



I. General Information

Course Name: Data Science and Data Engineering

Difficulty level (Beginner/Intermediate/ Advanced): Intermediate- Advanced

Knowledge Prerequisites:

- English B2-C1
- Programming skills – beginner -medium level
- Math and Statistics – good understanding of basic concepts
- 3^d - 4th year, Master and PhD Students

Number of total hours: 45

Course duration: 2 months (2 classes a week with duration of approx. 2,5 hours)

Case Study / Capstone Project / Laboratory project: Capstone Project

Examination form: Capstone project, Quiz Exam, Online Interview

Certification: Crystal Certificate

I. Curricula

Nº	Subjects/Lessons	Hours
		In-classroom
1.	Defining Data Science and its importance	0.5
2.	The fundamentals of descriptive statistics	1
3.	Sample or population data?	1
4.	Data Analytics Overview	0.5
5.	Introduction to Artificial Intelligence and Machine Learning. Feature Engineering	1.5
6.	Dataflow jobs and templates. Integration with other services. APIs.	1
7.	Measures of central tendency, asymmetry, and variability	0.5
8.	Probability distribution	1
9.	Hypothesis testing	1
10.	Type of Data Formats	1
11.	Big Data ecosystem, Frameworks and Algorithms	1
12.	Basic understanding of the Apache Hadoop ecosystem (Apache Hadoop, HDFS, HBase, Pig, Spark, Hive)	1
13.	R basics, Introduction and preliminaries	1
14.	Data structures in R, vectors, matrices lists and data frame	1.5
15.	Python Environment, Syntax, Data structures	1
16.	Mathematical and Scientific Computing with Python (NumPy & Scipy)	1.5
17.	Working with Data in R. Strings, Times and Dates in R	1.5
18.	R Programming fundamentals - Control Structures and writing functions	1
19.	Data Wrangling and Manipulation	1
20.	Data Manipulation with Pandas	1.5
21.	The fundamentals of regression analysis	1
22.	Dealing with categorical data. Residual Analysis	1.5
23.	Web Scraping with BeautifulSoup	1
24.	Text Mining	0.5
25.	Natural Language Processing with Scikit Learn	1
26.	Data Visualization in R (2D,3D and dynamic graphs)	1
27.	Statistics for Data Science-I Data manipulation DPLYR package	1



Nº	Subjects/Lessons	Hours
		In-classroom
28.	Statistics for Data Science-II Fitting probability distributions	0.5
29.	Data Visualization in Python using Matplotlib	1.5
30.	Visualizations for an Audience	1
31.	Time Series analysis, modeling and forecasting	2
32.	Regression Modeling, tree Based Methods	1
33.	Time Series Modelling	2
34.	Supervised Learning-Classification	1
35.	Supervised Learning-Classification methods (SVM)	1
36.	Unsupervised learning - Clustering techniques	1
37.	Unsupervised learning	1
38.	Ensemble Learning	1
39.	Recommender Systems	1
40.	BigQuery ML. Optimizations. Integration with other services	1
41.	Apache Beam: concepts (PCollections, Transforms, ParDo, Runners), pipeline I/O, schemas, windowing, triggers etc.	1
42.	Concepts (buckets and objects). Choosing the right storage class (standard, nearline, coldline, archive). Access control (best practises), signed URLs.	1
Total Hours in Teaching Lessons		45
43.	Case Study Projects – Automotive, Finance, Energy, ... etc.	~10



Teaching Lessons – Online / Offline / Blended Classroom



Teaching Lessons – Online / Offline / Blended Classroom