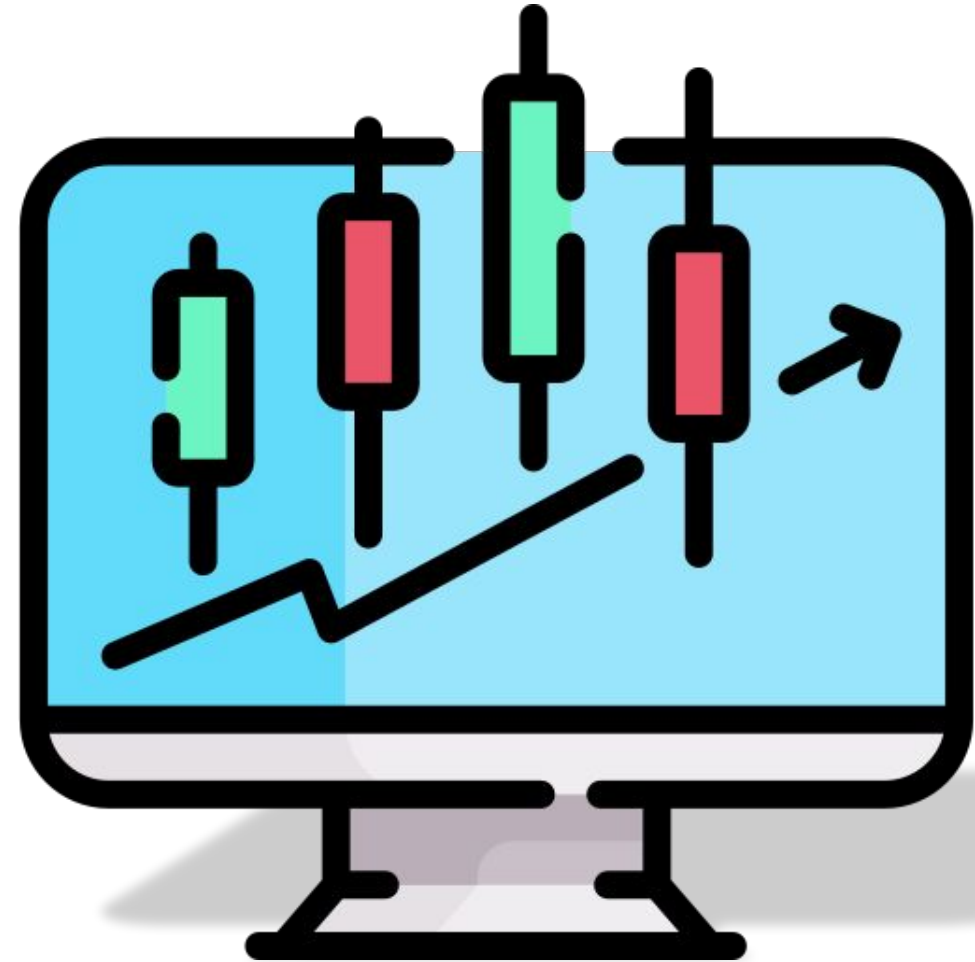


Statistics and Probability



Introduction to Statistics



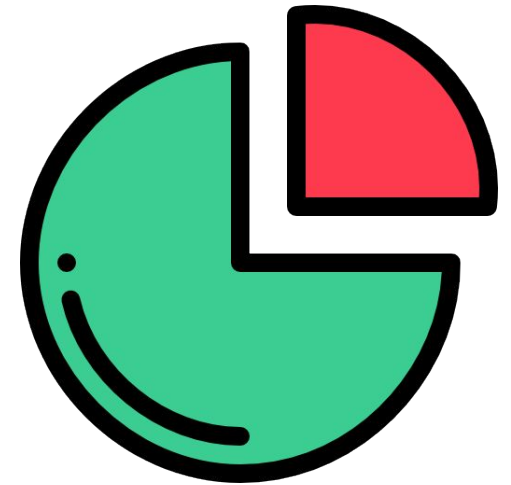
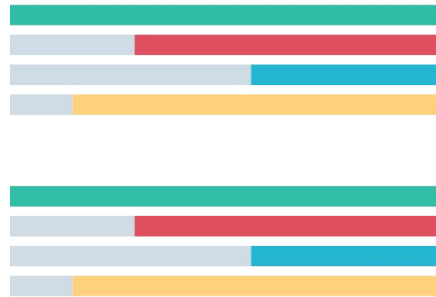


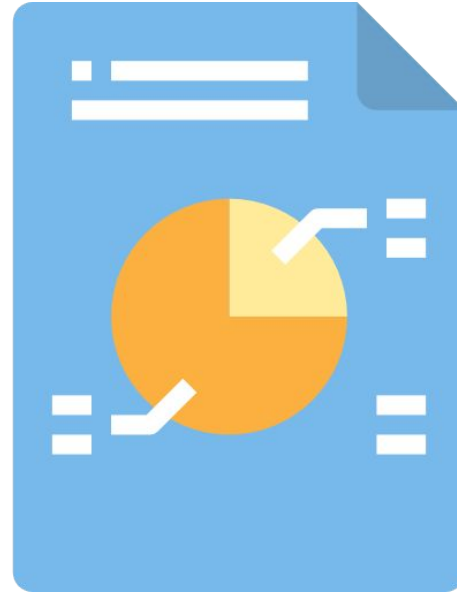
What is Statistics?

What is Statistics?

What?

Statistics is a branch of Mathematics that deals with collection, analyzing, and interpreting large amounts of data.

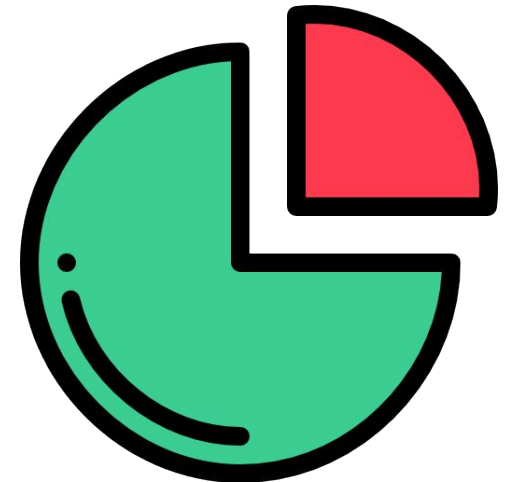
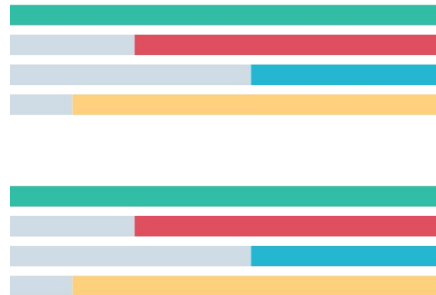


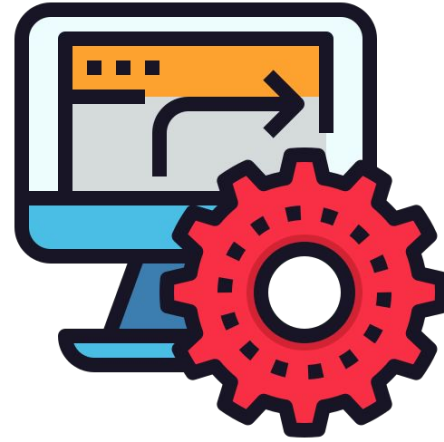


Why is **Statistics** important?

Why is Statistics important?

Statistics allows us to derive knowledge from large datasets and this knowledge can then be used to make predictions, decisions, classifications etc.

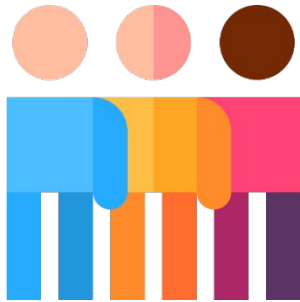




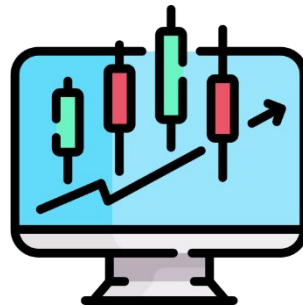
Where is **Statistics** used?

Where is Statistics used?

Statistics are used in various fields, some of them are:



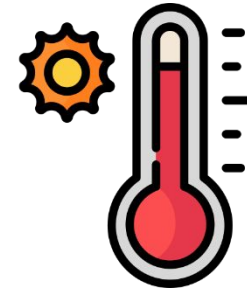
Medical Research



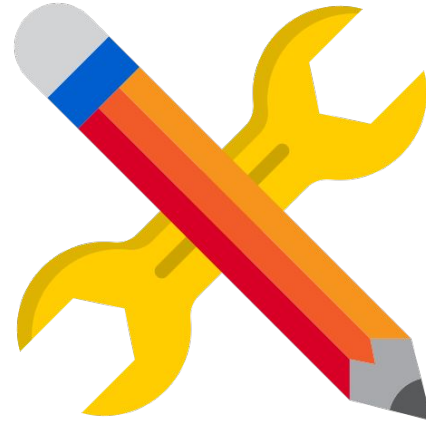
Stock Market



**Sales
Projection**

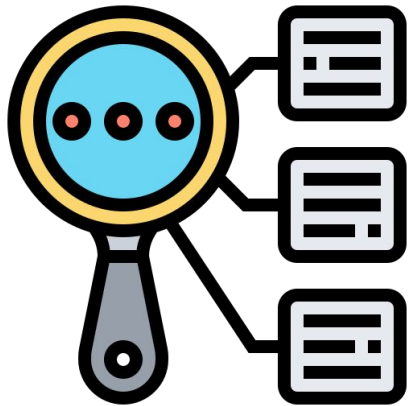


**Weather
Forecasting**



Sampling

Sampling is the process of collecting data to perform analysis on

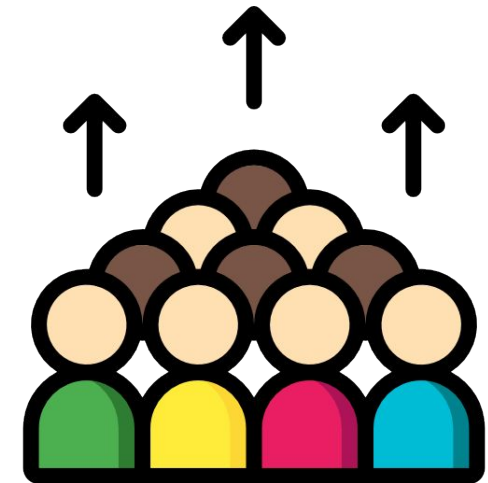
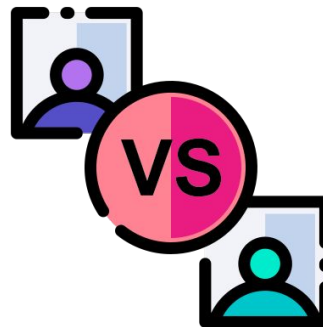


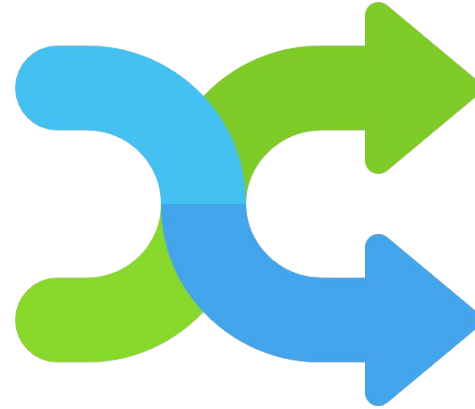


Sample vs Population

Sample vs Population

Population is the entire dataset such as the whole population of a country, **Sample** is subset of that population which is analyzed to make inferences

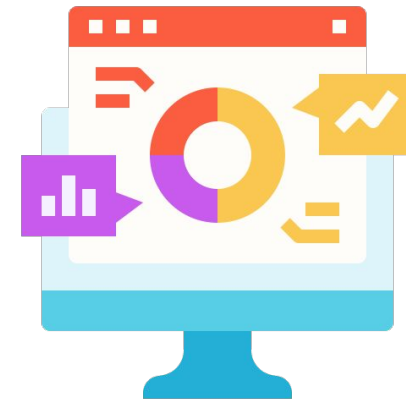


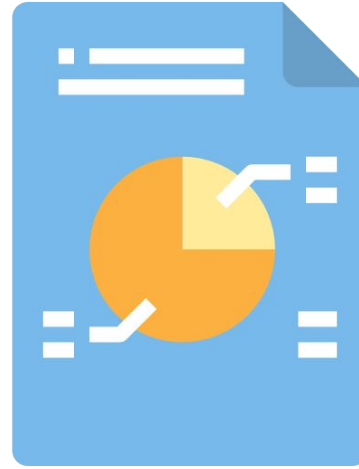


Random Sampling

Random Sampling

Random Sampling is the process of selecting a subset / sample from a population in such a way that every data point is equally likely to be included in the sample

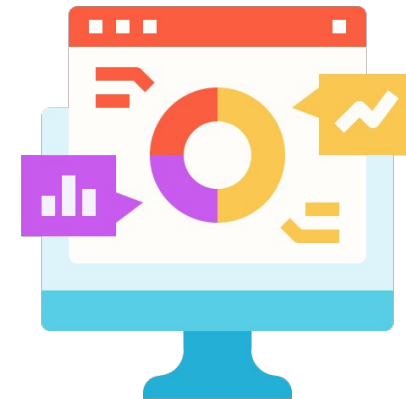
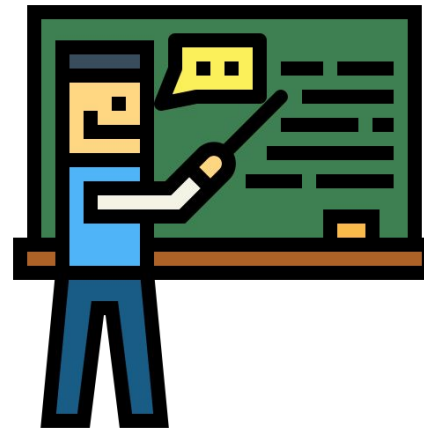


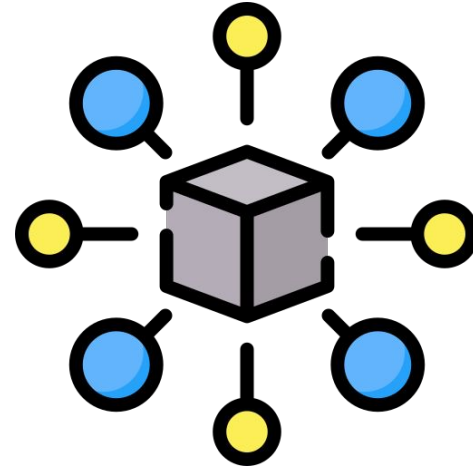


Stratified Sampling

Stratified Sampling

Stratified Sampling is the process of dividing your samples into layers or groups and then performing random sampling for each group





Central Tendencies

Central Tendency is used to indicate where does the middle or center of the distribution of our data lies

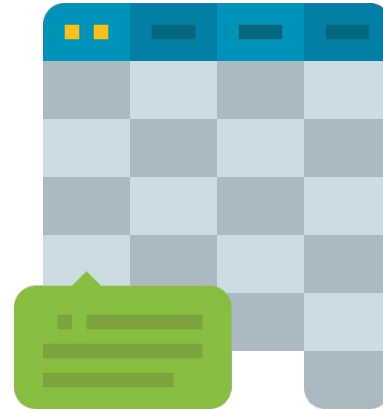


Mean

Mean is the average of the data. In simpler terms it's the sum of values divided by total number of values. It's represented by Greek letter Sigma



Mean



Mode

Mode is used to indicate the most frequent data point, in other words the one which occurs most number of times



Mode

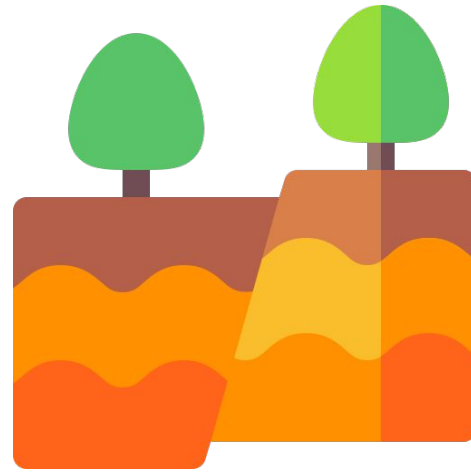


Median

Median is the middle of the data. If the data is arranged in ascending order then the data element which occurs right at the center is the median

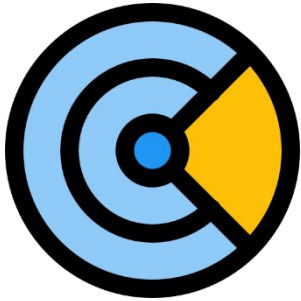


Median

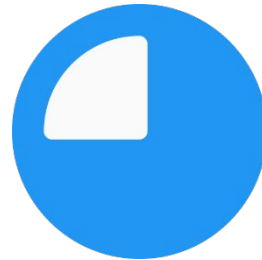


Variation

Variation in statistics is used to show how data is dispersed, or spread out. Several measures of variation are used in statistics.



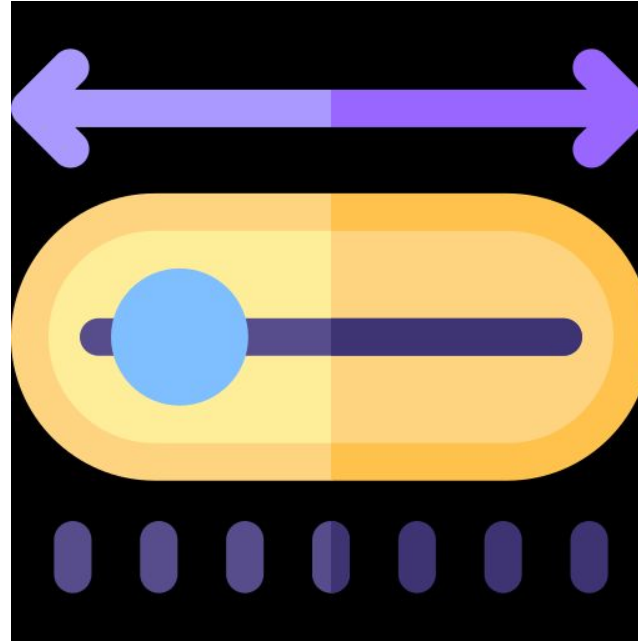
Range



Quartiles

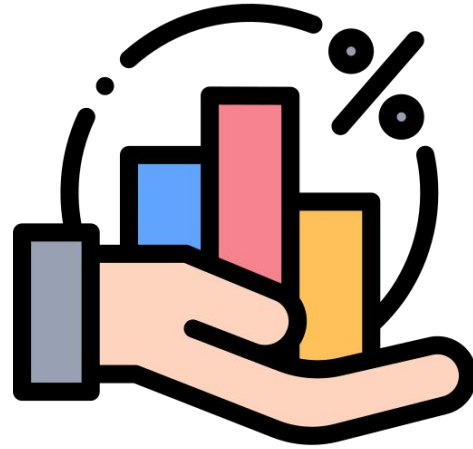


Variance



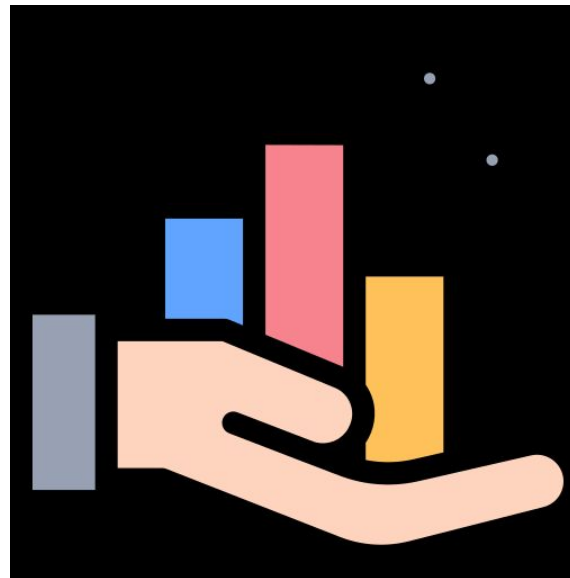
Range

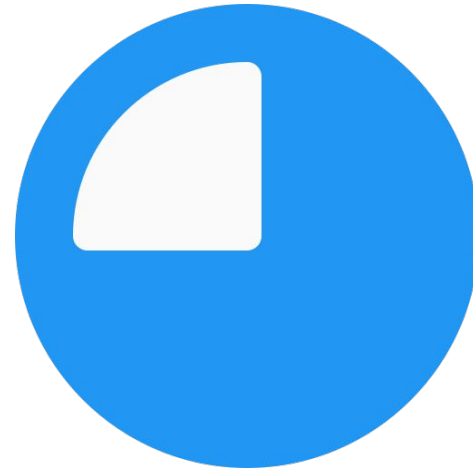
Range is the difference between the highest and the lowest values in our dataset. Range tells us the distance between the lowest and highest values in our data



Percentiles

Percentiles are scores that are used to describe a value below which some Observations fall.
E.g.: If X is at 70th Percentile it mean 70% of other data points from our sample are below X





Quartiles

Quartiles

Quartiles are used to break the data into 4 parts so as to better find the spread of data in a way that is less influenced by outliers.

Quartiles are expressed in percentiles. 1st Quartile is 25th Percentile, 2nd Quartile is 50th Percentile (Median) and 3rd Quartile is 75th Percentile



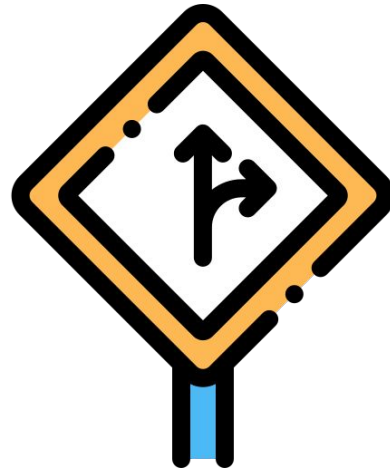


Interquartile Range (IQR)

Interquartile Range (IQR)

Interquartile Range (IQR) is the difference between the lower and upper quartile. This gives us a better idea of the range of data.

$$\begin{aligned} \text{Interquartile range} \\ = Q3 - Q1 \end{aligned}$$



Standard Variance and Standard Deviation

Standard Variance and Standard Deviation

Standard Variance measures how far a set of numbers are spread out from their average value.

Standard Deviation is used to express the magnitude by which the members of a group differ from the mean value for the group.

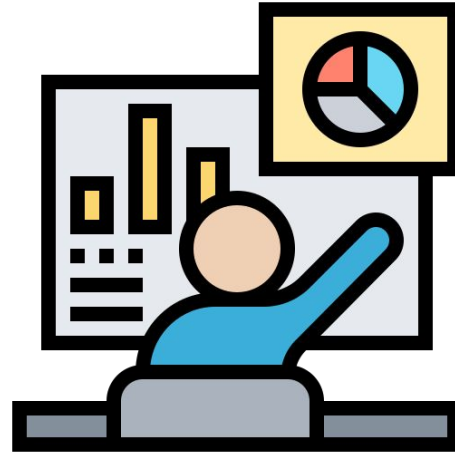
Standard Deviation is the square root of **Standard Variance**.

Sample Variance

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

Sample Standard Deviation

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$



Correlation

Correlation is a term that is a measure of the strength of a **linear relationship** between **two quantitative variables**

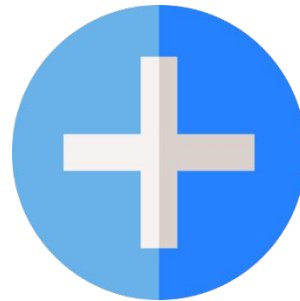
$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$



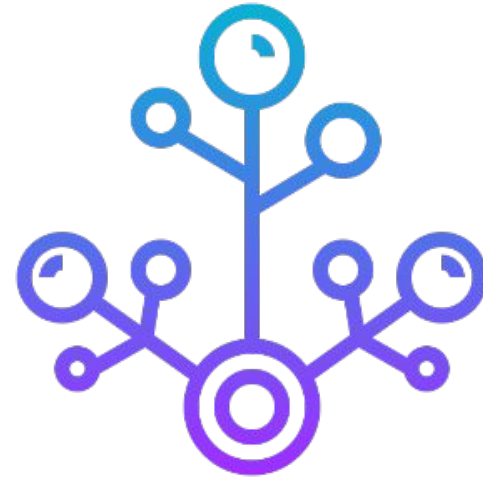
Positive Correlation

Positive Correlation

Positive Correlation is a term that is used to describe a **positive linear relationship** between **two quantitative variables**



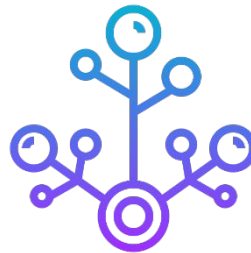
**Positive
Correlation**



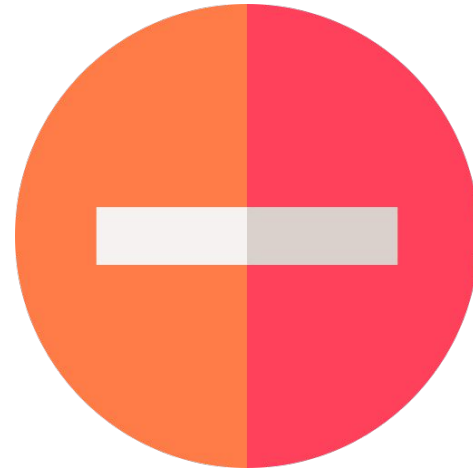
No Correlation

No Correlation

No Correlation is a term used to describe no **linear relationship** between **two quantitative variables**



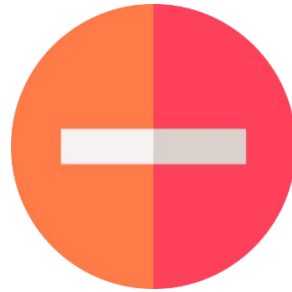
**No
Correlation**



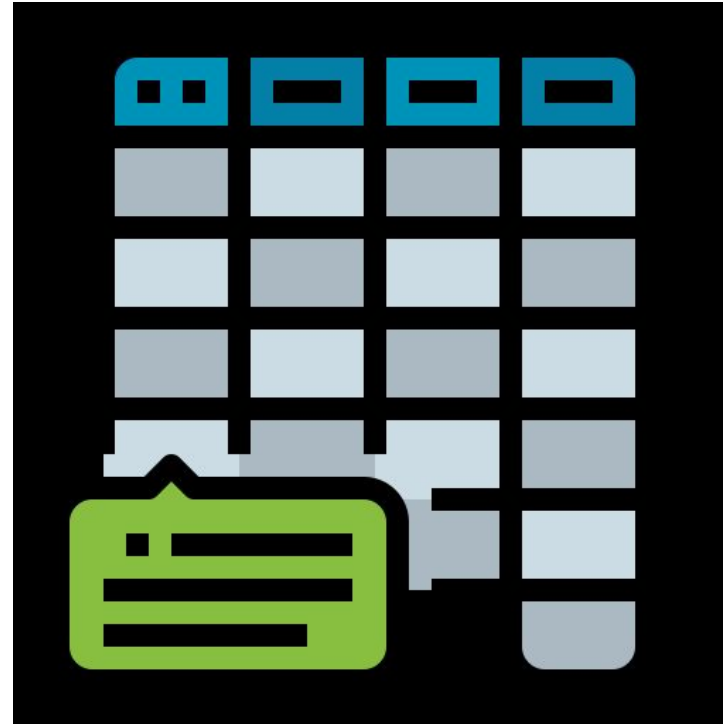
Negative Correlation

Negative Correlation

Negative Correlation is a term that is used to describe the strength of a **Negative linear relationship** between **two quantitative variables**



**Negative
Correlation**



Tables

A way of presenting statistical data through a systematic arrangement of the numbers describing some mass phenomenon or process

A statistical table may be regarded as representing a subject and predicate. The meaning of each number is indicated by the headings of the corresponding row and column.



Charts

A statistical graph or chart is defined as the pictorial representation of statistical data in graphical form. The statistical graphs are used to represent a set of data to make it easier to understand and interpret statistical information.

A statistical graph or chart is defined as the pictorial representation of statistical data in graphical form. The statistical graphs are used to represent a set of data to make it easier to understand and interpret statistical information.

Lets list down the types of charts

A statistical graph or chart is defined as the pictorial representation of statistical data in graphical form. The statistical graphs are used to represent a set of data to make it easier to understand and interpret statistical information.

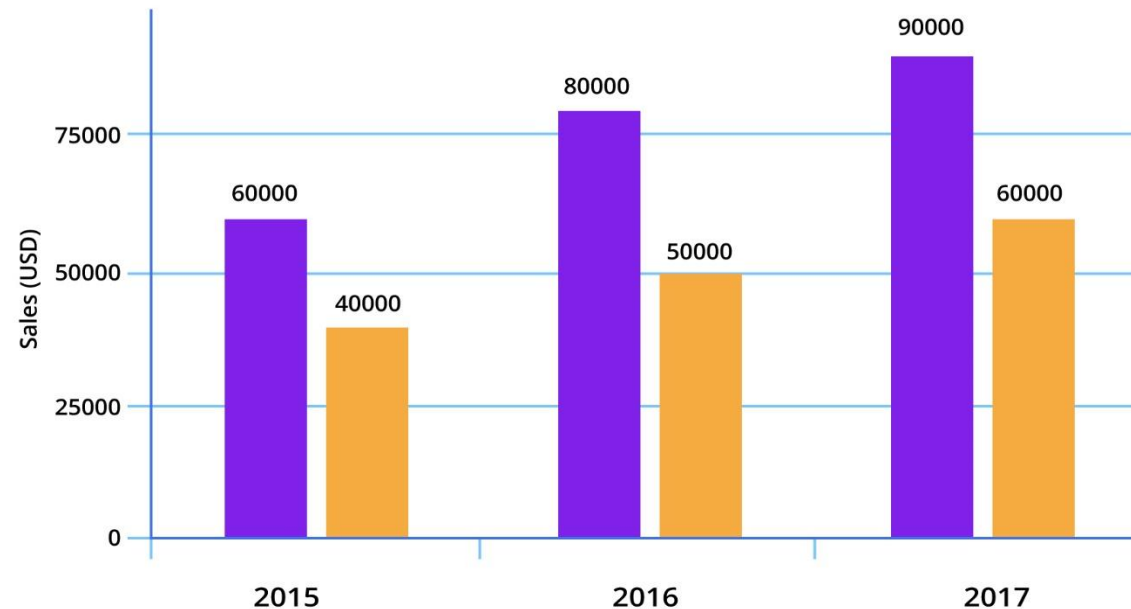
Lets list down the types of charts

Types Of Charts

- 1.Bar chart
- 2.Histogram
- 3.Pie chart
- 4.Box chart
- 5.Line Graph
- 6.Scatter plot

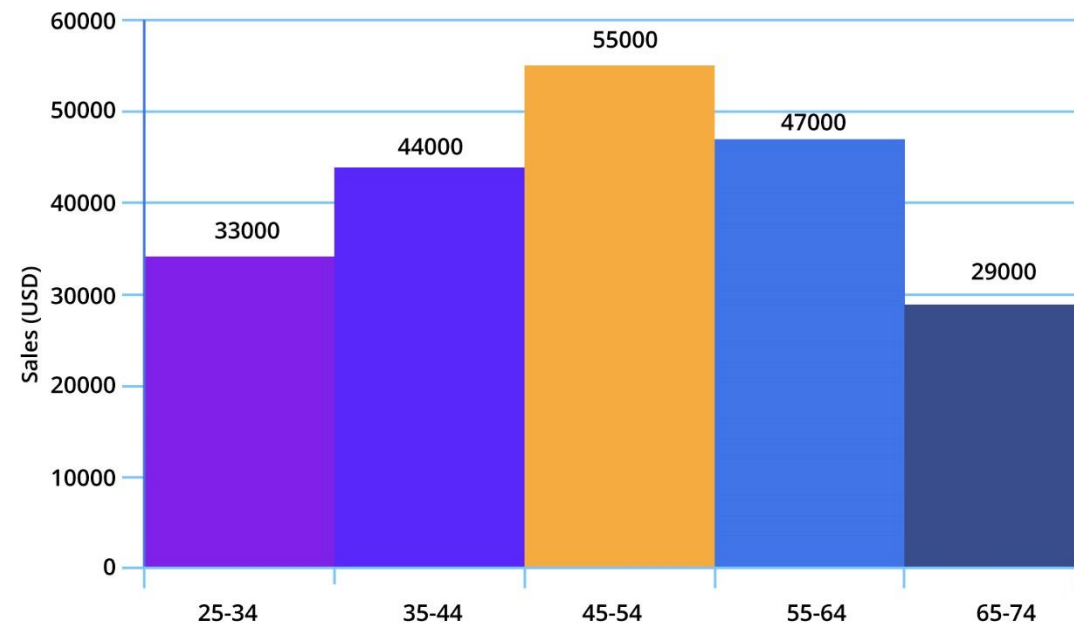
1. Bar chart

Bar charts are among the most frequently used chart types. As the name suggests a bar chart is composed of a series of bars illustrating a variable's development. Given that bar charts are such a common chart type, people are generally familiar with them and can understand them easily



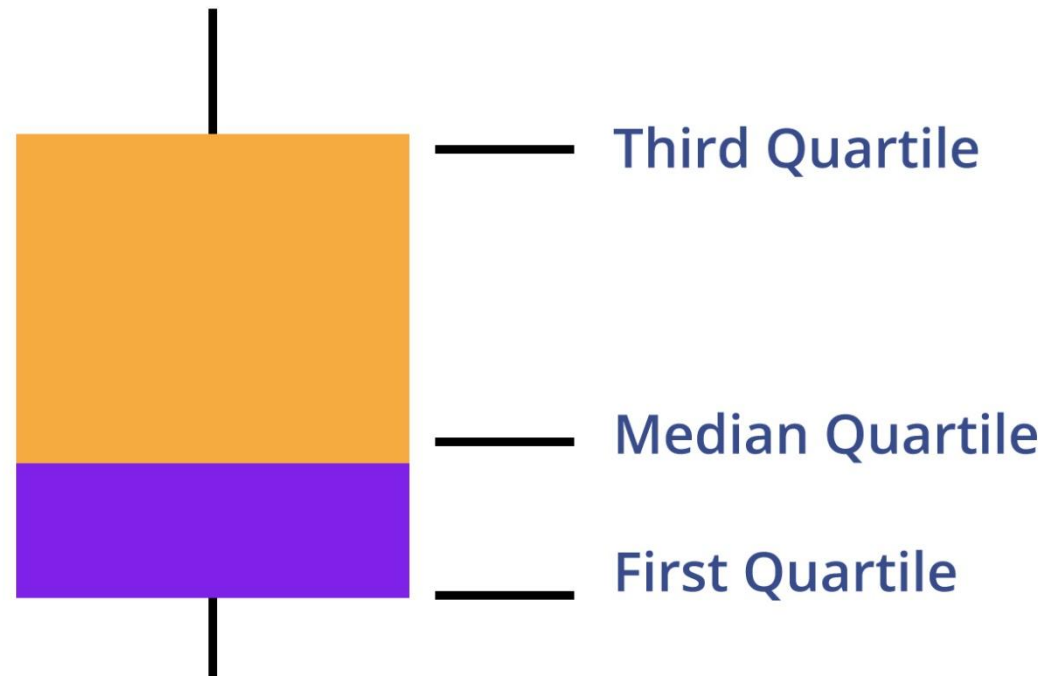
2. Histogram

A series of bins showing us the frequency of observations of a given variable. The definition of histogram charts is short and easy..



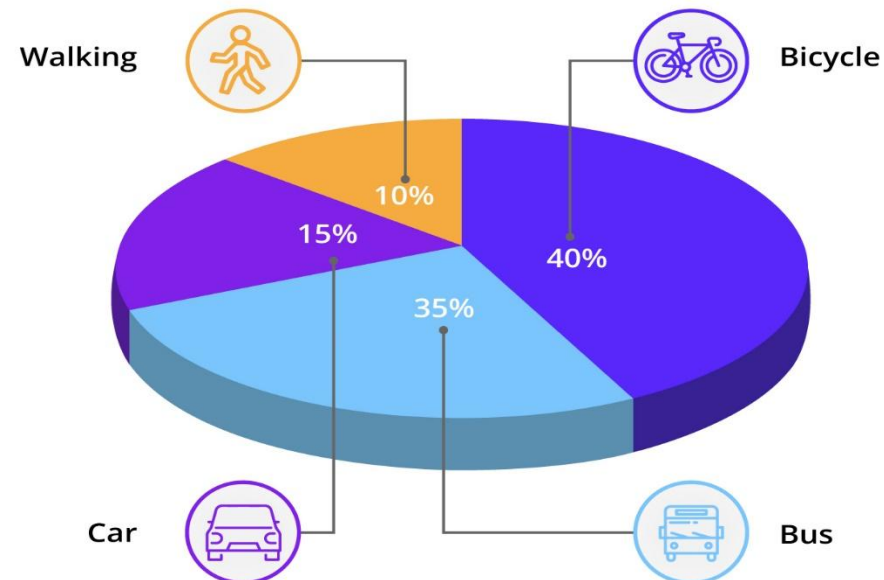
3. Box chart

Box plot, also called the box-and-whisker plot: a way to show the distribution of values based on the five-number summary: minimum, first quartile, median, third quartile, and maximum.



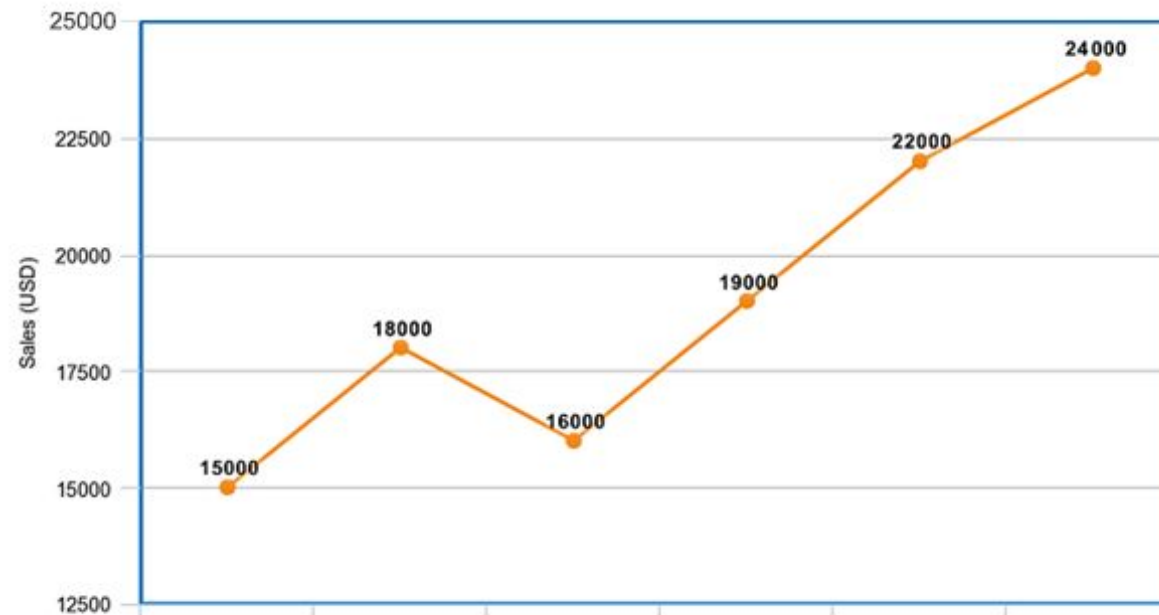
4. Pie chart

A pie chart is a circular graph divided into slices. The larger a slice is the bigger portion of the total quantity it represents.



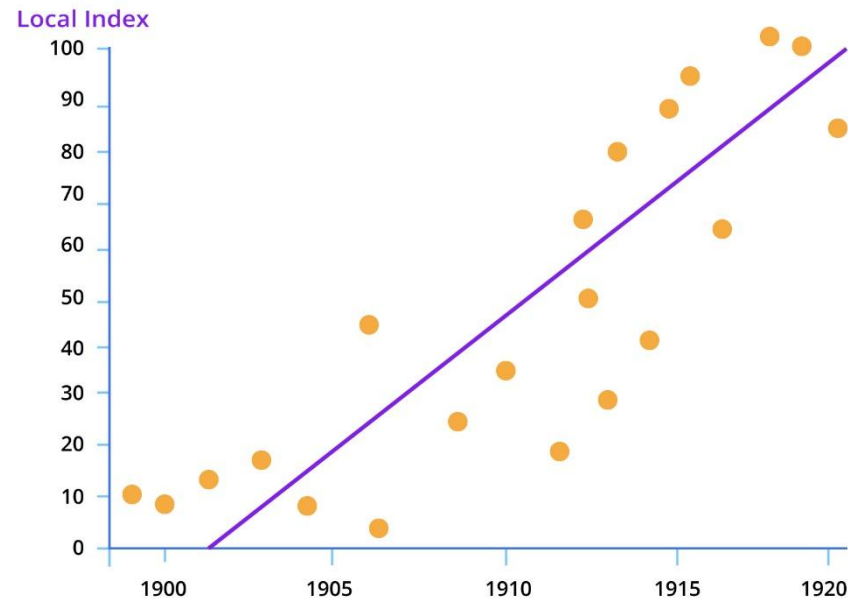
5. Line chart

A line chart is, as one can imagine, a line or multiple lines showing how single, or multiple variables develop over time. It is a great tool because we can easily highlight the magnitude of change of one or more variables over a period.



7. Scatter plot

A scatter plot is a type of chart that is often used in the fields of statistics and data science. It consists of multiple data points plotted across two axes. Each variable depicted in a scatter plot would have multiple observations. If a scatter plot includes more than two variables, then we would use different colors to signify that.





Probability

Probability defines the likelihood of occurrence of an event

Probability defines the likelihood of occurrence of an event

Probability can be defined as the ratio of the number of favorable outcomes to the total number of outcomes of an event.

Probability defines the likelihood of occurrence of an event

Probability can be defined as the ratio of the number of favorable outcomes to the total number of outcomes of an event.

Probability can be defined as the ratio of the number of favorable outcomes to the total number of outcomes of an event.

For an experiment having 'n' number of outcomes, the number of favorable outcomes can be denoted by x. The formula to calculate the probability of an event is as follows.

$$\text{Probability(Event)} = \frac{\text{Favorable Outcomes}}{\text{Total Outcomes}}$$



India: +91-7847955955

US: 1-800-216-8930 (TOLL FREE)



sales@intellipaate.com



24/7 Chat with Our Course Advisor