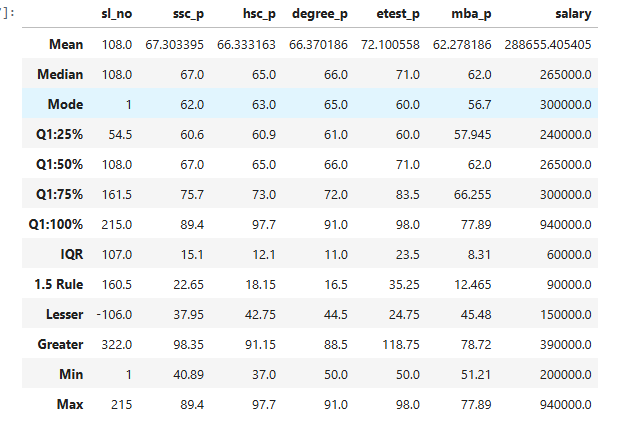
The given Placement.csv contains the students details with Gender, salary and all marks.



1. **CENTRAL TENDENCY:**

Above output of Mean, Median and Mode for all the

1. Students who had studied in SSC\_P got average mark of 67.30 and the 67-mark obtained by 62 Students.
2. Students who had studied in HSC\_P got average mark of 66.33 and the 65-mark obtained by 63 Students.
3. Students who had studied in degree\_P got average mark of 66.37 and the 66-mark obtained by 65 Students.
4. Students who had studied in etest\_P got average mark of 72.1 and the 71-mark obtained by 60 Students.
5. Students who had studied in mba\_P got average mark of 62.2 and the 62-mark obtained by 56 Students.

**Based on this analysis, if student obtains the above marks in each exam will get 265000 as salary**

1. **PERCENTILE:**

**SSC\_P**

Ssc\_p marks percentile of 25 is 60.6, and 50 percentile of mark is 67, which is 7% of increase.

And the same, 75 percentile of mark is 75.7. therefore it increments for 8.7% of mark

From this each quadrant of percentile gives 7 to 8 % of Mark increases.

But 75 to 100 percentile gives 14% increase in the marks obtained by students.

Hence SSC marks obtained by students are in increment format with as average of 8% increase.

**HSC\_P**

1. HSC\_p marks percentile of 25 is 60.9, and 50 percentile of mark is 65, which is 4% of increase.

And the same, 75 percentile of mark is 73. therefore, it increments for 8% of mark

From this each quadrant of percentile varies from 4 to 8 % of Mark increases.

But 75 to 100 percentile gives 25% increase in the marks obtained by students.

Hence SSC marks obtained by students are in increment format with as average of 8% increase till 3rd quadrant. Maximum marks obtained percentile is 25

1. **IQR:**

Outlier Detection is balance between sensitivity and stringency, aiming to identify outliers that are meaningfully different from rest of the data without being overly sensitive.

To detect outliers because it most closely follows Gaussian distribution. As result the method dictates that any data point that’s 1.5 points below the lower bound quartile or above the upper bound quartile is an outlier.

IQR is the column represents the Outlier.

Lesser and Greater are the maximum values in the range

1. **FINDING OUTLIERS**

hsc\_p:

42.75 37 60.9 65 73 91.5 97.75

Lesser and Greater Outlier is present

ssc\_p:

37.95 40.89 60.6 67 75.7 98.35 89.4

Greater outlier is present

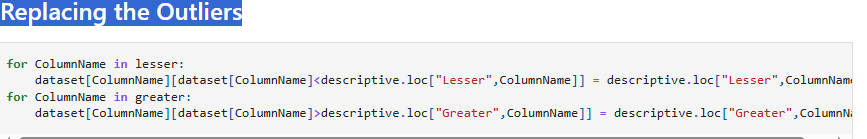
Degree\_p maximum is 91.0 and greater is 88.5 so Greater Outlier is present

Salary max =940000.0 and greater is 390000.0, so outlier is present.

1. **REPLACING THE OUTLIERS**

Outliers can be replaced with Lesser and Greater Value after comparing with dataset columns.

Which is Min and Max value replaced with Lesser and Greater value



1. **FREQUENCY:**

Repetition is the Frequency

Unique Values:

Frequency: How many repetitions happened for each data

Relative Frequency: Same repetition in the percentage format

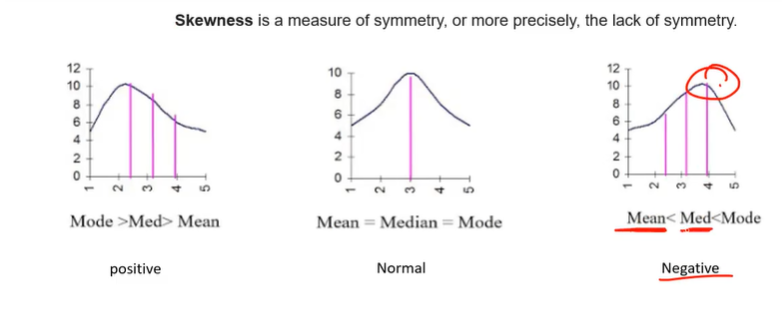
Cumulative Frequency: Summation of the frequency and the total has to be equals to 1.

1. **HISTOGRAM:**

Graphical View of Frequency is Histogram. Y-axis is Count and X-axis the dataset.

1. **SKEWNESS:**

It is the measure of symmetry. And it is defined with Mode, Mean and Median.



1. **KURTOSIS:**

Measure of the peakness or convexity of a curve

1. Platykurti:
   1. If a kurtosis graph is less than 3
2. MesoKurti:
   1. If Kurtosis equals with 3
3. Laptokurti:
   1. If Kurtosis greater than 3



As per the table, All the values in columns are less than 3 which is **Platykurti**