# DATA ANALYSIS STEP BY STEP

DATA, DATA, DATA.. there is a lot of data being gathered by all sort of channels and any organization's success, to an great extent, depends on how well they analyse such data and at what time.

**So what is the Data Analysis –**

"Data analysis is a process of

- inspecting,
- cleansing,
- transforming, and
- modeling data

with the goal of discovering **useful information, informing conclusions, hidden patterns, unknown correlations and supporting decision-making**. "

So in simple terms Data analysis is a process for obtaining raw data and converting it into information useful for decision-making by users. Data is collected and analyzed to answer questions, test hypotheses or disprove theories.

Lets understanding it from an example assume an IOT device located on a Toll Gate is collecting number of cars passing by that specific Toll Gate . It collects the registration number of the car and time of passing the traffic light. Now further assume, that there were 100,000 cars passed in a day.

Now !! what information this data can give, it depends on the user or management objective but there are following possibilities.

- No of cars passed during certain time slots of the day to understand traffic behavior
- Linkup the plate number of owner attributes and analyse further
    - What was the ratio of taxi to owned cars (if high taxi may be public transport can be planned)
    - How many were private cars and how many were company cars for business and personal.
    - If similar device and data is available at toll exit, can analyse the data to understand average speed by type of cars.

Lets Look at another example of an e-commerce business which sells 100s of products each day on its website. There will be atleast 3,000 transactions in a month and whopping 36,000 transactions in a year. Following can be achieved through data analytics.

1. What are most sold products on the platform.
2. What are most profitable products on platform and if they are amongst the most sold products.
3. Are there any products which are most sold but not so profitable?
4. Are there any products which are selling at loss?
5. Is there any product for whose sales is limited due to any controllable constraints?
6. What type of customers are buying product? What are opportunities !

Manish Gupta

7. What is the average price per order !
8. What products mostly brought together or by same customers?
   And much more depending on business development needs

Ok!! Now that we understand how data analysis can be powerful for business lets look at various steps and techniques performed for data analytics.

**Before we understand what is data analytics, Lets try to Understand the Decision making Process in brief related to data. For this part of book we will look into only decision making needs related to data. Strategic decision making process will be discussed another time.**

Decision Making points from a data set can be grouped into following

**Answer How**

- How Much or How Many of sales, products, customers, etc
- How long (delivery time, period sales, shelf life, sale time from purchase time)

**Answer WH Family**

- Who (customer profiles and segments),
- What (product segments),
- Where (Geo Segments) and
- When (date, timestamp month)

**Now that we understand what various possible decision-making points are, lets look at the process of making such decisions starting from collection till interpretation of data into information.**

**Key Steps of Data Analysis**

a. **Assess Data requirements**

The data is necessary as inputs to the analysis, which is specified based upon the requirements of those directing the analysis or customers (who will use the finished product of the analysis).

Following must be kept in mind while assessing data requirements.

1. Specific data variables (decision points) which can be
   a. numerical quantity of products, revenue, costing or
   b. categorical i.e category of products, regions, type of customer

2. Data can be collected to the basic transaction level
3. Data must be segmented and referenced to common sets for example customer profile can be collected separately and should be referenced in transactions so that repetitive data can be reduced.

Manish Gupta

### b. Data collection

Data is collected from a variety of sources. The requirements may be communicated by analysts to custodians of the data, such as information technology personnel within an organization.

The data may also be collected from

1. sensors in the environment, such as traffic cameras, satellites, recording devices, etc.
2. obtained through interviews,
3. downloads from online sources or reading documentation.

### c. Data processing

Data initially obtained must be processed or organised for analysis. For instance, these may involve placing data into rows and columns in a table format (i.e., structured data) for further analysis, such as within a spreadsheet or statistical software.

### d. Data cleaning

Once processed and organized, the data may be

- incomplete,
- contain duplicates, or
- contain errors.

Data cleaning is the process of preventing and correcting these errors.

Following processes and checks are usually applied while cleaning data.

a. record matching,
b. duplication, and
c. column segmentation.
d. Such data problems can also be identified through a variety of analytical techniques.
e. For example, with financial information, the totals for revenue can be matched against total revenue reported in financial statements.
f. Textual data spell checkers can be used to lessen the amount of mistyped words, but it is harder to tell if the words themselves are correct.

### e. Perform Data Analysis methodologies

**There are two main type of Data Analysis methods.**

**Quantitative Analysis**

1. **Descriptive Analysis -** Describe the main features of a large collection of data.
2. **Confirmatory Analysis** - Confirm or negate a hypothesis.
3. **Exploratory Analysis** - Find previously unknown relationships in the data.

Manish Gupta

4. **Inferential Analysis -** Use a smaller sample of data to learn something about a bigger population.
5. **Causal Analysis -** Find out what happens to one variable when you change another.

**Event Series Analysis**

facilitate searching for patterns across multiple event records and datasets

**f. Communication & Evaluation**

Data visualization to understand the results of a data analysis.

Once the data is analyzed, it may be reported in many formats to the users of the analysis to support their requirements.

**The users may have feedback, which results in additional analysis. As such, much of the analytical cycle is iterative.**

Now that we get the brief understanding of data analysis concepts and steps, we will continue to learn detailed methodologies and techniques using Excel.



Manish Gupta