cacttus Education

DATA EDU

Training

+383 (0) 38 600 237 www.cacttus.education apply@cacttus.education

cacttus Education

Data Edu *Modules*

M1 Statistician

M2 **M3** Statistician or Machine Learning Specialist

M1 **M4** Data Analyst

M1 **M2** **M3** **M4** Data Specialist

Items	Module 1	Module 2	Module 3	Module 4
Module Name	Statistical Analysis With SPSS	Practical Data Science with Python	Machine Learning	Data Visualization
Target Group	Candidate/ Young Professionals	Candidate/ Young Professionals	Candidate/ Young Professionals	Candidate/ Young Professionals
Hours	**30 hours**	**30 hours**	**30 hours**	**30 hours**
Key Topics	Variables, Hypothesis Testing, Data Cleaning, Exploratory Data Analysis, Advanced Statistical Analysis	Basic Python Data Structures, Python, Syntax, Jupiter Notebook, Data Manipulation Libraries such as Pandas, Numpy, Metploitlib, Decision Trees, Random Forest, Time Series	Regression, Classification, Logistic & k-nearest neighbors algorithm	Reporting, Data Modelling, Power BI Dashboards, Creative Visuals (Canva)
Credits	**3**	**3**	**3**	**3**

cacttus Education

Pricing

Price for the whole program (without and with VAT): **1038 EUR/1224 EUR**
Minimum number of candidates in the group: **15**

- By finishing all modules (M1-M4), candidates will be able to work as a **Data Scientist.**
- By finishing Module 1 and Module 4, candidates will be able to work as **Data Analyst.**
- By finishing Module 1, candidates will be able to work as **Statisticians or Researchers.**
- By finishing Module 2 and 3, candidates will be able to work as **Machine Learning Specialists.**

Module	Title
M1: Statistical Analysis with SPSS	**Statistician or Researcher**
M2: Practical Data Science with Python	
M3: Machine Learning	**Statistician or Researcher**
M4: Data Visualization	
M2: Practical Data Science with Python	**Machine Learning Specialist**
M3: Machine Learning	
M1: Statistical Analysis with SPSS	**Data Analyst**
M4: Data Visualization	

cacttus Education

Pricing

Module	Title	Original Price	Price with Discount
M1: Statistical Analysis with SPSS	**Data Scientist**	1,038.40 €	**980,00€**
M2: Practical Data Science with Python			
M3: Machine Learning			
M4: Data Visualization			

cacttus Education

Course Outline

Module 1: Statistical Analysis With SPSS
Who is it for: Students/Young Professionals
Prerequisite: None

Inferential Statistics

- Introduction to Inferential Statistics
- Probability Distribution
- Sample Size, Margin of Error, Population size
- Data Visualization: The Matplotlib and Seaborn libraries
- Estimations
- The Central Limited Theorem

Introduction to SPSS

- Overview of SPSS
- Navigating the SPSS Interface
- Creating and Managing Data Files

Data Management and Preparation

- Data Cleaning
- Data Transformation
- Data Merging

Descriptive Statistics

- Basic descriptive statistics
- Graphical Representation
- Crosstabulations

Advanced Statistical Analysis

- T- test
- ANOVA
- Linear Regression
- Chi Square
- Factor Analysis
- Multiple regression
- Logistic Regression

Reporting and Output Management

- Graphical Presentation
- Managing outputs
- Exporting files
- Practical Applications with SPSS

Module 2: Practical Data Science
Who is it for: Students/Young Professionals
Prerequisite: None
Topics covered:

Python for Data Science Foundations

- Numpy and Pandas: Operations and functions to work with data
- Pandas Dataframes & Series: Operations and applications
- Data Visualization: The Matplotlib and Seaborn libraries
- Practical Applications

Decision Trees

- Introduction to Decision Trees
- The Power of Decision Trees and their advantages
- Classification: The Main Idea
- Building a Decision Tree from data
- Misclassification and error criterion
- Decision Trees for categorical data
- Defining and optimizing splits, Entropy, Information Gain, Greedy algorithm for the split
- Practical Applications of Decision Trees

Introduction to Bagging and Random Forest

- The Bias-Variance tradeoff
- Overfitting and Pruning of Decision Trees
- Ensemble Learning
- Reduction in Variance
- Bagging, Bootstrapping and Random Forests with examples
- Sampling features at every node and their effects
- Extensions to the above processes
- Practical Applications of Random Forests

Time Series

- Introduction to Time Series and domains of Time Series analysis
- Time Series Implementations
- Stationarity in data and its importance
- Testing stationarity and Transformations to get stationary series
- Autocorrelation, Methods for Time Series, AR, ARMA, controlled series
- Estimation of AR models, Similarity of MA, data dependence
- Practical Applications of Time Series

Module 3: Machine Learning
Who is it for: Students/Young Professionals
Prerequisite: Either module 2 completed or prior knowledge in Python, Times Series, and Decision Trees and Random Forest.
Topics Covered:

Introduction to Supervised Learning: Regression

- Linear Regression
- Performance Assessment: Estimating parameter means and confidence intervals for predictions
- Practical Application
- Model Evaluation: Cross-validation and bootstrapping
- Practical Application

Key Concepts in Regression

- Prediction vs. Modelling
- Assumptions Behind Regression
- Overfitting and Regularization
- Bias-Variance Tradeoff
- Cross-validation
- Bootstrapping
- Practical Application

Introduction to Supervised Learning: Classification

- Classification Overview
- Gaussian Models
- Bayesian Formulation
- Logistic Regression
- Performance Assessments
- Other Classification Algorithms: K-Nearest Neighbors (K-NN)
- Practical Application

Module 4: Data Visualization
Who is it for: Students/Young Professionals
Prerequisite: None

Reporting

- Introduction to Data Report
- Structuring a Data Report
- Presenting Data in Reports
- Practical Applications

Power BI

- Power BI Desktop
- Power BI Query
- Power BI Dashboards
- Publishing and Sharing
- Practical Applications in Power BI

Canva

- Design Principles
- Using Canva elements
- Preparing Infographics
- Exporting and working with Text

+383 (0) 38 600 237 www.cacttus.education apply@cacttus.education