Percentile, Mean, Median, Mode

Mean / Average :-

Mean / Average is extracted from the whole data given by dividing the sum of all numbers and / the numbers.

It is affected by the outliers aswell.

Incomes :-

5000, 6000, 4000, 7500, 8000, 7000

    Average = 6250

Now, if outlier comes into it like

5000, 6000, 4000, 7500, 8000, 7000, 10000000

    Average = 1.4300000

Here you can see one value which is not in the range of other values affected the complete other dataset.

Median

This problem can be solved by the use of median.

The central value from the given dataset.

4000 5000 6000 | 7000 | 7500 8000 10 Million

4000 5000 6000 | 7000 8000 | 7500 8000 10 Million

Handling missing values

Median values are used to handle the missing values in the given dataset.
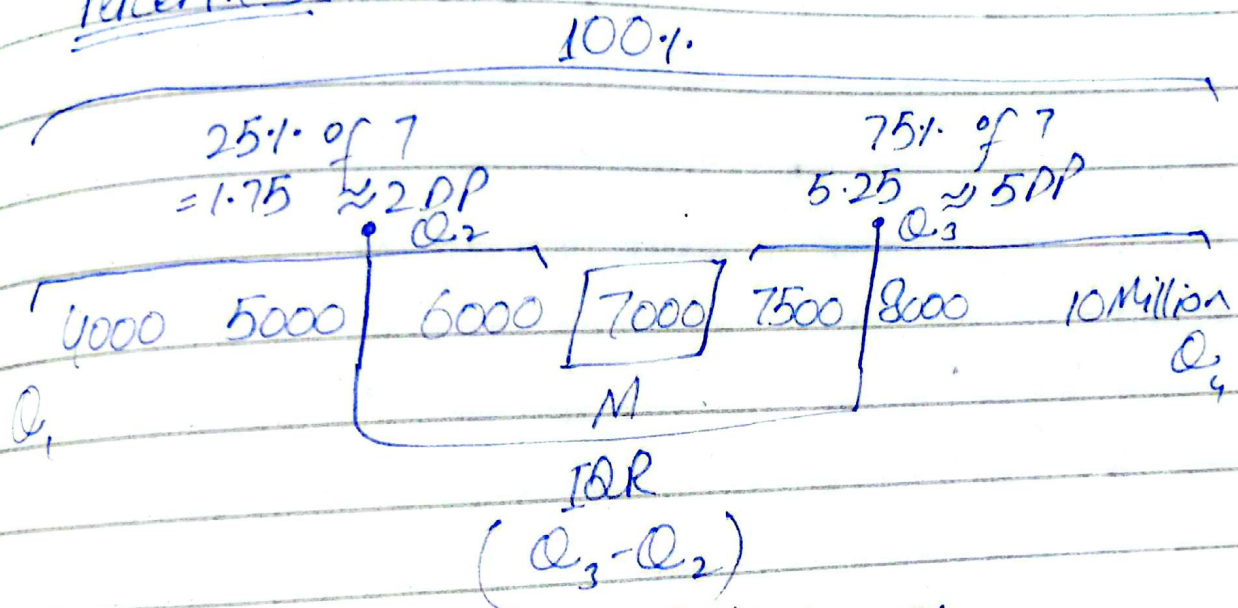
Mean and median uses In Data Sciences

① Descriptive Analysis (like for opening a car show room)

② Data Cleaning (Filling NAN values)

Outliers :-

"Outlier is a datapoint that is very different than rest of the datapoints."

Percentiles

$$100\%$$

| 25% of 7 | | | | | 75% of 7 | |
| $=1.75 \approx 2DP$ | | | | | $5.25 \approx 5DP$ | |

Q_2 Q_3

4000   5000 | 6000 $[7000]$ 7500 | 8000   10Million

Q_1 M Q_4

IQR

$(Q_3 - Q_2)$

Percentile Used in Data Sciences

① Outliers Removal.

② General Data Analysis.

Mode

"Mode means most frequently occurring value in the dataset."