

PRACTICE QUESTIONS

DASCA

Principal Data Scientist

SHARED BY [u/Rustic-Ibis-6342](#) on aiexamcerts

EDITION	Free Set 2
QUESTIONS	10
SHARED	October 15, 2026
EXAM	DASCA Principal Data Scientist

Shared by [u/Rustic-Ibis-6342](#) on aiexamcerts

10 practice questions for the DASCA Principal Data Scientist exam, each followed by its answer and a short explanation. More free sets and the full practice exam: www.aiexamcerts.com/dasca/principal-data-scientist. Spotted a mistake? Email support@aiexamcerts.com with the exam name and the question number.

1. Executives want a quarterly revenue forecast that accounts for seasonality and promotions. Which modeling family fits this need?
- A Association rule mining of single transactions in the point-of-sale system
 - B Unsupervised topic modeling of customer emails to group common themes
 - C Image segmentation of product photographs for the online catalog pages
 - D Time series models with seasonal and exogenous regressors

Answer: D

Time series models capture seasonality and trend over time, and exogenous regressors such as promotion flags let the forecast respond to planned events. Topic modeling, image segmentation, and association rules address different data and questions.

2. Data drift and concept drift are the same phenomenon, so one monitoring metric detects both.
- A True
 - B False

Answer: B (False)

Data drift is a change in the distribution of inputs, while concept drift is a change in the relationship between inputs and the target. They can occur independently and need different detection methods, so the statement is false.

3. Which metric is most appropriate for ranking quality in a recommender that presents ten items per page?
- A Count of unique items shown across all users
 - B Mean squared error on raw item rating predictions made for each user
 - C Silhouette score on the item embedding clusters
 - D Normalized discounted cumulative gain at rank ten

Answer: D

Normalized discounted cumulative gain at rank ten rewards placing relevant items near the top of the visible list, which matches how users see a ten-item page. Rating error, cluster quality, and item counts do not measure the order of what users see.

4. A cross-validation variant that keeps the proportion of each class the same in every fold is called stratified _____.

Answer: k-fold (also accepted: k fold, kfold)

Stratified k-fold keeps class proportions consistent across folds, which stabilizes estimates for imbalanced classification. Plain k-fold can produce folds with very few minority examples, making the scores for those folds unreliable.

5. Which two practices reduce the risk of false conclusions in A/B testing? (Choose two.)

- A Stopping the test as soon as one variant shows a lower p-value
- B Fixing the sample size and primary metric before the test begins
- C Testing dozens of metrics and reporting only the significant ones
- D Running the test only during the holiday period when traffic is highest
- E Correcting for multiple comparisons when several variants or metrics are evaluated

Answer: B, E

Fixing the sample size and primary metric in advance prevents result-chasing, and correcting for multiple comparisons controls false positives across many tests. Peeking until significance appears, reporting only significant metrics, and atypical test periods all inflate false conclusions.

6. A pipeline must process terabytes of log data nightly. Which design choice best supports scalable, repeatable processing?

- A Exporting the logs to spreadsheets that each analyst opens and aggregates manually
- B A single workstation running an interactive notebook by hand each morning
- C Distributed processing over partitioned data with idempotent, scheduled jobs
- D Sampling one hundred random lines and assuming the rest of the logs match them

Answer: C

Partitioned distributed processing scales with data volume, and idempotent scheduled jobs can be rerun safely after a failure without duplicating results. Manual spreadsheet aggregation, a hand-run workstation notebook, and sampling cannot handle terabytes reliably or repeatably.

7. Which governance structure most reliably moves analytics projects from pilot to production?

- A A rule that every project must use the same notebook template for all work
- B A single analyst who approves projects without consulting any business owners
- C Quarterly reviews that only track the quality of presentation slide designs
- D A cross-functional steering group with business, data, and engineering owners

Answer: D

A steering group with named business, data, and engineering owners ensures each project has a sponsor, data access, and operational support before launch. A lone approver, slide reviews, and notebook templates do not assign the responsibilities that production requires.

8. Data quality is solely the responsibility of the IT department, so business owners need no role in data governance.
- A True
 - B False

Answer: B (False)

Governance frameworks assign accountability for data quality and definitions to business owners, while IT supports the technical controls. Data quality cannot be delegated to IT alone. The statement is false.

9. Which framing best converts a model's error rate into business terms?
- A The model has a log-loss of 0.31 on the validation fold of the data
 - B About 40 valid customers in every 1,000 applications would be declined
 - C The F1 score improved by 0.02 compared with the previous model version in use
 - D The learning rate was set to 0.001 during the training of the model

Answer: B

Stating how many valid customers would be wrongly declined per thousand applications gives executives a concrete cost they can weigh against the benefit. Log-loss, F1 change, and learning rate are technical measures that leaders cannot easily translate into impact.

10. A model used for hiring must be explained to candidates. What is the strongest governance approach?
- A Rely on the candidate to ask the right questions about how the model works
 - B Keep the model's inner logic secret to protect the vendor's intellectual property rights
 - C Explain the decision in general terms only, without any specifics for the candidate
 - D Document decision factors, test for disparate impact, and offer human review

Answer: D

Documenting the decision factors, testing for disparate impact, and providing a route to human review combine transparency with accountability. Secrecy, vague explanations, and placing the burden on candidates each fail that standard.