

Career Track • Curriculum

# Data Scientist – Focus Python

Qualification for the Job Role of a **Data Scientist**





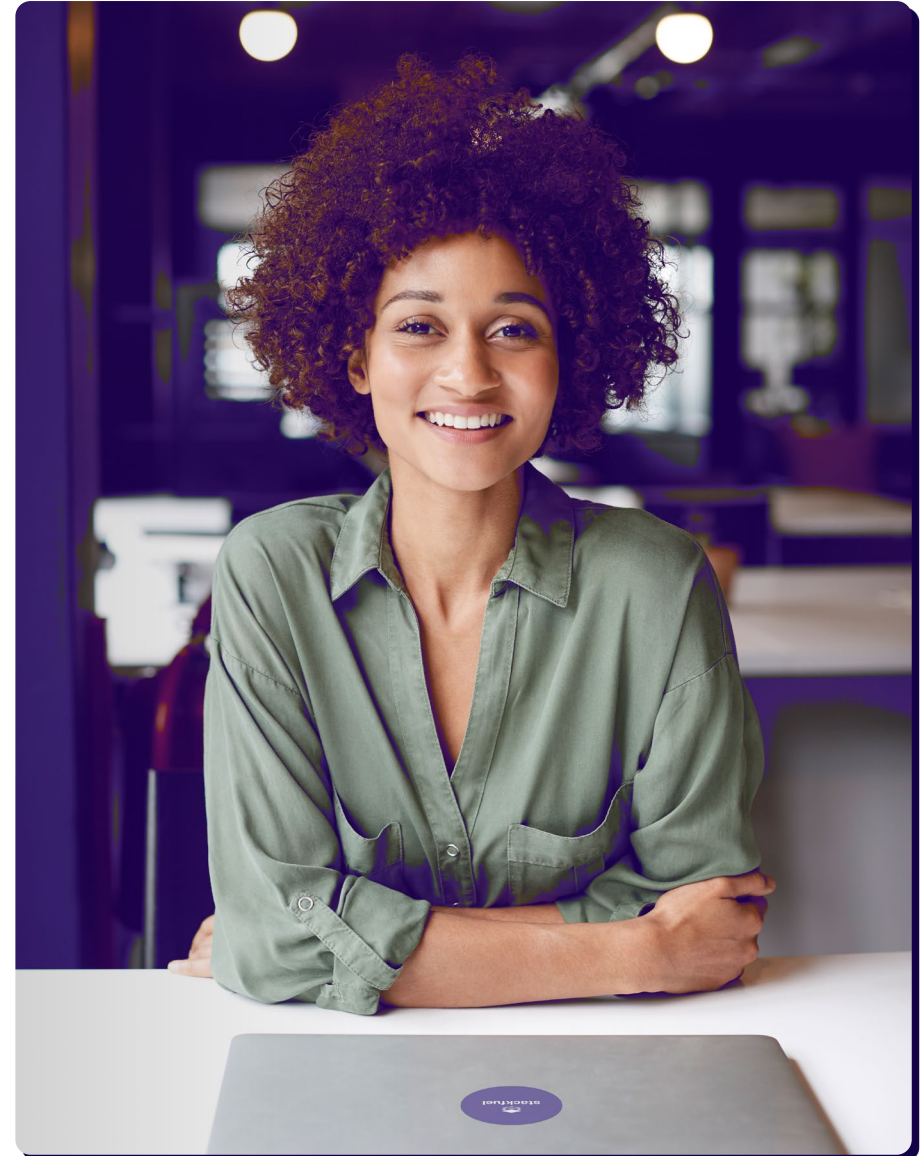
## Why this course?

Data Science is a professional field with incredible potential that will continue to grow significantly in the coming years. The demand for qualified data scientists is very high, as there is a serious shortage of skilled workers in this area. Last but not least, data science courses are becoming increasingly relevant and in demand. The job role of data scientist is considered one of the most desirable and in-demand careers across various industries.

In the age of digital transformation, data science skills are as unique as they are sought-after: expertise in data analysis, programming skills in Python, and using machine learning and deep learning.

With the certified online course to become a Data Scientist with the focus on Python, you will enter the world of data science. You will be able to generate, verify and interpret predictive models based on data in order to communicate the model results efficiently. You will gain advanced skills in the Python programming language and machine learning. Furthermore, you will finish the program with a final project, so that upon successful completion of the career path, you will be qualified for the job role of a data scientist or another analytical role.

More than **96.000** IT-specialists  
are currently being sought in Germany.  
Data scientists and big data experts are  
among the most in-demand qualifications.



## Course content.

The goal of this hands-on course is to implement data models with Python to predict different business scenarios. You will learn to implement and optimize supervised and unsupervised machine learning algorithms. You will also build skills in common visualization methods to qualify for the job role of data scientist upon completing the Data Scientist course.

### **In a nutshell:**

- > How to independently retrieve, clean, and filter data
- > How to explore and analyze data using descriptive statistics
- > How to develop and verify complex prediction models
- > Deep dive into your Python programming skills
- > How to build data models to predict business use cases
- > How to develop machine learning algorithms
- > Best practices for effective data visualization
- > Data storytelling methods

## At a glance.

Duration

**108 hours (4.5 months)**

Type

**Online course**

Structure

**4 modules + 1 final project**

Language

**German, English**

Level

**Mid level**

Completion

**Certificate of completion**

Prerequisites

**A good knowledge of programming skills in Python and common modules (Pandas, Matplotlib) is required for the course.**

Target group

**The mid level course is suitable for anyone who uses Python as a programming language, wants to analyze data and create predictions based on this data in order to make data-driven decisions. You should have a basic motivation for statistics, logical thinking and machine learning. This course is also suitable for career changers.**



# Module overview.

## Module 0

### Preparation

#### Objective:

- > Refresh knowledge of Python and mathematical principles

#### Description:

- > Participants carry out analysis and data manipulation in Python using the pandas and Matplotlib packages

### Chapter 1: Data Analytics with Python

Participants get to know our interactive programming environment - the Data Lab - and brush up on key programming and Python basics for data processing with Pandas, data visualization with Matplotlib and Seaborn, and database querying with SQLAlchemy.

### Chapter 2: Linear Algebra

Participants become familiar with the mathematical background of data science algorithms and learn the basic concepts of linear algebra. Participants use the NumPy package to perform calculations with vectors and matrices.

### Chapter 3: Probability Distributions

Participants learn more about the statistical background of data science algorithms. They explore important statistical concepts and learn about discrete and continuous distributions.



# Module overview.

## Module 1

### Machine Learning Basics

#### Objective:

- > Solving supervised and unsupervised machine learning problems with sklearn

#### Description:

- > Participants create data science workflows with sklearn, evaluate their model performance using appropriate metrics, and learn about the overfitting problem

#### Chapter 1: Supervised Learning (Regression)

Participants learn how to use the Python package sklearn with linear regressions. They also learn about the regression model's assumptions and evaluating the predictions generated. Participants learn about the bias-variance trade-off, regularization, and various metrics of model quality.

#### Chapter 2: Supervised Learning (Classification)

Participants are introduced to classification algorithms using the k-Nearest-Neighbors algorithm and learn to evaluate the algorithm and assess classification performance. They optimize the parameters of their model and pay attention to dividing the data into training and evaluation sets.

#### Chapter 3: Unsupervised Learning (Clustering)

Participants learn about the k-Means algorithm as an example of an unsupervised learning algorithm. They critically examine the algorithm's assumptions and performance metrics. Then they take a brief look at an alternative to k-Means clustering.



## Module overview.

### Module 1

#### Machine Learning Basics

##### Chapter 4: Unsupervised Learning (Dimensionality Reduction)

Participants learn how to reduce the dimension of their data using Principal Component Analysis (PCA) and use PCA to generate uncorrelated features from the original data. In this context, they explore the topic of feature engineering in more detail and new features are generated from the old ones.

##### Chapter 5: Outlier Detection

Participants learn about different approaches to identifying outliers and understand how to deal with these unusual data points. They use robust measures and models to minimize the impact of outliers.



### Module 2

#### Deep Dive Supervised Learning

##### Objective:

- > Expanding the data science toolkit

##### Description:

- > Participants deepen their knowledge of data classification models. In doing so, they also expand their skills in collecting and preparing data

##### Chapter 1: Data Gathering

Participants learn to gather data by mining web pages and PDF documents. They structure collected text data using regular expressions so that they can use it together with familiar algorithms.

##### Chapter 2: Logistic Regression

Participants learn a second classification algorithm: logistic regression. They use new performance metrics to evaluate results and learn how to prepare non-numeric data for their models.



## Module overview.

### Module 2

#### **Deep Dive Supervised Learning**

##### **Chapter 3: Decision Trees and Random Forests**

Participants learn about the decision tree as an easy-to-interpret model. They combine multiple models in an ensemble to improve the predictions of their model. They also learn methods to deal with unbalanced categories.

##### **Chapter 4: Support Vector Machines**

Participants learn about a final classification algorithm - Support Vector Machines (SVMs) and examine the behavior of different kernels for SVMs. They also learn the typical steps of Natural Language Processing (NLP) and work through an NLP scenario using bag-of-words models.

##### **Chapter 5: Neural Networks**

Participants are introduced to artificial neural networks and learn more about deep learning, to create a multilayer artificial neural network and apply it to a data set.

### Module 3

#### **Advanced Topics in Data Science**

##### **Objective:**

- > Independent application of simple and complex modeling

##### **Description:**

- > Participants gain confidence in solving data science problems and learn to communicate results competently

##### **Chapter 1: Visualization and Model Interpretation**

Participants learn important methods for interpreting and visualizing machine learning models. By using model-agnostic methods for interpretation, they learn to derive and communicate insights into the functioning of their models.

##### **Chapter 2: Spark**

Participants learn why working with distributed storage systems is relevant. Using the Python package PySpark, they learn how to read distributed databases, perform big data analyses and use known machine learning algorithms on distributed systems.



## Module overview.

### Module 3

#### **Advanced Topics in Data Science**

##### **Chapter 3: Exercise Project**

Participants work on a prediction problem using a larger data set and independently apply their data science skills from cleaning the data set to interpreting the model. Participants receive feedback on their approach to solving the problem in a project consultation with StackFuel's mentoring team.

##### **Chapter 4: Final Project**

Participants are given another larger dataset to analyze independently and solve with less assistance than they received for the practice project. Participants receive feedback on their solution approach in an individual project consultation with the StackFuel mentoring team.



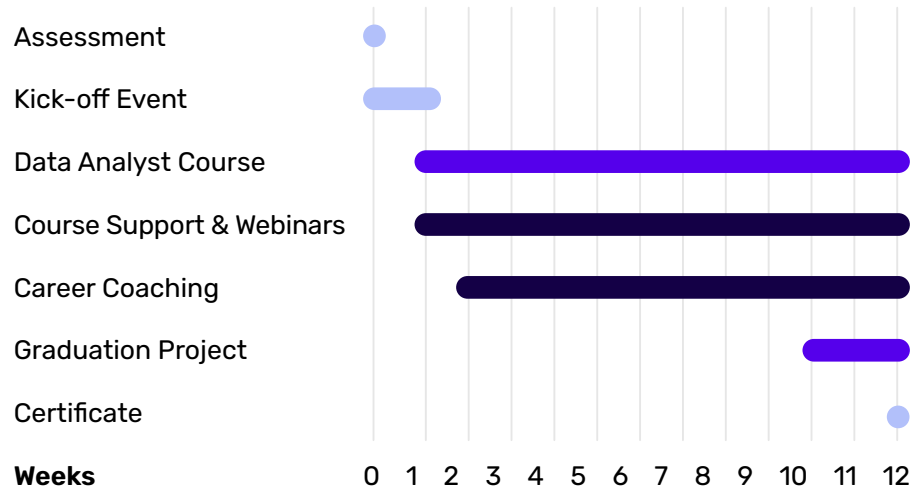


## About StackFuel.




We are your strategic learning partner, suitable for every career level and professional orientation.

Whether you're changing careers, an employee or a manager, our certified and state-supported training courses in data, AI and programming will keep you up to date with the latest technologies.

## Course schedule.



\*Sample course schedule for our Data Analyst course.

 Onboarding & offboarding  Training contents  Mentoring & support

## Training philosophy.



### Online & flexible

Do your course part-time or full-time and learn 100% online in your browser on your PC or laptop at home.



### Hands-on practice with real-life projects

In our Data Lab you will write your own algorithms with industrial data sets in interactive exercises and coding challenges.



### Mentoring & career coaching

Your personal mentoring team will accompany you with coaching, feedback, and weekly group webinars.



### Certified & eligible

As a certified training provider, you can get our courses fully subsidized by the job center and the employment agency.

## Facts.

**91%**  
completion  
rate

**80%**  
interactive  
exercises

**20%**  
expert videos  
& text lessons

**+150.000**  
learning hours in the Data Lab

**4,3 von 5**  
recommendation rate



## Sponsorship opportunities.

### Installments or part payment

Use our installments or part payment options to spread out the costs of your course over several months so you can remain financially flexible.

### Education voucher

With an [education voucher](#) (Bildungsgutschein in German), you can get your course financed up to 100% by the Jobcenter or the Employment Office, if you are currently out of work or looking for employment.

### Training opportunities act

If you are working, you can get your course partially or completely funded by your employer thanks to the [Training Opportunities Act](#) (Qualifizierungschancengesetz in German) - regardless of your qualifications, age or the size of the company.

### StackFuel scholarships

We regularly award various [scholarships](#) for our courses, to promote more diversity in the field of data. We want to encourage more people to take an interest in programming, and more women to work in data roles.



## Support & mentoring.

### Coaching

- > Assessment
- > Final project and evaluation
- > 1:1 project feedback consultation
- > Official certificate of completion

### Personal mentoring

- > Kick-off session
- > Webinars with other course participants
- > Support via email or phone
- > Online forum

### Career services

- > CV and application coaching
- > Talent pool and career intros
- > Data community
- > Career events



# Ready to enroll?

Get in touch with our **consultants**  
and build up the **skills** you need!

For Individuals



**Maria Schwenke**

Education Advisor

**Book free consultation**

For Businesses



**Ginesh Koottakara**

Head of Sales

**Book free demo**

August 2023

**Version:** 23/08/21

**Publisher:** StackFuel GmbH

**Web:** [www.stackfuel.com](http://www.stackfuel.com)

**Email:** [info@stackfuel.com](mailto:info@stackfuel.com)

**Phone:** +49 (0)30 544 533 420

Errors and omissions excepted.

© 2023 StackFuel GmbH