Table Creation:

Dataset:



Table\_Name=pd.DataFrame(index=[“Row Name1”,”Row Name2”],columns=quan)

descriptive=pd.DataFrame(index=["Mean","Median","Mode"],columns=quan)

| **sl\_no** | **ssc\_p** | **hsc\_p** | **degree\_p** | **etest\_p** | **mba\_p** | **salary** |
| --- | --- | --- | --- | --- | --- | --- |
| **Mean** | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| **Median** | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| **Mode** | NaN | NaN | NaN | NaN | NaN | NaN | NaN |

Table is been created but there is no value on it , so we need to insert value in each places.

**For adding single value**

descriptive["hsc\_p"]["Mean"]=0

| **sl\_no** | **ssc\_p** | **hsc\_p** | **degree\_p** | **etest\_p** | **mba\_p** | **salary** |
| --- | --- | --- | --- | --- | --- | --- |
| **Mean** | NaN | NaN | 0 | NaN | NaN | NaN | NaN |
| **Median** | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| **Mode** | NaN | NaN | NaN | NaN | NaN | NaN | NaN |

Insert values in the table:

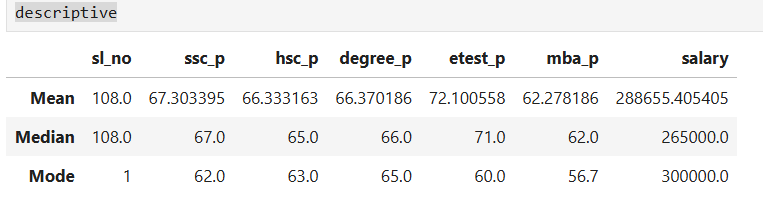
for columnName in quan:

descriptive[columnName]["Mean"]=dataset[columnName].mean()

descriptive[columnName]["Median"]=dataset[columnName].median()

descriptive[columnName]["Mode"]=dataset[columnName].mode()[0]

Final values:



Mean = Mean = add up all the given values, then divide by no of samples

10th, 12th, degree, mba 🡪 There is no much difference on their performance apart from the entrance test result.

Median = the middle value of a set of data

Median value is also similar as mean

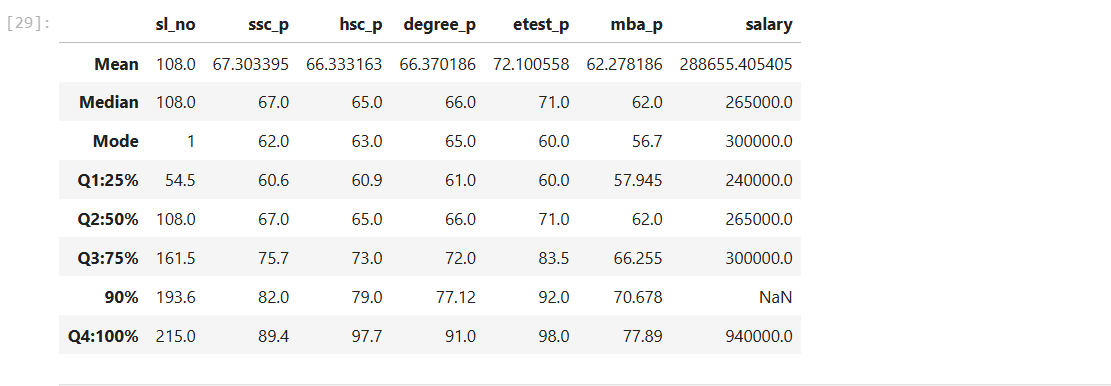
Median excludes outlier(unwanted data that will affect the middle value, so we use median for calculating salary types of data.

Mode = the value which occurs more often in a provided set of data values.

60 mark are the most occurrence of the given values.

Observation: Students information shows that they are average students.

**Percentile**:



Formula:

I=(k/100)\*(n+1)

I = The position in the sample, K = Percentage, n= no of sample

25% 🡪 ssc\_p 🡪 The first 25% of total people is getting 67% of marks

50% 🡪When we look at 50% of the percentile, the mark difference is only 6.3%

75% 🡪The marks difference between 50% and 75% is 8.3, it little bit good progress than the percentile between 25 and 50%.

90% 🡪Again the mark difference is getting reduced when we check the 90%. It is 6.3% only.

100% 🡪 The mark difference between 90 and 100 is 7.4%.

Similarly we can identify all the values related to percentile.

Hsc\_p🡪

25% 🡪 The first 25% of the students scored 60.9 marks.

Students scored only 4.9 marks difference between 25% and 50%

Marks difference between 50% and 75% is only 8 marks

24.3 marks difference is made between 75% and 100%

Degree\_p🡪

25% 🡪 The first 25% of the students scored 61 marks.

Students scored only 5 marks difference between 25% and 50%

Marks difference between 50% and 75% is only 6 marks

19 marks difference is made between 75% and 100%

**etest\_p**

25% 🡪 The first 25% of the students scored 60 marks.

Students scored only 11 marks difference between 25% and 50%

Marks difference between 50% and 75% is only 12.5 marks

14.5 marks difference is made between 75% and 100%

**mba\_p**

25% 🡪 The first 25% of the students scored 58 marks.

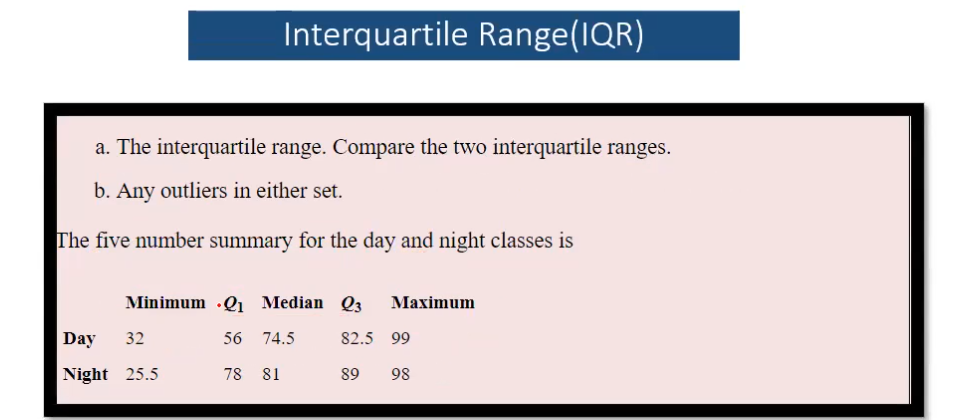
Students scored only 4 marks difference between 25% and 50%

Marks difference between 50% and 75% is only 4.25 marks

11 marks difference is made between 75% and 100%

Singnificantly the students are scored better in **etest,** as the progress of each percentile an**d 100** percentile is scored 98

**IQR**:



**Day time**

Q1 = 56, Q3 = 82.5

Finding IQR = Q3-Q1 = 82.5 – 56 = 26.5

Lesser Interquartile range = Q1-1.5\*IQR = 56 – (1.5\*26.5) = 56 -39.75 = 16.25

**Lesser IQR = 16.25**

Day minimum Temperature is 32, so there is **no lesser outlier in the given temperature**.

**Greater Outlier**

Q3+1.5\*IQR = 82.5+(1.5\*26.5)=82.5+39.75 = 122.25

Day Maximum Temo is 99, there is **no maximum outlier in the given day temperature**.

**Night Time**

Q1 = 78, Q3 = 89

IQR = Q3-Q1 = 89-78 = 11

**Lesser IQR** = Q1-1.5\*IQR = 78 – (1.5\*11) = 78-16.5 = 61.5

Lesser IQR = 61.5

Night miminum temp is 25.5, minimun temp should not be lesser than 61.5. then we need to replace the minimum temperature as 61.5

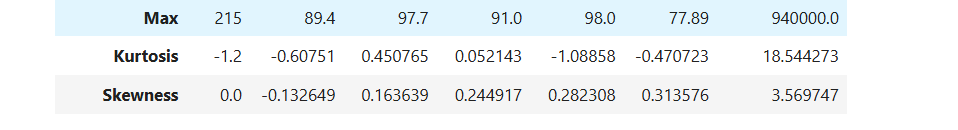
**Greater Outlier**

Q3+1.5\*IQR = 89+(1.5\*11) = 89+16.5 = 105.5

Greater Outlier is 105.5 , our maximim night temp is 98, so there is no greater outlier is present in the given temp.

**SKEWNESS**





Positive 🡪 Mode>Median>Mean

Negative🡪 Mean>Medain>Mode

Normal🡪 Mean=Median=Mode

Sl.no= skewness 🡪 Normal

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Sl.No | ssc\_p | hsc\_p | degree\_p | etest\_p | mba\_p | salary |
| Values | 0 | -0.132649 | 0.163639 | 0.244917 | 0.282308 | 0.313576 | 3.569747 |
| Skewness | Normal | Negative | Positive | Positive | Positive | Positive | Positive |
|  | Mean= Median= Mode | Mean> Medain> Mode | Mode> Median> Mean | Mode> Median> Mean | Mode> Median> Mean | Mode> Median> Mean | Mode> Median> Mean |

Positive Skewness 🡪 Example of Degree\_p 🡪 Most of the students got peak(Mode) in the beginning of the data.. means mode curve will fall in the first portion of the graph.

Positive skewness shows the good mark scoring students are there.

Negative skewness🡪 Example of ssc\_p🡺 Most of the students peak value falls on the later part of the graph.

It shows the no of low scoring students are more.

**KURTOSIS**:

Kurtosis<3(Kurtosis is lessthan 3) the length will be broarder🡪 Platykurtic

Kurtosis=3(Kurtosis is equal to 3) the length is lesser than the platykurtic 🡪 Misokurtic

Kurtosis>3(Kurtosis is greaterthan 3) the length is thinner than other 2 kutosis🡪Leptokurtic

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Sl.No | ssc\_p | hsc\_p | degree\_p | etest\_p | mba\_p | salary |
| Values | -1.2 | -0.60751 | 0.450765 | 0.052143 | -1.08858 | -0.470723 | 18.544273 |
| Kurtosis | Normal | Negative | Positive | Positive | Positive | Positive | Positive |
|  | Platykurtic | Platykurtic | Platykurtic | Platykurtic | Platykurtic | Platykurtic | Leptokurtic |

Platykurtic🡪 hsc\_p 🡪 lot of good scoring students are more, so that the curve breadth is longer.

Leptokurtic🡪 Salary 🡪Good scoring students counts are less, so that the curve breadth is less.