

Big Data Analytics Foundations for Technical Professionals

Python, SQL, statistics and data storytelling

FA-0589 · 15 days

Data & Analytics · Software Development

Build a foundation in Python data workflows, SQL, descriptive statistics, introductory machine learning and analytical communication.

Introduction

This fifteen-day foundation course introduces technical professionals to programming and analytical work without requiring prior coding experience. It progresses through Python, descriptive statistics, data preparation and visualisation, introductory model examples, SQL and data storytelling.

Guided projects use small prepared datasets and illustrative applications. Malaysian contexts may be used where relevant, but exercises do not imply access to a particular business's confidential records. Machine learning, NLP and time-series work are introductory; the programme builds prerequisites for further big-data study rather than claiming distributed-systems or data-science mastery.

Learning outcomes

The course teaches participants to:

- Write basic Python programs using data structures, conditionals, loops and functions.
- Compute and interpret descriptive statistics and basic distributions.
- Clean and manipulate data with pandas and visualise it with Matplotlib or Seaborn.
- Explore basic regression, classification, clustering, NLP and time-series examples.
- Query a prepared SQL Server database using SELECT, joins and aggregates.
- Create a small analytical narrative or dashboard that states its evidence and limitations.

Prerequisites

- Basic computer literacy and file-management skills.
- Basic mathematics; no prior programming experience is required.
- A suitable Python/notebook environment and access to a prepared SQL Server learning database.
- Internet access and approved public or synthetic datasets for the exercises.

Training outline

1. Days 1–3 — Introduction to Python

Day 1 — Environment, Syntax and Variables

- Python's role in data analysis and technical workflows.
- Setting up Python Environment
- Python Syntax and Variables
- A small 'Hello' or text-game exercise.

Day 2 — Strings and Collections

- Strings and Text Manipulation
- Lists and Dictionaries
- A simple rule-based food-recommendation exercise using fictional sample data; distinguish it from a trained recommendation model.

Day 3 — Control Flow and Functions

- Conditional Statements
- Loops
- Functions
- A small weather-display application using prepared data.

2. Days 4–5 — Descriptive Statistics

Day 4 — Data Types and Central Tendency

- Types of Data

- Measures of Central Tendency
- Analyse synthetic exam-score data.

Day 5 — Variability and Distribution

- Measures of Variability
- Data Distribution
- Describe a synthetic health-style dataset without clinical conclusions or identifiable records.

3. Days 6–11 — Python for Analytics

Day 6 — Data Manipulation with pandas

- Reading and Writing Data
- Data Cleaning
- Clean a prepared public or synthetic demographic dataset.

Day 7 — Visualisation

- Introduction to Matplotlib and Seaborn
- Visualise illustrative tourism statistics and check chart labels and scales.
- Explore a licensed public comparison dataset, such as historical aggregate election statistics; describe rather than claim to forecast outcomes.

Days 8–9 — Introductory Machine Learning

- Classification and Regression
- Clustering
- Illustrative property-price regression using prepared data and held-out evaluation; it is not a valuation service or guaranteed forecast.
- Separate training and held-out evaluation data; fit learned preprocessing on training data only.

Days 10–11 — NLP and Time-Series Introductions

- Introductory text processing with NLTK and a selected classification or analysis demonstration.
- Time-series concepts and a small time-respecting analysis example.
- Analyse historical or synthetic market-style trends and explain uncertainty; no trading or return claims.

4. Days 12–13 — SQL Essentials

Day 12 — SQL Server and Basic Queries

- Introduction to Microsoft SQL Server and relational data.
- SQL Syntax and Data Types
- SELECT Queries
- Query a fictional business learning database.

Day 13 — Joins and Aggregates

- JOIN Operations
- Aggregate Functions
- A small SQL-backed application or game using prepared tables.

5. Days 14–15 — Data Storytelling

Day 14 — Narratives and Visual Evidence

- Principles of evidence-led storytelling.
- Data Narratives and Visualization
- Build a narrative around a public or synthetic economic-trend dataset and state limitations.

Day 15 — Dashboards and Presentation

- Introduction to Dashboards
- Compare selected dashboard tools available in the learning environment.
- Create a small public-health-style dashboard from approved historical or synthetic data; clearly label dates and source limitations.
- Present the selected project, its data-quality constraints and appropriate next steps.