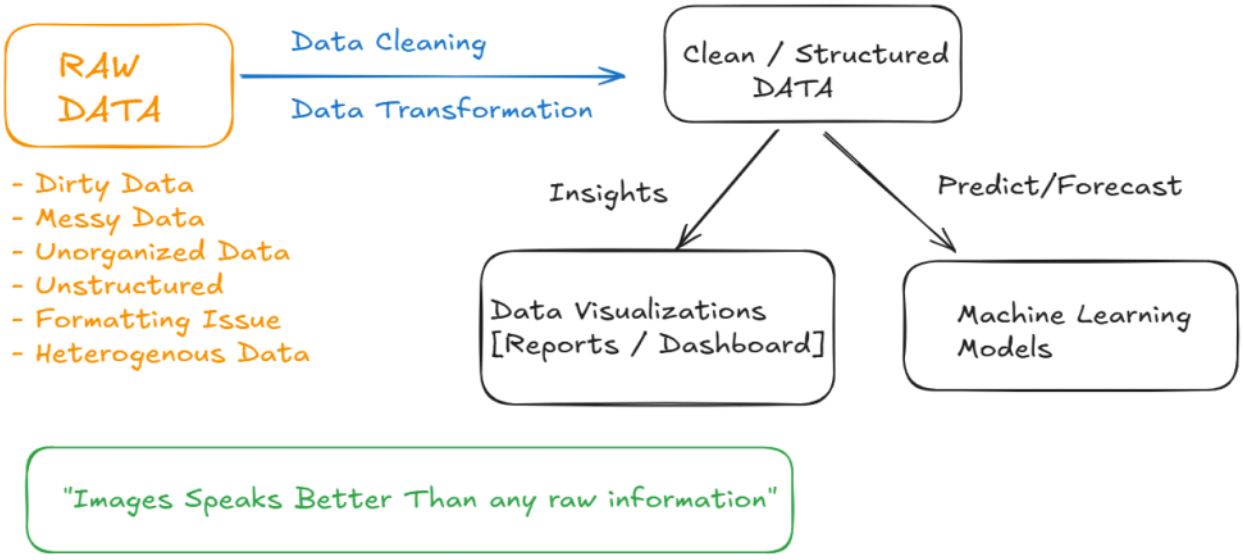
📊 Introduction & Lifecycle in Data Analytics

- ✅ Understand what data analytics is and how it impacts our world.
- ✅ Learn about career opportunities in the data domain.
- ✅ Distinguish between types of data and types of data analytics.
- ✅ Understand the lifecycle of data analysis through relatable examples
- ✅ Learn data cleaning methods
- ✅ Identify tools used at each stage of the data lifecycle

Guidelines To Follow

- Session needs to be very interactive.
- It should be having high energy.
- Unmute yourself whenever needed.
- 0/1 - Binary Answer
- 0 [Not Clear , Repeat, Wait].
- 1 [Done, Clear, We can move ahead].
- Recorded Format
- Notes with provided to you on Platform after T + 5 hours.
- Feedback Form

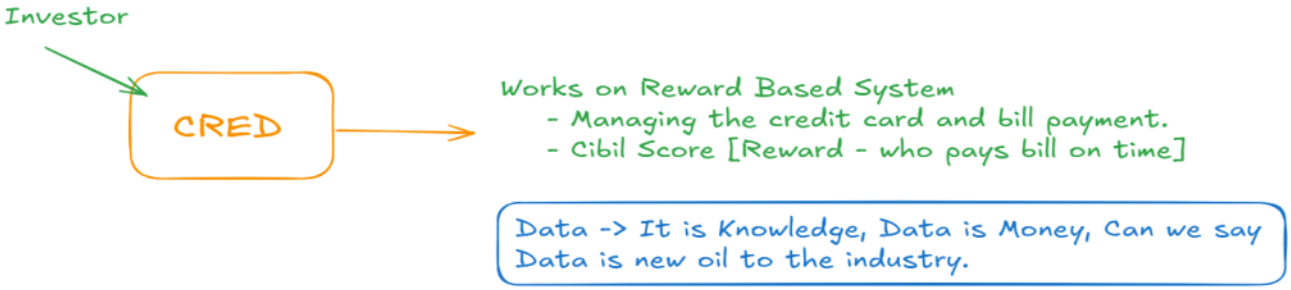
Data - Raw facts or figures -> Transformed -> Insights.



"Images Speaks Better Than any raw information"

What is Data Analytics

- Data Analyst is collecting, cleaning, analyzing, interpretating the raw data into smart decision.



Real World Examples.

- E-commerce Recommendation:
  - Search History
  - Purchase History
  - User Behaviour
- Amazon Prime, Netflix, Hot star:
  - Watch History
  - Recommendation on search.
- Uber / Ola finding the nearest driver. [Google Map API]
- Banking System - Cibil Score, Risk Analysis on Loan.

Tools & Technologies

Analytical Mindset

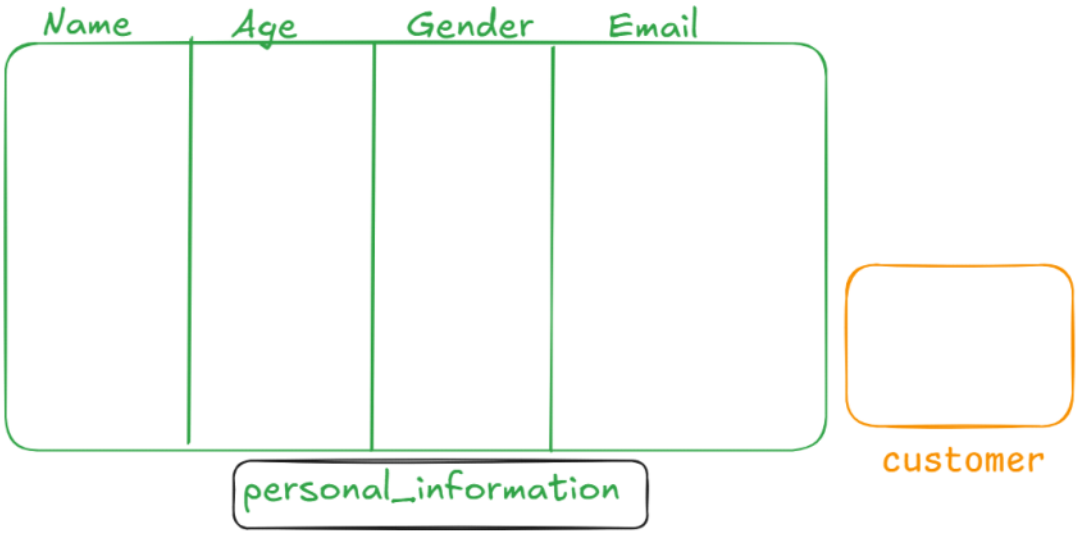
- EXCEL, Google Sheet -> Handling Spreadsheet.
- Power BI [Microsoft], Quick Sight[Amazon], Looker Data Studio [Google], Tableau[Salesforce] -> Data Visualization

Power Query Editor

- MySQL, SQLite, PostgreSQL, Oracle, SQL Server -> Structured Dataset [Database]
- Python / R Language - Programming Language

Why Python? → Libraries [Numpy , Pandas] - Data Transformation
[Matplotlib, plotly, seaborn] - Data Visualization

- It is easy to understand
- It is having highest number of libraries that are specially design for Data Industry.
- It is having an awesome community.



user defined

All Columns

```
SELECT * FROM personal_infromation;
```

Read

user_defined

```
SELECT Name,Age FROM personal_infromation;
```

Specific Column Name

Syntax

```
SELECT * FROM customer;
```

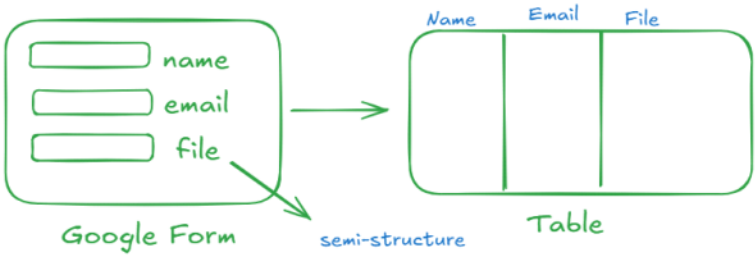
Syntax - It is a predefined Keywords.

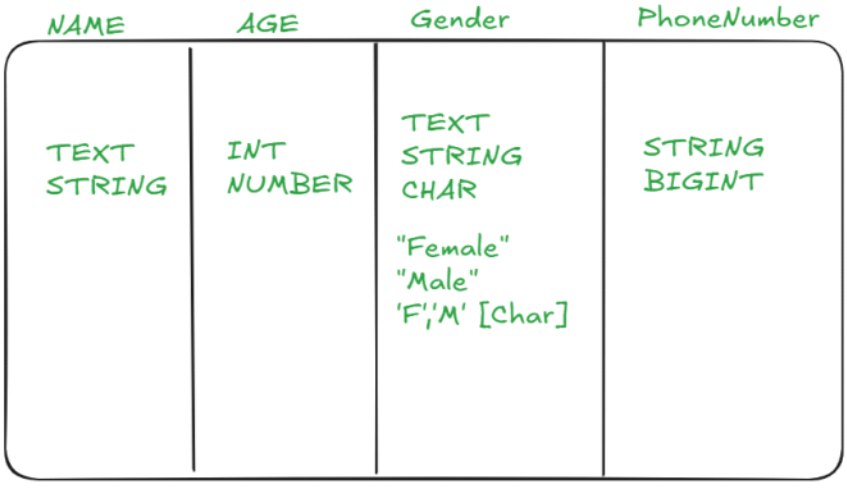
MySQL

- SELECT
- JOIN
- WHERE
- FROM
- ORDER BY
- UPDATE
- CASE
- GROUP BY
- HAVING
- LEFT/RIGHT
- CROSS JOIN

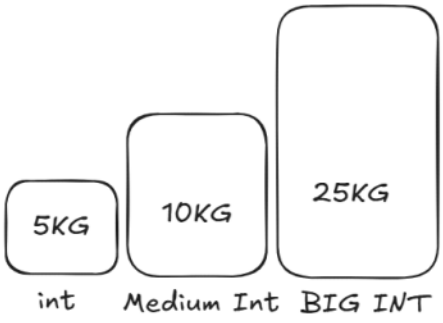
Types of Data By Structure

- Structured Data
- Semi - Structured Data
- Unstructured Data





Structured Data - Organized



Data types - helping us to store a numeric value.

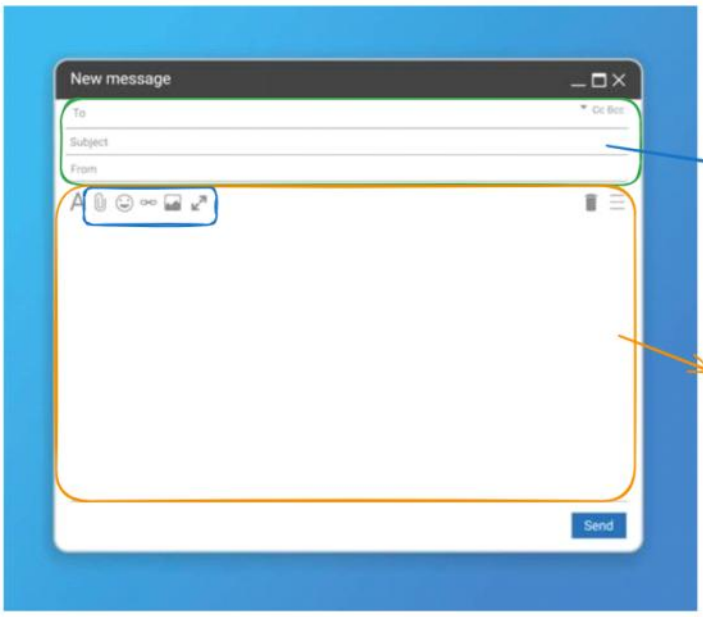
Homogeneous - [1,2,3,4,5,6,7,8,9,10,11] - having same data type

Heterogeneous - [1,2,3,1.99,2.11,'K',"Coding",True] - having same data type

char - 'a', 'b', 'c'
String - Combination of characters to make a meaningful words.

Semi-Structured - Mix of Structured + Unstructured.

Gmail



Unstructured → No fixed structure/format , hard to process

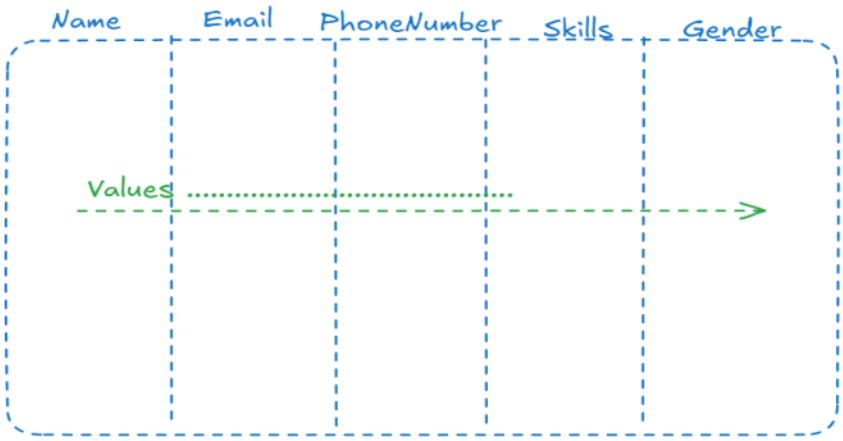
Social Media - Images, Post, Videos, Voices

What happen when we extract the data from internet/web?

- HTML , JSON , XML

Parsing

```
{   Key                Value
    name : "Krishna Madan"
    Email : "krishnamadan77@gmail.com"
    Phone Number: XXXXXXX328
    Skills : [Python, SQL, Power BI, Excel]
    Gender : 'M'
    Address: 'XYZ Street - Pin Code'
    Country: India
}
```



Types of Data Analytics [DDPP]

- Descriptive Analytics [What Happened?]

  - Analysis on Past Performance
  - Historical Data.

- Diagnostic Analytics [Why did it Happened?]

  - Find the root cause of the issue
  - Star Bucks

-Predictive Analytics [What will Happened?]

  - Forecasting future outcomes by historical data
  - Amazon
  - Banking : Detecting Frauds

-Prescriptive Analytics [What should we do?]

  - Recommend actions based on Analysis.
  - If sales is declining from the last few years?
    - Prescriptive Analytics says :
      - Launch a new product.
      - Change the marketing strategy.
      - Give Loyalty offers.
      - Take customer Feedbacks
      - Do influencer marketing.

"Knowing how to analyze data = Power 🔥 "

LifeCycle of Data Analysis - 7 Stages

1. Objective

Define the problem Clearly.

- Where are the high income People and luxury Homes available in Goa.

2. Get the relevant Data: [Collecting the data from various sources]

- Area Name [Cover Maximum area to avoid biased insights]
- Coordinates - Longitude , Latitude.
- Houses Prices
- Average Income
- Market Area Nearby
- Facilities / Amenities

3. Understanding the Data

- Check The Data Types of each column
- Do we cover All region?
- Are all columns relevant?

4. Data Cleaning/ EDA [Exploratory Data Analysis]

- Magic and hard works happens...
- 60% - 70 % of the time spend here.

A. Merging Tables.

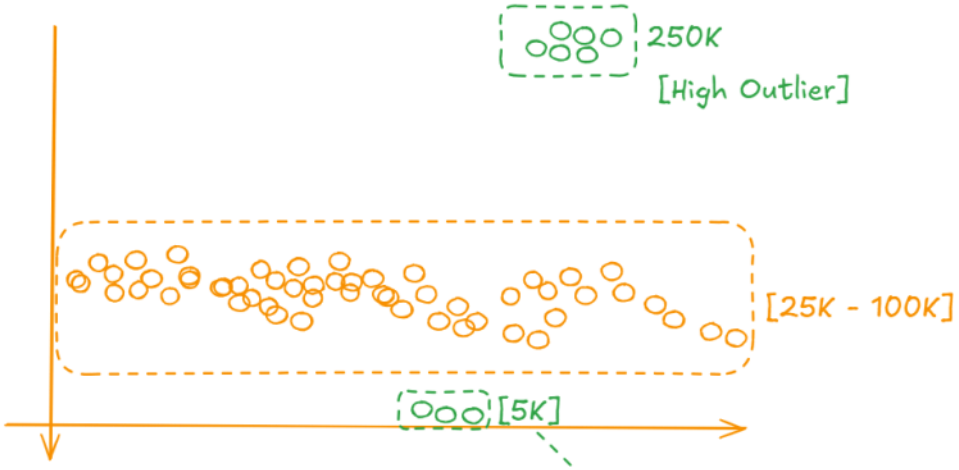
B. Handling the Data Types

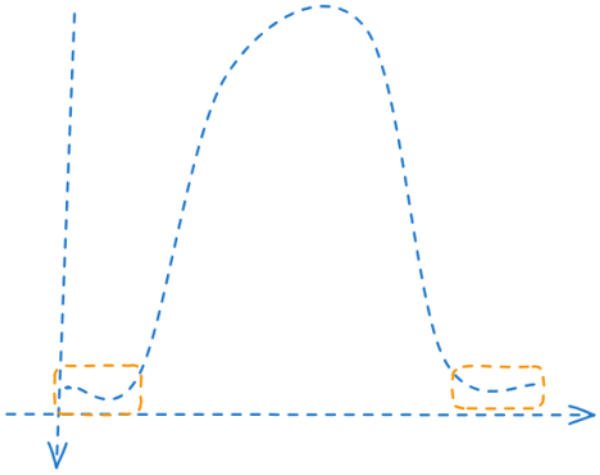
- Date Columns -> Data Format [dd/mm/yyyy] India/UK
  -> Date Format [mm/dd/yyyy] US
- Currency -> $ -> USA. Euro-> Europe, India -> ₹ Rupees

C. Handling the Missing Value

- Drop the Rows [if it is very less in number]
- Categorical Data -> *NA / Unknown*
- Numerical Data -> [Mean,Median,Mode]

D. Handling the Outliers





E. Ensuring Consistency

- Goa , goa, GoA -> 'Goa' [3 different Filters] [Fix Them]
- $50K , ₹500000 -> Change it only one currency.
- Consistency with the date Format.

F. Removing Unnecessary Columns

Extra Columns = Extra Confusion

- Remove anything that doesn't serve the goal.

G. Removing Duplicates

- Make sure to remove duplicated otherwise your final visuals may show inflation.

5.Data Analysis

- Do the income is directly proportional to the high housing price.

6.Visualization

- Turn numbers into visuals.
- Pie Chart - Composition
- Bar Chart / Column Chart - Comparison
- Cards / Matrix to show Financials.
- Slicer - To apply filters.
- Map Visuals - To show insights related to territories.
- And so on.....

7.Communication

- Tell a story backed with the data.
- Clear Visuals
- Less Jargons
- Actionable insights
- Complete context [avoid missing details].

HomeWork

- Watch Money Ball in Netflix.
- Watch at least few episode of Shark Tank.
  [KPI - Key Performing Indicators].