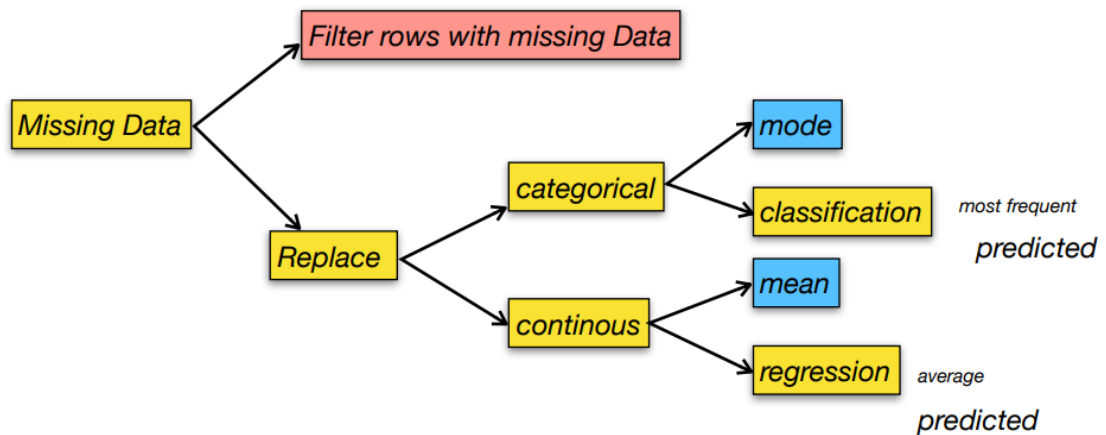


Missing Data Treatment



To Treat Missing Data there are two main ways,

- 1-Filter rows with missing Data
 - Deleting the rows which is having missing Data
 - The no. of rows you are deleting alongwith this you are deleting experience of that model.
 - This is not recomended way to treat the missing Data

- 2.Replace Missing Data
(The Replace method has a drawback it has extreme effect on higher value or lower value causes improper bounce best replacement in data)
 - A) For Continous Columns - There two way to treat missing data for continous columns
 - a) By Mean of remaining values
 - Just taking a mean of remaining values
 - and replace the missing value by that mean
 - b) By Prediction using KNN Regression
 - Using KNN Algorithm we can find out value for missing value

 - B) For Categorical Columns - There two way to treat missing data for continous columns
 - a) By Mode of remaining values
 - Just taking a Mode of remaining values and replace the missing value by that mode
 - b) By Prediction using KNNClassification
 - Using KNN Algorithm we can findout best values for missing data

Read Dataset

```

In [1]: import pandas as pd
        from pandas import read_csv
        from warnings import filterwarnings
        filterwarnings("ignore")
        A=pd.read_csv("C:/Users/Admin/Desktop/Data set/Cars93.csv")
        A.head()
  
```

```
Out[1]:
```

	id	Manufacturer	Model	Type	Min.Price	Price	Max.Price	MPG.city	MPG.highway	Air
0	1	Acura	Integra	Small	12.9	15.9	18.8	25	31	
1	2	Acura	Legend	Midsize	29.2	33.9	38.7	18	25	Drive
2	3	Audi	90	Compact	25.9	29.1	32.3	20	26	Passenger
3	4	Audi	100	Midsize	30.8	37.7	44.6	19	26	
4	5	BMW	535i	Midsize	23.7	30.0	36.2	22	30	

5 rows × 28 columns

Check for Missing Values

```
In [2]: A.isna().sum()
```

```
Out[2]: id                0
Manufacturer            0
Model                  0
Type                   0
Min.Price              0
Price                  0
Max.Price              0
MPG.city               0
MPG.highway            0
AirBags                4
DriveTrain             0
Cylinders              0
EngineSize             0
Horsepower             0
RPM                    0
Rev.per.mile           0
Man.trans.avail        0
Fuel.tank.capacity     0
Passengers             0
Length                 0
Wheelbase              0
Width                  0
Turn.circle            0
Rear.seat.room         2
Luggage.room           11
Weight                 0
Origin                 0
Make                   0
dtype: int64
```

```
In [3]: #There are 4 Missing Values in AirBags Columns
#There are 2 Missing Values in Rear.seat.room Columns
#There are 11 Missing Values in Luggage.room Columns
```

Treatment for Missing Values

```
In [4]: # Lets starts with AirBags Column
```

```
In [5]: A[["AirBags"]].dtypes
```

```
Out[5]: AirBags    object  
dtype: object
```

```
In [6]: #AirBags is an Categorical type
```

Missing Values Treatmemnt with Mode

```
In [7]: x=A.AirBags.mode()[0] #Replace missing values with Mode
```

```
In [8]: A.AirBags = A.AirBags.fillna(x)
```

```
In [9]: A.AirBags.isna().sum()
```

```
Out[9]: 0
```

Missing Values Treatmemnt with Mean

```
In [13]: A["Rear.seat.room"].dtypes
```

```
Out[13]: dtype('float64')
```

```
In [14]: #Rear.seat.room is in float format so we will use mean method to replace missing va
```

```
In [17]: X=A["Rear.seat.room"].mean()
```

```
In [18]: A["Rear.seat.room"]=A["Rear.seat.room"].fillna(X)
```

```
In [19]: #Lets check for missing value now in Rear.seat.room
```

```
In [20]: A['Rear.seat.room'].isna().sum()
```

```
Out[20]: 0
```

```
In [21]: A['Luggage.room'].dtype
```

```
Out[21]: dtype('float64')
```

```
In [22]: #Luggage room is in float format so we will use mean method to replace missing valu
```

```
In [23]: X=A['Luggage.room'].mean()
```

```
In [24]: A['Luggage.room']=A['Luggage.room'].fillna(X)
```

```
In [27]: A['Luggage.room'].isna().sum()
```

```
Out[27]: 0
```

```
In [28]: #Now check again the status of Origina Dataset for Missing Values get replaced or n
```

```
In [29]: A.isna().sum()
```

```
Out[29]: id                0
Manufacturer            0
Model                  0
Type                   0
Min.Price              0
Price                  0
Max.Price              0
MPG.city               0
MPG.highway            0
AirBags                0
DriveTrain             0
Cylinders              0
EngineSize             0
Horsepower             0
RPM                    0
Rev.per.mile           0
Man.trans.avail        0
Fuel.tank.capacity     0
Passengers             0
Length                0
Wheelbase              0
Width                  0
Turn.circle            0
Rear.seat.room         0
Luggage.room           0
Weight                0
Origin                 0
Make                   0
dtype: int64
```

```
In [ ]:
```