Data Cleaning with Python Training Course

Professional Development Training Course Outline

Category	Business Intelligence	**Duration**	5 Days
Base Fee	$1,500 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

Data is the backbone of modern business decision-making, yet raw datasets are often messy, incomplete, and inconsistent. Effective data cleaning is essential for accurate analysis, predictive modeling, and actionable insights. Data Cleaning with Python Training Course empowers professionals to transform disorganized datasets into reliable, structured information using Python libraries and tools. Participants will develop hands-on skills in handling missing data, detecting outliers, standardizing formats, and automating cleaning processes to enhance data quality for analytics and reporting.

Python has emerged as the leading language for data manipulation due to its versatility, simplicity, and rich ecosystem of libraries such as pandas, NumPy, and OpenRefine. By mastering Python-based data cleaning techniques, professionals can streamline workflows, improve analytical accuracy, and deliver strategic insights faster. This course is designed for individuals seeking to elevate their data handling skills, improve data-driven decision-making, and support organizational efficiency through reliable data preparation practices.

Programme Curriculum

Data Cleaning with Python Training Course

Introduction

Data is the backbone of modern business decision-making, yet raw datasets are often messy, incomplete, and inconsistent. Effective data cleaning is essential for accurate analysis, predictive modeling, and actionable insights. Data Cleaning with Python Training Course empowers professionals to transform disorganized datasets into reliable, structured information using Python libraries and tools. Participants will develop hands-on skills in handling missing data, detecting outliers, standardizing formats, and automating cleaning processes to enhance data quality for analytics and reporting.

Python has emerged as the leading language for data manipulation due to its versatility, simplicity, and rich ecosystem of libraries such as pandas, NumPy, and OpenRefine. By mastering Python-based data cleaning techniques, professionals can streamline workflows, improve analytical accuracy, and deliver strategic insights faster. This course is designed for individuals seeking to elevate their data handling skills, improve data-driven decision-making, and support organizational efficiency through reliable data preparation practices.

Course Objectives

1. Master Python libraries for data cleaning including pandas and NumPy
2. Handle missing, duplicate, and inconsistent data efficiently
3. Apply advanced data transformation and standardization techniques
4. Perform exploratory data analysis for data quality assessment
5. Detect and manage outliers and anomalies in datasets
6. Automate repetitive data cleaning tasks using Python scripts
7. Integrate data cleaning processes with data visualization tools
8. Understand data validation, integrity, and normalization techniques
9. Develop robust workflows for cleaning large-scale datasets
10. Optimize datasets for machine learning and predictive modeling
11. Gain proficiency in real-world case studies of messy data scenarios
12. Build reusable data cleaning pipelines for organizational applications
13. Implement best practices for documentation and reproducibility in data cleaning

Organizational Benefits

- Improved decision-making with accurate data
- Reduced errors in reporting and analytics
- Enhanced operational efficiency through automated cleaning processes
- Better compliance with data quality standards
- Streamlined workflows for data-intensive projects
- Increased productivity for data teams
- Reduced time spent on manual data cleaning
- Improved predictive model accuracy
- Strengthened data governance practices
- Facilitated faster insights for strategic initiatives

Target Audiences

- Data Analysts
- Business Analysts
- Data Scientists
- Python Developers
- Machine Learning Engineers
- Database Administrators
- Research Analysts
- Professionals in IT and business intelligence

Course Duration: 5 days

Course Modules

Module 1: Introduction to Data Cleaning and Python

- Overview of data cleaning concepts and importance
- Introduction to Python for data manipulation
- Understanding datasets, data types, and structures
- Key Python libraries for cleaning and transformation
- Common challenges in raw data
- Case Study: Cleaning customer survey datasets

Module 2: Handling Missing Data

- Identifying missing values in datasets
- Techniques for imputing or removing missing data
- Handling nulls in categorical and numerical fields
- Using pandas for missing data handling
- Impact of missing data on analytics
- Case Study: Cleaning sales transaction data

Module 3: Removing Duplicates and Inconsistencies

- Identifying duplicate rows and records
- Techniques to remove or consolidate duplicates
- Standardizing inconsistent entries
- Handling inconsistent formats across datasets
- Ensuring uniqueness and consistency
- Case Study: Deduplicating employee records

Module 4: Data Transformation and Standardization

- Data normalization and scaling techniques
- Formatting dates, strings, and numeric fields
- Converting data types for consistency
- Handling categorical data transformation
- Applying mapping and encoding methods
- Case Study: Standardizing product catalog data

Module 5: Outlier Detection and Management

- Identifying outliers in numerical datasets
- Techniques to handle or remove outliers
- Understanding impact of outliers on analysis
- Using statistical methods for outlier detection
- Visualizing outliers with Python tools
- Case Study: Detecting anomalies in financial datasets

Module 6: Automating Data Cleaning with Python

- Writing reusable Python scripts for cleaning
- Automating repetitive data cleaning tasks
- Using functions and loops for efficiency
- Scheduling data cleaning workflows
- Logging and error handling in scripts
- Case Study: Automating cleaning of daily sales data

Module 7: Data Validation and Integrity

- Understanding data validation rules
- Checking data for integrity and accuracy
- Techniques for ensuring consistent data entry
- Using Python to validate and audit datasets
- Best practices for data quality assurance
- Case Study: Validating inventory management data

Module 8: Preparing Data for Analysis and Machine Learning

- Cleaning datasets for analytics and visualization
- Optimizing data for machine learning models
- Feature engineering for predictive analysis
- Integrating cleaned data with BI tools
- Documentation and reproducibility of cleaned data
- Case Study: Preparing dataset for predictive sales modeling

Training Methodology

- Interactive instructor-led sessions with practical exercises
- Hands-on labs using real-world datasets
- Step-by-step guided Python scripts and projects
- Group discussions and peer learning for problem-solving
- Assignments with real-time feedback
- Case study analysis to apply techniques in practical scenarios

Register as a group from 3 participants for a Discount

Send us an email: info@fineskilltrainingcenter.org or call *+254769199797*

Certification

Upon successful completion of this training, participants will be issued with a globally- recognized certificate.

Tailor-Made Course

We also offer tailor-made courses based on your needs.

Key Notes

a. The participant must be conversant with English.

b. Upon completion of training the participant will be issued with an Authorized Training Certificate

c. Course duration is flexible and the contents can be modified to fit any number of days.

d. The course fee includes facilitation training materials, 2 coffee breaks, buffet lunch and A Certificate upon successful completion of Training.

e. One-year post-training support Consultation and Coaching provided after the course.

f. Payment should be done at least a week before commence of the training, to Fineskill Training Center account, as indicated in the invoice so as to enable us prepare better for you.

Upcoming Scheduled Sessions

Start Date	End Date	Location	Price
23 Nov 2026	27 Nov 2026	Online	$1,000
23 Nov 2026	27 Nov 2026	Physical	$1,500
23 Nov 2026	27 Nov 2026	Physical	$3,000
23 Nov 2026	27 Nov 2026	Physical	$3,000
23 Nov 2026	27 Nov 2026	Physical	$6,000
23 Nov 2026	27 Nov 2026	Physical	$4,500
23 Nov 2026	27 Nov 2026	Physical	$6,000
23 Nov 2026	27 Nov 2026	Physical	$6,000

Ready to enroll or request a customized program? Book online or contact us:

Register Online Now

Email: info@fineskilltrainingcenter.com | Website: www.fineskilltrainingcenter.com