Measure of Central Tendency

That describe the centre or typical value of the dataset

Measure of central tendency is consist of 3 value

Mean – Average value of the data

Mean will not provide the accurate result if outlier data available,it will ommit the Outlier dataset

Mean calculation

1,2,3,4,5=1+2+3+4+5/5

Ans :3

Median : Midpoint of the data

Median will provide result if outlier data available

Mean formula : n+1/2

Mean calculation:

1,2,3,4,5

It pick mid value of the dataset

Mode: it provide the repeated values in the dataset

Mean Median Mode code details

Central Tendency (Meanmedianmode)

```
In [53]: univariate
```

Out[53]:

	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Mean	67.3034	66.3332	66.3702	72.1006	62.2782	288655
Median	67	65	66	71	62	265000
Mode	62	63	65	60	56.7	300000

Short description of the Mean Median Mode

Mean

In the placement docs average of the student in 10,12th,and graduate score is above average,but in the entrance test the results is good, so the most of the student in the class will get a placement.

Median

As we compare to the mean and median, both of the result provide the almost same.so this dataset will working for the both dataset

Mode

As per mode result most of the student in the class get the below result frequently. This also not a bad result,as we see mean median mode result.most will work very consistently for this dataset.

Mode	62	63	65	60	56.7	300000

Measure of Location of the data

Percentile :

Percentile tells about the value exist within the range, dividing whole dataset into four parts in terms of percentage.,percentile will not find the category dataset

25%,50%,75%,100%

Before finding the percentage we have to allocate original dataset by ascending order to fnd the value.

Percentile formula: i=k/100(n+1)

I is the index

K is the percentile

N is the total no of data points.

Percentile will comes under numpy directory

Percentile code details:

In [23]:
```
#after insert Mean,median,mode in the dataset, now we have to input datas in the dataframe
univariate=pd.DataFrame(index=["Mean","Median","Mode","25%","50%","75%","100%"],columns=Quan)
# columnName will take data from the input data
for columnName in Quan:
    univariate[columnName]["Mean"]=dataset[columnName].mean()
    univariate[columnName]["Median"]=dataset[columnName].median()
    univariate[columnName]["Mode"]=dataset[columnName].mode()[0]
#we can use this code to find thepercentile, but in the data is any missing values found numpy will not take that values.
#ex below
univariate[columnName]["25%"]=np.percentile (dataset[columnName],25)
```

In [24]:
```
univariate
```

Out[24]:

	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Mean	67.3034	66.3332	66.3702	72.1006	62.2782	288655
Median	67	65	66	71	62	265000
Mode	62	63	65	60	56.7	300000
25%	60.6	60.9	61	60	57.945	NaN
50%	NaN	NaN	NaN	NaN	NaN	NaN
75%	NaN	NaN	NaN	NaN	NaN	NaN
100%	NaN	NaN	NaN	NaN	NaN	NaN

In [29]:
```
univariate=pd.DataFrame(index=["Mean","Median","Mode","25%","50%","75%","100%"],columns=Quan)
# columnName will take data from the input data
for columnName in Quan:
    univariate[columnName]["Mean"]=dataset[columnName].mean()
    univariate[columnName]["Median"]=dataset[columnName].median()
    univariate[columnName]["Mode"]=dataset[columnName].mode()[0]
#we can use this code to find thepercentge but in the data is any missing values,numpy will not take that values.
#univariate[columnName]["25%"]=np.percentile (dataset[columnName],25)
    univariate[columnName]["25%"]=dataset.describe()[columnName]["25%"]
    univariate[columnName]["50%"]=dataset.describe()[columnName]["50%"]
    univariate[columnName]["75%"]=dataset.describe()[columnName]["75%"]
    univariate[columnName]["100%"]=dataset.describe()[columnName]["max"]
```

In [30]:
```
univariate
```

Out[30]:

	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
Mean	67.3034	66.3332	66.3702	72.1006	62.2782	288655
Median	67	65	66	71	62	265000
Mode	62	63	65	60	56.7	300000
25%	60.6	60.9	61	60	57.945	240000
50%	67	65	66	71	62	265000
75%	75.7	73	72	83.5	66.255	300000
100%	89.4	97.7	91	98	77.89	940000

Report

This report shows the Average score value of student in the class

For 12th 25% of student got 60% of the marks will get the salary of 2,40,000.00

For 12th 50% of student got 65% of the marks will get the salary of 2,65,000.00

For 12th 75% of student got 73% of the marks will get the salary of 3,00,000.00

For 12th 100% of student got 97.7% of the marks will get the salary of 9,40,000.00

But when comparing the data from 75% - 100% the difference of percentage increasing rapidly and same for Salary.

Interquartile range(IQR)

To know the outlier range present in the dataset

IQR formula

IQR = Q3-Q1

IQR=356 -

Lesser outlier & Greater Outlier

Lesser Outlier

=q1-1.5*IQR

Outlier value should not go the below value of Lesser Value

Greater Outlier

Q3+1.5*IQR

Outlier value should not go the above value of greater value

10=

In Machine learning outlier and missing values not exists

Mean will take the outlier

Median will omit the outlier

Either remove the outlier or replace the outlier

By using IQR we will replace the outlier

Interquartile Range(IQR) exercise for manual calculation

A, The interquartile range.compare the two interquartile ranges.

B, Any outliers in either set.

The five number summary for the day and night classes is

	Minimum	Q1	Median	Q3	Maximum
Day	32	56	74.5	82.5	99
Night	25.5	78	81	89	98

To find the lesser outlier and greater outlier in the above dataset,1st we have to find IQR

QR formula is Q3-Q1

Day IQR = 56-82.5

IQR for Day is = 26.5

Night IQR = 78 – 89

IQR for night is =11

We have to find lesser & greater outlier for day and night

Lesser outlier for day formula

q1-1.5*IQR

Q1=56

IQR = 26.5

1.5*26.5 =39.75

56-39.75 = 16.25

Lesser Outlier for day = 16.25

As per above dataset we didn't find the lesser outlier

Greater outlier for day

formula

Q3+1.5*IQR

Q3= 82.5

82.5 +39.75 =122.25

Greater outlier for day is 122.25

As per above dataset we didn't find the greater outlier,so in the given dataset no outlier available

Lesser outlier for night

IQR = 11

q1-1.5*IQR

78-1.5*(11)

78-16.5

Lesser outlier for night=61.5

As per above dataset we find the lesser outlier for Night

Greater outlier for night

IQR = 11

Q3+1.5*IQR

89+1.5*11

89+16.5 = 105.5

Greater outlier for night =105.5

As per above dataset we didn't find the greater outlier, so in the given dataset no outlier available