000000
75%	11950.000000	70.000000	87000.000000	110.000000	1.000000
max	32500.000000	80.000000	243000.000000	192.000000	1.000000

In [7]:

```
# to fill missing values on age
carsdata['Age'].mean() #55.67
carsdata['Age'].fillna(carsdata['Age'].mean(),inplace=True)
```

In [8]:

```
# mising value of km
carsdata['KM'].median() #63634.0
carsdata['KM'].fillna(carsdata['KM'].median(),inplace=True)
```

In [9]:

```
# mising value of hp
carsdata['HP'].median() #101.41
carsdata['HP'].fillna(carsdata['HP'].mean(),inplace=True)
```

Imputing missing values of ‘FuelType’ Series.value\_counts()

In [10]:

```
#for categorical value
carsdata['FuelType'].value_counts()
```

Out[10]:

```
Petrol      1177
Diesel       144
CNG          15
Name: FuelType, dtype: int64
```

In [11]:

```
carsdata['FuelType'].fillna(carsdata['FuelType'].value_counts().index[0],inplace=True)
```

In [12]:

```
# mode
carsdata['MetColor'].mode()
```

Out[12]:

```
0      1.0
dtype: float64
```

In [13]:

```
carsdata['MetColor'].fillna(carsdata['MetColor'].mode()[0],inplace=True)
```

In [14]:

```
carsdata.isnull().sum()
```

Out[14]:

```
Price      0
Age         0
KM          0
FuelType   0
HP          0
```