



DATA SCIENTIST COURSE

October 2020

CRYSTAL SYSTEM

Email: talents@crystal-system.eu

Website: www.crystal-system.eu

Tel: +40 757 536 864 ; +40 753 030 130

Agias Fylaxeos & Zinonos Rossidi nr. 2

CY - 3082 Limassol, Cyprus



CONTENTS

ABOUT THE COURSE _____	3
KEY FEATURES OF DATA SCIENTIST COURSE _____	4
ABOUT CRYSTAL TALENT SOLUTIONS _____	5
DATA SCIENTIST LEARNING PATH _____	6
WHO SHOULD ENROLL IN THE COURSE? _____	7
COURSE CURRICULA _____	8
Module 1. Data Science in a business context _____	8
Module 2. Math and Statistics fundamentals for data Science _____	8
Module 3. R Programming for Data Science _____	9
Module 4. Data Science with R _____	9
Module 5. Python programming for Data Science _____	10
Module 6. Data Science with Python _____	10
Module 7. Artificial Intelligence and Machine Learning _____	11
Module 8. Tableau for Data Science _____	11
Module 9. Big Data Hadoop and Spark Developer _____	12
Module 10. Data Science Case Study Project _____	12
CERTIFICATES _____	13
CRYSTAL TALENT SOLUTIONS EXECUTIVE TEAM _____	14

ABOUT THE COURSE

Crystal System Data Scientist Course (DSC) is made in collaboration with the best domain experts and academic professionals, it is aligned with the market business requirements, and supported by the research conducted by Crystal in Data Science Talents skills (annex).

Crystal DSC builds and accelerates Talents career in Data Science providing them with the world-class training and skills required to become successful in this domain, as well as prepare the industry ready Data Science Talents.

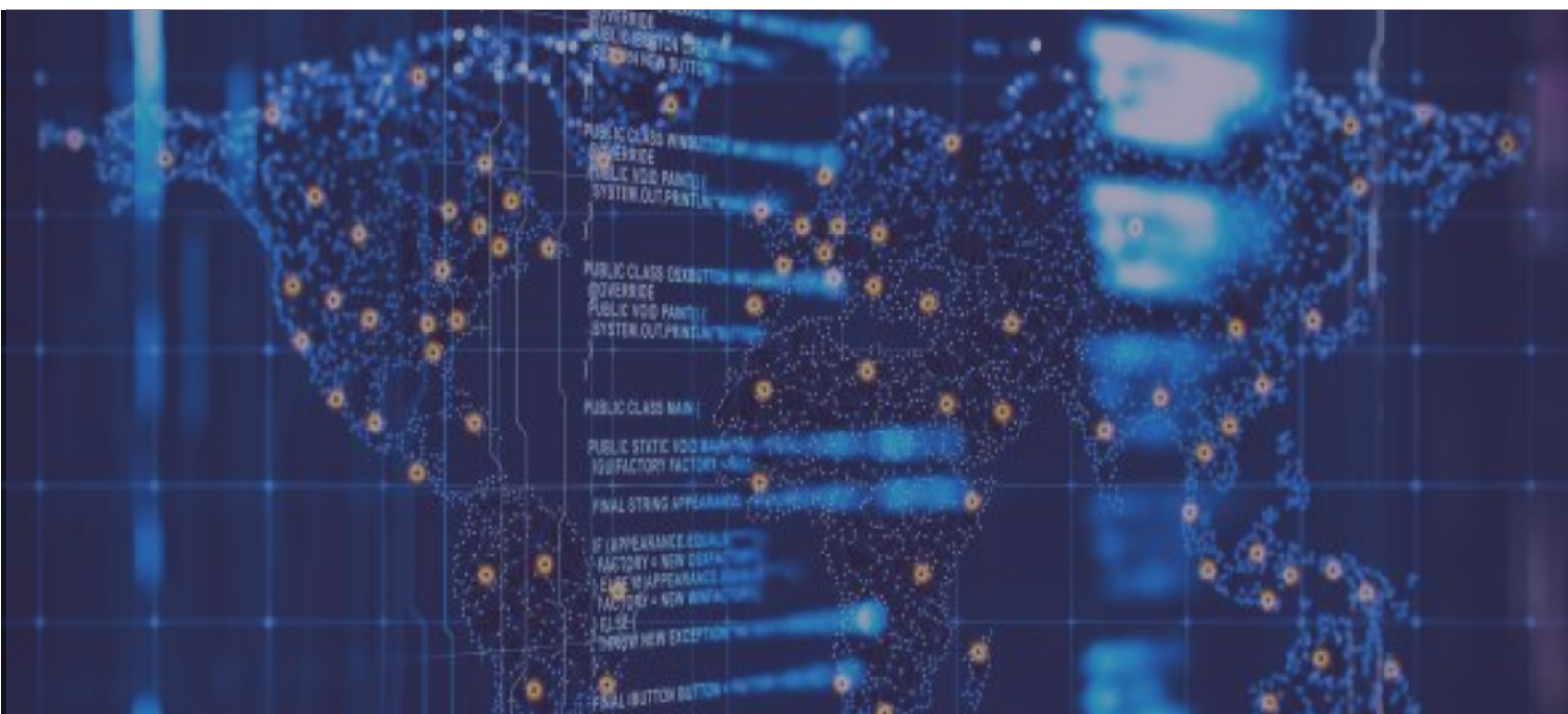
Crystal DSC offers extensive training on the most in-demand Data Science skills with hands-on exposure to key tools and technologies.

Crystal DSC helps Talents to become an industry ready Data Scientist Talents by deep diving into the nuances of data interpretation, mastering powerful programming skills and focusing on the business acumen as an essential element.

Course duration: 11 weeks, 150 hours (84 hours in classrooms, 56 self-studies (+10 Capstone project)), 3 classes per week with 2 hours duration each.

Start date: October 2020

Type: online or mixed (onsite + online)



KEY FEATURES OF DATA SCIENTIST COURSE



Curricula and Course materials created in collaboration with Data Science Experts, AI academic professionals, business partners and research driven approach



Data Science Expert Trainers holding Phd in Computer Science and Statistics, university instructors experience



45+ Live sessions + 24/7 free access to e-learning platform with training materials



Case Study based on real-life projects and with elective data sets: Energy, Banking, Retail, Automotive and others



Certificates given based on 3 Level Exam results validation: 1- Case Study, 2 - Exam quizz, 3 - Experts Team interview



Hiring opportunities for the best Students and valuable prizes

ABOUT CRYSTAL TALENT SOLUTIONS

In the Information Era, **knowledge capital** is the most valuable asset and the key element for success; expanding **qualified education** is the only feasible way to generate **long-term economic growth**.

This is the philosophy that Crystal System operates with and the reason why a strong and coherent emphasis on education is central to our activity. We think that only by the aid of **talented collaborators** it is possible to maximize performance and obtain the **best results**, making a difference for our customers. To achieve that, we emphasize the importance of **cognitive skills** and we developed a **unique system of training and recruitment** that starts from the academic world.

Crystal Talent Solutions is a business unity of Crystal System, and periodically organize training sessions in leading Universities of East Europe.

Thanks to our courses, Students have the possibility to gain practical knowledge in the most up-to-date Information Technologies and they move their first steps in the IT consulting career



DATA SCIENTIST LEARNING PATH

Data Science in a business context

Math and Statistics fundamentals for Data Science

R Programming for Data Science

Data Science with R

Python programming for Data Science

Data Science with Python

Artificial Intelligence and Machine Learning

Tableau for Data Science

Big Data Hadoop and Spark Developer

Data Science Case Study Project

WHO SHOULD ENROLL IN THE COURSE?

- **Beginners or recent Graduates** in Bachelors and Master's Degree
- **PhD holders** in Math, Statistics, Computer Science, and others
- **Professionals** including: IT Professionals, Analytics Managers, Banking and Finance Professionals others who wishes to **accelerate their career** in Data Science, update their knowledge and skills with the **last domain technologies**, create an **added value on the top of their expertise**



COURSE CURRICULA

Module 1. Data Science in a business context

- 1.1. Defining Data Science and its importance
- 1.2. What Does a Data Science Professional Do?
- 1.3. Data Science for Business
- 1.4. Lifecycle of Data Science through Use Cases
- 1.5. Data Science People

Module 1. Learning outcomes

- ✚ Gain fundamental knowledge of what is Data Science and what do Data Science specialists do;
- ✚ Learn about Data Science in a business context and what is the future of Data Science;
- ✚ Understand Data Science applications and discover some use cases for Data Science

Module 2. Math and Statistics fundamentals for data Science

- 2.1. Introduction to data types
- 2.2. Sample or population data?
- 2.3. The fundamentals of descriptive statistics
- 2.4. Measures of central tendency, asymmetry, and variability
- 2.5. Contingency tables, covariance and correlation
- 2.6. Discrete probability Distributions
- 2.7. Continuous probability Distributions
- 2.8. Point Estimators and estimates
- 2.9. Confidence intervals, determining sample size: advanced topics
- 2.10. Hypothesis testing: Introduction
- 2.11. Hypothesis testing and decision making
- 2.12. Hypothesis testing: Goodness of Fit and Independence
- 2.13. The fundamentals of regression analysis
- 2.14. Assumptions for linear and logistic regression analysis
- 2.15. Dealing with categorical data
- 2.16. Residual Analysis
- 2.17. Regression analysis: model Building

Module 2. Learning outcomes

- ✚ Understand the fundamentals of math and statistics for DS; understand and work with various data types; sampling techniques; data representation and visualization; calculate and interpret numerical parameters which represent the data such as: the measures of central tendency, asymmetry, and variability; interpret and analyze contingency tables;
- ✚ Calculate and interpret covariance and correlation;
- ✚ Understand probability rules;
- ✚ Use Bayes Theorem;
- ✚ Distinguish and work with different types of probability distribution;
- ✚ use of Central Limit Theorem;
- ✚ Understand and use point estimation and confidence intervals;
- ✚ Explore and perform hypothesis testing;
- ✚ Make data-driven decisions;
- ✚ Understand the concept of regression analysis;
- ✚ Carry out regression analysis; Use and understand dummy variables;
- ✚ Understand the concepts needed for data science

Module 3. R Programming for Data Science

- 3.1. R basics, Introduction and preliminaries
- 3.2. Data structures in R, vectors, matrices lists and data frame
- 3.3. R Programming fundamentals: Control Structures and writing functions
- 3.4. Working with Data in R
- 3.5. Strings, Times and Dates in R

Module 3. Learning outcomes

- ✚ Installation and introduction to R environment, Learn about math operators, variables, and strings.
- ✚ Vectors, factors and vector operations;
- ✚ Create and save objects in R;
- ✚ Gain fundamental knowledge on arrays and matrices, lists, and data frames;
- ✚ Import and export data;
- ✚ Get understanding on Control Structures (if, if-else, for, while, repeat, ect). Writing your own functions, objects, loop functions and Debugging Tools;
- ✚ Understand and work on strings and dates in R

Module 4. Data Science with R

- 4.1. Introduction to Business Analytics
- 4.2. Introduction to R Programming
- 4.3. Data Structures
- 4.4. Data Visualization (2D,3D and dynamic graphs)
- 4.5. Statistics for Data Science-I Data manipulation DPLYR package
- 4.6. Statistics for Data Science-II Fitting probability distributions
- 4.7. Regression Modeling, tree Based Methods
- 4.8. Supervised Learning-Classification methods (SVM)
- 4.9. Unsupervised learning: Clustering techniques
- 4.10. Time Series analysis, modeling and forecasting

Module 4. Learning outcomes

- ✚ Gain a foundational understanding of business analytics;
 - ✚ Install R-studio, and workspace setup, and learn about the various R packages and their use;
 - ✚ R programming and understand how various statements are executed in R;
 - ✚ Gain an in-depth understanding of data structure used in R and learn to import/export data in R;
 - ✚ Define, understand and use the various apply functions and DPLYR functions;
 - ✚ Understand and use the various graphics packages (ggplot) in R for data visualization;
 - ✚ Gain a basic understanding of various statistical concepts;
 - ✚ Understand and use hypothesis testing method to drive business decisions;
 - ✚ Understand and use linear, non-linear regression models, tree based methods, support vector machines (SVM) and classification techniques for data analysis;
 - ✚ Learn and use the various association rules and Apriori algorithm;
 - ✚ Learn and use clustering methods including K-means, DBSCAN, and hierarchical clustering.
- Analyze, model and forecast time series (Arima, Sarima, ETS, exponential smoothing and hybrid models). Accuracy of the models and forecasting packages in R.

Module 5. Python programming for Data Science

- 5.1. Python Basics
- 5.2. Python Data Structures
- 5.3. Python Programming Fundamentals
- 5.4. Working with Data in Python
- 5.5. Working with NumPy arrays

Module 5. Learning outcomes

- ✚ Write your first Python program by implementing concepts of variables, strings, functions, loops, conditions;
- ✚ Understand the nuances of lists, sets, dictionaries, conditions and branching, objects and classes;
- ✚ Work with data in Python such as reading and writing files, loading, working, and saving data with Pandas

Module 6. Data Science with Python

- 6.1. Data Science Overview
- 6.2. Data Analytics Overview
- 6.3. Statistical Analysis and Business Applications
- 6.4. Python Environment Setup and Essentials
- 6.5. Mathematical Computing with Python (NumPy)
- 6.6. Scientific computing with Python (Scipy)
- 6.7. Data Manipulation with Pandas
- 6.8. Machine Learning with Scikit-Learn
- 6.9. Natural Language Processing with Scikit Learn
- 6.10. Data Visualization in Python using matplotlib
- 6.11. Web Scraping with BeautifulSoup

Module 6. Learning outcomes

- ✚ Gain an in-depth understanding of Data Science processes, data wrangling, data exploration, data visualization, hypothesis building, and testing. You will also learn the basics of statistics. Install the required Python environment and other auxiliary tools and libraries.
- ✚ Understand the essential concepts of Python programming such as data types, tuples, lists, dicts, basic operators and functions.
- ✚ Perform high-level mathematical computing using the NumPy package and its vast library of mathematical functions. Perform scientific and technical computing using the SciPy package and its sub-packages such as Integrate, Optimize, Statistics, IO, and Weave.
- ✚ Perform data analysis and manipulation using data structures and tools provided in the Pandas package.
- ✚ Gain expertise in Machine Learning using the Scikit-Learn package. Gain an in-depth understanding of supervised learning and unsupervised learning models such as linear regression, logistic regression, clustering, dimensionality reduction, K-NN and pipeline.

Module 7. Artificial Intelligence and Machine Learning

- 7.1. Introduction to Artificial Intelligence and Machine Learning
- 7.2. Data Wrangling and Manipulation
- 7.3. Supervised Learning
- 7.4. Feature Engineering
- 7.5. Supervised Learning-Classification
- 7.6. Unsupervised learning
- 7.7. Time Series Modelling
- 7.8. Ensemble Learning
- 7.9. Recommender Systems
- 7.10. Text Mining

Module 7. Learning outcomes

- ✚ Master the concepts of supervised and unsupervised learning, recommendation engine, and time series modeling.
- ✚ Gain practical mastery over principles, algorithms, and applications of Machine Learning through a hands-on approach that includes working on four major end-to-end projects and 25+ hands-on exercises.
- ✚ Acquire thorough knowledge of the statistical and heuristic aspects of Machine Learning.
- ✚ Implement models such as support vector machines, kernel SVM, naive Bayes, decision tree classifier, random forest classifier, logistic regression, K-means clustering and more in Python.
- ✚ Validate Machine Learning models and decode various accuracy metrics. Improve the final models using another set of optimization algorithm, which include Boosting & Bagging techniques.
- ✚ Comprehend the theoretical concepts and how they relate to the practical aspects of Machine Learning

Module 8. Tableau for Data Science

- 8.1. Getting Started with Tableau
- 8.2. Working with Tableau
- 8.3. Creating Charts
- 8.4. Adding calculations to your workbook

- 8.5. Mapping data in Tableau
- 8.6. Visualizations for an Audience

Module 8. Learning outcomes

- ✚ Grasp the concepts of Tableau Desktop 10, become proficient with statistics and build interactive dashboards.
- ✚ Master data sources and data blending, create data extracts and organize and format data. Master arithmetic, logical, table and LOD calculations and ad-hoc analytics.
- ✚ Become an expert on visualization techniques such as heat map, tree map, waterfall, Pareto, Gantt chart and market basket analysis.
- ✚ Learn to analyze data using Tableau Desktop as well as clustering and forecasting techniques.
- ✚ Gain command of mapping concepts such as custom geocoding and radial selections.
- ✚ Master Special Field Types and Tableau Generated Fields and the process of creating and using parameters.
- ✚ Learn how to build interactive dashboards, story interfaces and how to share your work.

Module 9. Big Data Hadoop and Spark Developer

- 9.1. Introduction to Big Data and Hadoop Ecosystem
- 9.2. MapReduce and Sqoop
- 9.3. Type of Data Formats
- 9.4. Apache Pig
- 9.5. Basics of Apache Spark
- 9.6. Implementation of Spark Applications
- 9.7. Spark Optimization Techniques
- 9.8. Spark Algorithm

Module 9. Learning outcomes

- ✚ Master the concepts of the Hadoop framework and its deployment in a cluster environment.
- ✚ Understand how the Hadoop ecosystem fits in with the data processing lifecycle.
- ✚ Learn to write complex MapReduce programs Describe how to ingest data using Sqoop and Flume.
- ✚ Get introduced to Apache Spark and its components
- ✚ List the best practices for data storage.
- ✚ Explain how to model structured data as tables with Impala and Hive

Module 10. Data Science Case Study Project

Each Student will choose a data set coherent with his business interest (Energy, Banking, Automotive, Oil and Gas) to do a Case Study Project and practice knowledge and skills gained in Module 1-9.

Final Exam to certify course finalists

Expert Team Interview with each course finalist to create his/her Data Scientist profile

CERTIFICATES



Upon completion of the Crystal Data Science Course Crystal System will certify each Student with the following certificates:

Certificates of Completion are given to Students with less than 70% completed Final Exam.

Certificates of Participation are given to Students with less than 70% completed Final Exam



CRYSTAL TALENT SOLUTIONS EXECUTIVE TEAM



Frank Nagorschel (Director Marketing & Talent Solutions)



frank.nagorschel@crystal-system.eu



+40 753 030 130



Maria Cerednicenco (Education Program Manager)



maria.cherednicenco@crystal-system.eu



+40 757 536 864



Stefan Cojocar (Talents Relationship Leader)



stefan.cojocar@crystal-system.eu



+40 738 999 850