

**CMA INTER**

**GROUP-2**

**BDA**

**FAST TRACK BOOK**

**(SYLLABUS-2022)**

*“WINNERS ARE NOT THOSE WHO NEVER FAIL BUT THOSE WHO NEVER QUIT.”*

*- DR. APJ ABDUL KALAM*

**TARGET**  
Academy

---

---

**BDA BOOSTER BOOK**

**(GROUP – 02)**

---

---

**CONTACT: 90595 29540, 90106 76997.**

**INDEX(CMA INTER GROUP-2 NEW SYLLABUS)**  
**BUSINESS DATA ANALYTICS (PAGE 1 – 24)**

S.NO	CHAPTER NAME	PAGE NUMBER
1	DATA SCIENCE FOR BUSINESS DECISION- MAKING-INTRO	1 – 7
2	DATA PROCESING ORGANISATION, CLEANING AND VALIDATION	8 – 16
3	DATA ANALYSIS AND MODELLING	17 – 23
4	DATA PRESENTATION VISUALISATION AND GRAPHICAL PRESENTATION	24 – 27

**TARGET**  
Academy



# TARGET

## Academy

---

---

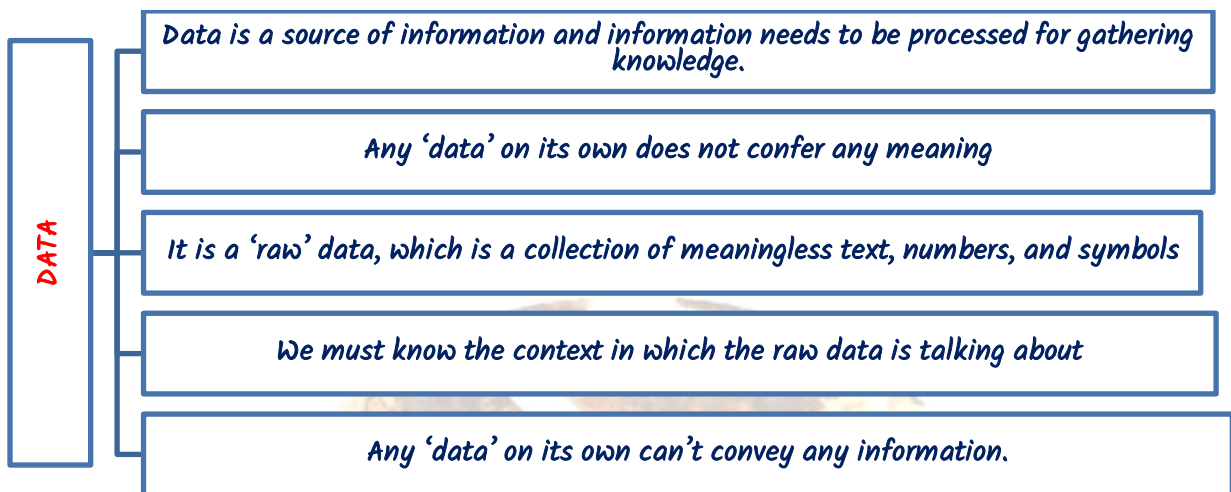
### **DATA SCIENCE FOR BUSINESS DECISION- MAKING-INTRO**

---

---

CONTACT: 90595 29540, 90106 76997.

DATA SCIENCE FOR BUSINESS DECISION- MAKING-INTRO



34

36

39

42

44

**INFORMATION**

Data needs to be processed for gathering information. Most commonly, we take the help of computers and software packages for processing data

INFORMATION = DATA + CONTEXT

34,36,39,42,44 =Temperature in HYD in last 4 weeks

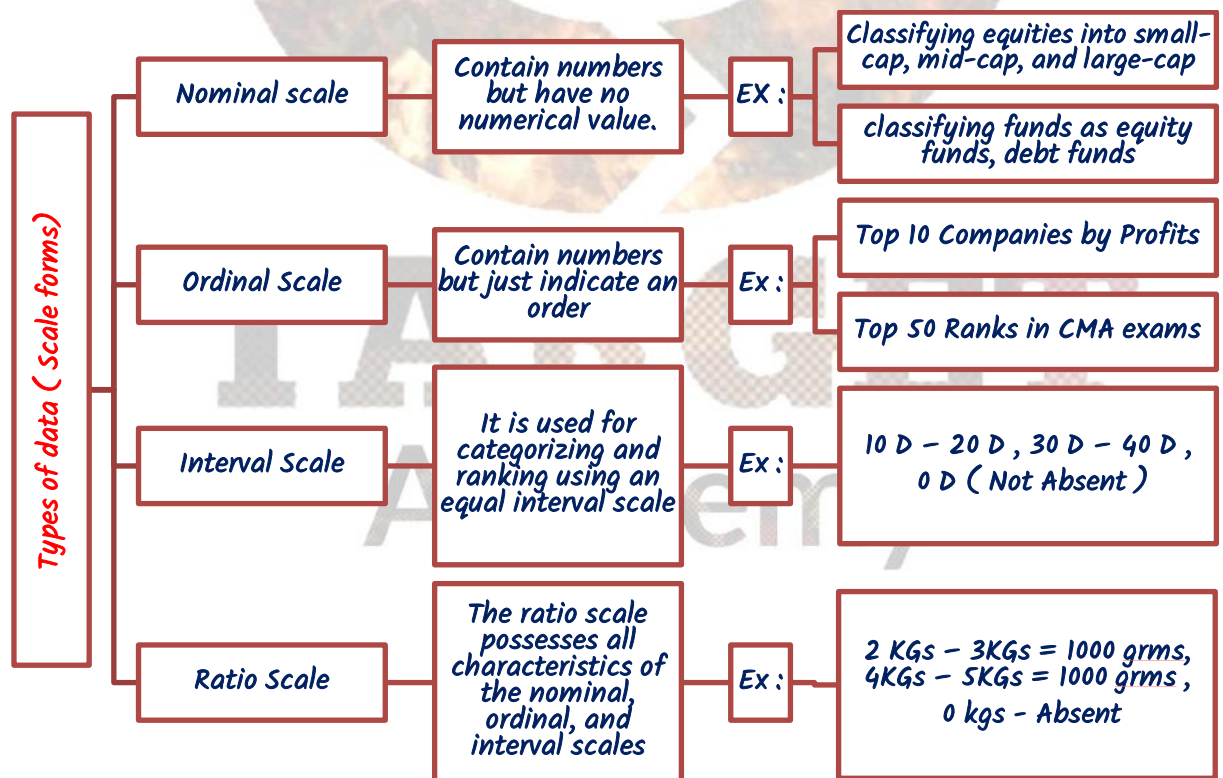
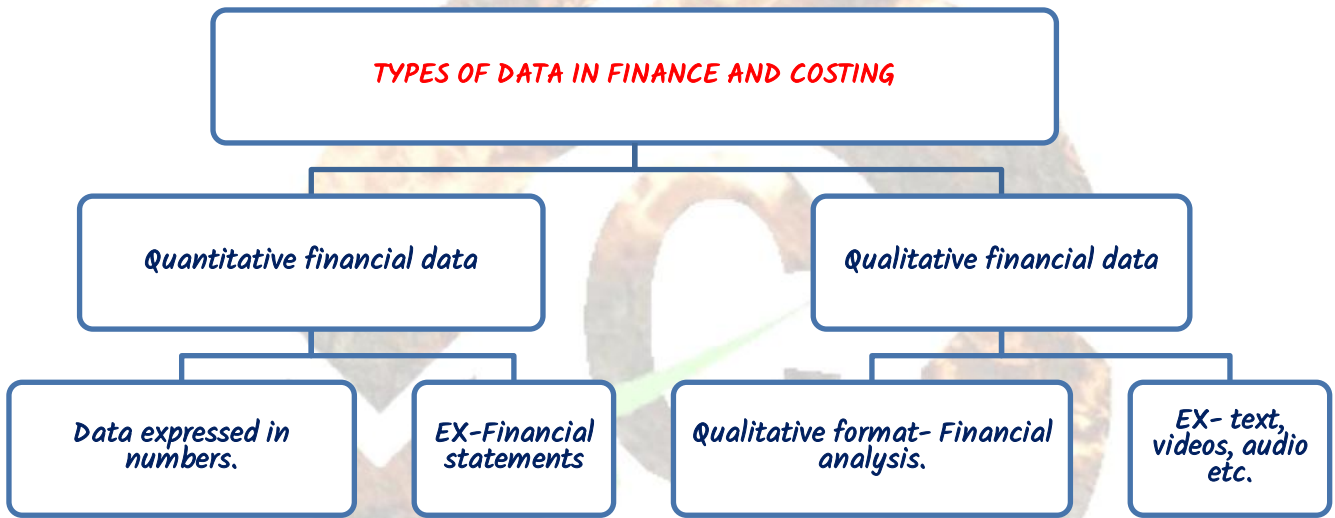
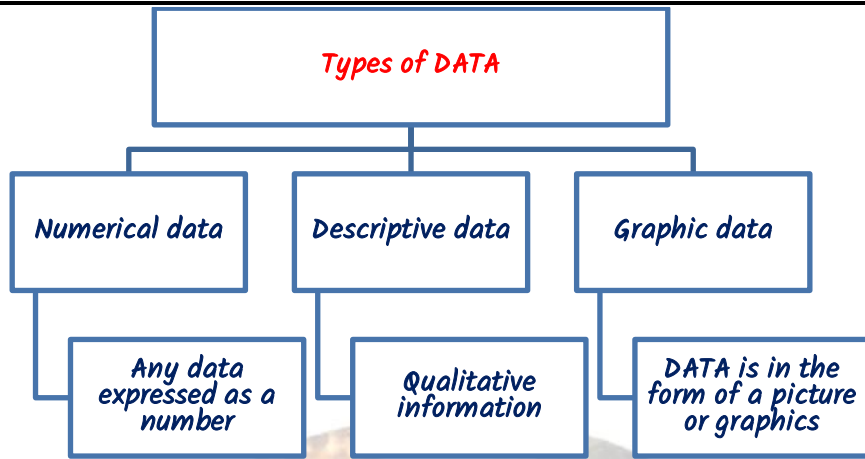
**KNOWLEDGE**

When these 'information' is used for solving a problem, we say it's the use of knowledge

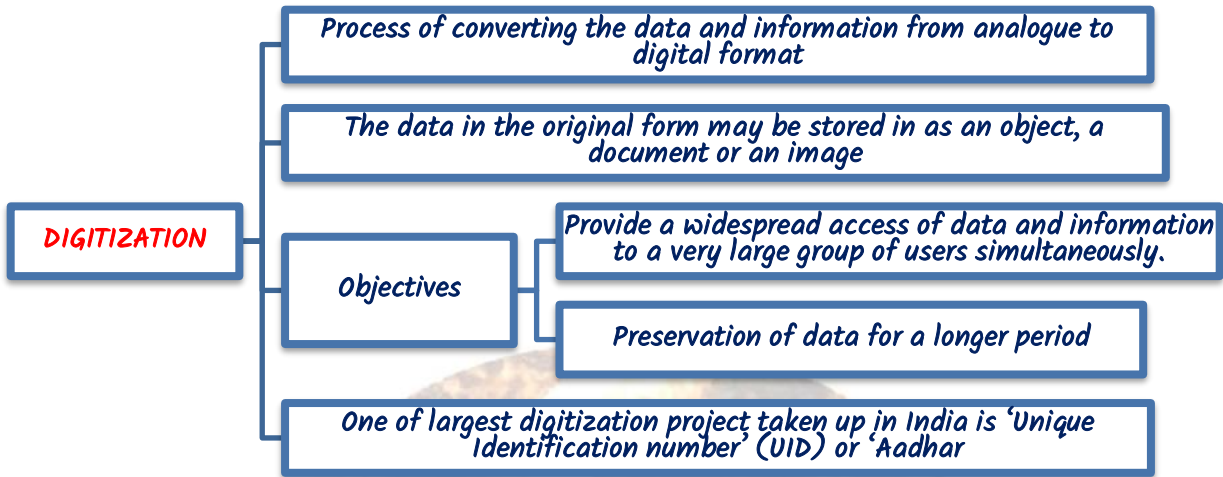
KNOWLEDGE = INFORMATION + APPLICATION OF IT

34,36,39,42,44 =Temperature in HYD in last 4 weeks

Purchase of AC's - Increases

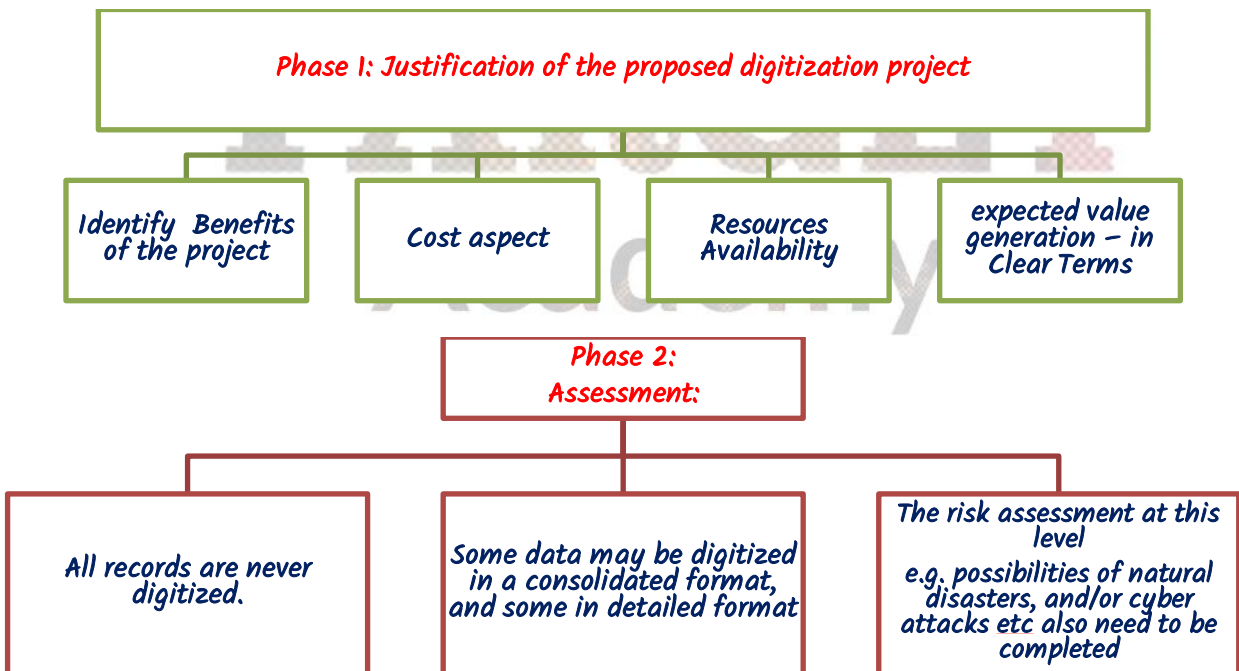


**DIGITIZATION OF DATA AND INFORMATION:**



**Why we digitize**

- Digitized records may be accessed by more than one person simultaneously.
- It becomes easier to reuse the data
- Requires less physical storage space
- Helps in work processing



**Phase 3: Planning:**

Successful execution of digitization project needs PERFECT planning

There are several stages for planning

complete the digitization

It may be done

In-House

on-demand

Outsource

in batches.

**Phase 4: Digitization activities**

Upon the completion of assessment and planning phase, the digitization activities start

**Phase 5: Processes in the care of records**

Once the digitization of records is complete, there are few additional requirements arise which may be linked to administration of records.

Additional requirements of data

The permission for accession of data,

intellectual control (over data),

classification (if necessary), and

upkeeping and maintenance of data

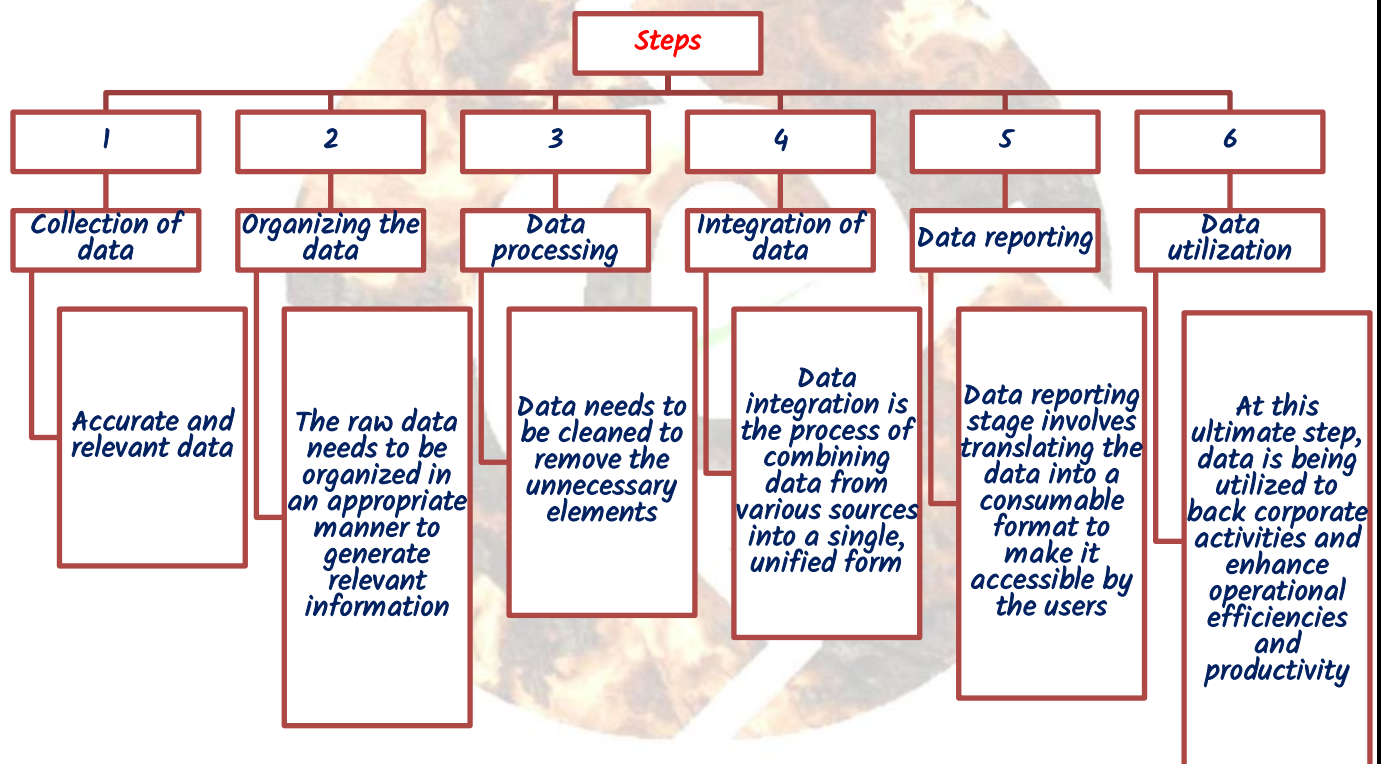
**Phase 6: Evaluation**

Once the digitization project is updated and implemented, the final phase should be a systematic Evaluation

**TRANSFORMATION OF DATA TO DECISION RELEVANT INFORMATION:**

To make the data turn into user friendly information, it should go through six core steps:

1. Collection of data
2. Organizing the data
3. Data processing
4. Integration of data
5. Data reporting
6. Data utilization



**PROFESSIONAL SCEPTICISM REGARDING DATA:**

Data analytics helps make decisions, but don't blindly trust it; look deeper for hidden insights.

New tools in finance give unique insights but also make skepticism tricky. More data means more red flags to manage, and false positives are a risk.

Skepticism is important but can strain client relationships and budgets. Finding a balance between skepticism and using data efficiently is key.

Ethical concerns about collecting and using data are important today. Data ethics involves doing the right thing with personal information.

Following 5 simple principles can help handle data ethically.

*Five basic principles*

*Ownership:*

*Personal info belongs to the person, so getting consent before collecting it is a must, especially for financial data.*

*Transparency:*

*Be clear about why you're collecting user data, like using cookies for tracking, and state how financial data will be used.*

*Privacy:*

*Keep collected data safe and don't share personally identifiable info publicly. Use security measures like encryption.*

*Intention:*

*Don't gather unnecessary data or use it for unethical purposes.*

*Outcomes:*

*Ensure data analysis doesn't harm anyone unintentionally, like through unfair impacts on certain groups*



**TARGET**  
Academy

THE END



# TARGET

## Academy

---

---

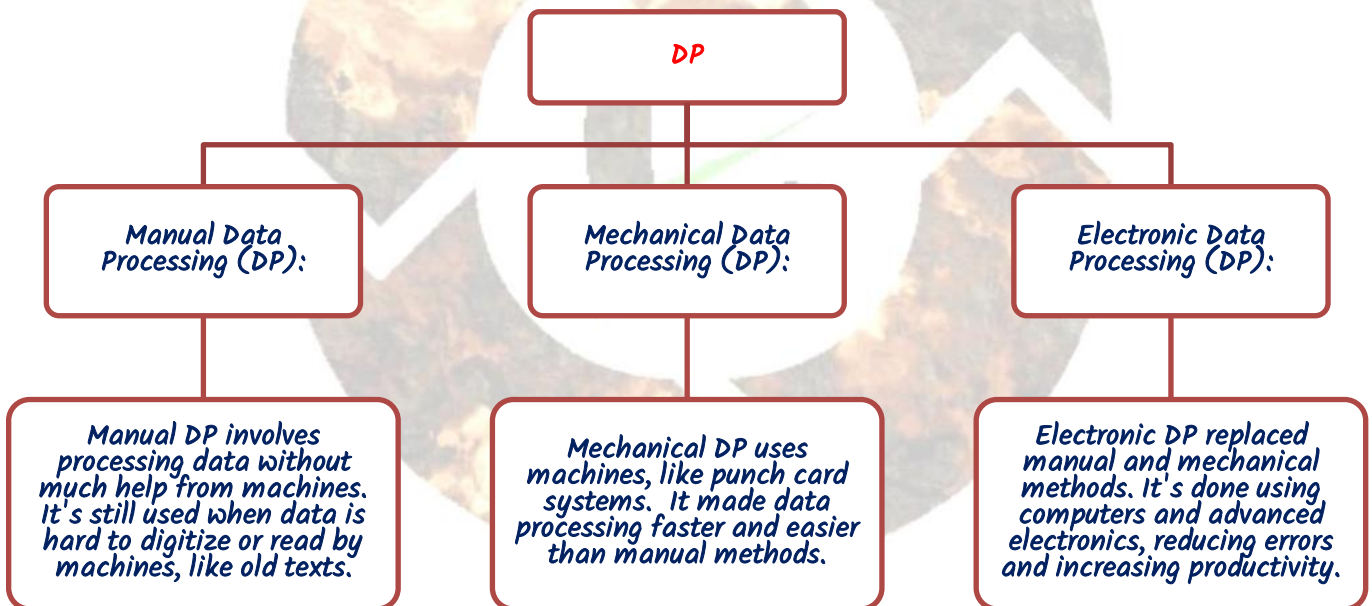
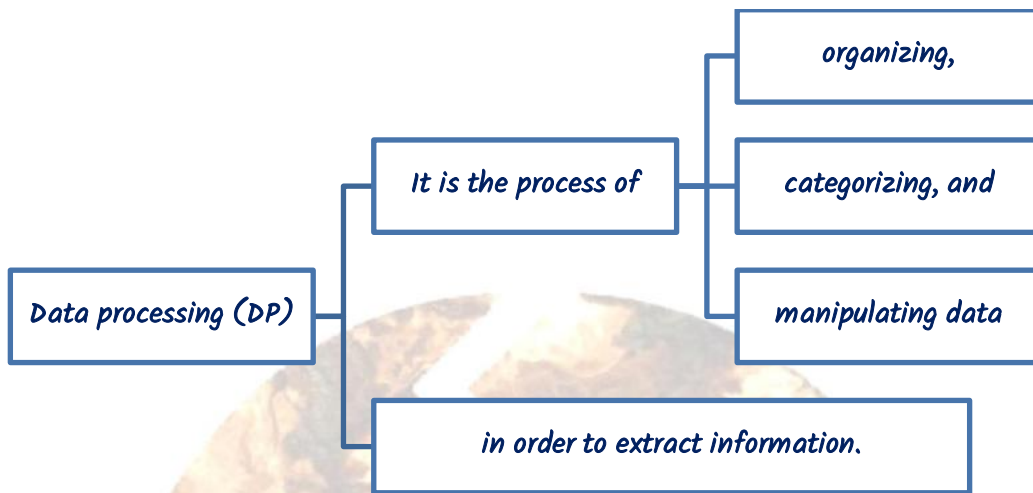
**DATA PROCESING  
ORGANISATION, CLEANING AND  
VALIDATION**

---

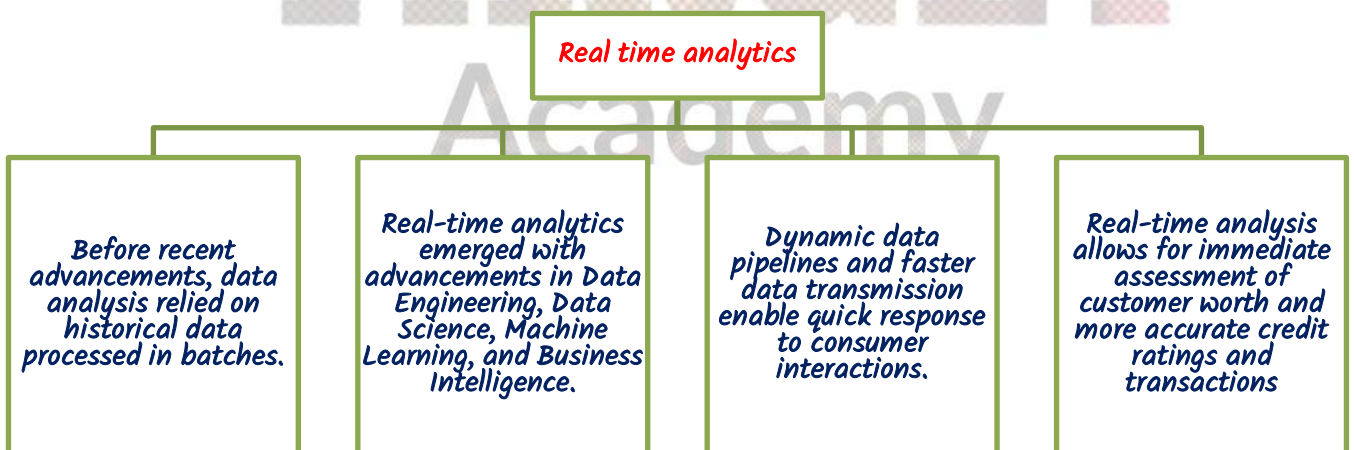
---

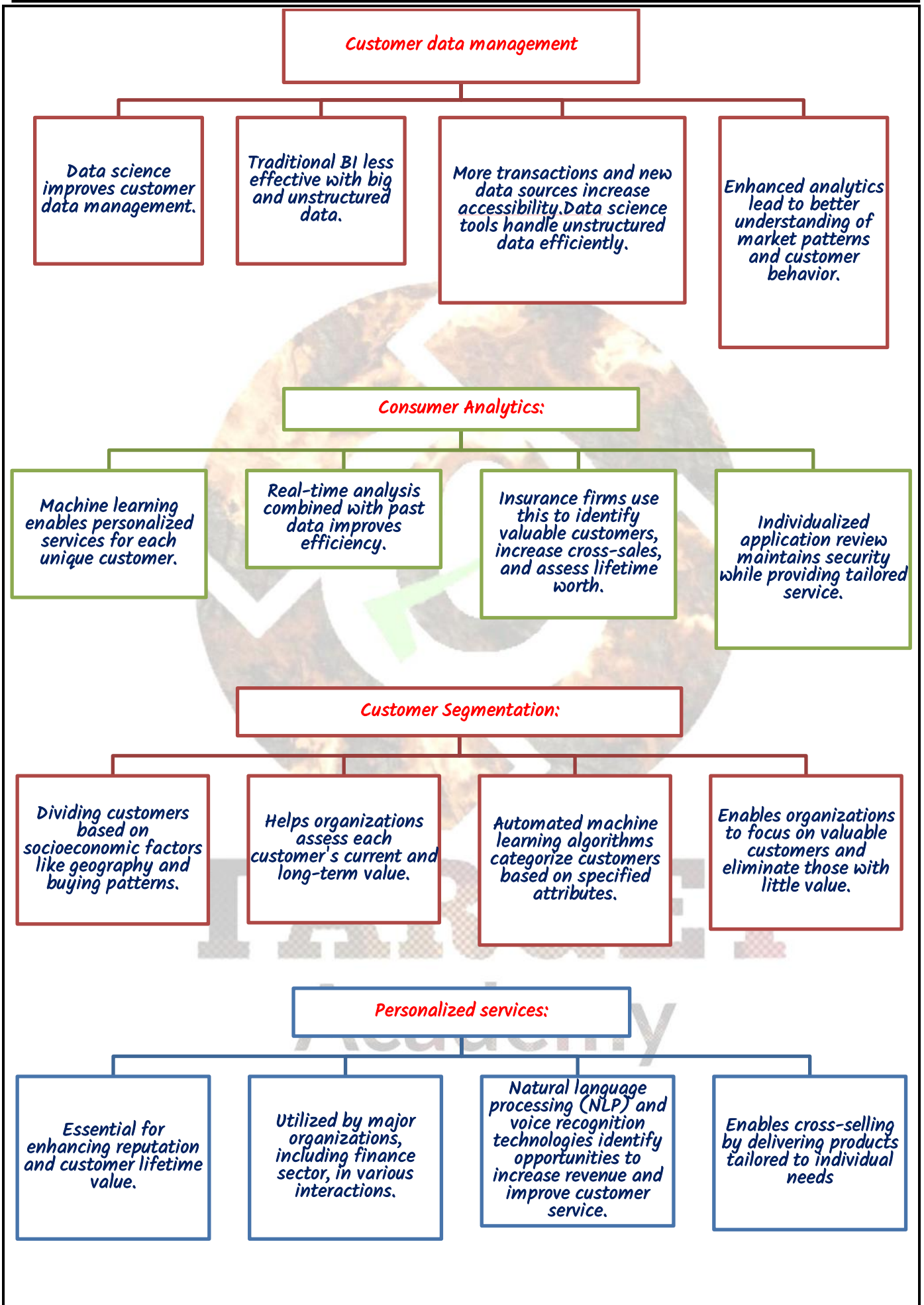
CONTACT: 90595 29540, 90106 76997.

DATA PROCESING ORGANISATION, CLEANING AND VALIDATION



THE ELEVEN SIGNIFICANT AREAS WHERE DATA SCIENCE PLAY IMPORTANT ROLE:





**Advanced customer service:**

Data science enables real-time evaluation of customer interactions.

Natural language processing provides practical financial advice during conversations.

Customer support agents can cross-sell or up-sell efficiently while addressing inquiries.

Knowledge from each interaction informs future interactions, enhancing system efficacy over time

**Predictive Analytics:**

Uses existing data to forecast future financial trends.

Machine learning techniques learn from pre-processed data.

Deep learning achieves superior performance without human preparation.

Helps in stock market pricing by predicting future trends for investment decisions

**Fraud Detection:**

Crucial for financial institutions to mitigate risks associated with rising transactions.

Advanced technologies like AI and machine learning enhance early detection of fraud.

Real-time analysis allows for immediate response to suspicious behavior.

Protects clients, banks, and insurance carriers by terminating transactions and notifying relevant parties promptly

**Anomaly Detection:**

Crucial for spotting unusual activities in bank accounts.

Data science provides real-time insights but can't predict individual incidents like fraud.

Helps identify unlawful insider trading before significant harm occurs.

Utilizes algorithms like Recurrent Neural Networks and Long Short-Term Memory models.

Next-generation designs like Transformers enhance anomaly detection capabilities

**Algorithmic Trading:**

Utilizes algorithms to make stock market trade suggestions.

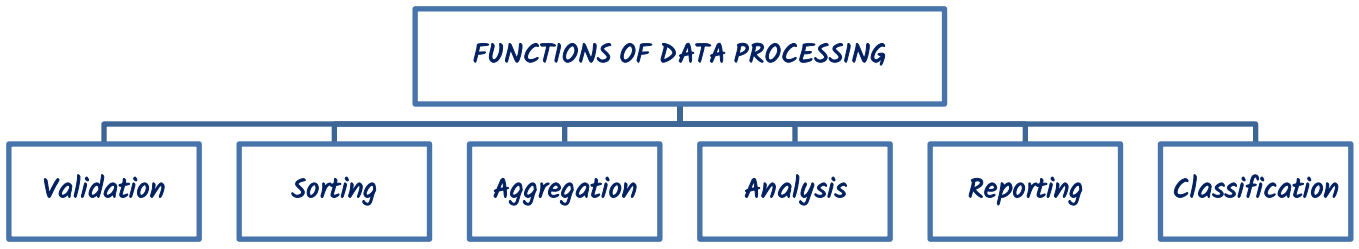
Reduces risk of loss due to human indecision and error

Developed using Reinforcement Learning, where model learns from penalties and rewards.

Allows for increased frequency of trades based on facts and learned behavior.

Trades only when profit opportunity aligns with its rule set.

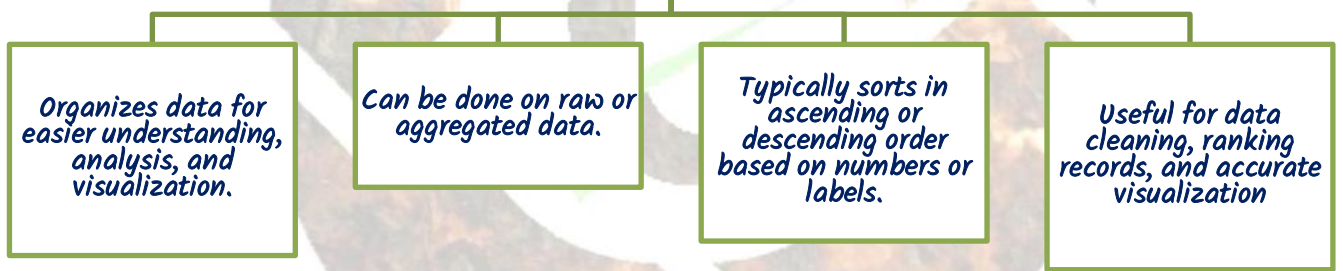
**FUNCTIONS OF DATA PROCESSING:**



**Validation:**



**Sorting:**



**Analysis:**

Data analysis is described as the process of cleaning, converting, and modelling data to obtain actionable business intelligence

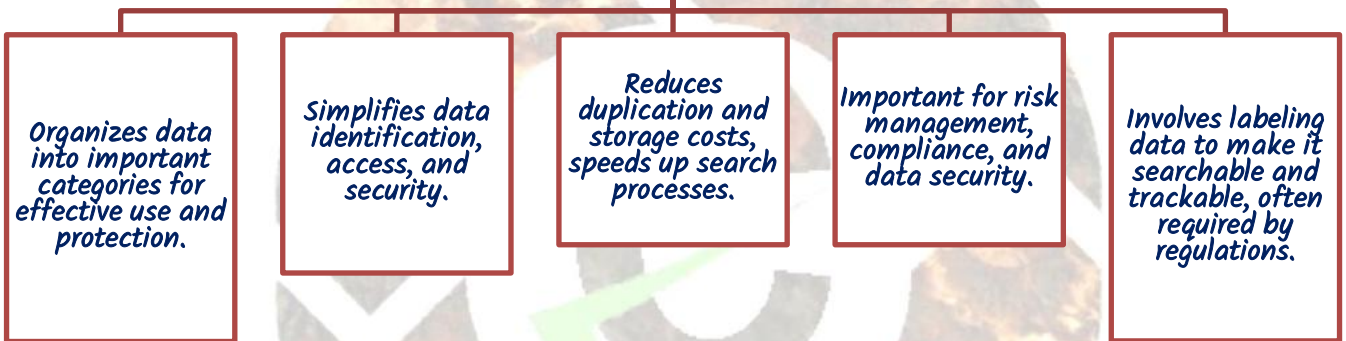
**Reporting:**



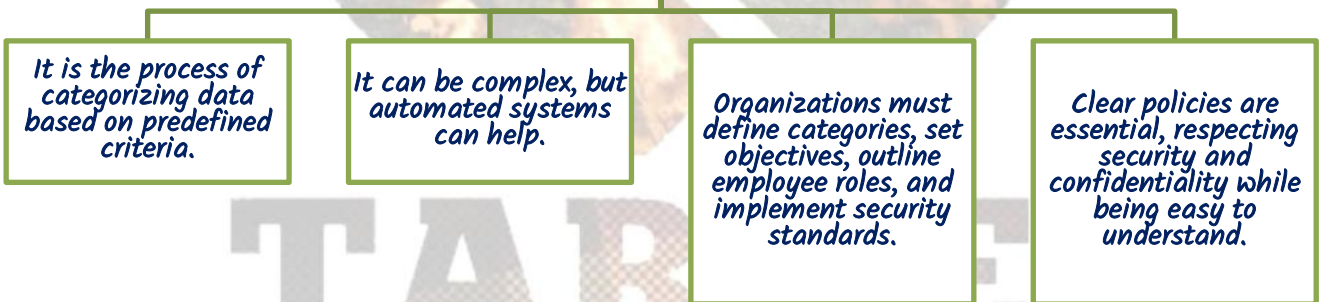
**Aggregation:**



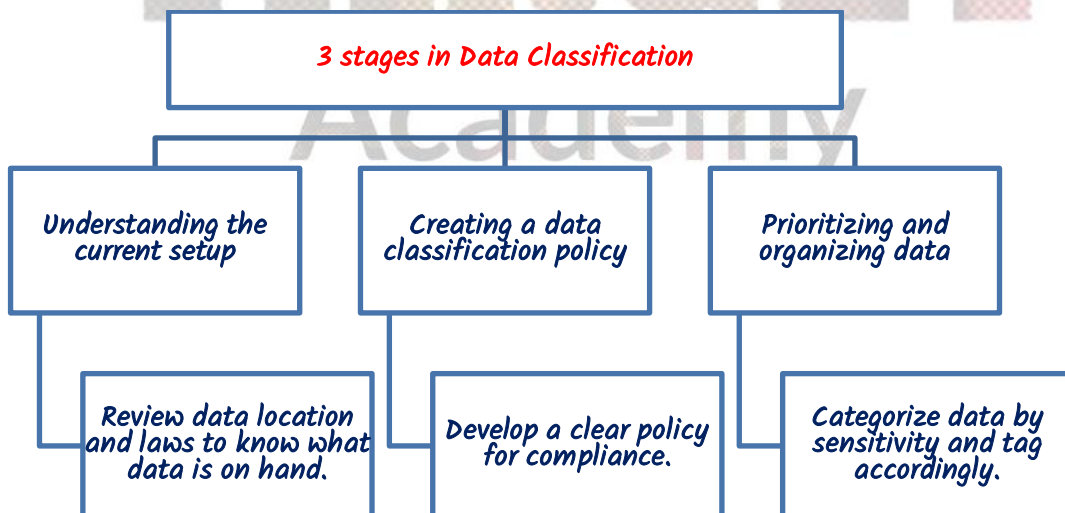
**Classification**



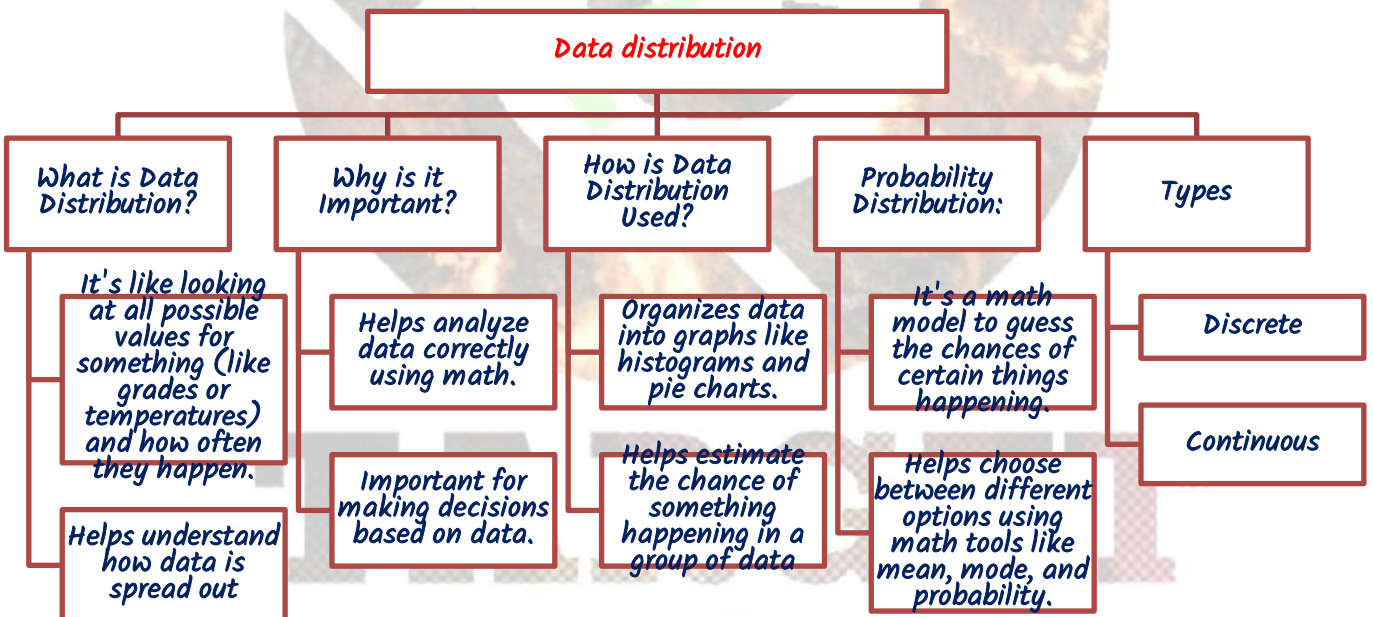
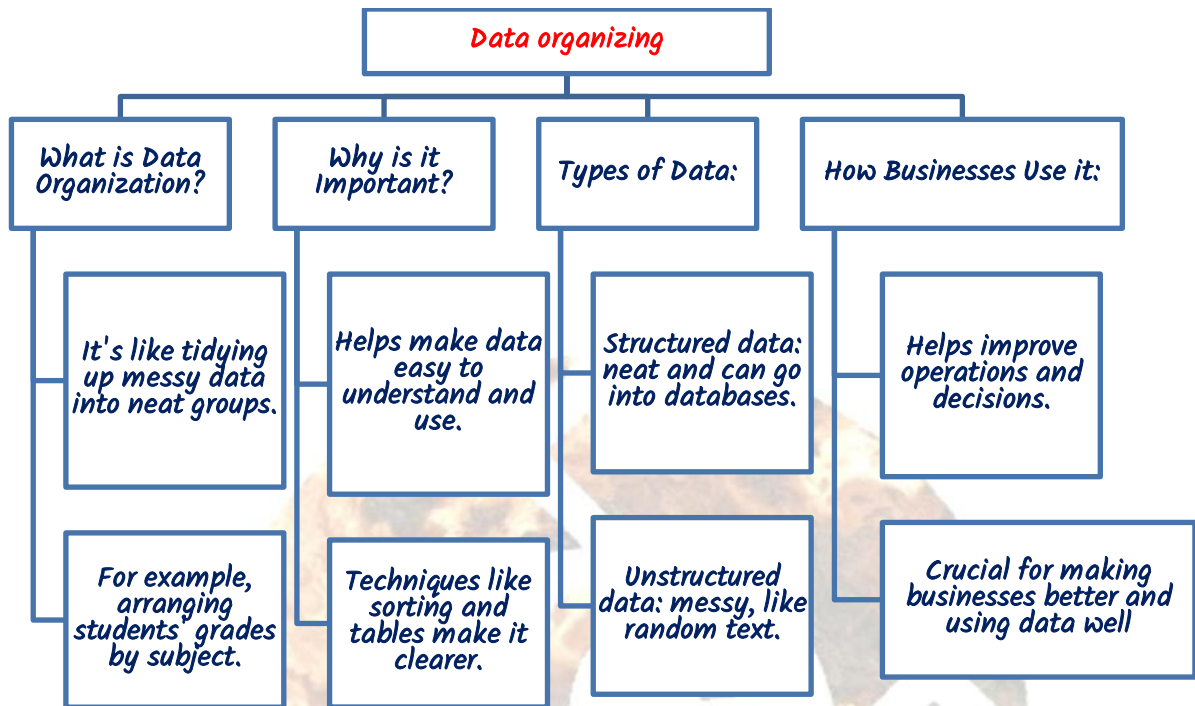
**Data classification**



**3 stages in Data Classification**



**DATA ORGANISATION AND DISTRIBUTION:**



Academy

**Discrete Distributions:** Countable data with a finite number of possible values.  
Examples: Rolling dice, counting heads.

**Types:**

a) Binomial: Counts successes/failures in experiments like coin tosses.

b) Poisson: Counts events happening in a specific time period, like flaws or accidents.

c) Hypergeometric: Counts successes in trials without replacement from a large population.

d) Geometric: Counts the probability of the first success.

**Continuous Distribution** Definition: Unlimited data points on a continuous scale.

**Types:**

i) Normal: Bell-shaped curve, with values clustered around the center.

ii) Lognormal: Logarithm of variable follows a normal distribution.

iii) F Distribution: Tests equality of variances between two populations

iv) Chi-Square: Squaring and summing independent standard normal variables.

v) Exponential: Represents consistent failure rates, closely related to Poisson.

vi) T-Student: Symmetrical bell-shaped distribution used for testing hypotheses and confidence intervals when standard deviation is unknown

**Data Cleaning:**

- Definition:**  
*Correcting or removing inaccurate, duplicate, or irrelevant data from a dataset.*
- Importance:**  
*Ensures trustworthy outcomes and algorithms despite apparent accuracy.*
- Different from Data Transformation:**  
*Cleaning removes irrelevant data, while transformation changes data format or structure.*

**Steps:**

- i) Removal of duplicate and irrelevant information.*
- ii) Fix structural errors like typos or inconsistent naming.*
- iii) Filter unwanted outliers for better data analysis.*
- iv) Handle missing data by either dropping observations or filling in missing values based on other observations.*
- v) Validation and Quality Assurance to ensure data makes sense, adheres to regulations, and supports working hypotheses*

**Data Validation:**

- Importance:** *Crucial for accurate outcomes in data management processes.*
- Purpose:** *Ensures data accuracy, clarity, and specificity before use.*

**Types:**

- Data Type Check:** *Validates correct data types (e.g., numeric values).*
- Code Check:** *Confirms values match a set of options or specific formatting requirements.*
- Range Check:** *Verifies input falls within specified range (e.g., latitude and longitude).*
- Format Check:** *Ensures data follows predetermined format (e.g., date columns).*
- Consistency Check:** *Verifies data entered in a consistent manner (e.g., delivery date later than shipment date).*
- Uniqueness Check:** *Ensures unique items in a database (e.g., PAN or email ids).*

**THE END**



# TARGET

## Academy

---

---

### **DATA ANALYSIS AND MODELLING**

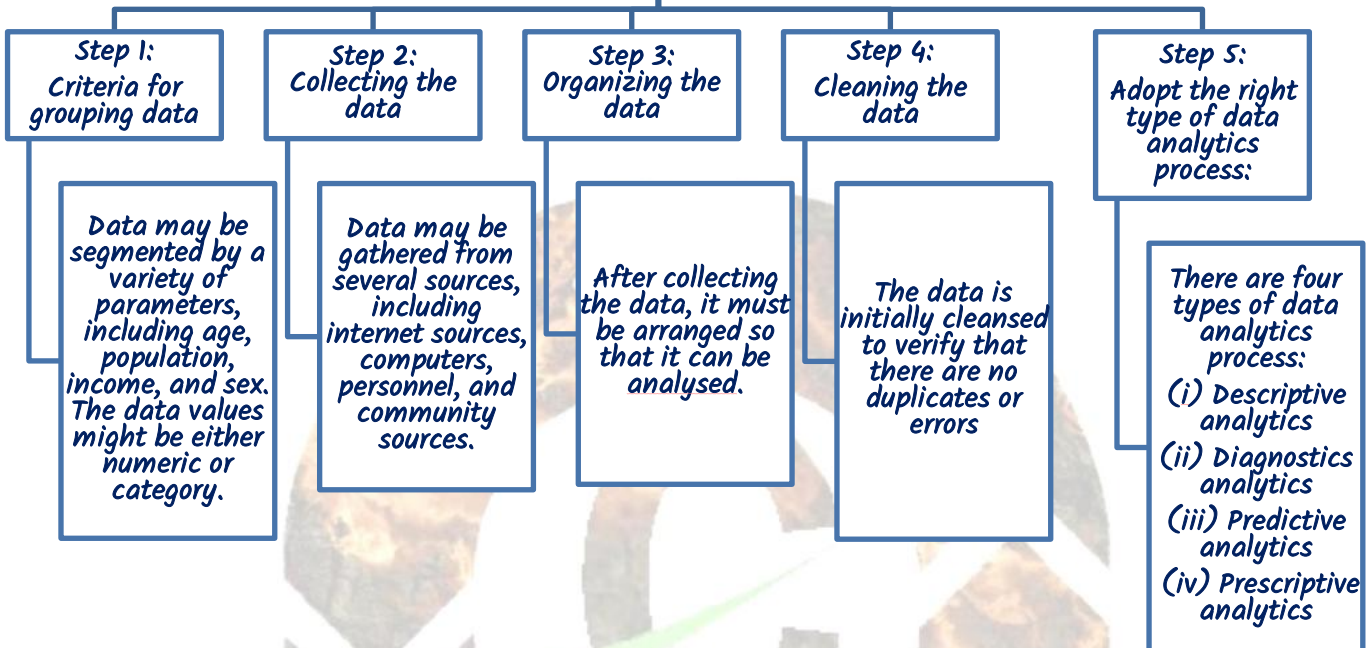
---

---

CONTACT: 90595 29540, 90106 76997.

DATA ANALYSIS AND MODELLING

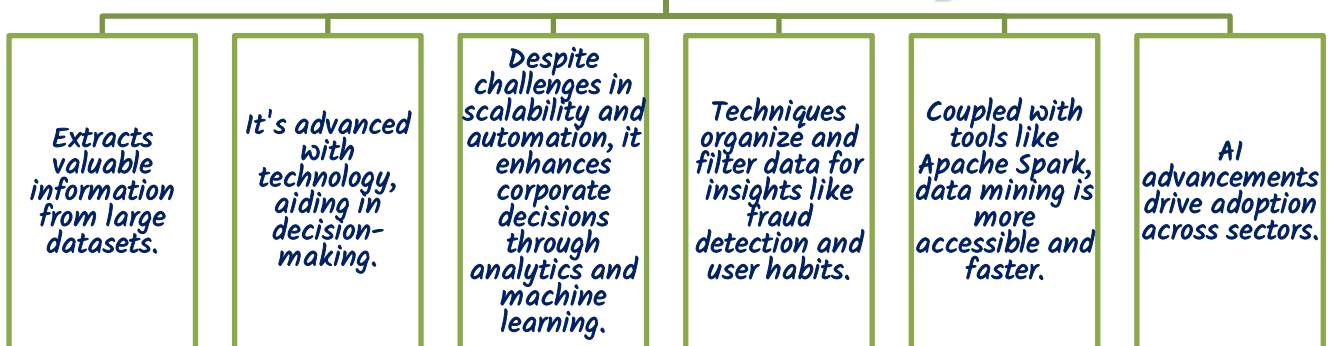
Process of data analytics

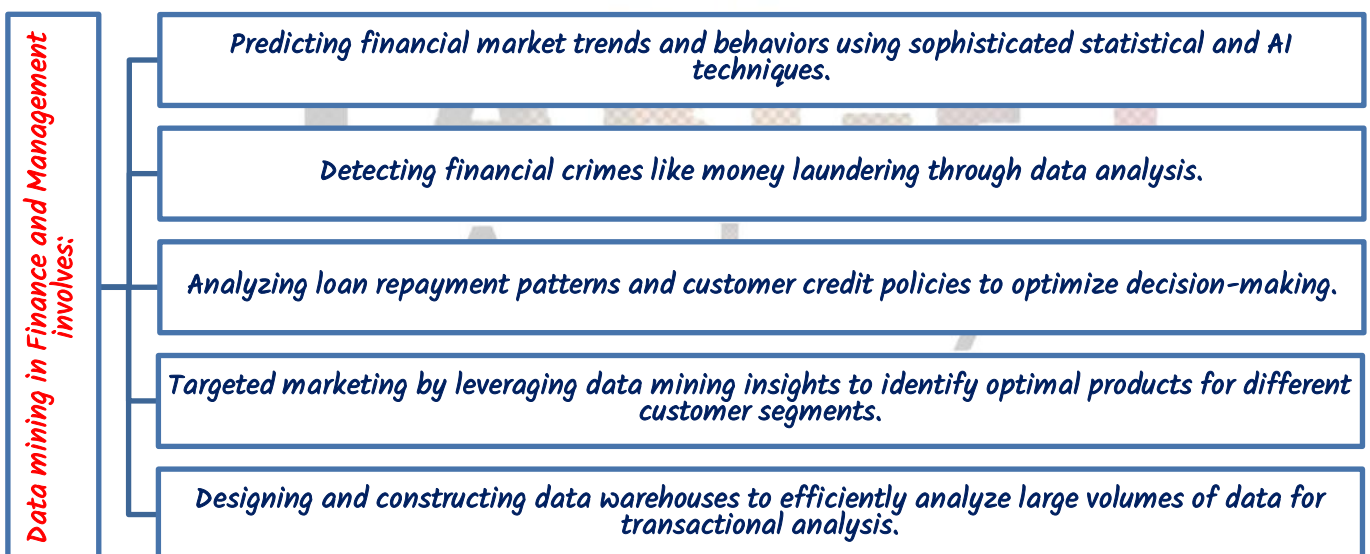
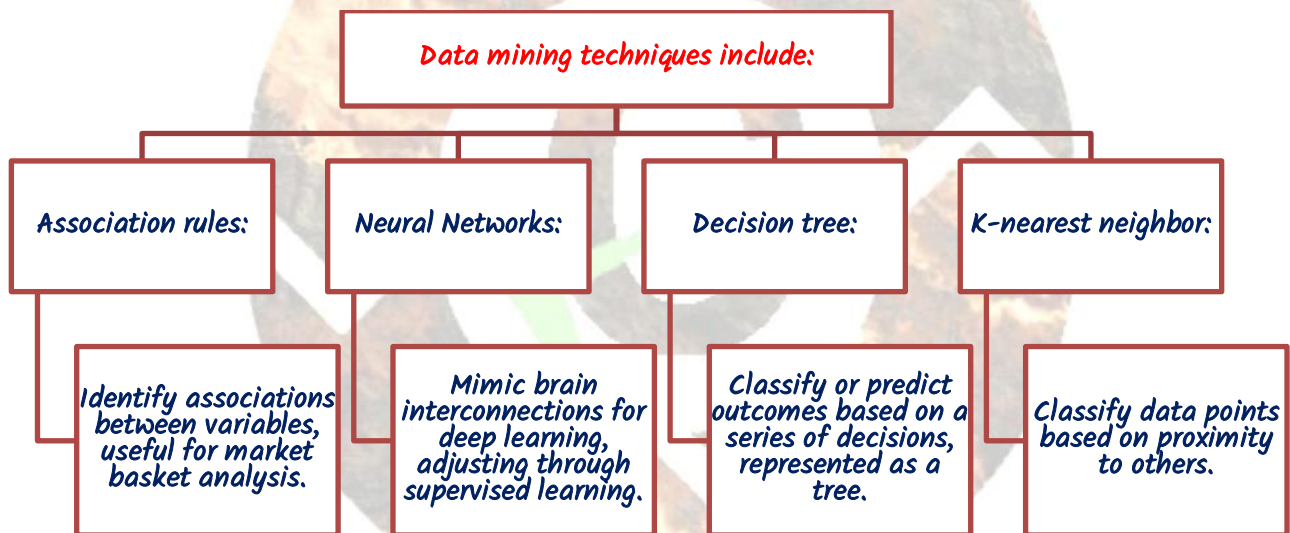
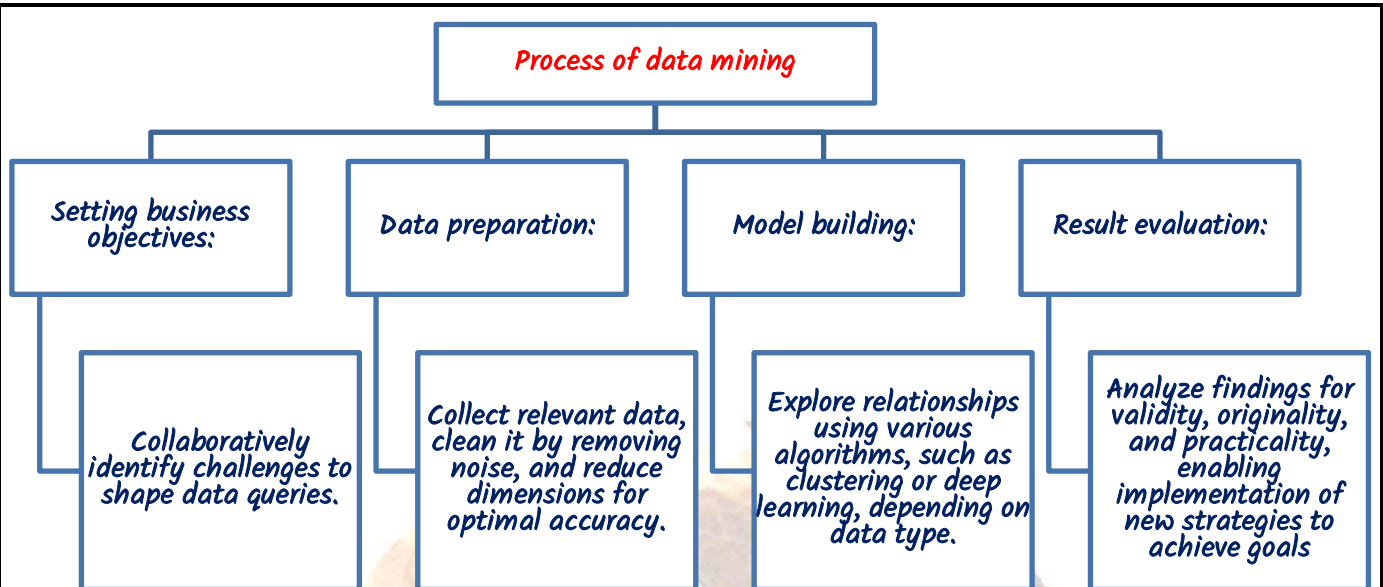


Benefits of data analytics include:



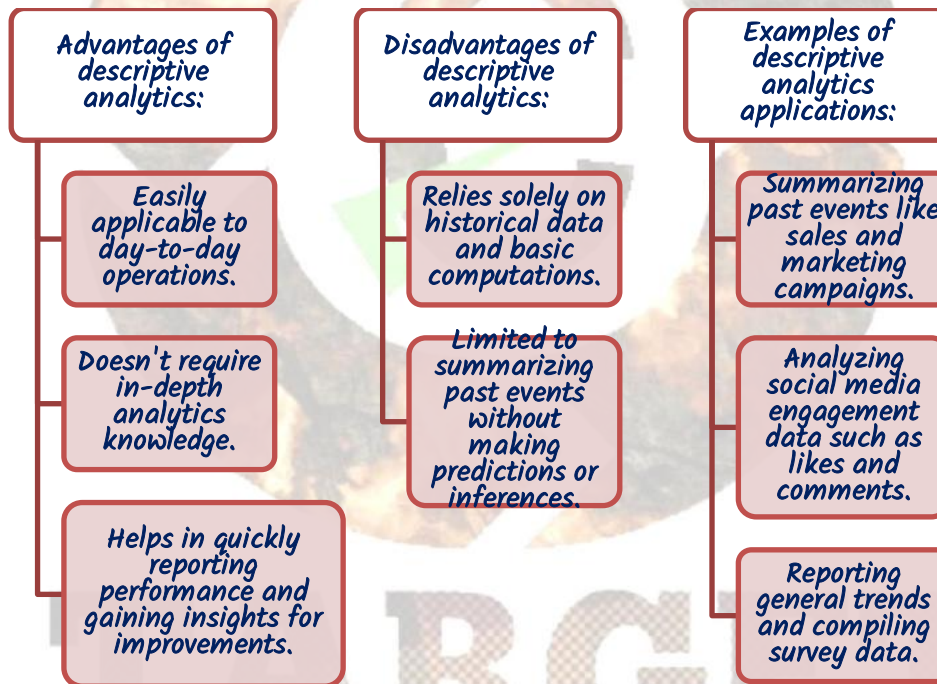
Data mining, or knowledge discovery in data (KDD)





**ANALYTICS AND MODEL BUILDING(DESCRIPTIVE, DIAGNOSTIC, PREDICTIVE, PRESCRIPTIVE):**

**Descriptive analytics:**



**Diagnostic analytics:**

*Asks "Why did this occur?" by examining data thoroughly.*

*Goes beyond descriptive analytics by revealing the rationale behind outcomes.*

*Involves data discovery, drill-down, data mining, and correlations.*

*Helps derive value from data by asking relevant questions and conducting in-depth analyses.*

*Requires adaptable BI and analytics platforms.*

**Examples of diagnostic analytics:**

*Determine the cause of a problem or identify a positive trend.*

*Set up analysis to identify root causes or patterns.*

*Use techniques like linear regression to identify relationships.*

**Predictive Analytics:**

*Focuses on forecasting future events based on past data patterns.*

*Utilizes probability, data mining, statistical modeling, and machine learning algorithms.*

*Helps in proactive decision-making and strategic planning.*

**Advantages:**

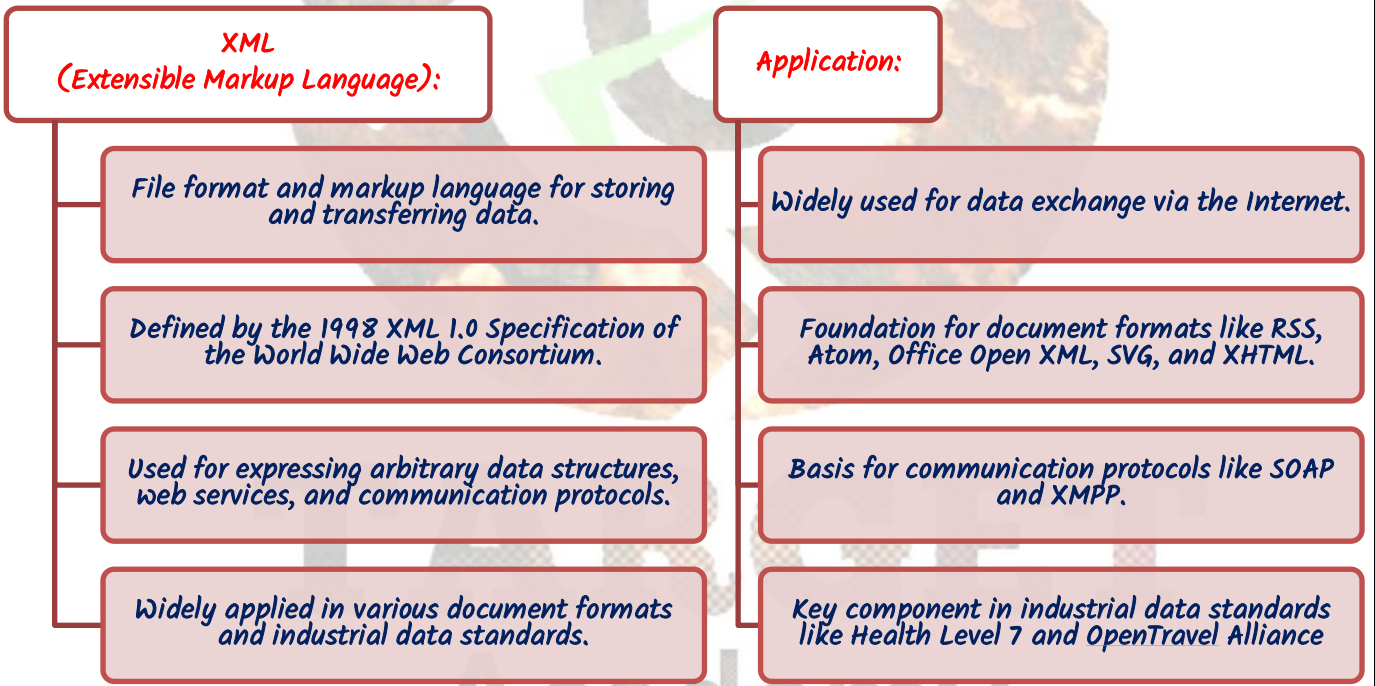
*Enhances effectiveness, customer service, fraud detection, and risk mitigation.*

*Requires historical data availability.*

**Example:**

*Healthcare industry predicts patient deterioration to reduce emergency hospitalizations.*

*Used in e-commerce, sales, human resources, IT security, and healthcare for various predictive purposes.*



**Extensible Business Reporting Language (XBRL):**

- Data description language for exchanging standard corporate data.
- Based on XML, enables automated interchange and trustworthy extraction of financial data.
- Organizes data using tags, ensuring uniformity and integrity.
- Provides expanded contextual information on financial documents.

**Benefits:**

- Reduces chance of erroneous data entry and increases reliability.
- Simplifies and automates report preparation, reducing expenses. Accelerates decision-making for financial entities.
- Facilitates publication of analyst and investor reports.
- Offers unparalleled access, comparison, and analytic capabilities.

**Business Intelligence (BI):**

- Includes business analytics, data mining, visualization, tools, and best practices for data-driven decision-making.
- Modern BI systems promote self-service analysis, controlled data on reliable platforms, empowered users, and rapid insight delivery.
- Originated in the 1960s as a method for disseminating information across enterprises and evolved into distinct products and services.
- Procedures include data mining, reporting, performance metrics, descriptive analytics, querying, statistical analysis, data visualization, visual analysis, and data preparation.



THE END



# TARGET

## Academy

---

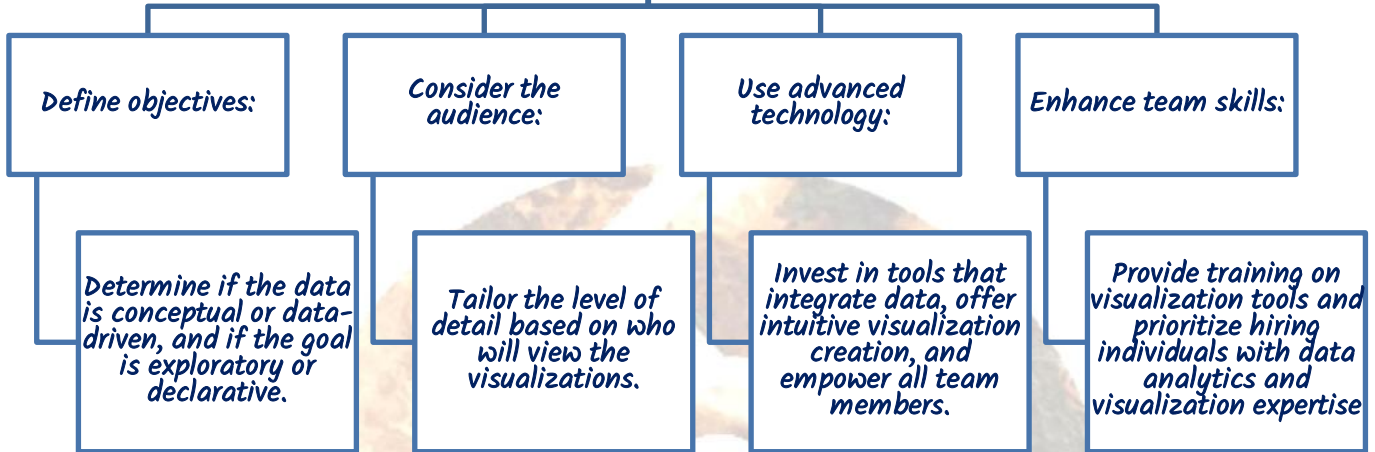
**DATA PRESENTATION  
VISUALISATION AND  
GRAPHICAL PRESENTATION**

---

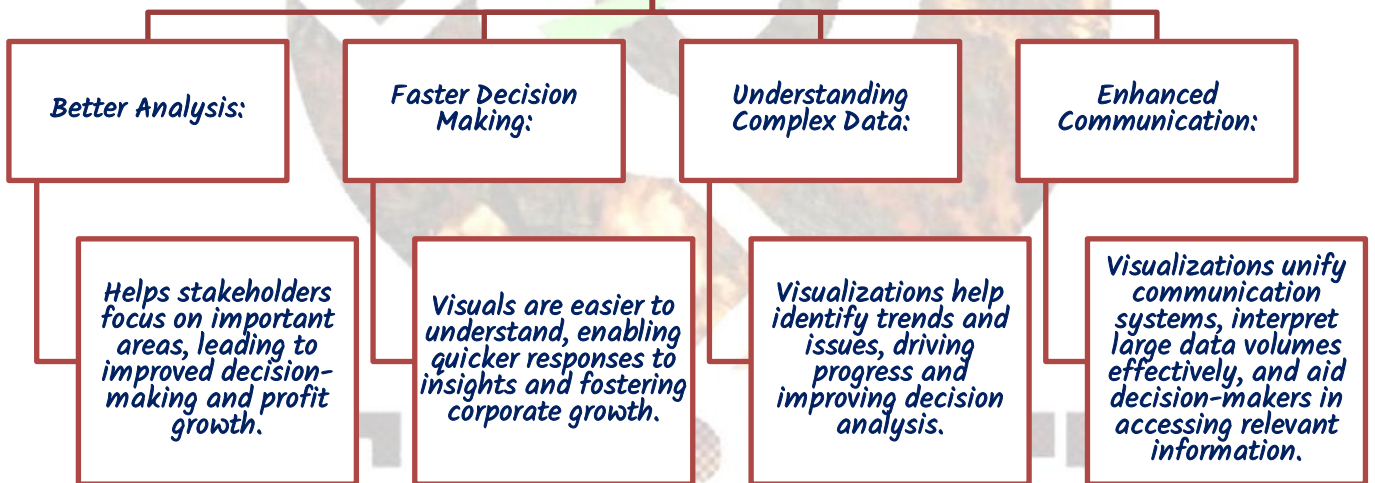
CONTACT: 90595 29540, 90106 76997.

DATA PRESENTATION VISUALISATION AND GRAPHICAL PRESENTATION

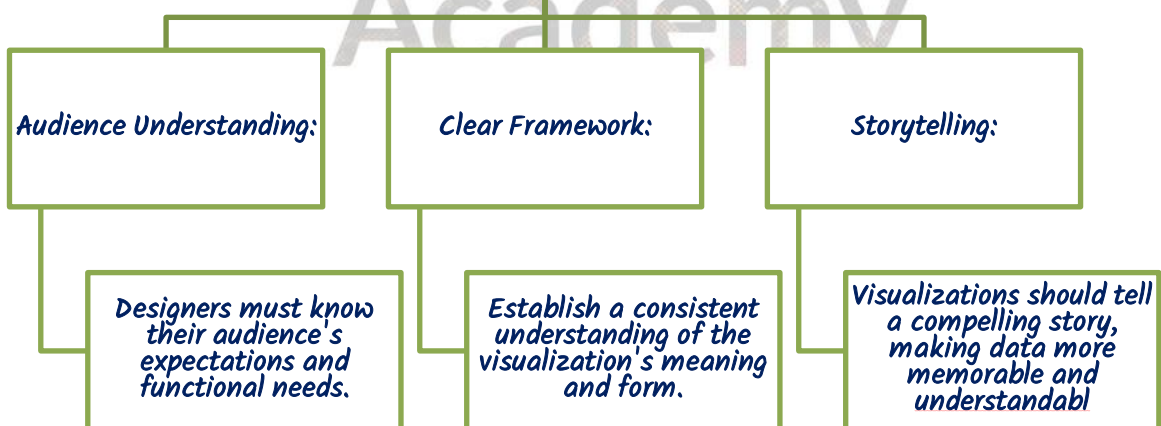
*Doing data Visualisation in the right way*



OBJECTIVE AND FUNCTION OF DATA PRESENTATION



*Common errors in data visualization include*



**DPA- Data Presentation Architecture**

*DPA is a crucial skill set that combines data science with visualization, communication, and change management to provide actionable business intelligence.*

*Its objectives include efficiently conveying relevant data to each audience participant and minimizing unnecessary complexity*

**To design a report using data visualization:**

- Find a compelling story within the data by arranging it, identifying outliers, and examining relationships.
- Create a narrative that engages the reader, providing context and highlighting significant insights.
- Choose appropriate visualizations that enhance comprehension, considering design principles like form, size, color, and labeling.
- Follow consistent visual language and style guidelines, whether for internal or external consumption.
- Publicize the report to share valuable insights beyond the organization

**Tableau**

*It is a powerful data visualization tool for creating interactive graphs, charts, and maps.*

*It connects to various data sources and allows quick visualization creation.*

*Tableau Desktop generates static visualizations, while Tableau Public is a free version with limitations.*

*Learning Tableau takes effort, but there are resources available. It's a crucial tool for data scientists*

**Microsoft Power BI**

*It is a business intelligence data visualization tool.*

*It supports reporting, self-service analytics, and predictive analytics.*

*Users can generate reports and share insights within their organization, making it a crucial tool for consolidated data reporting.*

**Microsoft Excel**

*It is a user-friendly data visualization tool with many options like scatter plots, bar charts, and more.*

*It helps illustrate relationships between datasets and variables, useful for analyzing various types of data for research and planning purposes*

**QlikView**

*It is a tool that helps users make faster, informed decisions by speeding up analytics and uncovering business insights.*

*It's widely used by enterprises and allows mixing data sources with various visualizations like tables, graphs, and charts.*

*With a user-friendly interface, it lets users input data from different sources without coding, making it easy to learn and understand.*



**TARGET**  
Academy

THE END