

# Unprocessed vs. Processed Data

# Introduction

- Definition of data:-Data refers to raw facts, figures, or statistics collected from various sources

# Importance of data in decision-making

- Improves Data Accuracy
- Enhances Predictive Analytics
- Optimizes Efficiency
- Better Business Insights

# What is Unprocessed Data?

- Also known as raw data
- Data in its original form
- Examples: Sensor readings, Election survey responses

# Characteristics of Unprocessed Data

- Unstructured and unorganized
- Can contain errors and redundancies

# Sources of Unprocessed Data

- Databases
- IoT devices
- Social media

# Challenges of Unprocessed Data

- Inconsistency
- Data redundancy
- Large volume and complexity

# What is Processed Data?

- Data that has been cleaned, formatted, and structured
- Ready for analysis and decision-making
- Examples: Reports, dashboards, structured tables



# Characteristics of Processed Data

- Organized and structured
- Free from errors
- Meaningful and usable

# Importance of Data Processing

- Enhances data accuracy
- Helps in decision-making
- Improves efficiency

# Data Processing Steps

- Data collection
- Data cleaning
- Data transformation
- Data analysis
- Data visualization

# Techniques for Processing Data

- Measure of Central Tendency(Mean,Median,Mode)
- Delete missing value rows

# Tools Used for Data Processing

- Excel
- SQL
- Python (Pandas, NumPy)
- Data visualization tools (Tableau, Power BI)

# Challenges in Data Processing

- Large data volume
- Data security concerns
- Processing time and costs

## Automation in Data Processing

- Automated ETL (Extract, Transform, Load) and ELT(Extract,Load,Transform)

# Data Processing in Real Life

- Finance
- Marketing
- Manufacturing
- Healthcare



# Common Issues in Unprocessed Data

- Missing values
- Duplicates
- Inconsistent formatting
- Outliers and noise
- Incorrect data types

# Effect on Data Visualization

- Charts and graphs may misrepresent information
- Inconsistent formatting leads to misinterpretation
- Example: Incomplete time-series data

# Impact on Business Decision-Making

- Poor data quality leads to incorrect business strategies

# Impact on Machine Learning Models

- Poor model accuracy
- Overfitting or underfitting

# Data Preprocessing Steps

- Handling missing values
- Removing duplicates
- Detecting outliers

# Outlier Detection and Treatment

- Boxplot method

# Impact of Clean Data on Decision Making

- More accurate insights
- Better forecasting

# Impact of Clean Data on Machine Learning

- Higher model accuracy
- Reduced bias
- Faster training and evaluation



# Definition of Data Anomalies

- What are data anomalies?
- Any irregularity in data that affects accuracy and reliability