Data Science with Python

Analyse, visualise and model data in Python, from Pandas to machine learning and Spark.

DURATION	FORMAT	LEVEL	LABS
60 hours	Onsite, online or hybrid	Intermediate	Cloud sandbox included

About this programme

An end-to-end data science programme in Python: NumPy and Pandas for analysis, Matplotlib, Seaborn and Plotly for visualisation, scikit-learn for machine learning, and introductions to Spark, TensorFlow and generative models, taught through exercises and projects.

Who it is for: Analysts, engineers and graduates moving into data science

What your team will be able to do

- Set up a Python data science environment with Anaconda and Jupyter
- Clean, reshape, join and aggregate data with NumPy and Pandas
- Build clear static, interactive and geographic visualisations
- Train and evaluate regression, classification, clustering and recommender models
- Process large datasets with PySpark and build basic neural networks

Before you start

- Basic programming in any language
- School-level mathematics and statistics

Course outline

1. Python for data science

- Environment set-up: Anaconda, Jupyter notebooks and an IDE
- Python crash course: syntax, OOP concepts and modules
- Optional: virtual environments
- Python exercises and solutions

2. Data analysis with NumPy and Pandas

- NumPy arrays, indexing and operations
- Pandas Series and DataFrames
- Missing data, group-by, merging, joining and concatenating
- Data input and output
- Exercises: salaries and e-commerce purchases datasets

3. Data visualisation

- Matplotlib fundamentals
- Seaborn: distribution, categorical, matrix and regression plots
- Pandas built-in plotting
- Interactive charts with Plotly and Cufflinks
- Geographical plotting with choropleth maps

4. Machine learning with scikit-learn

- Linear and logistic regression
- K-nearest neighbours, decision trees and random forests
- Support vector machines
- K-means clustering and principal component analysis
- Recommender systems and NLP, each with a project

5. Big data and deep learning

- Big data and Spark overview
- Setting up PySpark locally and on AWS EC2
- RDD transformations and actions with lambda expressions
- Neural network theory and TensorFlow basics
- MNIST with a multi-layer perceptron

6. Introduction to generative AI

- Generative models: GANs and variational autoencoders
- Training, loss functions and evaluation metrics for generative models
- Text, image, audio and video generation, and style transfer
- Limitations, ethics, interpretability and applications

Sample schedule

8 days onsite, or 15 half-day sessions online. We fit days and session lengths to your team's calendar.

Day 1	8 h	Python for data science
Day 2	8 h	Python for data science (continued); Data analysis with NumPy and Pandas
Day 3	8 h	Data analysis with NumPy and Pandas (continued); Data visualisation
Day 4	8 h	Data visualisation (continued); Machine learning with scikit-learn
Day 5	8 h	Machine learning with scikit-learn (continued)
Day 6	8 h	Machine learning with scikit-learn (continued); Big data and deep learning
Day 7	8 h	Big data and deep learning (continued); Introduction to generative AI
Day 8	4 h	Introduction to generative AI (continued)

Assessment and certification

How learning is checked

- A short knowledge check on day one and again at the end, so you can see each person's progress
- Graded hands-on labs in every module, reviewed by the trainer
- A final exercise that applies the programme to a problem like your team's own

Certificate and exams

- Participants who complete the programme receive a Cognitive Networks AI certificate of completion.

Formats and cohort size

- Onsite: At your office or a venue you choose, in full days.
- Online, live: Instructor-led over video, usually in half-day sessions so people can keep up with work.
- Hybrid: Some days in a room together, the rest online. Useful when a team works across cities.

Cohort size: 15 to 25 people per trainer. Larger teams run as parallel cohorts on the same outline, so everyone learns the same thing.

Pricing: We quote per cohort, based on duration, format and number of cohorts. There is no per-seat price list: tell us your team and we send a proposal within one business day.

Bring this programme to your team

Request a proposal at cognitivenetworks.ai/contact?programme=data-science-python

reachus@cognitivenetworks.ai · Bangalore +91 72043 05303 · Dallas +1 (303) 827-7415