Data Munging and Wrangling with Pandas/dplyr in Research Training Course

Professional Development Training Course Outline

Category	Research and Data Analysis	**Duration**	5 Days
Base Fee	$1,500 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

In today's data-driven world, researching sensitive topics—such as gender-based violence, mental health, minority rights, or health disparities—requires not only ethical handling but also advanced data management skills. Data Munging and Wrangling with Pandas/dplyr Training Course equips learners with the tools to extract, clean, transform, and structure complex and delicate datasets using Python’s Pandas and R’s dplyr libraries. The course emphasizes data integrity, confidentiality, bias minimization, and analytical accuracy, making it ideal for researchers working with critical, confidential, or socially impactful datasets.

By integrating real-world case studies, hands-on coding sessions, and contextual learning, this course empowers participants to handle noisy, incomplete, and ethically challenging data with precision. Whether you're a public health researcher, social scientist, journalist, or data analyst, this course provides an essential bridge between technical data processing and socially responsible research practices.

Programme Curriculum

Data Munging and Wrangling with Pandas/dplyr Training Course

Introduction

In today's data-driven world, researching sensitive topics—such as gender-based violence, mental health, minority rights, or health disparities—requires not only ethical handling but also advanced data management skills. Data Munging and Wrangling with Pandas/dplyr Training Course equips learners with the tools to extract, clean, transform, and structure complex and delicate datasets using Python’s Pandas and R’s dplyr libraries. The course emphasizes data integrity, confidentiality, bias minimization, and analytical accuracy, making it ideal for researchers working with critical, confidential, or socially impactful datasets.

By integrating real-world case studies, hands-on coding sessions, and contextual learning, this course empowers participants to handle noisy, incomplete, and ethically challenging data with precision. Whether you're a public health researcher, social scientist, journalist, or data analyst, this course provides an essential bridge between technical data processing and socially responsible research practices.

Course Objectives

1. Understand the principles of ethical research with sensitive data.
2. Gain proficiency in data wrangling using Pandas and dplyr.
3. Identify and mitigate biases and outliers in sensitive datasets.
4. Learn best practices for data anonymization and privacy protection.
5. Automate data cleaning workflows for reproducibility.
6. Master handling of missing, incomplete, or corrupted data.
7. Utilize descriptive statistics for exploratory data analysis.
8. Apply grouping, filtering, and summarization techniques for insights.
9. Visualize sensitive data using safe, aggregated plots.
10. Conduct data validation and integrity checks.
11. Learn cross-platform coding (Python & R) for wrangling sensitive datasets.
12. Develop custom functions and pipelines for repeatable wrangling.
13. Apply knowledge to real-life ethical data case studies.

Target Audience

1. Public Health Researchers
2. Social Science Academics
3. NGO & Policy Analysts
4. Journalists & Investigative Reporters
5. Clinical Researchers
6. Data Analysts working with surveys
7. Graduate Students & Scholars
8. Government & Development Agency Researchers

Course Duration: 5 days

Course Modules

Module 1: Introduction to Sensitive Data Research

- Understanding ethical concerns in sensitive research
- Examples of sensitive topics in real-world studies
- Legal frameworks: GDPR, HIPAA, local laws
- Risks and responsibilities of data handling
- Introduction to anonymization techniques
- **Case Study:** Handling data on domestic violence reports

Module 2: Introduction to Pandas and dplyr

- Setting up environments in Python (Pandas) and R (dplyr)
- DataFrames: creation, structure, and access
- Importing and exporting datasets (CSV, Excel, JSON)
- Syntax comparisons: Pandas vs. dplyr
- Choosing tools based on data context
- **Case Study:** Comparing health survey data in Pandas and dplyr

Module 3: Cleaning and Preparing Sensitive Data

- Removing duplicates, fixing structural issues
- Formatting dates, strings, and numeric values
- Handling incorrect or inconsistent labels
- Dealing with missing and null data
- Creating cleaning scripts for reproducibility
- **Case Study:** Preprocessing mental health survey data

Module 4: Exploratory Data Analysis for Sensitive Data

- Descriptive statistics with context
- Detecting trends while avoiding disclosure
- Boxplots, histograms, and safe aggregations
- Identifying and addressing outliers
- Masking identifiable information
- **Case Study:** Analyzing depression trends without violating privacy

Module 5: Advanced Data Wrangling Techniques

- Merging and joining data from multiple sources
- Filtering by conditions relevant to sensitive cases
- Grouping and summarizing for subpopulations
- Reshaping data with pivot/melt/gather/spread
- Writing reusable wrangling functions
- **Case Study:** Combining hospital and community data on HIV

Module 6: Data Privacy and Anonymization

- Types of identifiers and quasi-identifiers
- De-identification, pseudonymization, and k-anonymity
- Balancing utility with privacy
- Tools for data masking and encryption
- Ensuring ethical publication practices
- **Case Study:** Publishing anonymized gender-based violence data

Module 7: Workflow Automation and Reproducibility

- Building wrangling pipelines in Pandas and dplyr
- Using Jupyter Notebooks and RMarkdown
- Version control and documentation
- Parameterizing scripts for multiple datasets
- Best practices in collaborative research coding
- **Case Study:** Reproducing a data wrangling pipeline for refugee camp data

Module 8: Final Project and Integrated Case Study

- Selecting and importing a sensitive dataset
- Applying full wrangling and munging pipeline
- Producing ethical EDA and summary report
- Peer reviewing anonymization decisions
- Presenting results using safe visualizations
- **Capstone Case Study:** Wrangling data on school bullying and self-harm

Training Methodology

- Hands-on exercises and live coding using Pandas and dplyr
- Ethical scenario simulations with guided discussions

- Group case study reviews to apply theoretical frameworks
- Cross-platform labs (Python and R) for code diversity
- Interactive Q&A and peer critique sessions

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

Start Date	End Date	Location	Price
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