Data Analyst

Certification Specification (Version 2024.1)

Certification Purpose

This certification verifies that individuals have acquired the knowledge and skills required of entry-level data analysts.

The awarded certification will be valid for two years.

Target Audience

The Data Analyst Certification is intended to be used by students and potential employers for demonstrating the level of skill required for employment into an entry-level role. Specific stakeholders are:

- **Data Analysts** aiming to demonstrate that they have acquired the skills required for an entry-level position through a course of training/education
- **Hiring managers** wanting to confirm the skill level of job applicants
- **Students** wanting to measure their achievement in a learning programme
- **Employers** wanting to confirm the current skill level of their teams

Domains and Competencies

In addition to the competencies required for the associate level (see Data Analyst Associate Specification), Data Analysts will be required to demonstrate competency in five domains:

- Data Management
- Exploratory Analysis
- Statistical Experimentation
- Data Communication
- Business Acumen

The following sections break down each domain into the detailed knowledge, skills and abilities (KSAs) that will be required to be demonstrated, along with technologies required where relevant.

Data Management

1. Perform standard data import, joining and aggregation tasks using R or Python
    1.1. Import data from flat files into R or Python.
    1.2. Import data from databases into R or Python.
    1.3. Aggregate numeric, categorical variables and dates by groups using R or Python.
    1.4. Combine multiple tables by rows or columns using R or Python.
    1.5. Filter data based on different criteria using R or Python.

Exploratory Analysis

1. Calculate metrics to effectively report characteristics of data and relationships between features using Python or R
    1.1. Calculate measures of center (e.g. mean, median, mode) for variables using R or Python.
    1.2. Calculate measures of spread (e.g. range, standard deviation, variance) for variables using R or Python.
    1.3. Calculate skewness for variables using R or Python.
    1.4. Calculate missingness for variables and explain its influence on reporting characteristics of data and relationships in R or Python.
    1.5. Calculate the correlation between variables using R or Python.

2. Create data visualizations in R or Python to demonstrate the characteristics of data
    2.1. Create and customize bar charts using R or Python.
    2.2. Create and customize box plots using R or Python.
    2.3. Create and customize line graphs using R or Python.
    2.4. Create and customize histograms graph using R or Python.

3. Create data visualizations in R or Python to represent the relationships between features
    3.1. Create and customize scatterplots using R or Python.
    3.2. Create and customize heatmaps using R or Python.

3.3. Create and customize pivot tables using R or Python.

Statistical Experimentation

1. Apply sampling methods to data
    1.1. Distinguish between different types of random sampling techniques and apply the methods using R or Python
    1.2. Sample data from a statistical distribution (e.g. normal, binomial, Poisson, exponential, etc.) using R or Python
    1.3. Calculate a probability from a statistical distribution (e.g. normal, binomial, Poisson, exponential, etc.) using R or Python

2. Implement methods for performing statistical tests
    2.1. Run statistical tests (e.g. t-test, ANOVA test, chi-square test) using R or Python.
    2.2. Analyze the results of statistical tests from R or Python.

Data Communication

1. Frame, convey, and summarize stories using data
    1.1. Employ techniques in data storytelling to propose findings and relay solutions to business stakeholders
2. Employ multiple tactics (written and verbal) to communicate to business leaders
    2.1. Deliver a verbal presentation addressing the business goals, outcomes and recommendations
    2.2. Provide a written explanation of findings and/or reasoning for selecting approaches

Business Acumen

1. Collect relevant information, detect patterns, observe and interpret data
    1.1. Describe business goals
    1.2. Explain how solution addresses the business problem
    1.3. Provide recommendations for future action to be taken based on the outcome of the work done

2. Benchmark, monitor, and evaluate business processes

    2.1. Define metrics that can be used by the business in the future to measure success in solving the problem

    2.2. Evaluate metrics using the existing data to provide a baseline measure for the problem

Exam Format

To earn the Data Scientist Associate Certification, candidates will have:

- To complete 3 exams, or 2 exams if they currently hold the associate level certification
- A total of 30 days to complete all of the exams
- To pass the exams in the order specified
- To pass each exam before progressing to the next
- Two attempts to pass each exam within the 30 days

All exams are offered in English only.

Exams will be taken online. Candidates will need a laptop or computer with internet access.

Exam	Type	Number of items	Time Available
DA101	Timed	60	2 hours
DA201	Timed	45	2 hours
DA601P	Practical	NA	NA

Or for holders of the Data Science Associate Certification:

Exam	Type	Number of items	Time Available
DA201	Timed	45	2 hours
DA601P	Practical	NA	NA

Results

Results for the timed exams will be displayed upon completion of the exam. Results will include a score for each domain as well as an overall score. The overall score determines the final pass/fail decision.

Results for the timed exams are calculated as an average of the scores for each domain included in the exam.

Results for the practical exam will be available within 14 days of submission. Candidates will be emailed to inform them that their results are available to review. The practical exam will be a pass/fail exam. All criteria must be passed. No scores will be given.

Re-taking

Candidates who are unsuccessful in any component will have to wait 14 days before they can attempt the certification again. They will be informed of the domains where they were unsuccessful. They will have to complete all exams again, including any that they may have passed on a previous attempt.

Unscored Items

Candidates may be presented with one unscored item for each domain & technology combination during a timed exam. These items are included to validate them for use in future exams. They will not count towards the final score received. They will not be highlighted during the test taking process.

Exam Content

The following sections outline the content breakdown for each exam. Each section contains an overview of the number of items that we will ask for each domain within an exam. There is also a breakdown of the overall pool size and the split of the types of questions contained within each domain.

DA101

The first exam for the Data Analyst certification will cover the competencies of the associate level (please refer to the associate level specification for the detailed competencies and KSAs).

The exam contains Data Management in SQL, Statistical Experimentation and Exploratory Analysis theory and Exploratory Analysis in SQL.

Domain	Technology	Number of competencies	Items per Competency	Total Items in Exam
Exploratory Analysis	Theory	2	7-8	15
Exploratory Analysis	SQL	1	15	15
Data Management	SQL	3	3-5	15

Statistical Experimentation	Theory	1	15	15
				60

DA201

The second exam for the Data Analyst Certification will cover all of the remaining domains and competencies.

Candidates will have to select either Python or R before starting the exam and all domains will use the same technology.

Domain	Technology	Number of competencies	Items per Competency	Total Items in Exam
Statistical Experimentation	Python or R	2	7-8	15
Data Management	Python or R	1	15	15
Exploratory Analysis	Python or R	3	5	15
				45

DA601P

The practical exam will test KSAs in the following domains:

- Data Management
- Exploratory Analysis
- Data Communication
- Business Acumen

The exam must be completed in Python or R.

Candidates will be given a single real-world scenario. They will be required to:

- submit a written report intended for a technical audience
- deliver a spoken presentation intended for a non-technical audience

Submissions will be graded by human graders against the following criteria:

Competency	Sufficient	Insufficient
Data Management		
Assess data quality and	Has validated all variables	Has not conducted all the

perform validation tasks	and where necessary has performed cleaning tasks to result in analysis-ready data.	required checks and/or has not cleaned the data. May have removed data rather than performed cleaning tasks.
Exploratory Analysis		
Create data visualizations in R or Python to demonstrate the characteristics of data and represent the relationships between features	Has created at least two different visualizations of single variables (e.g. histogram, bar chart, single boxplot) Has created at least one visualization including two or more variables (e.g. scatterplot, filled barchart, multiple boxplots) Has used visualizations that support the findings being presented	Has used the same visualization throughout. Has not included graphics to represent single variables and relationships. Has not used visualizations that support the findings being presented.
Data Communication		
Employ multiple tactics (written and verbal) to communicate to business leaders	For each analysis step, has provided a written explanation of their findings and/or reasoning for selecting approaches Has delivered a verbal presentation addressing the business goals, outcomes and recommendations	Has not provided a written summary for each step Has not delivered a verbal presentation
Business Acumen		
Collects relevant information, detects patterns, observes and interprets data	Has described at least one of the business goals of the project Has explained how their work has addressed the business problem Has provided at least one recommendation for future action to be taken based on the outcome of the work done	Has not identified any business goals Has not explained how their work has addressed the business problem Has not provided any recommendations for future actions

Benchmarks, monitors, and evaluates business processes	Has defined a metric that can be used by the business in the future to measure success in solving the problem Has evaluated the metric using the existing data to provide a baseline measure for the problem	Has not identified a metric to compare performance to the business problem or has not shown the metric with the current data

Item Formats

The following information provides examples of the item formats used across the timed and practical exams and their purpose.

Timed Exams

The timed exams will use items of three formats:

- Multiple choice
- Fill-in-the-blank
- Typing

Practical Exam

The practical exam will require submission of a published workbook.

Test takers will need to create a workbook from their Certification dashboard that will be shared with members of the grading team and DataCamp Certification team admins for grading purposes.

On submission of their published workbook, candidates will be able to use our recording tool to record their submission. It is not required to submit presentation materials. It is not possible to submit recordings in any other format.

Submissions will be automatically allocated for grading at random to a member of the grading team. A proportion of submissions will be allocated to multiple graders as part of our quality assurance process.

An example of a practical exam can be viewed here.