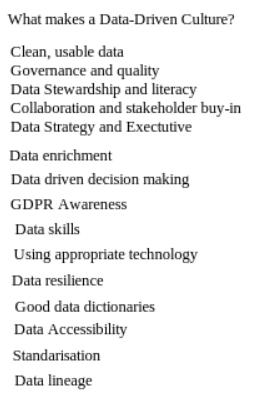
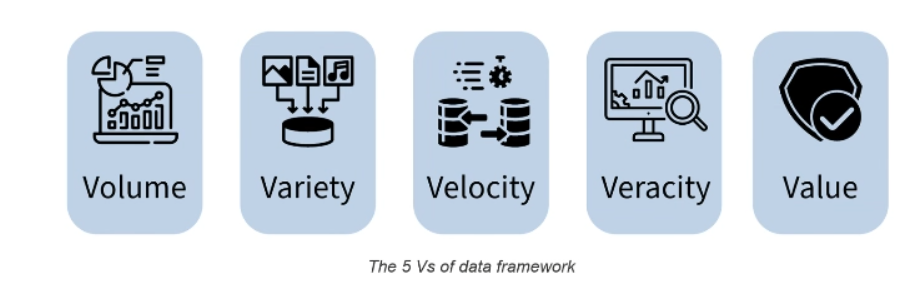
# ****Data Fundamentals****

**29/01/2025**



* **Data-Driven Culture**
  + The session focused on building a data-driven culture within organizations.
  + Key elements of a data-driven culture include:
    - Clean, usable data
    - Data governance and quality
    - Data stewardship and responsibility
    - Data literacy across the organization
    - Collaboration and stakeholder buy-in
    - Executive support and data strategy



* **Data Engineering Role**
  + Data engineers play a crucial role in enabling businesses to harness the power of big data.
  + They design, build, and maintain infrastructure and systems for data collection, storage, and processing.
  + Data engineers ensure data quality and reliability, which is essential for data-driven decision-making.
* **Big Data vs. Small Data**
  + **Big Data**: Characterized by high volume, variety, velocity, veracity, and value. Examples include social media platforms analyzing billions of user interactions daily.
  + **Small Data**: Smaller in volume, often structured, and easier to analyze. Examples include personal fitness trackers recording daily steps and calories.
* **Data Formats and Standards**
  + Common data formats include JSON, CSV, and XML.
  + Cloud computing standards like AWS Well-Architected Framework and OpenAPI Specification were discussed.
  + Regulatory requirements such as GDPR (General Data Protection Regulation) and ISO 27001 (Information Security Management) were highlighted.

#### ****Best Practices in Data Engineering****

* **Scalability and Reliability**
  + Data systems must handle growing data volumes without sacrificing performance.
  + Redundancy and disaster recovery plans are essential for safeguarding data.
* **Data Documentation**
  + Proper documentation ensures that data is understandable and usable by others.
* **Data Ethics**
  + Ethical guidelines must be followed to ensure responsible data usage.

#### ****Activity: Employee Churn Decision Tree****

* **Objective**: Build a decision tree for predicting employee churn.
* **Key Factors Identified**:
  + Salary
  + Job satisfaction
  + Work hours
  + Career development
  + Manager and colleague relationships
  + Flexibility and work environment
* **Outcome**: The activity helped participants understand how to identify key decision points and factors that influence employee retention.

### Key Takeaways:

1. **Engagement is Key**: Active participation through polls, reactions, and discussions enhances the learning experience.
2. **E-Learning is Essential**: Completing e-learning modules before sessions helps in better understanding and engagement.
3. **Data-Driven Culture**: Building a data-driven culture requires clean data, governance, literacy, and stakeholder buy-in.
4. **Data Engineering Role**: Data engineers are crucial in designing and maintaining systems that enable data-driven decision-making.
5. **Big Data vs. Small Data**: Understanding the characteristics of big data (volume, variety, velocity, veracity, value) is essential for data engineers.
6. **Best Practices**: Scalability, reliability, documentation, and ethics are key best practices in data engineering.
7. **Learning Journal**: Reflecting on learning through a journal hosted on GitHub is a valuable practice for continuous improvement.