Statistical Genetics and Bioinformatics for Researchers Training Course

Professional Development Training Course Outline

Category	Research and Data Analysis	**Duration**	10 Days
Base Fee	$3,000 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

Handling missing data is one of the most critical challenges in modern data science and analytics. Data Imputation Techniques for Missing Data Training Course equips learners with in-depth knowledge and hands-on experience in dealing with incomplete datasets using statistical, machine learning, and AI-powered imputation methods. This training ensures data integrity, improves model accuracy, and supports effective decision-making by addressing data gaps in real-world datasets.

With the explosive growth of data-driven industries, mastering missing data management and data imputation strategies has become essential. This course blends theory with practice, providing participants with tools to apply predictive modeling, ML-based imputations, multiple imputation techniques, and advanced analytics. Whether you're a data analyst, machine learning engineer, or a researcher, this course will help you optimize datasets and achieve better model performance.

Programme Curriculum

Data Imputation Techniques for Missing Data Training Course

Introduction

Handling missing data is one of the most critical challenges in modern data science and analytics. Data Imputation Techniques for Missing Data Training Course equips learners with in-depth knowledge and hands-on experience in dealing with incomplete datasets using statistical, machine learning, and AI-powered imputation methods. This training ensures data integrity, improves model accuracy, and supports effective decision-making by addressing data gaps in real-world datasets.

With the explosive growth of data-driven industries, mastering missing data management and data imputation strategies has become essential. This course blends theory with practice, providing participants with tools to apply predictive

modeling, ML-based imputations, multiple imputation techniques, and advanced analytics. Whether you're a data analyst, machine learning engineer, or a researcher, this course will help you optimize datasets and achieve better model performance.

Course Objectives

1. Understand the impact of missing data on predictive analytics and model performance.
2. Identify different types of missing data: MCAR, MAR, and MNAR.
3. Apply statistical methods for data imputation (mean, median, mode).
4. Explore machine learning techniques for imputing missing values.
5. Implement KNN imputation in Python and R.
6. Learn Multiple Imputation using MICE.
7. Use deep learning for advanced data imputation (Autoencoders, GANs).
8. Conduct missing data analysis using pandas, scikit-learn, and TensorFlow.
9. Evaluate imputation methods using cross-validation.
10. Develop data cleaning pipelines in real-time systems.
11. Improve data quality in business intelligence dashboards.
12. Handle missing time series data using forward and backward fill.
13. Design imputation workflows for healthcare, finance, and social sciences.

Target Audience

1. Data Scientists
2. Machine Learning Engineers
3. Statisticians
4. Health Informatics Professionals
5. Social Science Researchers
6. Business Intelligence Analysts
7. Software Developers
8. Graduate Students in Data-Related Fields

Course Duration: 10 days

Course Modules

Module 1: Introduction to Missing Data

- Define missing data types
- Consequences of missing values
- Real-world implications
- Overview of imputation
- When to drop vs. impute
- **Case Study: E-commerce data with missing customer demographics**

Module 2: Mechanisms of Missingness (MCAR, MAR, MNAR)

- MCAR: Missing completely at random
- MAR: Missing at random
- MNAR: Not missing at random
- Techniques to identify patterns
- Statistical testing for mechanisms
- **Case Study: Health survey data analysis**

Module 3: Descriptive Analysis for Missing Data

- Visualizing missing data with heatmaps

- Using Python’s missingno and seaborn
- Pattern analysis using correlation
- Exploratory Data Analysis (EDA)
- Missing data summary statistics
- **Case Study: Customer churn dataset EDA**

Module 4: Simple Statistical Imputation

- Mean, median, and mode imputation
- Imputation limitations and bias
- When simple imputation is appropriate
- Handling categorical data
- Using pandas and R base functions
- **Case Study: Educational test scores analysis**

Module 5: Advanced Statistical Methods

- Regression imputation
- Stochastic regression
- Expectation Maximization (EM)
- Linear models for imputation
- Dealing with collinearity
- **Case Study: Real estate price prediction dataset**

Module 6: K-Nearest Neighbors (KNN) Imputation

- Theory behind KNN imputation
- Choosing optimal k-values
- Handling numerical/categorical variables
- Implementation in scikit-learn
- Performance benchmarking
- **Case Study: Telecom usage data**

Module 7: Multiple Imputation with MICE

- Overview of chained equations
- Iterative modeling in MICE
- Using fancyimpute and statsmodels
- Diagnostics and convergence
- Randomness vs. determinism
- **Case Study: Financial income survey**

Module 8: Time Series Data Imputation

- Missing data in time series
- Forward fill and backward fill
- Interpolation techniques
- Seasonal adjustment methods
- Smoothing and ARIMA-based methods
- **Case Study: Weather station data imputation**

Module 9: Deep Learning Imputation Methods

- Autoencoders for imputation
- GANs and missing data generation
- Neural network-based inference

- TensorFlow and PyTorch workflows
- Performance and limitations
- **Case Study: Medical imaging dataset**

Module 10: Handling Missing Data in Big Data

- Apache Spark data imputation
- Distributed imputation pipelines
- Efficient memory and CPU use
- Streaming data challenges
- Fault-tolerant designs
- **Case Study: IoT sensor data pipeline**

Module 11: Data Imputation in Healthcare

- Dealing with incomplete patient records
- HIPAA and ethical considerations
- Clinical trial imputation strategies
- EHR system integration
- Validity and reproducibility
- **Case Study: Hospital admission dataset**

Module 12: Data Imputation for Business Intelligence

- Dashboard readiness
- Imputing for KPIs
- Ensuring data consistency
- Real-time data sync
- Linking imputation with ETL
- **Case Study: Sales forecasting dashboard**

Module 13: Imputation and Predictive Modeling

- Preprocessing pipeline design
- Impact on classification accuracy
- Data leakage prevention
- Model testing after imputation
- Feature engineering synergy
- **Case Study: Credit scoring models**

Module 14: Evaluation and Validation

- Cross-validation with imputed datasets
- Comparing models with/without imputation
- Quantitative evaluation metrics
- Benchmarking various methods
- Reporting reproducibility
- **Case Study: Bank loan application data**

Module 15: Ethics, Transparency, and Best Practices

- Bias and fairness in imputation
- Transparency in model design
- Explainable AI and imputation
- Documenting decisions
- Regulatory compliance (e.g., GDPR)

- **Case Study: Government census data**

Training Methodology

- Interactive lectures with real-world datasets
- Hands-on coding labs in Python and R
- Case study analysis and presentations
- Peer discussion and collaborative exercises
- Quiz assessments and feedback sessions
- Capstone project with real-time imputation task

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	04 Dec 2026	Online	$2,000
23 Nov 2026	04 Dec 2026	Physical	$3,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$6,000

Start Date	End Date	Location	Price
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$9,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$12,000

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

Register Online Now

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