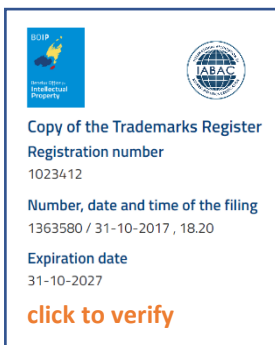




# Certified Data Scientist (CDS)

## CODE: DS1050

## Syllabus and Examination



*The International Association for Data Science Certification (IABAC®) is a globally recognized professional association dedicated to growing and enhancing the field of applied Data Science and Business Analytics.*

*IABAC™ founding principles are based on Edison Data Science Framework (EDSF), a European commission initiative, with the goal of aligning data science skills to industry requirements. IABAC was founded in the year 2017 as an association and, in the year 2018, registered as The Netherlands B.V (equivalent of English Private Limited).*

***KVK-nummer (Commercial Registration Number): 73234893 (click to verify)***

---

Copyright © IABAC 2017-2020.

IABAC®, IABAC logo®, DSF®, CDS®, CAIE®, CNLPE®, CCVE®, CDSHR®, CDSFIN®, CVAE®, CDLE®, CDSHC®, CDSMKT®, CDS®, DSCMGR®, AICE®, CDSD® and CMLÉ® are registered trademarks of IABAC BV, The Netherlands. Registration Number (KVK-nummer) :73234893. Reproduction of any part of this material requires written permission of IABAC BV.

All rights reserved. EN\_CDS\_2020\_Syllabus\_V2.0

# CONTENTS

---

1	Introduction	3
2	Course Syllabus	3
2.1	Data Science Foundation	3
2.2	Python for Data Science	3
2.3	Statistics for Data Science	3
2.4	Visual Analytics Foundation	3
2.5	SQL for Data Science	3
2.6	Machine Learning Associate	4
2.7	Machine Learning Expert	4
2.8	Time Series Foundation	4
2.9	Model Deployment	5
2.10	Deep Learning Foundation	5
3	Examination	6
3.1	Pre-requisite qualifications	6
3.2	Material permitted	6
3.3	Exam Duration and Format	6
3.4	Exam mode	6
3.5	Pass Criteria	6
3.6	Results Timeline	6
3.7	Certificate Issuance	6
4	IABAC® Knowledge Areas Mapping	7
5	Bloom's Taxonomy Reference	8

---

Copyright © IABAC 2017-2020.

IABAC®, IABAC logo®, DSF®, CDS®, CAIE®, CNLPE®, CCVE®, CDSHR®, CDSFIN®, CVAE®, CDLE®, CDSHC®, CDSMKT®, CDS®, DSCMGR®, AICE®, CDSD® and CMLE® are registered trademarks of IABAC BV, The Netherlands. Registration Number (KVK-nummer) :73234893. Reproduction of any part of this material requires written permission of IABAC BV.

All rights reserved. EN\_CDS\_2020\_Syllabus\_V2.0

# 1 INTRODUCTION

---

This document is intended to provide information on Certified Data Scientist (CDS-DS1050) certification for registered training providers to structure the course curriculum as per IABAC syllabus guidelines and for individuals, who are preparing for IABAC CDS certification exam.

## 2 COURSE SYLLABUS

---

### 2.1 DATA SCIENCE FOUNDATION

- Introduction to Data Science
- Data Science vs Business Analytics vs Big Data
- Classification of Business Analytics
- Data Science Project Workflow
- Various Roles in Data Science
- Application of Data Science in various industries

### 2.2 PYTHON FOR DATA SCIENCE

- Introduction to Data Science with Python
- Python Basics: Basic Syntax, Data Structures
- Data objects, Math, Comparison Operators, Condition Statements, loops, lists, tuples, dicts, functions
- Numpy Package
- Pandas Package
- Python Advanced: Data Mugging with Pandas
- Python Advanced: Visualization with Matplotlib
- Exploratory Data Analysis: Data Cleaning, Data Wrangling
- Exploratory Data Analysis: Case Study

### 2.3 STATISTICS FOR DATA SCIENCE

- Introduction to Statistics
- Harnessing Data
- Exploratory Analysis
- Distributions
- Hypothesis & Computational Techniques
- Correlation & Regression

### 2.4 VISUAL ANALYTICS FOUNDATION

- Visual Analytics Basics
- Basic Charts, Plots

### 2.5 SQL FOR DATA SCIENCE

- Install SQL packages and Connecting to DB

---

Copyright © IABAC 2017-2020.

IABAC®, IABAC logo®, DSF®, CDS®, CAIE®, CNLPE®, CCVE®, CDSHR®, CDSFIN®, CVAE®, CDLE®, CDSHC®, CDSMKT®, CDS®, DSCMGR®, AICE®, CDSD® and CMLE® are registered trademarks of IABAC BV, The Netherlands. Registration Number (KVK-nummer) :73234893. Reproduction of any part of this material requires written permission of IABAC BV.

All rights reserved. EN\_CDS\_2020\_Syllabus\_V2.0

- RDBMS (Relational Database Management) Basics
- Basics of SQL DB, Primary key, Foreign Key
- SELECT SQL command, WHERE Condition
- Retrieving Data with SELECT SQL command and WHERE Condition to Pandas DataFrame.
- SQL JOINS
- Left Join, Right Joins, Multiple Joins

## 2.6 MACHINE LEARNING ASSOCIATE

- Machine Learning Introduction
- What is ML? ML vs AI. ML Workflow, Statistical Modelling of ML. Application of ML
- Machine Learning Algorithms
- Popular ML algorithms, Clustering, Classification and Regression, Supervised vs Unsupervised.
- Choice of ML
- Supervised Learning
- Simple and Multiple Linear Regression, KNN, and more
- Linear Regression and Logistic Regression
- Theory of Linear regression, hands on with use cases
- K-Nearest Neighbour (KNN)
- Decision Tree
- Naïve Bayes Classifier
- Unsupervised Learning: K-Means Clustering

## 2.7 MACHINE LEARNING EXPERT

- Advanced Machine Learning Concepts
- Tuning with Hyper parameters
- Random Forest – Ensemble
- Ensemble Theory, Random Forest Tuning
- Support Vector Machine (SVM)
- Simple and Multiple Linear Regression, KNN
- Natural Language Processing (NLP)
- Text Processing with Vectorization, Sentiment Analysis with Text Blob, Twitter Sentiment Analysis.
- Naïve Bayes Classifier
- Naïve Bayes for Text Classification, New Articles Tagging
- Artificial Neural Network (ANN)
- Basic ANN network for Regression and Classification
- TensorFlow Overview
- Deep Learning Intro

## 2.8 TIME SERIES FOUNDATION

- What is a Time-Series?
- Trend, Seasonality, Cyclical and Random
- White Noise
- Auto Regressive Model (AR)
- Moving Average Model (MA)

---

Copyright © IABAC 2017-2020.

IABAC®, IABAC logo®, DSF®, CDS®, CAIE®, CNLPE®, CCVE®, CDSHR®, CDSFIN®, CVAE®, CDLE®, CDSHC®, CDSMKT®, CDS®, DSCMGR®, AICE®, CDS® and CMLE® are registered trademarks of IABAC BV, The Netherlands. Registration Number (KVK-nummer) :73234893. Reproduction of any part of this material requires written permission of IABAC BV.

All rights reserved. EN\_CDS\_2020\_Syllabus\_V2.0

- ARMA Model
- Stationarity of Time Series
- ARIMA Model – Prediction Concepts
- ARIMA Model Hands on with Python
- Case Study Assignment on ARIMA

## 2.9 MODEL DEPLOYMENT

- Basics of Application Program Interface (API)
- API basics, loosely Coupled Architecture
- Installing Flask
- Installation and configuring Flask and cross domain authentication.
- End to End ML project with API Deployment
- Complete Project Flow with API Deployment and assessing through the website

## 2.10 DEEP LEARNING FOUNDATION

- Introduction to Deep learning
- What is Deep Learning?
- Various Deep Learning models in practice and applications.
- Convolutional Neural Network CNN Intro
- Case Study: Keras–TensorFlow Image Classification
- CNN hands on application for classification of images of Cats and Dogs

## 3 EXAMINATION

---

### 3.1 PRE-REQUISITE QUALIFICATIONS

1. Certified Machine Learning Expert Course (CMLE) or Demonstrable competence at CMLE level
2. Recommended essential knowledge in
  - a. Mathematics: Calculus, Statistics, Linear Algebra, Probability
  - b. Machine Learning and Python/R Programming
3. Training: Though formal training is not mandatory; it is recommended to attend IABAC® registered course through Registered Education Partners

### 3.2 MATERIAL PERMITTED

1. The examination is an 'open book'
2. Candidates can refer to any material

### 3.3 EXAM DURATION AND FORMAT

1. Exam format is through a Project Submission
2. The assessment duration is 8 hours.
3. The project is graded for three areas: Project Summary with recommendations, Machine Learning model performance and Exploratory Data Analysis

### 3.4 EXAM MODE

1. Project needs to be submitted at IABAC project submit page, as per exam guidelines
2. Any copied work, ideas, concepts or a piece of text needs to be marked with reference as per IABAC project plagiarism guidelines

### 3.5 PASS CRITERIA

1. The candidate needs to score assessment grade A+, A, B+, B, C+, C as a PASS Criteria
2. The candidate will be awarded grade F in case of failing to meet the pass criteria
3. The results will be declared after validation with the project guidelines

### 3.6 RESULTS TIMELINE

1. The preliminary results are usually released within **9 days** of the exam date
2. The official results are usually released within **15 days** from the exam date

### 3.7 CERTIFICATE ISSUANCE

- IABAC® e-certificate will be issued through the candidate's registered email
- The e-certificate is digital verifiable at <https://www.iabac.org/verify-certificate>
- The candidate has a license to share digital certificate validation in professional networking portals such as [www.linkedin.com](http://www.linkedin.com)
- The candidate has a license to print physical copy (hardcopy) of the certificate

## 4 IABAC® KNOWLEDGE AREAS MAPPING

Knowledge Area	Syllabus Details	Bloom's Index
<p><b>KAG1-DSDA:</b> Data Analytics group including Machine Learning, Statistical Methods, and Business Analytics</p>	<ul style="list-style-type: none"> <li>● Case Study on Statistical Analysis</li> <li>● Curating the Data and performing, Discrete Mathematics, Probabilistic Reasoning</li> <li>● Statistical Methods, including Descriptive Statistics, Exploratory Data Analysis (EDA) and Confirmatory Data Analysis (CDA)</li> <li>● Case Study &amp; Creating Machine Learning Model</li> <li>● With detailed implementation of algorithms: Artificial Intelligence, Natural Language Processing</li> <li>● Knowledge Representation and Reasoning</li> <li>● Data mining and Knowledge Discovery</li> <li>● Text Analysis, Data Mining, Text Analytics including Statistical, Linguistic, and Structural techniques to analyse Structured and Unstructured Data</li> <li>● Creating Predictive Forecasting Models</li> <li>● Decision Analysis and Decision Support Systems</li> <li>● Data Mining</li> </ul>	6
<p><b>KAG2-DSENG:</b> Data Science Engineering group including Software and Infrastructure Engineering</p>	<ul style="list-style-type: none"> <li>● Set Up Infrastructure and Big Data applications</li> <li>● Computer networks for high-performance computing and Big Data Infrastructure</li> <li>● Cloud enabled applications development</li> <li>● Modelling and Simulation</li> <li>● Modelling and Simulation theory and techniques (general and domain oriented)</li> <li>● Large scale modelling and simulation systems</li> <li>● Set up Big Data (Data Science) applications design</li> <li>● Programming languages for Big Data analytics: R, Python, Others</li> <li>● Models and languages for complex interlinked Data Presentation and Visualisation</li> </ul>	5
<p><b>KAG3-DSDM:</b> Data Management group including Data Curation, Preservation and Data Infrastructure</p>	<ul style="list-style-type: none"> <li>● Creating Database Models and Data Curation</li> <li>● Data Modelling, Databases and Database Management Systems, Data Models and Query Languages, Database Administration</li> <li>● Set up Data Management and Enterprise Data Infrastructure</li> <li>● Data management, including Reference and Master Data, Data Warehousing and Business Intelligence, Data Storage and Operations</li> <li>● Data Archives/Storage Compliance and Certification Metadata, Linked Data, Provenance</li> </ul>	4

Copyright © IABAC 2017-2020.

IABAC®, IABAC logo®, DSF®, CDS®, CAIE®, CNLPE®, CCVE®, CDSHR®, CDSFIN®, CVAE®, CDLE®, CDSHC®, CDSMKT®, CDS®, DSCMGR®, AICE®, CDSD® and CMLE® are registered trademarks of IABAC BV, The Netherlands. Registration Number (KVK-nummer) :73234893. Reproduction of any part of this material requires written permission of IABAC BV.

All rights reserved. EN\_CDS\_2020\_Syllabus\_V2.0

	<ul style="list-style-type: none"> <li>Data Infrastructure, Data Management and Organisation</li> <li>Research Data Infrastructure, Open Science, Open Data, Open Access, Data Infrastructure Compliance and Certification, Ethical Principle and Data Privacy</li> </ul>	
<b>KAG4-DSRM:</b> Scientific and Research Methods group	<ul style="list-style-type: none"> <li>Scientific/Research Methods</li> <li>Research Methodology, Paradigms and Research Cycle, Modelling and Experiment Planning</li> <li>Data Selection and Quality Evaluation</li> <li>Use case analysis: Research Infrastructures and projects, Research Data Management Plan and Ethical Issues</li> </ul>	6
<b>KAG5-DSBPM:</b> Business Process Management group	<ul style="list-style-type: none"> <li>Business Process Management</li> <li>Business Processes and Operations, Project Scope and Risk Management</li> <li>Business Analysis - Organisation and Management</li> <li>Business Analysis - Planning and Monitoring</li> <li>Requirements Analysis and Design Definition</li> <li>Requirements Life Cycle Management (from inception to retirement) Solution Evaluation and Improvements Recommendation</li> <li>Business Analysis and Enterprise Organisation</li> <li>Agile Data Driven Methodologies, Processes and Enterprises</li> <li>Use case analysis: Business and Industry</li> </ul>	4
<b>KAG6-DSDK:</b> Data Science Domain Knowledge group includes domain specific knowledge	<ul style="list-style-type: none"> <li>Applied Data Science use cases in Domains, HR, Retail, Fraud Analytics, Finance Trends, Health Care, Infrastructure Management</li> </ul>	2

## 5 BLOOM'S TAXONOMY REFERENCE

Bloom's Learning Index	Description
1	Remembering: Recall or retrieve previous learned information.

Copyright © IABAC 2017-2020.

IABAC®, IABAC logo®, DSF®, CDS®, CAIE®, CNLPE®, CCVE®, CDSHR®, CDSFIN®, CVAE®, CDLE®, CDSHC®, CDSMKT®, CDS®, DSCMGR®, AICE®, CDSD® and CMLE® are registered trademarks of IABAC BV, The Netherlands. Registration Number (KVK-nummer) :73234893. Reproduction of any part of this material requires written permission of IABAC BV.

All rights reserved. EN\_CDS\_2020\_Syllabus\_V2.0



2	Understanding: Comprehending the meaning, translation, interpolation, and interpretation of instructions and problems. State a problem in one's own words.
3	Applying: Use a concept in a new situation or unprompted use of an abstraction. Applies what was learned in the classroom into novel situations in the workplace.
4	Analysing: Separates material or concepts into component parts so that its organizational structure may be understood. Distinguishes between facts and inferences.
5	Evaluating: Make judgments about the value of ideas or materials.
6	Creating: Builds a structure or pattern from diverse elements. Put parts together to form a whole, with emphasis on creating a new meaning or structure.

---

Copyright © IABAC 2017-2020.

IABAC®, IABAC logo®, DSF®, CDS®, CAIE®, CNLPE®, CCVE®, CDSHR®, CDSFIN®, CVAE®, CDLE®, CDSHC®, CDSMKT®, CDS®, DSCMGR®, AICE®, CDSD® and CMLE® are registered trademarks of IABAC BV, The Netherlands. Registration Number (KVK-nummer) :73234893. Reproduction of any part of this material requires written permission of IABAC BV.

All rights reserved. EN\_CDS\_2020\_Syllabus\_V2.0