Open Source Tools for Data Science 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

The rise of open-source tools has transformed the landscape of data science research, providing researchers, analysts, and institutions with powerful, cost-effective, and customizable platforms. Open Source Tools for Data Science Research Training Course is designed to equip participants with in-demand skills in data manipulation, statistical modeling, data visualization, machine learning, and reproducible research. Leveraging popular tools such as Python, R, Jupyter Notebooks, Git, and Apache Spark, this course provides a practical, hands-on learning experience essential for academic and professional success in today’s data-driven environment.

With the growth of big data, AI integration, and collaborative research platforms, understanding open-source data science tools has become crucial across sectors. Whether you're engaged in academic research, public policy, healthcare analytics, or business intelligence, mastering these tools enhances productivity, supports open science principles, and ensures compliance with FAIR (Findable, Accessible, Interoperable, and Reusable) data practices. This course will help participants navigate the open-source ecosystem and apply these tools to real-world research problems using industry-standard best practices.

Programme Curriculum

Open Source Tools for Data Science Research Training Course

Introduction

The rise of open-source tools has transformed the landscape of data science research, providing researchers, analysts, and institutions with powerful, cost-effective, and customizable platforms. Open Source Tools for Data Science Research Training Course is designed to equip participants with in-demand skills in data manipulation, statistical modeling, data visualization, machine learning, and reproducible research. Leveraging popular tools such as Python, R, Jupyter Notebooks, Git, and Apache Spark, this course provides a practical, hands-on learning experience essential for academic and professional success in today’s data-driven environment.

With the growth of big data, AI integration, and collaborative research platforms, understanding open-source data science tools has become crucial across sectors. Whether you're engaged in academic research, public policy, healthcare analytics, or business intelligence, mastering these tools enhances productivity, supports open science principles, and ensures compliance with FAIR (Findable, Accessible, Interoperable, and Reusable) data practices. This course will help participants navigate the open-source ecosystem and apply these tools to real-world research problems using industry-standard best practices.

Course Objectives

1. Understand the fundamentals of open-source tools in data science.
2. Master data wrangling using Python and R.
3. Build interactive visualizations with libraries like ggplot2 and Plotly.
4. Perform statistical analysis and hypothesis testing in R.
5. Use Jupyter Notebooks for collaborative and reproducible research.
6. Integrate version control with Git and GitHub.
7. Apply machine learning techniques using Scikit-learn and TensorFlow.
8. Implement big data analytics using Apache Spark and Hadoop.
9. Automate workflows with open-source scripting.
10. Ensure data integrity with reproducibility and data provenance tools.
11. Analyze real-world datasets using open science frameworks.
12. Enhance cloud-based research using open-source platforms.
13. Conduct ethical and FAIR-aligned data science research.

Target Audiences

1. Academic researchers in STEM and social sciences
2. Data scientists seeking cost-effective research tools
3. Graduate and postgraduate students
4. Policy analysts and government researchers
5. Healthcare data analysts
6. IT professionals transitioning into data science
7. Open science advocates
8. Research and development teams in NGOs

Course Duration: 5 days

Course Modules

Module 1: Introduction to Open Source Data Science Ecosystem

- Overview of open-source principles
- Importance in scientific research
- Comparison with proprietary tools
- Key platforms: Python, R, Jupyter
- Licensing and collaboration models
- **Case Study**: Transitioning a university lab from Excel to open-source tools

Module 2: Data Wrangling with Python and R

- Data import/export (CSV, JSON, Excel)
- Cleaning and preprocessing techniques
- Handling missing values
- Data transformation with dplyr and Pandas
- Integrating SQL with open-source tools

- **Case Study**: Cleaning a national survey dataset for analysis

Module 3: Data Visualization and Reporting

- Visualization principles and best practices
- Creating plots using ggplot2 and seaborn
- Interactive dashboards with Plotly and Dash
- Reporting with RMarkdown and Jupyter
- Exporting visuals for publication
- **Case Study**: Visualizing climate change trends in East Africa

Module 4: Statistical Analysis Using R

- Descriptive statistics and data summaries
- Hypothesis testing and confidence intervals
- Regression analysis
- ANOVA and multivariate statistics
- Reproducible statistical reports
- **Case Study**: Statistical analysis of patient outcome data

Module 5: Machine Learning with Open-Source Frameworks

- Introduction to machine learning concepts
- Supervised and unsupervised learning
- Implementing models in Scikit-learn and TensorFlow
- Cross-validation and model evaluation
- Feature engineering techniques
- **Case Study**: Predicting student performance using education data

Module 6: Big Data Analytics with Apache Spark

- Understanding big data architecture
- Working with PySpark and R with Sparklyr
- Distributed data processing techniques
- Streaming and real-time analytics
- Integrating with cloud data sources
- **Case Study**: Analyzing mobile phone data for migration patterns

Module 7: Version Control and Collaboration

- Basics of Git and GitHub
- Branching, merging, and pull requests
- Managing project repositories
- Collaborating on research code
- Open science and data sharing
- **Case Study**: Building a collaborative GitHub repository for a journal article

Module 8: Reproducibility and FAIR Research Practices

- Importance of reproducible research
- Workflow automation with Make and Snakemake
- Documenting data and code
- Applying FAIR principles
- Sharing research with Zenodo, Figshare
- **Case Study**: Publishing a fully reproducible academic project with Zenodo DOI

Training Methodology

- Hands-on coding sessions and exercises
- Guided case studies using real-world datasets
- Peer-to-peer collaboration and group activities
- Weekly project-based assignments
- Expert-led webinars and Q&A forums
- Access to a GitHub-based learning repository

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

Start Date	End Date	Location	Price
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