

DATA ANALYSIS :-

- Process of gathering, cleaning, analysing, mining data., interpreting results, report findings.
- Find correlations & patterns b/w data points.
Helps understand past performance, & validate course of action.

4 Primary types of Data Analysis:-

1) Descriptive Analysis :-

- Helps analyse what happened over a given period of time, by summarising past data.
Ex:- Tracking past performance by cash flow analysis.

2) Diagnostic Analysis :-

- Helps analyse Why did it happen? for any event.
- Takes insight from Descriptive Analysis, to find more details about the outcome.
Ex:- Knowing cause of sudden increase in sales in a place with no marketing.

3) Predictive Analysis :-

- ⇒ Helps analyse "What will happen next?"
- ⇒ Historical Data & Trends are used in prediction.

Ex:- Used in risk assessment and sales forecast.

4) Prescriptive Analysis :-

- ⇒ Helps analyse "What should be done about it?"
- ⇒ Analyses past decisions and events to estimate the likelihood of different outcomes.
- ⇒ Ex:- Self-driving cars: analyse environment to make decisions about speed, lane-change, route, etc.

Key Steps in Data Analysis Process:-

- 1) Understanding the Problem & Desired Result:
 - ⇒ Problem needs to be solved, & target to achieve should be clearly understood.
 - ⇒ Know where we are, and where we want to be, precisely, is important before beginning of data analysis.

2) Setting a clear metric :-

⇒ Deciding what exactly should be measured.

Ex:- Whether products sold per quarter, or products sold in festive season should be measured.

3) Gathering Data :-

⇒ Identifying the data we need, the potential data sources to pull data from, and the best tools / techniques to gather data.

4) Cleaning Data :-

⇒ Fixing quality issues in data (that may lead to inaccuracy)

⇒ Standardizing data from multiple sources.

⇒ Data cleaning is done for missing/incomplete values and outliers.

Ex:- In demographic data, a single data, say with 150 yrs age is an outlier.

5) Analyzing & Mining Data :-

⇒ Manipulate data to identify trends, recognise correlations, find patterns and variations.

6) Interpreting Results :-

⇒ Evaluating defendability of results.

⇒ Understanding circumstances for which the analysis may not hold true.

7) Presenting findings/results :-

⇒ Presenting the results in clear and impactful way is important.

⇒ Dashboards, charts, reports, maps, case studies—ways of presenting data.

④ The Data Analysis we do naturally?

⑤ In the morning, we see news & weather, & we know what the diurnal temperature would be. Based on it, we decide, whether to take bath, wear what type of clothes, etc.

⇒ Before going to some place on tour, we see the weather at that place, and take action as to what type of clothes & things we should pack and take with ourselves.

Data Analytics vs. Data Analysis

The terms **Data Analysis** and **Data Analytics** are often used interchangeably, including in this course.

However it is important to note that there is a subtle difference between the terms and meaning of the words *Analysis* and *Analytics*. In fact some people go far as saying that these terms mean different things and should not be used interchangeably. Yes, there is a technical difference...

The dictionary meanings are:

Analysis - detailed examination of the elements or structure of something

Analytics - the systematic computational analysis of data or statistics

Analysis can be done without numbers or data, such as business analysis psycho analysis, etc. Whereas *Analytics*, even when used without the prefix "Data", almost invariably implies use of data for performing numerical manipulation and inference.

Some experts even say that *Data Analysis* is based on inferences based on historical data whereas *Data Analytics* is for predicting future performance. The design team of this course does not subscribe to this view, and you will see why later in the course as you become familiar with the terms like *predictive analytics*, *prescriptive analytics*, etc.

Responsibilities of a Data Analyst:

- 1) Acquiring Data
- 2) Creating queries to extract required data:-
from databases & data collec. systems.
- 3) Filtering, cleaning, standardisation & reorganising data.
- 4) Use statistical tools to interpret datasets.
- 5) Use statistical techniques to identify patterns & correlations in data.
- 6) Preparing reports and charts – that convey the trends and patterns.
- 7) Create app. documentation to demonstrate the data analysis process being conducted.

•) Technical Skills :-

- ⇒ MS Excel or Google Sheets.
- ⇒ Proficiency in statistical analysis & visualization tools, and softwares like:- IBM Cognos, IBM SPSS, Oracle Visual Analyser, MS Power BI & Tableau.
- ⇒ Atleast 1 prog. lang. proficiency :- Python, R, C++, Java, MATLAB.
- ⇒ SQL, Relational & non-relational DB
- ⇒ Ability to extract data from repositories :- Data Marts, Data Warehouses, Data Lakes & Data Pipelines.
- ⇒ Familiarity with Big Data processing tools:- Hadoop, Hive, Spark.

•) Functional Skills :-

- ⇒ Proficiency in Statistic: Analyze & validate data, identify logical errors.
- ⇒ Analytical Skills: Research & interpret data, make forecasts.
- ⇒ Problem-Solving Skills: Come up with solutions.
- ⇒ Probing Skills: Identify & define the problem statement & desired outcome in standard way.
- ⇒ Data Visualisation Skills: To effectively present the data based on target audience, type of data, type of analysis.
- ⇒ Project Management Skills: Manage the process, people, dependencies & timeline of analysis.

➲ Soft-Skills :-

- ➲ Compelling & convincing storytelling ability.
- ➲ Work collaboratively with business and teams.
- ➲ Efficient communication, presentation ability.
- ➲ Discover challenges & facing them.
- ➲ Induction of future based on pattern recognition & past experience.

➲ Qualities:-

- 1) Natural curiosity, 2) Attention to detail,
- 3) Enjoy working with computers.
- 4) Love numbers and info

➲ Applications of Data Analysis :-

- i) What data should be shared in TV advertisements of a company.
- ii) A diabetes patient analysing his/her diet and medicines based on his/her sugar-test reports.
- iii) Financial reporting.
- iv) Satellite geolocation data to predict development & urbanisation of areas.