


# Getting & Cleaning Data

**Duration:** 5 Days

**Language:** en

**Course Code:** PI2-108

## Objective

Upon completion of this course, participants will be able to:

- Understand how to source data from different platforms and databases.
- Learn techniques for cleaning and preparing data for analysis.
- Develop skills in identifying and handling missing, inconsistent, and duplicate data.
- Gain expertise using tools like Excel, Python, and R for data cleaning.
- Learn to transform raw data into structured, accurate, and ready-to-use formats.
- Apply best practices in data preparation to ensure high-quality, reliable datasets.

## Audience

This course is intended for

- **Data Analysts and Data Scientists** : Professionals responsible for collecting and cleaning data to prepare for analysis and reporting.
- **Business Intelligence Professionals** : Individuals who work with data to support decision-making and need clean, structured data for accurate insights.
- **Researchers** : Those who collect data from surveys, experiments, or other research methods must ensure its accuracy and reliability.
- **IT Professionals and Database Managers** : Individuals tasked with integrating, maintaining, and managing large datasets in enterprise environments.
- **Anyone Working with Data** : Individuals who handle data daily and want to improve their data preparation and cleaning skills.

## Training Methodology

This course combines theoretical insights with practical experience to equip participants with essential data collection and cleaning skills. Through instructor-led lectures, hands-on exercises with real-world datasets, and interactive group discussions, participants will explore data sourcing, cleaning, and transformation techniques. Tool-based tutorials in Excel, Python, and R will provide practical training, while case studies will challenge participants to develop solutions for real-world data issues. Regular assessments and feedback will reinforce learning and ensure a thorough grasp of key concepts.

## Summary

Data is the cornerstone of informed decision-making in today’s data-driven world, but raw data is often messy, incomplete, or unstructured. The "Getting and Cleaning Data" training course is designed to help professionals understand the essential processes of collecting, preparing, and cleaning data to ensure it is accurate, complete, and ready for analysis. Whether working with large or smaller, more focused datasets, having clean and reliable data is critical to ensuring meaningful insights and accurate results.

This course provides participants with the knowledge and skills to effectively source data from various platforms, including databases, APIs, web scraping, and flat files like Excel or

CSV. Participants will learn how to handle missing data, incorrect formats, duplicates, and outliers—common challenges encountered when working with raw data. In addition, the course will cover best practices in data cleaning and preparation, including using tools like Excel, SQL, Python, and R.

By the end of the course, participants will be able to collect and clean data efficiently, transforming raw data into a usable format that can be trusted for analysis. This course is ideal for data analysts, researchers, business intelligence professionals, and anyone working with large or complex datasets.

## Course Content & Outline

**Section 1: Introduction to Data Sourcing and Cleaning**

- Overview of the Data Lifecycle: From Collection to Cleaning
- Why Clean Data Matters: The Importance of Accuracy and Consistency
- Common Data Issues: Missing Values, Duplicates, Outliers, and Inconsistencies

**Section 2: Sourcing Data from Different Platforms**

- Collecting Data from Databases Using SQL
- Extracting Data from APIs and Web Scraping Techniques
- Importing and Exporting Data from Excel, CSV, and Flat Files
- Handling Real-Time Data Streams and Integrating Data Sources

**Section 3: Data Cleaning Fundamentals**

- Handling Missing Data: Techniques for Imputation and Removal
- Identifying and Removing Duplicates
- Dealing with Outliers and Incorrect Formats
- Standardising Data for Consistency

**Section 4: Data Transformation and Preparation**

- Converting Unstructured Data into Structured Formats
- Transforming Data Using Excel Functions, Python, and R
- Aggregating, Merging, and Joining Datasets

- Normalising and Scaling Data for Analysis

**Section 5: Tools and Techniques for Data Cleaning**

- Cleaning Data in Excel: Functions and Tools for Data Validation
- Data Cleaning with Python: Pandas and Numpy Libraries
- Using R for Data Cleaning: dplyr and tidyr Libraries
- Automating Data Cleaning Tasks with Scripts and Macros

**Section 6: Best Practices in Data Cleaning**

- Creating Data Dictionaries and Documentation
- Ensuring Data Quality with Validation Rules
- Maintaining Data Integrity Throughout the Cleaning Process
- Continuous Monitoring and Iterative Data Cleaning

**Section 7: ase Studies and Practical Applications**

- Real-world Data Cleaning Examples from Various Industries
- Solving Complex Data Cleaning Challenges
- Applying Data Cleaning Techniques to Your Organisation’s Data

## Certificate Description

Upon successful completion of this training course, delegates will be awarded a Holistique Training Certificate of Completion. For those who attend and complete the online training course, a Holistique Training e-Certificate will be provided.

Holistique Training Certificates are accredited by the British Assessment Council (BAC) and The CPD Certification Service (CPD), and are certified under ISO 9001, ISO 21001, and ISO 29993 standards.

CPD credits for this course are granted by our Certificates and will be reflected on the Holistique Training Certificate of Completion. In accordance with the standards of The CPD Certification Service, one CPD credit is awarded per hour of course attendance. A maximum of 50 CPD credits can be claimed for any single course we currently offer.

## Categories

AI, Data and Visualisation, IT & Computer Application, Technology

## Tags

Data, Data cleaning

## Related Articles



**Data Cleaning Best Practices: Unlocking The Power Of Accurate Data**

Learn the importance of data cleaning and its benefits in enabling accurate decision-making. Discover best practices to implement an effective data cleansing strategy and overcome common data issues.