CERTLABZ CERTIFICATION BOARD

Certified Data Analytics Specialist (CDAS)

Ten Sample Questions

These ten questions are written at the difficulty of the live exam. Six carry a figure, a data table, a configuration or a code listing; four are plain scenario questions, as on the live paper. Every one is a scenario a working data analyst would meet, and none of them appears in the live question bank.

The live exam is multistage adaptive: how you score on one stage selects the difficulty of the next, so no two candidates sit the same paper. It is a performance-based exam: each form can carry up to 5 to 12 performance-based questions alongside the multiple-choice questions, which means you build, sequence or complete something rather than only choosing from a list. Answers and full explanations follow each question here; the live exam does not disclose item-level answers.

Question 1 — Domain 4: Visualization and Dashboards — DATA

Emma Wilson must pick a chart for each communication goal.

Goal	Data
Show monthly revenue over two years	One value per month
Compare sales across 8 regions	One value per region
Show relationship of ad spend to sales	Two numeric variables
Show share of 3 product lines	Three parts of a whole

Table 1.1 Goals and the data available.

Which set of chart choices is best?

A. Pie for the trend, line for regions, bar for the relationship, and a scatter plot for the three product-line shares shown.

B. Line for the trend, bar for the regions, scatter plot for the relationship, and a pie for the three product-line shares.

C. Scatter for the trend, pie for regions, line for the relationship, and a bar chart for the three product-line shares shown.

D. Bar for the trend, scatter for regions, pie for the relationship, and a line chart for the three product-line shares shown.

Correct answer: B
A line chart shows a trend over time, a bar chart compares values across categories such as regions, a scatter plot shows the relationship between two numeric variables, and a pie chart suits a few parts of a whole such as three product lines. The other options misassign these standard chart-to-intent pairings.

Question 2

Domain 4: Visualization and Dashboards

CONFIG

James Carter reviews a bar chart of quarterly results.

```
chart: bar
y-axis: starts at 90 (not 0), ends at 100
bars: Q1 92, Q2 94, Q3 95, Q4 96
caption: "Huge growth this year!"
```

Listing 2.1 The chart configuration.

Why is this chart misleading, and what fixes it?

A. It is fine as drawn, because starting the axis at 90 zooms in on the values and makes the small changes easier to read.

B. The truncated y-axis starting at 90 exaggerates small differences; start the axis at zero so the modest change is shown honestly.

C. The problem is only the caption wording, and rewriting it while keeping the axis starting at 90 fully resolves the misleading effect.

D. The chart should be a pie chart instead, because a pie would represent the four quarterly values more accurately than any bar chart.

Correct answer: B
A bar chart with a y-axis starting at 90 makes a rise from 92 to 96 look enormous when it is a few percent. Bars encode value by length, so they should start at zero to be honest. Fixing only the caption leaves the visual distortion, and a pie chart is wrong for comparing quarterly values.

Question 3

Domain 3: Exploratory Analysis and Statistics

DATA

Olivia Davis reviews four analytical claims.

Observation	Proposed claim
Ice cream sales and drownings both rise in summer	Ice cream causes drowning
A/B test: B better, p-value 0.3	B is proven better
Mean income up, median income down	Everyone earns more
Result from a sample of 6 people	Applies to all customers

Table 3.1 Observations and claims.

Which claim is sound?

A. The ice cream claim, because two things rising together over the summer months is strong evidence of a direct causal link.

B. None of the claims is sound; each ignores confounding, statistical significance, skew, or the unreliability of a tiny sample.

C. The A/B test claim, because B scored higher in the test and a higher score in a single test proves that B is the better option.

D. The mean income claim, because a rising mean income always means that every individual in the group is now earning more money.

Correct answer: B
Every claim is flawed: the ice cream and drowning link is driven by the confounder of hot weather, a p-value of 0.3 is not significant, a rising mean with a falling median signals skew rather than everyone earning more, and a sample of six is far too small to generalise. None is a sound conclusion.

Question 4 — Domain 2: Data Preparation, Cleaning and Quality — DATA

Liam Anderson profiles a customer table before analysis.

Finding	Detail
Customer name	"Acme", "ACME Inc", "Acme, Inc." for one firm
Age column	Some values are 999
Order date	Stored as text in places
Rows	Some exact duplicates

Table 4.1 Profiling findings.

Which set of fixes is correct?

A. Leave the name variants as they are, keep age 999 as a valid age, keep dates as text, and keep the duplicate rows for completeness.

B. Standardise the name variants, treat age 999 as an invalid placeholder, cast dates to a date type, and deduplicate the repeated rows.

C. Delete every customer with any name variant, delete the whole age column, delete all dates, and keep duplicates as separate customers.

D. Encrypt the customer names, replace all ages with zero, convert dates to random values, and merge duplicates into one giant row per table.

Correct answer: B
The name variants are inconsistent values to standardise and deduplicate, age 999 is an invalid placeholder to treat as missing, text dates should be cast to a proper date type, and exact duplicate rows should be removed. Leaving the issues distorts results, and the destructive options throw away or corrupt valid data.

Question 5 — Domain 3: Exploratory Analysis and Statistics — CONFIG

Ava Thompson summarises customer spend.

```
spend values: mostly $20 to $60
plus three customers at $50,000 each
mean spend  : $310
median spend: $42
```

Listing 5.1 Spend summary.

Which statistic better represents a typical customer, and why?

A. The mean of $310, because the mean always represents the typical customer more accurately than the median in any dataset.

B. The median of $42, because a few very large values pull the mean up, so the median better reflects a typical customer here.

C. The mean of $310, because including the three largest customers is essential to describing what a typical customer spends.

D. Neither statistic is useful, because with any variation in the data no single number can summarise customer spend at all.

Correct answer: B
A few $50,000 customers pull the mean up to $310, far above what almost anyone spends, while the median of $42 sits among the typical $20 to $60 values. With this skew, the median better represents a typical customer. The mean is still worth reporting alongside, but it is not the typical value.

Question 6

Domain 6: Governance, Ethics and Analytics Operations

CONFIG

An analyst is asked to adjust a report before a review.

```
request: "remove the two weeks where sales dipped
         so the trend looks like steady growth"
context: the dip is real, verified data
```

Listing 6.1 The request.

What is the right response?

A. Remove the two weeks as requested, because presenting a smoother, more positive trend is a reasonable way to reassure the reviewers.

B. Decline to delete valid data to change the story; present the real trend, and separately explain the dip and any context around it.

C. Remove the two weeks and also any other low points, so that the trend line is as clean and upward as it can possibly be made.

D. Replace the dip with the average of the surrounding weeks, since smoothing over a short dip is a harmless adjustment to the report.

Correct answer: B
Deleting real, verified data to manufacture a nicer trend misleads the reviewers and is an integrity failure. The analyst should present the true trend and explain the dip and its context. Removing low points or silently smoothing them over is the same dishonesty in a different form.

Question 7 Domain 1: Analytics Foundations and Lifecycle

A stakeholder says "just give me all the data on sales". What should the analyst do first, and why?

A. Export every sales table and hand it all over, because giving the stakeholder the complete raw data is the most helpful response.

B. Clarify the decision and specific question behind the request, so the analysis focuses on something useful rather than a data dump.

C. Immediately build ten different dashboards covering every angle of sales, so the stakeholder is certain to find what they need somewhere.

D. Decline the request outright, because a request phrased as broadly as "all the data on sales" can never be turned into useful analysis.

Correct answer: B
A vague request for "all the data" rarely helps anyone. The analyst adds value by clarifying the decision and the specific question behind it, so the analysis answers something that matters. A raw data dump, a scattershot of dashboards, or refusing the request all miss that core skill of framing the question.

Question 8 Domain 3: Exploratory Analysis and Statistics

A dashboard shows two variables that are strongly correlated. A colleague concludes that one must be causing the other. How should you respond?

A. Agree, because a strong correlation between two variables is reliable proof that one of them is directly causing the other one.

B. Explain that correlation does not imply causation; the link may be coincidental or driven by a confounding third variable to investigate.

C. Agree, but only if the correlation is above 0.9, because any correlation stronger than that threshold always guarantees a causal link.

D. Dismiss the correlation entirely, because correlations between variables are meaningless and can never suggest anything worth exploring.

Correct answer: B
A strong correlation can arise from coincidence or a confounding variable that drives both, so it does not prove causation. The right response is to explain this and investigate further, for example with a controlled experiment. No correlation threshold proves causation, but correlations are still useful signals worth exploring, not meaningless.

Question 9 Domain 5: Business Intelligence and Communication

An analyst presents twenty slides of statistical methodology to executives, who then ask "so what should we do?". What went wrong?

A. Nothing went wrong, because executives should be walked through the full statistical methodology before any recommendation is made.

B. The communication led with method instead of the decision-relevant insight and a clear recommendation for that executive audience.

C. The analysis itself must have been incorrect, because a correct analysis would never prompt an audience to ask what they should do.

D. The executives simply need training in statistics, and once they understand the methods the presentation will communicate effectively.

Correct answer: B
For an executive audience, communication should lead with the insight and a clear recommendation, keeping detailed method available but secondary. Twenty slides of methodology buries the answer, which is why they still had to ask what to do. The analysis may be fine; the problem is communicating at the wrong altitude, not the executives' statistics knowledge.

Question 10 Domain 6: Governance, Ethics and Analytics Operations

Two departments report different revenue for the same period from what should be the same data. What is the most likely cause, and the fix?

A. One department is deliberately falsifying its numbers, so the fix is to escalate the discrepancy as suspected fraud for investigation.

B. Different metric definitions or data sources; align on a single source of truth and an agreed definition so figures are consistent.

C. Revenue is inherently random from one report to the next, so small differences between departments are expected and need no fix.

D. The charts are drawn differently, so simply standardising the chart colours and layout between the two reports will reconcile the figures.

Correct answer: B
Conflicting numbers for the same period almost always trace to different metric definitions or different data sources, not fraud or randomness. The fix is governance: agree a single source of truth and a precise, shared definition so both departments compute the figure the same way. Chart styling has nothing to do with the underlying discrepancy.

Published by the CertLabz Certification Board, a division of CertLabz LLC. Companion to objectives document CDAS-001. https://certlabz.com/certifications/certified-data-analytics-specialist.html