



Module-1

Overview of data and it's overview

What is Data?

- **Data** is Plural word and its singular form is Datum.
- **Datum** is a Latin word which means something given.
- **Data is a collection of** –Text, Number, Dates, Symbol, Images

Types of Data

- **Structured Data**- In Excel (arranged in Row & Column), Easy to locate
- **Unstructured Data**- Difficult to locate, audio video, mail, Images

Data Formats

- CSV
- TXT
- JSON
- XLSX

Analytics in Market

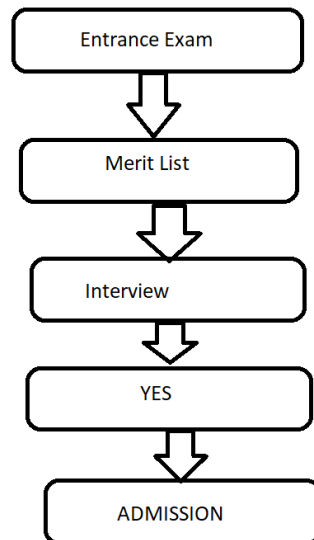
- Audio Analysis
- Text Analysis
- Video Analysis
- Number Analysis

What is Data Science

Data science is also science to extract hidden values from any data by applying scientific, statistical, mathematical (Algorithm) and computing techniques on it.

What is Algorithm

Algorithm is a step-by-step process.



Ex- Student wants to take admission in MBA collage through CAT
Then student needs to follow process like-

- Entrance exam
- Merit list
- Interview
- Admission

About Data Scientist's Skills

- Machine Learning
- Data Visualization
- Statistical Modeling & Analysis
- Data Wrangling
- Communication
- Computer Programming & database

About Data Analyst's Skills-

- MySQL

- Mongo DB
- Tableau
- Power BI
- Advanced Excel

What is Data Analytics

Analytics is the scientific process of discovering and communicating the meaningful patterns, which can be found in data.

What is Database

- A Database is an organized, structured collection of similar information.
- A Database can be stored on paper (manual) or in a computer(electronic).
- Data is organized in a database as file, record and fields.

DBMS-

- Database Management System (also known as DBMS) is a software for storing and retrieving users' data by considering appropriate security measures. It allows users to create their own databases as per their requirement.
- It consists of a group of programs which manipulate the database and provide an interface between the database. It includes the user of the database and other application programs.
- The DBMS accepts the request for data from an application and instructs the operating system to provide the specific data. In large systems, a DBMS helps users and other third-party software to store and retrieve data.

DBMS Softwares

- MySQL
- Microsoft Access
- Oracle
- PostgreSQL
- dBASE
- FoxPro
- SQLite
- IBM DB2
- LibreOffice Base
- MariaDB
- Microsoft SQL Server et

Data Set

- A dataset (or data set) is a collection of data. A dataset is organized into some type of data structure.
- Dataset might contain a collection of business data (names, salaries, contact information, sales figures, and so forth). Several characteristics define a dataset's structure and properties.

What is CLOUD

It's cheap and best technique to store our data like Document, Photo, File to the internet for the further use.

EX-

- iCloud- 5GB Free Storage
- Dropbox- 2GB Free Storage
- Google Drive- 15GB Free Storage
- Amazon- 5GB Free Storage
- Sky Drive- 7GB Free Storage

Steps of Data Science

- 1) **Acquisition** -Where to get Data, where to store data
- 2) **Exploration and Understanding** - How much information you can get from the Data
- 3) **Mugging, Wrangling, and Manipulation** - Data is almost never in the needed form for the desired analysis.
- 4) **Analysis and Modelling** –which method, Algorithm will we use
- 5) **Communicating and Operationalizing** – How will we present data

1) ACQUISITION

- Data acquisition has been understood as the process of gathering, filtering, and cleaning data before the data is put in a data warehouse or any other storage solution.
- Some sources from where you can acquire data.
 - Website data
 - Purchase data
 - Smart tv data
 - Social data
 - Offline data
 - Mobile data
 - 3rd Party data

2) EXPLORATION

- The second step is to an understanding of the data that how will use it
- How much information you can get from data
- In order to better explain the essence of the data, data exploration refers to the initial step of data processing in which data analysts use data visualization and mathematical methods to define dataset characterizations, such as scale, quantity, and precision.

3) MUGGING, WRANGLING

- Data is almost never in the needed form for the desired analysis.
- It is the initial step of preprocessing and refining raw data into content or formats better suited for analysis.

4) Analysis and Modeling

- Data modeling is a collection of methods and tools used to explain and analyze how data can be processed, modified, and maintained, and maintained be entity.
- Data analysis allows the analyst to extrapolate insights from huge bulks of data, it is done with the help of statistical and machine learning techniques.

5) COMMUNICATION (Data Visualization)

- At the end of the pipeline, we need to give the data in a compelling form and structure, sometimes to ourselves to inform the next iteration, and sometimes to a completely different audience.
- The data products produced can be a simple one-off report or a scalable web product that will be used interactively by millions.