Data Analyst Associate

Certification Specification (Version 2023.1)

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 Associate 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

Data Analyst Associates will be required to demonstrate competency in four domains:

- Data Management

- Exploratory Analysis

- Statistical Experimentation

- Data Communication

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 extraction, joining and aggregation tasks using SQL

    1.1. Aggregate numeric, categorical variables and dates by groups using PostgreSQL.

    1.2. Interpret a database schema and combine multiple tables by rows or columns using PostgreSQL.

    1.3. Extract data based on different conditions using PostgreSQL.

    1.4. Use subqueries to reference a second table (e.g. a different table, an aggregated table) within a query in PostgreSQL

2. Perform standard cleaning tasks to prepare data for analysis using PostgreSQL

    2.1. Match strings in a dataset with specific patterns using PostgreSQL.

    2.2. Convert values between data types in PostgreSQL.

    2.3. Clean categorical and text data by manipulating strings in PostgreSQL.

    2.4. Clean date and time data in PostgreSQL.

3. Assess data quality and perform validation tasks using PostgreSQL

    3.1. Identify and replace missing values using PostgreSQL.

    3.2. Perform different types of data validation tasks (e.g. consistency, constraints, range validation, uniqueness) using PostgreSQL.

    3.3. Identify and validate data types in a data set using PostgreSQL.

Exploratory Analysis

1. Calculate metrics to effectively report characteristics of data and relationships between features using PostgreSQL

    1.1. Calculate measures of center (e.g. mean, median, mode) for variables using PostgreSQL.

    1.2. Calculate measures of spread (e.g. range, standard deviation, variance) for variables using PostgreSQL.

    1.3. Calculate skewness for variables using PostgreSQL.

    1.4. Calculate missingness for variables and explain its influence on reporting characteristics of data and relationships in PostgreSQL.

1.5.   Calculate the correlation between variables using PostgreSQL.

2.   Read and analyze data visualizations to demonstrate characteristics of data

      2.1.   Distinguish between different types of data visualizations (e.g. bar chart, box plot, line graph, and histogram) in demonstrating the characteristics of data.

      2.2.   Interpret data visualizations (e.g. bar chart, box plot, line graph, and histogram)and summarize the characteristics of data.

3.   Read and analyze data visualizations to represent the relationship between features

      3.1.   Distinguish between different types of data visualizations (e.g. scatterplot, heatmap, and pivot table) in representing the relationships between features.

      3.2.   Interpret data visualizations (e.g. scatterplot, heatmap, and pivot table) and summarize the relationship between features.

Statistical Experimentation

1.   Describe statistical concepts that underpin hypothesis testing and experimentation

      1.1.   Define different statistical distributions (e.g. binomial, normal, Poisson, t-distribution, chi-square, and F-distribution, etc.).

      1.2.   Explain the statistical concepts in hypothesis testing (e.g. null hypothesis, alternative hypothesis, one-tailed and two-tailed hypothesis tests, etc.).

      1.3.   Explain the statistical concepts in the experimental design (e.g. control group, randomization, confounding variables, etc.).

      1.4.   Explain parameter estimation and confidence intervals.

Data Communication

1.   Present data concepts to small, diverse audiences

      1.1.   Explain findings and/or the reasoning for selecting approaches.

2.   Employ data visualization to support findings

      2.1.   Create charts using visualization tools.

      2.2.   Use visualizations that support the findings being presented.

Exam Format

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

- To complete 2 exams

- 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
DA501P	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 Associate certification will cover 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

DS501P

The practical exam will test KSAs in the following domains:
- Data Management
- Exploratory Analysis
- Data Communication

The exam can be completed using any tools and technologies but must be written up using DataCamp Workspace.

Candidates will be given a single real-world scenario. They will be required to:
- submit a written report intended for a technical audience

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

Competency	Sufficient	Insufficient
Data Management		
Assess data quality and perform validation tasks	Has validated all variables and where necessary has performed cleaning tasks to result in analysis-ready data.	Has not conducted all the required checks and/or has not cleaned the data. May have removed data rather than performed cleaning tasks.
Exploratory Analysis		
Create data visualizations in 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		
Presents data concepts to small, diverse audiences	For each analysis step, has provided a written explanation of their findings and/or reasoning for selecting approaches	Has not provided a summary for each step (data quality, exploratory analysis, model fitting and model evaluation)

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 workspace.

Test takers will need to create a workspace from their Certification dashboard that will be shared with members of the grading team and DataCamp Certification team admins for grading purposes. They can choose to use any tools or technologies to complete the work but must write it up using a Markdown DataCamp Workspace. If candidates wish to use R or Python they can choose this option when creating their Workspace.

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.