Faysal Aziz

Data Science for Beginners

Using python

FA-0157 · 5-day

Software Development · Data & Analytics

Introduction

Data Science for Beginners is a comprehensive 5-day training course designed specifically for engineers with little to no prior experience in data science. The course will provide a strong foundation in data science concepts, techniques, and tools, utilizing Python programming language and focusing on real-world applications. Throughout the training, participants will be engaged in hands-on activities using Google Colab as the primary IDE and Anaconda as a backup IDE. By the end of the course, attendees will be equipped with the necessary skills to identify, plan, and execute data-driven tasks and projects within their organizations.

Learning outcomes

Upon successful completion of the course, participants will be able to:

- Perform hands-on tasks using Python in Google Colab and Anaconda.
- Identify potential data-driven tasks and projects in their organization.
- Formulate a workflow for a data analytic project.
- Collect, store, and manage data sets.
- Perform simple exploratory analysis on data.
- Build data storytelling to communicate insights effectively.

Prerequisites

This course assumes no prior knowledge of data science or Python. However, participants should have

- Basic computer skills
- Familiarity with engineering concepts.

They also need to:

- Have unrestricted internet connectivity for using Google's Colab notebook.
- Alternatively, if they wish to use their own PC, they should have Administrative / root / sudo access to be able to install and create environments in Anaconda.
- Installation rights of 3rd party software using conda forge and / or pip.

Training outline

1. Day 1: Introduction to Data Science and Python

Introduction to data science

- Definition and history
- Data science process
- Roles and skill sets in data science
- Applications in engineering

Overview of Python programming language

- Python's role in data science
- Comparison with other programming languages

Setting up Google Colab and Anaconda

- Creating accounts
- Exploring the interface
- Installation and setup of Anaconda

Basic Python programming concepts

- Variables and data types
- Arithmetic and logical operations
- Conditional statements (if, elif, else)
- Loops (for, while)

Python data structures

- Lists
- Tuples
- Dictionaries
- Sets

- Hands-on exercises: Basic Python operations

2. Day 2: Data Collection and Storage

Introduction to data sources and types

- Structured data
- Semi-structured data
- Unstructured data

Data collection methods

- Web scraping
- APIs
- Databases

Python libraries for data collection

- Requests
- BeautifulSoup
- Pandas

Storing and managing data sets

- File formats (CSV, JSON, SQL)
- Data storage options (local, cloud)
- Data organization best practices

- Hands-on exercises: Collecting and storing data

3. Day 3: Data Preprocessing and Exploratory Data Analysis (EDA)

Introduction to data preprocessing and cleaning

- Importance of clean data
- Common data quality issues

Handling missing and inconsistent data

- Identifying missing data
- Imputation methods
- Data consistency checks

Data transformation and normalization techniques

- Log transformations
- Scaling and normalization
- Categorical encoding

Python libraries for data preprocessing

- Pandas

- NumPy

Introduction to EDA and visualization

- EDA process
- Descriptive statistics
- Data visualization types

Python libraries for EDA and visualization

- Pandas
- Matplotlib
- Seaborn

- Hands-on exercises: Data preprocessing and EDA

4. Day 4: Data Analysis Techniques

Introduction to descriptive and inferential statistics

- Measures of central tendency
- Measures of dispersion
- Correlation and covariance
- Hypothesis testing

Overview of machine learning algorithms

- Regression algorithms (linear, logistic)
- Classification algorithms (k-NN, decision trees, SVM)
- Clustering algorithms (k-means, DBSCAN)

Introduction to model selection and validation techniques

- Train-test split
- Cross-validation
- Model evaluation metrics

Python libraries for data analysis

- Scikit-learn
- Statsmodels

- Hands-on exercises: Building and evaluating machine learning models

5. Day 5: Data Storytelling and Project Workflow

Introduction to data storytelling and visualization best practices

- Storytelling concepts
- Audience considerations

- Visualization design principles

Building interactive visualizations with Python

- Plotly
- Bokeh

Crafting a data-driven narrative

- Structuring the narrative
- Choosing the right visualizations
- Telling a compelling story

Overview of project workflow and best practices

- CRISP-DM methodology
- Agile development in data science
- Collaboration and version control

Hands-on exercises: Building a data storytelling project

- Identifying a story from the data
- Creating interactive visualizations
- Assembling a data-driven narrative

- Course recap, Q&A, and next steps for further learning
- Review of key concepts and techniques covered in the course
- Q&A session to address any remaining questions or concerns

Resources for further learning and skill development

- Online courses and tutorials
- Books and articles
- Industry conferences and workshops

Networking and community involvement

- Data science meetups and events
- Online forums and discussion groups
- Social media and professional networks

With this detailed breakdown, each day of the 5-day training is organized into specific subtopics to ensure that participants gain a thorough understanding of data science concepts and techniques. This will help participants to confidently apply their new skills to real-world data-driven tasks and projects within their organizations.

Furthermore, each topic shall be followed up with short assessments to ensure that each learner is able to grasp the topics being covered. Apart from individual assignments, there shall also be group assignments.

The advantage of this meticulously crafted outline over others lies in its comprehensive coverage of essential data science concepts, techniques, and real-world applications specifically tailored for engineers. It guides participants step by step, ensuring a solid understanding of the subject matter even

for complete beginners. Furthermore, the trainers who will conduct this training possess a remarkable combined experience of over 30 years in the industry.

This wealth of knowledge and experience enables them to provide invaluable insights, practical examples, and industry best practices that go beyond textbook learning. As a result, participants will not only gain theoretical knowledge but also benefit from the trainers' expertise, enhancing their ability to apply data science skills effectively in their professional roles. This unique combination of a well-structured outline and seasoned trainers sets this training apart from others, delivering a high-quality learning experience that prepares participants for success in the rapidly evolving field of data science.

© 2026 Faysal Aziz · Kuala Lumpur