



CBA STANDARD

Data Analytics Professional

Technology · 6 weighted domains · 34 learning objectives

Version 7.0

This edition published September 2026

About this standard

This standard defines what competent practice consists of in the discipline set out under Scope below. It divides the discipline into weighted domains and states, for each domain, the learning objectives a competent practitioner meets. It is one of the CBA Standards, which the Corporate Business Alliance maintains across finance, accounting, management and technology.

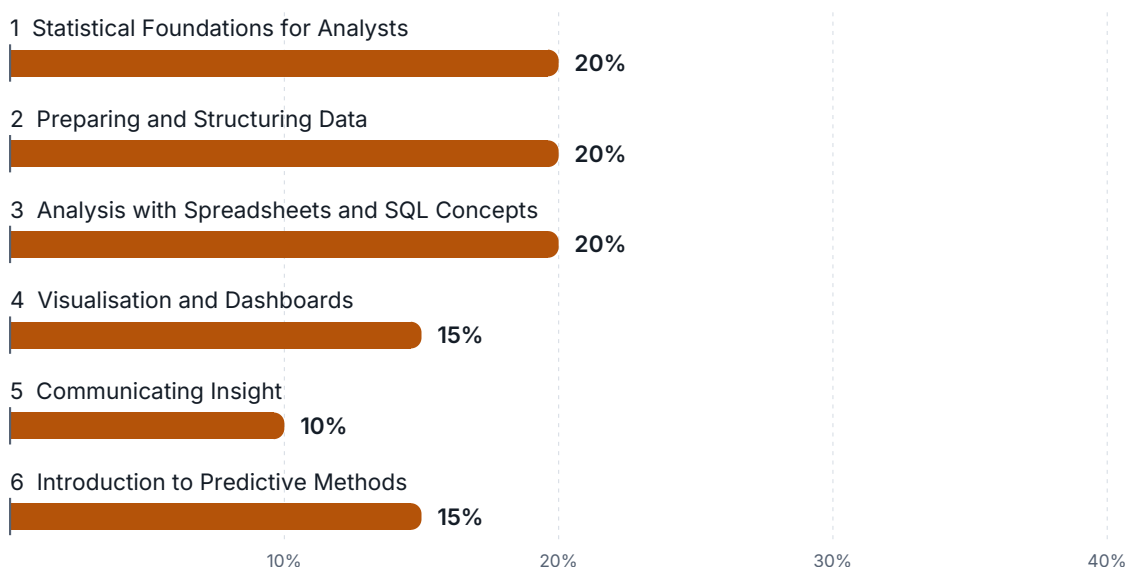
The standard is written for employers defining what a role requires, for training providers designing courses to a published specification, and for practitioners planning their own development.

Each domain carries a weight: its share of the standard, and of the examination for the CBA-DAP credential, which is assessed against it. The version number rises whenever a change would alter what competence in the discipline means, and every credential records the version its holder met.

Scope

A professional certification covering the full analytics workflow: statistical reasoning, data preparation, spreadsheet and SQL analysis, visualisation and dashboards, insight communication, and an introduction to predictive methods.

Weighting by domain



Domains and learning objectives

Domain 1

20% of the standard

Statistical Foundations for Analysts

- 1.1 Calculate and interpret measures of central tendency, dispersion and position, selecting the measure appropriate to the data's shape and scale.
- 1.2 Distinguish between population and sample, parameter and statistic, and correlation and causation in described business scenarios.
- 1.3 Interpret probability statements, distributions and expected values as they arise in everyday

analytical questions.

- 1.4 Evaluate the reliability of a sample-based conclusion using confidence intervals, margin of error and sample size reasoning.
- 1.5 Select the appropriate basic significance test for a described comparison and interpret its result, including what a p-value does and does not say.
- 1.6 Identify common statistical traps in analyses, including survivorship bias, Simpson's paradox, regression to the mean and misuse of averages.

Domain 2

20% of the standard

Preparing and Structuring Data

- 2.1 Identify data quality defects in a described or shown dataset, including duplicates, inconsistent categories, invalid values, and mixed data types.
- 2.2 Select the appropriate cleaning action for a defect, weighing deletion, correction, imputation and flagging, and state its effect on later analysis.
- 2.3 Distinguish between wide and long data layouts and select the reshaping needed for a stated analysis or chart.
- 2.4 Interpret relationships between tables, including primary and foreign keys, one-to-many and many-to-many structures, and the effect of granularity on aggregation.
- 2.5 Evaluate a described data collection or combination step for bias, leakage of meaning, or loss of information, including date, text and unit handling issues.
- 2.6 Select documentation and reproducibility practices that let another analyst retrace a preparation workflow.

Domain 3

20% of the standard

Analysis with Spreadsheets and SQL Concepts

- 3.1 Select the correct spreadsheet function or function combination (lookup, conditional aggregate, logical, text, date) for a described analytical task.
 - 3.2 Interpret what a given formula or SQL query returns, including its filtering, grouping and calculation logic.
 - 3.3 Distinguish between SQL clauses and their spreadsheet equivalents, including filtering rows versus filtering groups and joining versus looking up.
 - 3.4 Calculate summary results that a described pivot table or GROUP BY query would produce from a small dataset.
 - 3.5 Evaluate an analytical approach for correctness and efficiency, including common formula and query errors such as wrong join type, absolute versus relative references, and aggregation before filtering.
 - 3.6 Select appropriate verification steps for a formula, query or AI-generated draft of either, before its output is used in a decision.
-

Domain 4

15% of the standard

Visualisation and Dashboards

- 4.1 Select the chart type that best answers a stated analytical question, given the data types and comparison involved.
- 4.2 Identify design faults that distort or obscure data, including truncated axes, dual axes misuse, 3D effects, overplotting and rainbow category colouring.
- 4.3 Evaluate a described dashboard against its audience and decision purpose, including layout hierarchy, filter design and refresh expectations.
- 4.4 Interpret charts accurately, including log scales, stacked and 100 percent stacked bars, box plots and heat maps.
- 4.5 Select accessibility and labelling practices that make a visual readable for colour-blind viewers and international audiences.

Domain 5

10% of the standard

Communicating Insight

- 5.1 Select the framing of a finding that is accurate, decision-relevant and appropriately caveated for a stated audience.
- 5.2 Distinguish between observation, interpretation and recommendation in analytical statements.
- 5.3 Evaluate a written analytical claim for overreach, including generalising beyond the data, hiding uncertainty and burying assumptions.
- 5.4 Interpret stakeholder questions to identify the underlying decision and the analysis actually required.
- 5.5 Select an appropriate structure for an analytical narrative, including headline-first summaries and the role of appendices and reproducibility notes.

Domain 6

15% of the standard

Introduction to Predictive Methods

- 6.1 Distinguish between descriptive, diagnostic, predictive and prescriptive questions, and between supervised and unsupervised approaches.
- 6.2 Interpret the output of a simple linear regression, including slope, intercept, R-squared and residual patterns, in business terms.
- 6.3 Select an appropriate baseline forecasting approach for a described time series, including naive, moving average and trend-with-seasonality reasoning.
- 6.4 Calculate and interpret basic evaluation measures, including forecast error measures and classification accuracy, precision and recall from a small confusion matrix.
- 6.5 Identify overfitting, data leakage and unrepresentative training data as causes of models that fail after deployment.
- 6.6 Evaluate a described predictive claim or AI-generated model output for plausibility, fairness concerns and fitness for the stated decision.

Competence at award

What a holder of the CBA-DAP credential has demonstrated.

- Apply descriptive and inferential statistics correctly to business data, including distributions, variability, confidence intervals and common significance tests.
- Prepare raw data for analysis by profiling, cleaning, reshaping and joining datasets, and document the transformations so results are reproducible.
- Analyse datasets using spreadsheet functions, pivot summaries and conceptual SQL, selecting the right aggregation, filter and join logic for the question asked.
- Design charts and dashboards that match chart type to analytical purpose, avoid distortion, and support decision making for a defined audience.
- Communicate analytical findings as clear narratives and recommendations, stating assumptions, limitations and uncertainty honestly.
- Explain and evaluate introductory predictive methods, including regression, classification basics, time-series forecasting and how model quality is measured.
- Judge when and how to use AI assistance in an analytical workflow, and verify AI-produced formulas, queries and summaries before relying on them.

Recognition at Practitioner level

The CBA-DAP credential is awarded at Practitioner level of the CBA Competency Framework, which evidences that the holder can carry out the core tasks of the discipline.

Awarded on the examination and 4 assessed workbooks of applied work.

Who the standard is written for

Working analysts formalising their skills

Business, operations, finance or marketing analysts who learned on the job and want an independent credential confirming that their statistical reasoning, data preparation and reporting practice meet a professional standard.

Career changers entering analytics

Professionals moving from administration, accounting support, engineering or customer-facing roles who have built spreadsheet and basic SQL skills through self-study and need a structured, examined benchmark to present to employers.

Managers and specialists who commission analysis

Team leads, product owners and consultants who do not analyse data full time but must scope analytical work, question dashboards and forecasts intelligently, and judge whether conclusions placed in front of them are sound.

Appendix: glossary of technology terms

Terms used across the technology standards, as CBA defines them.

A/B test

A controlled experiment comparing two versions of something by assigning users at random to each, so that a difference in outcome can be attributed to the change.

Acceptable use policy

An organisation's rules on which AI tools staff may use, for what purposes, and with what data — including what must never be entered into an external service.

AI impact assessment

A structured assessment, made before deployment, of how an AI system could affect individuals, groups and the organisation, and of the controls needed in response.

AI risk classification

Assigning each AI system a risk level according to the consequence of it being wrong — who could be harmed, how seriously and how easily an error would be noticed — so that controls are proportionate.

AI system inventory

A register of every AI system an organisation uses or builds, recording what each does, the decisions it informs, its owner, its data and how its outputs are checked.

Algorithm

A defined sequence of steps for solving a problem or performing a calculation.

Algorithmic bias

Systematic error in a model's outputs that disadvantages particular groups, often arising from unrepresentative training data or from features that act as proxies for protected characteristics.

Anonymisation

Processing personal data so that individuals can no longer be identified by any means reasonably likely to be used. Truly anonymised data falls outside data protection law; pseudonymised data does not.

Artificial intelligence

Computer systems that perform tasks associated with human intelligence, such as recognising patterns, making predictions, understanding language or generating content.

Classification

A machine learning task that assigns inputs to categories, such as flagging a transaction as fraudulent or not.

Confusion matrix

A table comparing a classification model's predictions with actual outcomes, showing true and false positives and negatives.

Context window

The maximum amount of text, measured in tokens, that a language model can take into account at once. Material beyond it is not considered, which can cause instructions or facts to be lost in long tasks.

Correlation and causation

Correlation is a statistical association between two variables; causation means one brings about the other. Correlation alone does not establish causation.

Dashboard

A visual display of the key measures needed to monitor a process or objective, arranged so the most important information is read first.

Data cleansing

Detecting and correcting errors, duplicates, inconsistent formats and missing values in a dataset before analysis.

Data governance

The roles, policies and standards that determine who owns, may use and is accountable for an organisation's data, and how its quality and security are maintained.

Data lineage

The documented path of data from its source through each transformation to where it is used, so that a figure in a report can be traced back.

Data minimisation

The data protection principle that personal data must be adequate, relevant and limited to what is necessary for the purpose.

Data quality

The degree to which data is accurate, complete, consistent, timely and fit for its intended use.

Data visualisation

Representing data graphically — charts, maps, tables — so that patterns and comparisons can be understood quickly and accurately.

Data warehouse

A central store of structured data integrated from several sources and organised for reporting and analysis.

Descriptive analytics

Analysis that summarises what has happened, through measures, aggregations and reports.

Diagnostic analytics

Analysis that investigates why something happened, by drilling into data, comparing segments and testing explanations.

ETL

Extract, transform, load: the process of taking data from source systems, converting it into a consistent structure and loading it into a target store such as a data warehouse.

EU AI Act

Regulation (EU) 2024/1689, the European Union's law on artificial intelligence. It classifies AI systems by risk, prohibits certain practices and places obligations on providers and deployers of high-risk systems.

Explainability

The extent to which the reasons for an AI system's output can be understood by the people who rely on it or are affected by it.

Feature

An input variable used by a machine learning model, such as transaction amount or customer tenure.

Fine-tuning

Further training of a pre-trained model on a smaller, task-specific dataset to adapt its behaviour.

Generative AI

AI systems that produce new content — text, images, code, audio — in response to a prompt, based on patterns learned from their training data.

Hallucination

Output from a generative AI system that is fluent and plausible but factually wrong or unsupported, such as an invented figure or citation.

Human oversight

Arrangements that allow people to monitor, question and override an AI system. It is real only where the reviewer has the time, information and authority to intervene.

ISO/IEC 42001

The international standard, published in 2023, specifying requirements for an artificial intelligence management system within an organisation.

Large language model (LLM)

A generative AI model trained on large volumes of text to predict and produce language, used for drafting, summarising, answering questions and similar tasks.

Machine learning

A branch of AI in which systems learn patterns from data to make predictions or decisions, rather than following explicitly programmed rules.

Median

The middle value of a set of numbers arranged in order. Unlike the mean, it is not pulled by extreme values.

Model card

A short document describing an AI model's intended use, training data, performance, limitations and known risks.

Model drift

The decline in a model's accuracy over time as the data it meets in use departs from the data it was trained on.

NIST AI Risk Management Framework

A voluntary framework published by the US National Institute of Standards and Technology in 2023 to help organisations identify, assess and manage the risks of AI systems.

Outlier

A data point far from the others. It may be an error to correct or a genuine and important observation, and should be investigated rather than removed by default.

Overfitting

When a model learns the noise in its training data as well as the signal, performing well on that data but poorly on new data.

Personal data

Under UK and EU data protection law, any information relating to an identified or identifiable living individual.

Pivot table

A spreadsheet tool that summarises a table of records by grouping and aggregating them along chosen dimensions.

Precision and recall

Precision is the share of cases a model flags that are truly positive; recall is the share of truly positive cases the model flags. Improving one usually costs the other.

Predictive analytics

Analysis that estimates what is likely to happen, using statistical or machine learning models built on historical data.

Prescriptive analytics

Analysis that recommends a course of action, typically by combining predictions with optimisation or decision rules.

Prompt

The instruction and context given to a generative AI system. A well-specified prompt states the task, the relevant context, the constraints and the expected form of the output.

Pseudonymisation

Replacing identifying details with a code, so that data cannot be attributed to an individual without separately held information. Pseudonymised data remains personal data.

Red-teaming

Deliberately attempting to make an AI system fail — to produce harmful, wrong or prohibited output — in order to find weaknesses before users do.

Regression

A statistical or machine learning method for predicting a continuous value, such as sales, from one or more input variables.

Retrieval-augmented generation (RAG)

A technique in which relevant documents are retrieved and supplied to a language model with the prompt, so that its answer is grounded in that material rather than in its training alone.

Robustness

An AI system's ability to keep performing correctly when inputs are noisy, unusual or deliberately manipulated.

Shadow AI

The use of AI tools by staff without the organisation's knowledge or approval, often with data that should not leave it.

SQL

Structured Query Language, the standard language for querying and managing data in relational databases.

Standard deviation

A measure of how widely values are spread around their mean. Roughly two thirds of a normally distributed set lie within one standard deviation of the mean.

Statistical significance

An indication that an observed result would be unlikely if there were no real effect. It does not show that the effect is large or important.

Supervised learning

Machine learning from examples labelled with the correct answer, so that the model learns to predict the label for new cases.

Test set

Data held back from training and used only to measure how well a finished model performs on data it has not seen.

Token

The unit of text a language model processes — often a word or part of a word. Context limits and usage costs are measured in tokens.

Training data

The data from which a machine learning model learns. Its quality, coverage and representativeness largely determine the model's behaviour.

Unsupervised learning

Machine learning that finds structure in unlabelled data, such as grouping similar customers into segments.

Using this publication

Citation

Corporate Business Alliance (2026). Data Analytics Professional: CBA Standard, version 7.0.
London: Corporate Business Alliance.

Terms of use

© Corporate Business Alliance 2026. You may download, print and use this publication, and share it unaltered and free of charge, for personal study and within your organisation. It may not be sold, or republished in whole or in part, without the written permission of the Corporate Business Alliance.

Status

The CBA Standards and CBA credentials are the professional standards and certifications of the Corporate Business Alliance. CBA credentials are not regulated qualifications, and CBA is not accredited to ISO/IEC 17024; CBA claims no equivalence to the qualifications of any other body.

Publisher

Corporate Business Alliance Ltd

Registered office: 167–169 Great Portland Street, 5th Floor, London, W1W 5PF, United Kingdom
support@cba-uk.com · cba-uk.com