Skewness and Kurtosis Analysis

Dataset:

```
dataset
```

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status	salary
0	1	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55.0	Mkt&HR	58.80	Placed	270000.0
1	2	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed	200000.0
2	3	M	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No	75.0	Mkt&Fin	57.80	Placed	250000.0
3	4	M	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No	66.0	Mkt&HR	59.43	Not Placed	NaN
4	5	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.8	Mkt&Fin	55.50	Placed	425000.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
210	211	M	80.60	Others	82.00	Others	Commerce	77.60	Comm&Mgmt	No	91.0	Mkt&Fin	74.49	Placed	400000.0
211	212	M	58.00	Others	60.00	Others	Science	72.00	Sci&Tech	No	74.0	Mkt&Fin	53.62	Placed	275000.0
212	213	M	67.00	Others	67.00	Others	Commerce	73.00	Comm&Mgmt	Yes	59.0	Mkt&Fin	69.72	Placed	295000.0
213	214	F	74.00	Others	66.00	Others	Commerce	58.00	Comm&Mgmt	No	70.0	Mkt&HR	60.23	Placed	204000.0
214	215	M	62.00	Central	58.00	Others	Science	53.00	Comm&Mgmt	No	89.0	Mkt&HR	60.22	Not Placed	NaN

Quantitative Analysis:

```
dataset[quan]
```

	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
0	67.00	91.00	58.00	55.0	58.80	270000.0
1	79.33	78.33	77.48	86.5	66.28	200000.0
2	65.00	68.00	64.00	75.0	57.80	250000.0
3	56.00	52.00	52.00	66.0	59.43	NaN
4	85.80	73.60	73.30	96.8	55.50	425000.0
...	...	...	...	...	...	...
210	80.60	82.00	77.60	91.0	74.49	400000.0
211	58.00	60.00	72.00	74.0	53.62	275000.0
212	67.00	67.00	73.00	59.0	69.72	295000.0
213	74.00	66.00	58.00	70.0	60.23	204000.0
214	62.00	58.00	53.00	89.0	60.22	NaN

215 rows × 6 columns

Output:

```
descriptive.loc[['kurtosis','skew'], :]
```

	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
kurtosis	-0.60751	0.450765	0.052143	-1.08858	-0.470723	18.544273
skew	-0.132649	0.163639	0.244917	0.282308	0.313576	3.569747

Skewness Rule:
If skewness value is 0 means - Normal Skewness i.e symmetrical
If skewness value is < 0 means - Negative Skewness (Mode > Median > mean)
If skewness value is > 0 means - Positive Skewness (Mean > Median > Mode)

Kurtosis Rule:
If skewness value is 3 means - Mesokurtic (Normal)
If skewness value is < 3 means - Platykurtic (data points are present in high proximity to mean)
If skewness value is > 3 means - Leptokurtic (more chance of outliers)

As per the rule, please find the observation,

Univariate Analysis	**ssc_p**	**hsc_p**	**degree_p**	**etest_p**	**mba_p**	**salary**
Kurtosis	< 3 so it has fewer extreme values	< 3 so it has fewer extreme values	< 3 so it has fewer extreme values	< 3 so it has fewer extreme values	< 3 so it has fewer extreme values	>3 so it has more extreme values or more outliers
Skew	Negative so skewed to left side	Positive so skewed to right side	Positive so skewed to right side	Positive so skewed to right side	Positive so skewed to right side	Positive so skewed to right side

Skew Observation:

1. ssc_p - has skewed to left side which means, only few student got high marks
2. Others columns have skewed to the right side which means the majority of students got high marks. The skewness orders from low to high positive skewness

a. hsc_p < degree_p < etest_p < mba_p < salary

Kurtosis Observation:

1. Salary value is > 3 - it has more outliers or more extreme values - Leptokurtic (fat-tailed)
2. Other columns have values which are < 3 which means, the data has fewer extreme values - Platykurtic (thin-tailed). The kurtosis orders from low to high value
    a. etest_p < ssc_p < mba_p < degree_p < hsc_p

Overall,
As per skewness, Majority of students got low marks in ssc_p because it skewed to the left side. Majority of students got high marks in remaining subjects because it skewed to the right side

As per the kurtosis, Salary has extreme values or more outliers than other columns