**Histogram**

A histogram is the most commonly used graph to show frequency distributions. It looks very much like a bar chart, but there are important differences between them. To construct a histogram, the first step is to ["bin" (or "bucket")](https://en.wikipedia.org/wiki/Data_binning) the range of values— divide the entire range of values into a series of intervals—and then count how many values fall into each interval.

Use a histogram when:

* You want to see the shape of the data’s distribution, especially when determining whether the output of a process is distributed approximately normally
* Analyzing whether a process can meet the customer’s requirements
* Seeing whether a process change has occurred from one time period to another
* Determining whether the outputs of two or more processes are different
* You wish to communicate the distribution of data quickly and easily to others

Note: Histograms are sometimes confused with [bar charts](https://en.wikipedia.org/wiki/Bar_chart). In a histogram, each bin is for a different range of values, so altogether the histogram illustrates the distribution of values. But in a bar chart, each bar is for a different category of observations (e.g., each bar might be for a different population), so altogether the bar chart can be used to compare different categories.

**Skewness**

Skewness is a statistical measure that indicates the asymmetry of a distribution. It shows how data is spread out in relation to the mean, with three types of skewness:

Note: A high skew can reflect the presence of outliers or Kurtosis. In summary, skewness helps in understanding the shape and spread of data in a dataset

* Positive Skew: If the mean is greater than the mode, the skewness is positive. The tail on the right side is longer or fatter, indicating that the majority of data points are concentrated on the left.

Relationship: *Mean > Median > Mode*

* Negative Skew: If the mean is smaller than the mode, the skewness is negative. **Skewness below -1:**Strong left skewness (negative skew) with a long tail on the left side.

Relationship:*Mean < Median < Mode*

* Zero Skew:  If the mean is equal to the mode, the skewness is zero. The distribution is symmetrical, resembling a normal distribution. Skewness close to 0 (between -0.5 and 0.5): The distribution is approximately symmetric.

Relationship:*Mean = Median = Mode*

**Kurtosis**

**Kurtosis**is a measure of the tailedness of a distribution. Tailedness is how often outliers occur. **Excess kurtosis**is the tailedness of a distribution relative to a [normal distribution](https://www.scribbr.com/statistics/normal-distribution/).

Note: **Tails**are the tapering ends on either side of a distribution. They represent the probability or frequency of values that are extremely high or low compared to the mean. In other words, tails represent how often [**outliers**](https://www.scribbr.com/statistics/outliers/) occur.

* Distributions with medium kurtosis (medium tails) are [**mesokurtic**](https://www.scribbr.com/statistics/kurtosis/#mesokurtic). Have a kurtosis of (Normal) =3
* Distributions with low kurtosis (thin tails) are [**platykurtic**](https://www.scribbr.com/statistics/kurtosis/#platykurtic). Have a Kurtosis of (Low) < 3
* Distributions with high kurtosis (fat tails) are [**leptokurtic**](https://www.scribbr.com/statistics/kurtosis/#leptokurtic). Have a Kurtosis of (High) > 3

1. A **mesokurtic**distribution is medium-tailed, so outliers are neither highly frequent, nor highly infrequent.

* A kurtosis (Normal) equal to (=) 3
* An excess kurtosis equal to (=) 0

1. A **platykurtic**distribution is thin-tailed, meaning that outliers are infrequent. Platykurtic distributions have less kurtosis than a [normal distribution](https://www.scribbr.com/statistics/normal-distribution/). Platykurtic is sometimes called**negative kurtosis**, since the excess kurtosis is negative. In other words, platykurtic distributions have:

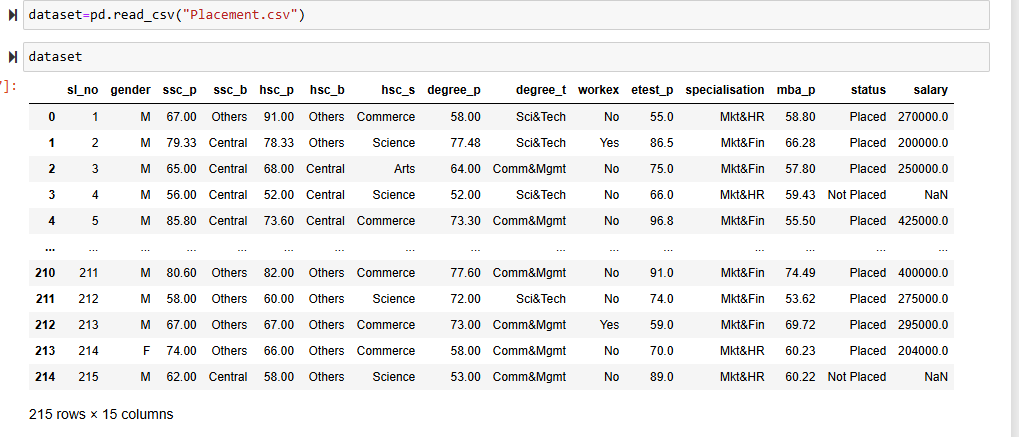
* A kurtosis (Low - Negative) of less than (<) 3
* An excess kurtosis of less than (<) 0

1. A **leptokurtic**distribution is fat-tailed, meaning that there are a lot of outliers. Leptokurtosis is sometimes called**positive kurtosis**, since the excess kurtosis is positive. Leptokurtic distributions are more kurtotic than a normal distribution. They have:

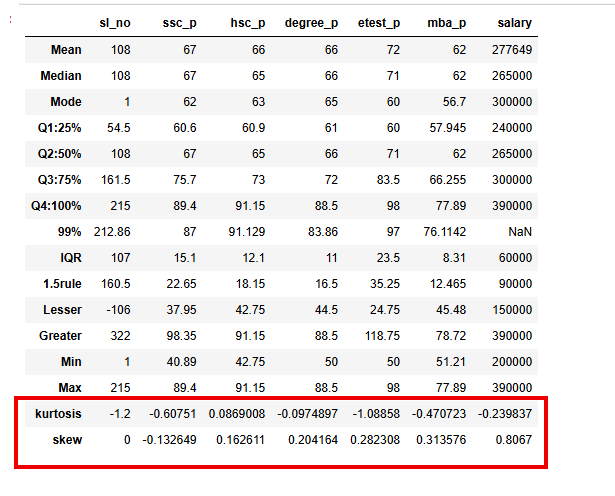
* A kurtosis (High -Positive) of more than (>) 3
* An excess kurtosis of more than (>) 0



**Dataset Used:**



Skewness- Kurtosis Ananlysis



**Kurtosis and Skew value analysis**

**ssc\_p**

Kurtosis value is less than 3: (-0.60751) - **platykurtic**distribution

Skew mean is greater than mode and skew is less than -1: (-.132649) – **negative skew**

Mean = 67 Mode= 62

**hsc\_p**

Kurtosis value is less than 3: (0.0869008) - **platykurtic**distribution

Skew mean is greater than mode: (0.162611) - **positive skew**

Mean = 66 Mode= 63

**degree\_p**

Kurtosis value is less than 3: (-0.0974897) - **platykurtic**distribution

Skew mean is greater than mode: (0.2041264) - **positive skew**

Mean = 66 Mode= 65

**etest\_p**

Kurtosis value is less than 3: (-1.08858) - **platykurtic**distribution

Skew mean is greater than mode: (0.282308) - **positive skew**

Mean = 72 Mode= 60

**mba\_p**

Kurtosis value is less than 3: (-0.470723) - **platykurtic**distribution

Skew mean is greater than mode: (0.282308) - **positive skew**

Mean = 62 Mode= 56.7

**Salary**

Kurtosis value is less than 3: (-0.239837) - **platykurtic**distribution

Skew mean is less than mode (0.8067) **– negative skew**

Mean = Mode= 277649 Mean - 300000