

Handling Missing Data

Missing data can lead to bias and inaccurate result in the model.

- It is important to address missing data by either dropping the missing observation or imputing the missing values using mean/median/mode.
- In python this can be done using methods of pandas library

```
DataFrame_name.fillna()  
DataFrame_name.dropna()
```

- There are multiple reasons for the missing value in dataset

- improper maintenance of data
- no proper format defined for the data
- human error
- user not given information intentionally

Ways to Handle it

- In imputing the missing values the above mentioned methods are used mostly in various ways. Some ways are mentioned below:

```
DataFrame_name.fillna(0) # missing/Nan values are replaced with 0  
DataFrame_name.fillna( ) # any object or number or any series of number can be  
used to replace the missing value  
  
DataFrame_name.fillna(method="ffill") # for filling the next value ## mostly used  
to handle the missing data of date format  
DataFrame_name.fillna(method="bfill",inplace=True) # for filling the previous  
value  
# inplace to fill in-place, no new column will be created
```