

5-Day Training Agenda

Data Collection Techniques (Avoiding Problems with Incorrect and Biased Data)

Duration: 5 Days (8 Hours per Day)

Day 1: Foundations of Data Collection

Introduction to Data Collection

- What is Data Collection?
- Importance of High-Quality Data
- Data Lifecycle
- Types of Data (Structured, Semi-Structured, Unstructured)
- Quantitative vs. Qualitative Data

Sources of Data

- Internal Data Sources
- External Data Sources
- Primary vs. Secondary Data
- First-Party, Second-Party, and Third-Party Data

Data Collection Methods

- Surveys and Questionnaires
- Interviews
- Observations
- Experiments
- Focus Groups
- Sensor and IoT Data
- Transactional Systems
- Web Analytics

Data Collection Planning

- Defining Objectives
- Identifying Target Population

- Sampling Basics
- Selecting Appropriate Collection Methods

Practical Exercise

- Identify suitable data collection methods for different business scenarios.
-

Day 2: Ensuring Data Accuracy and Quality

Understanding Data Quality

- Dimensions of Data Quality
- Accuracy
- Completeness
- Consistency
- Timeliness
- Validity
- Uniqueness

Common Data Collection Errors

- Human Errors
- Manual Entry Mistakes
- Duplicate Records
- Missing Data
- Incorrect Formats
- Inconsistent Values

Designing Better Data Collection Forms

- Required vs Optional Fields
- Validation Rules
- Drop-down Lists
- Standardized Formats
- Lookup Values
- Mandatory Checks

Data Validation Techniques

- Range Validation
- Format Validation
- Cross-field Validation
- Business Rule Validation
- Referential Integrity

Practical Exercise

- Improve a poorly designed data collection form.
-

Day 3: Identifying and Avoiding Bias in Data Collection

Understanding Data Bias

- What is Data Bias?
- Why Bias Matters
- Impact on Analytics and AI Models

Types of Bias

- Selection Bias
- Sampling Bias
- Confirmation Bias
- Observer Bias
- Measurement Bias
- Response Bias
- Recall Bias
- Survivorship Bias

Reducing Bias

- Representative Sampling
- Random Sampling
- Stratified Sampling
- Balanced Data Collection

- Blind and Double-Blind Studies
- Neutral Question Design

Ethical Data Collection

- Privacy Principles
- Informed Consent
- Data Minimization
- Confidentiality
- Responsible Data Usage

Practical Exercise

- Identify different biases in real-world datasets.
-

Day 4: Cleaning, Monitoring, and Improving Collected Data

Data Profiling

- Understanding Data Characteristics
- Frequency Analysis
- Pattern Analysis
- Outlier Detection
- Missing Value Analysis

Data Cleaning Techniques

- Handling Missing Values
- Removing Duplicates
- Standardizing Formats
- Correcting Invalid Entries
- Data Transformation
- Data Normalization

Monitoring Data Quality

- Data Quality Metrics
- Quality Dashboards

- Data Quality KPIs
- Continuous Monitoring

Root Cause Analysis

- Finding Sources of Poor Data
- Process Improvements
- Preventive Controls

Practical Exercise

- Clean and improve a real dataset using Excel or Power BI.
-

Day 5: Best Practices, Governance, and Capstone Project

Data Governance Fundamentals

- Data Ownership
- Roles and Responsibilities
- Data Stewardship
- Data Standards
- Metadata Management

Data Collection Best Practices

- Standard Operating Procedures
- Documentation
- Audit Trails
- Version Control
- Continuous Improvement

AI and Automated Data Collection

- APIs
- Web Forms
- Mobile Data Collection
- IoT Devices
- OCR

- AI-Assisted Data Capture

Case Studies

- Customer Feedback Collection
- Healthcare Data Collection
- Manufacturing Quality Data
- Retail Sales Data
- HR Employee Surveys

Capstone Project

- Define a business problem
 - Select the most appropriate data collection method
 - Design a data collection plan
 - Identify potential sources of bias
 - Develop strategies to improve data quality
 - Present findings and recommendations
-

Hands-on Activities Throughout the Course

- Designing surveys and questionnaires
 - Creating data collection templates
 - Validating data using Excel
 - Identifying incorrect and biased data
 - Cleaning datasets
 - Data profiling exercises
 - Sampling practice
 - Case study discussions
 - Group activities
 - Final capstone presentation
-

Target Audience

- Business Analysts
- Data Analysts
- Power BI Professionals
- Excel Users
- Data Stewards
- Data Engineers
- AI/ML Beginners
- Quality Assurance Teams
- Project Managers
- Anyone involved in collecting, managing, or analyzing business data