**Introduction**

Data analysis is a process used to analyze, clean, transform, and model data to discover useful information, draw conclusions, and support decision making. Data analysis has versatile and diverse approaches covering a variety of techniques under various names in different business, science and social science fields.  
Data integration is the pioneer of data analysis. Data analysis is closely related to data visualization and data distribution. The term data analysis is sometimes used as a synonym for data modeling.

**Data Analysis Process**

Data analysis is a process used to obtain raw data and turn them into useful information for users to make decisions. Data is collected first and then analyzed to answer questions, test hypotheses or reject theories. Data analysis has several stages. The stages are repetitive.

* **Data Requirements:-**

The data required as input to the analysis is selected based on the requirements of the analyst or the customers who will use the result of the analysis. Data can be numerical or categorical.

* **Data Collecting:-**

Data can be collected from a variety of sources. Data can be collected by surrounding sensors such as traffic cameras, satellites, recording devices. It is also possible to use interviews, downloads from online resources or documentation.

* **Data processing:-**

Data initially obtained must be processed or edited for analysis. For example, these can be placed in rows and columns in a table format for further analysis such as a spreadsheet or statistical software.

* **Data Cleaning:-**

Data may be incomplete, duplicate, or contain errors. The need for data cleaning will result from problems with obtaining and storing data. Data cleansing is the process of preventing and correcting these errors. Data cleansing tasks include record matching, detecting data inaccuracy, overall quality of existing data, deduplication, and column segmentation. Such data problems can also be detected through a variety of analytical techniques. Unusual amounts above or below certain threshold values can be examined. Quantitative data methods for outlier detection can be used to remove possible incorrectly entered data.

* **Exploratory Data Analysis (EDA):-**

Various mathematical formulas or models called algorithms can be applied to data to determine relationships between variables, such as correlation or causality. Inference statistics include techniques used to measure relationships between specific variables.  
Analysts can try to create models that describe the data to simplify the analysis and communicate results.