**Data Analysis**

Extracting the meaningful knowledge from the data you can gather.

It has two components

1 Domain Knowledge : Expertise in a field

2 Technology

ETL

Eg Informatica

Extract Transfer Load

1 Extracting Data from Data Sources

2 Transform the data

3 Load the data in databases

Business Intelligence (BI)

In real time it is used to make reports

Power BI , Tablu are some tools

Data Scientist

A person who extracts the data and analyse it, is called a data analyst ( Tech part me expert )

A data scientist is the one who performs

Data acquisition : Gathering data

Data preparation : Transform data

Data Mining : Extracting meaningful insight from data

Data Modelling : Building models based on analysis using ML Techniques

Model Maintenance : Updating the data

**Data has two types**

**1 Structured data** : Human readable format, highly organized , Tabular format

Eg Excel,RDBMS systems

**2 Unstructured data** : Unorganized,unformatted,Can't be organized,not in Tabular format

Eg Emails,Social Media Posts, Documents,Images,Audio

**Examples of RDBMS Systems to store the organized data**

My SQL

Sql Server

Postgres

Oracle

Data is again of two Categories

**1 Categorical Data (Qualitative)**

Non-Numeric data , Textual data

After doin mathematical calculations of it makes no meaningful data

Eg Gender,Marital Status, Vehicle, Educational Levels,Residential city

**2 Numerical Data ( Quantitative)**

Numeric Data that can be measured or counted

Jispr kuch mathematical aggregations perform ho

Sakte h and it makes meaning

Discrete Numerical Data : Exact Whole numbers

Continuous Numerical Data : In decimals

Categorical Data Measurement levels

**Nominal :** Type of Data that represents categories or labels vd no inherent order or ranking among them (Sorted Data)

Eg Colors , Fruit Types, Vehicle Makes,Days etc

**Ordinal :** Vd specific ordering or ranking but no equal spacing (Unsorted Data)

Eg Job Status,Economic Status ,Health,Ratings etc

Numerical Data Measurement Levels

**Interval :** Equal Spacing

EG Temperature,IQ scores Year on the Gregorian Calendar, Latitude and Longitude,Time Duration, Weight in pounds, GPA score , Stock Price

**Ratio** : True zero hoga

Eg : Height , Weight (kg), Distance(meters,miles Km) ,Age(years),Income(dollars),Amount Of Dollars(moles),Reaction time (sec)

TEMPORAL DATA(TIME-SERIES DATA)

Data is captured over the period of time with time stamped

Eg Stock price,Daily Temperature,Hourly Websites traffic data,Heart rate

SPATIAL DATA(GEOSPATIAL DATA)

Data with representation objects and phenomenon in physical state

Eg GPS, Satellite images, Road Websites