

Data Science Developer Foundations and Certification Preparation

Python-based data preparation, statistics and introductory machine learning

FA-0720 · 7 days

Data & Analytics · Software Development

Introduction

Build a foundation in Python-based data science, from data preparation and visualisation to selected supervised and unsupervised models. The seven-day programme combines guided implementation, statistical interpretation and a review exercise using small datasets.

The original source refers to the Global Tech Council Certified Data Science Developer credential. This training does not award that certification, guarantee employment or imply that every algorithm can be mastered in one week; current assessment terms must be checked with the provider.

Learning outcomes

- Use foundational Python, NumPy and pandas to work with tabular data.
- Prepare and visualise sample datasets and explain key statistical summaries.
- Distinguish regression, classification, clustering and dimensionality-reduction goals.
- Implement selected model examples and evaluate them with suitable data separation and metrics.
- Recognise recommendation, association-rule and time-series approaches and their assumptions.
- Identify modelling limitations and prepare for further practice or separately administered certification assessment.

Prerequisites

- Basic computer, file-management and internet skills and high-school-level mathematics.
- A supported Python 3 environment with compatible packages or a supplied notebook setup; approved installation rights only where needed.
- For remote delivery, a reliable connection and conferencing equipment; a second screen is helpful, not compulsory.
- A Google account is needed only if the chosen lab uses Colab; no unrestricted root access or fixed attendance-based certification promise is implied.

Training outline

1. Day 1 — Programming and data-science context

- Set up Python and packages; variables, collections, loops, functions, scope and basic object-oriented concepts.
- Use exceptions and simple text/regular-expression examples; consult documentation rather than memorise history.
- Introduce R concepts through a brief comparison; the hands-on progression is primarily Python, not a full second-language course.

2. Day 2 — Arrays, tables and statistics

- NumPy arrays, matrices and selected operations; pandas DataFrames and CSV import/export.
- Descriptive statistics, distributions and basic probability; linear-algebra concepts needed for the examples.
- Introduce inferential statistics and hypothesis tests with assumptions and interpretation limits.
- Practical work: inspect and summarise a small dataset.

3. Day 3 — Preprocessing and visualisation

- Missing values, categorical encoding, scaling and outlier decisions.
- Separate training and evaluation data before fitting transformations; use an appropriate pipeline to reduce leakage.
- Create selected Matplotlib or Seaborn charts and check whether they support the analytical question.
- Build a documented preparation template and inspect its effect on the sample data.

4. Day 4 – Regression and evaluation

- Simple, multiple and polynomial regression: intuition, suitable problem types and assumptions.
- Introduce decision-tree and random-forest regression and compare selected implementations.
- Choose meaningful evaluation measures, compare a baseline and recognise overfitting.
- Review results without implying causal explanation or guaranteed predictive performance.

5. Day 5 – Classification

- Compare logistic regression, trees, random forests, SVM, naïve Bayes and k-nearest neighbours at an introductory level.
- Implement selected classifiers; broader algorithms use worked demonstrations.
- Use a confusion matrix and appropriate metrics; consider class imbalance and data limitations.
- Review validation and model settings without using held-out data to choose preprocessing.

6. Day 6 – Unsupervised and recommendation methods

- K-means and hierarchical clustering; inspect scaling, cluster assumptions and interpretation.
- PCA for unsupervised dimensionality reduction; distinguish supervised linear discriminant analysis.
- Content-based and collaborative recommendation approaches; assess a small supplied example.
- Apriori and market-basket association rules; distinguish association from causation.

7. Day 7 – Time series, consolidation and preparation

- Time-series structure, basic ARIMA concepts and chronological evaluation using a small example.
- Review statistical assumptions, model residuals and limitations.
- Complete a bounded analysis exercise and document preparation, evaluation and interpretation.
- Consolidate concepts with practice questions; consult current provider scope and exam policies before separately registering.

Certification context

The course preserves the source's certification-preparation context but is not represented as the provider's current self-paced product or an official credential award. Confirm current exam scope, access and retake conditions directly with Global Tech Council.