CERT-PASS

# AWS AWS Certified Machine Learning Engineer – Associate MLA-C01

Free Practice Questions Preview

Here are 35 sample questions to help you get started. Unlock the full exam to access all 1060+ questions with detailed explanations.

## Question 1 : Content Domain 1: Data Preparation for Machine Learning (ML)

**An education platform is preparing an ML solution for ticket priority. An education platform stores 600 million events of curated tabular data in Amazon S3. Analysts commonly read a small subset of columns with Athena. Which format should the ML engineer choose?**

**A. Write the curated dataset in Apache Parquet format.**

B. Write one JSON object per row without compaction.

C. Store the analytical dataset as newline-delimited text only.

D. Write the curated dataset as uncompressed CSV files.

**Answer: A**
A is correct. Parquet is columnar and supports efficient compression and column pruning for analytical reads. The strongest distractor is D. CSV is portable, but it usually scans more bytes for column-oriented analytics.

## Question 2 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A regulated workload at an industrial manufacturer has the following constraint. A data lake contains S3 objects partitioned by event date. The engineer needs a serverless SQL query to validate a sample before training. Which service should be used?**

A. Use Amazon OpenSearch Service as the default SQL engine for S3.

**B. Use Amazon Athena to query the data in Amazon S3.**

C. Use Amazon Redshift provisioned clusters for a one-time inspection.

D. Use AWS Glue DataBrew only to run SQL queries.

**Answer: B**
B is correct. Athena provides serverless SQL querying directly over S3 data. The strongest distractor is C. A provisioned Redshift cluster can work for warehouse workloads, but it adds unnecessary management for a serverless S3 inspection.

## Question 3 : Content Domain 1: Data Preparation for Machine Learning (ML)

**An engineer is troubleshooting an AWS ML workload for an energy company. The ML team must catalog raw S3 data and run repeatable Spark-based transformations. Which AWS service should it select?**

A. Use Amazon Athena as the primary distributed ETL runtime.

B. Use AWS Lake Formation as the Spark execution engine.

**C. Use AWS Glue and the AWS Glue Data Catalog.**

D. Use Amazon Data Firehose as the metadata catalog.

**Answer: C**
C is correct. AWS Glue provides managed ETL and integrates with the Glue Data Catalog. The strongest distractor is A. Athena is useful for SQL queries, but it is not the primary distributed ETL runtime for this requirement.

## Question 4 : Content Domain 1: Data Preparation for Machine Learning (ML)

**The next release at a property-management platform includes a change to the ML workflow. Several training instances require concurrent POSIX-style access to shared feature files. Which storage category is appropriate?**

A. Use Amazon DynamoDB to emulate POSIX file locking.

B. Use a single Amazon EBS volume attached to one instance.

C. Use Amazon S3 Glacier Deep Archive as the active training mount.

**D. Use a shared file system such as Amazon EFS or Amazon FSx that matches the throughput requirements.**

**Answer: D**
D is correct. EFS or FSx can provide shared file-system access for training workloads. The strongest distractor is B. A single EBS volume is not the normal choice for concurrent shared file access across multiple training instances.

## Question 5 : Content Domain 1: Data Preparation for Machine Learning (ML)

**The platform team at a sports analytics company is improving a workflow for customer clickstream events. The ML platform needs windowed aggregations over an event stream before persisting features. Which managed service should it choose?**

**A. Use Amazon Managed Service for Apache Flink.**

B. Use AWS Glue crawlers on a one-minute schedule.

C. Use AWS DataSync to calculate streaming windows.

D. Use SageMaker batch transform for each event.

**Answer: A**
A is correct. Managed Service for Apache Flink is designed for stateful streaming transformations and windowed aggregations. The strongest distractor is B. Glue crawlers discover metadata; they do not perform stateful event-stream processing.

## Question 6 : Content Domain 1: Data Preparation for Machine Learning (ML)

**An engineer is troubleshooting an AWS ML workload for a cybersecurity vendor. The team wants a managed stream-delivery path into S3 and does not need complex stateful processing. What is the best fit?**

A. Use Amazon EBS snapshots as the delivery mechanism.

**B. Use Amazon Data Firehose to deliver the stream to Amazon S3.**

C. Use Amazon Managed Service for Apache Flink for every simple delivery use case.

D. Use AWS DataSync for event-by-event stream ingestion.

**Answer: B**
B is correct. Amazon Data Firehose is suited to managed stream delivery to destinations such as S3 with minimal operational effort. The strongest distractor is C. Flink is powerful for stateful processing, but it is unnecessarily complex when the main need is managed delivery.

## Question 7 : Content Domain 1: Data Preparation for Machine Learning (ML)

**During a design review at a telecom operator, the team identifies the following requirement. A low-cardinality feature contains non-ordinal values such as product category. Which encoding is appropriate for a linear model?**

A. Drop the feature automatically because it is categorical.

B. Apply label encoding and treat the integer values as ordered magnitudes.

**C. Apply one-hot encoding.**

D. Replace each category with a random hash and assume the hash magnitude is meaningful.

**Answer: C**
C is correct. One-hot encoding represents low-cardinality non-ordinal categories without introducing a false ranking. The strongest distractor is B. Direct label encoding can incorrectly imply an ordinal relationship.

## Question 8 : Content Domain 1: Data Preparation for Machine Learning (ML)

**The platform team at a video-streaming platform is improving a workflow for marketing-response events. A data scientist wants to profile a dataset and create reusable preprocessing steps visually inside a SageMaker-oriented workflow. What is the best fit?**

A. Use Amazon Athena only as the visual transformation interface.

B. Use AWS Config rules to impute missing values.

C. Use SageMaker Model Monitor to clean the training data interactively.

**D. Use Amazon SageMaker Data Wrangler.**

**Answer: D**
D is correct. SageMaker Data Wrangler supports exploration, profiling, and data transformation for ML preparation. The strongest distractor is C. Model Monitor evaluates production data and model behavior; it is not the interactive preparation tool for this requirement.

## Question 9 : Content Domain 1: Data Preparation for Machine Learning (ML)

**An engineer is troubleshooting an AWS ML workload for a video-streaming platform. A data analyst needs a visual, no-code-oriented tool to clean tabular data before handing it to the ML team. Which service is appropriate?**

**A. Use AWS Glue DataBrew.**

B. Use AWS Glue Data Quality only as the visual cleaning tool.

C. Use Amazon QuickSight to alter source records.

D. Use AWS Lake Formation to normalize feature values.

**Answer: A**
A is correct. Glue DataBrew provides visual data preparation and cleaning recipes. The strongest distractor is B. Glue Data Quality validates data rules, but it is not the primary visual cleaning interface.

## Question 10 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A cost and reliability review at a property-management platform raises the following question. A team wants to reduce training-serving skew by reusing governed features in batch training and low-latency inference. What should it use?**

A. Use AWS Secrets Manager as the offline feature repository.

**B. Use SageMaker Feature Store with the appropriate online and offline stores.**

C. Use Amazon Athena query history as the online feature store.

D. Store features only in notebook-local CSV files.

**Answer: B**
B is correct. SageMaker Feature Store supports managed feature reuse for online and offline access patterns. The strongest distractor is D. Notebook-local files do not provide a governed shared feature repository or production online access.

## Question 11 : Content Domain 1: Data Preparation for Machine Learning (ML)

**During a design review at an online bank, the team identifies the following requirement. The team must create a managed workforce-based labeling job for training images. Which AWS service is designed for this need?**

A. Use Amazon Rekognition labels as a guaranteed human-reviewed dataset.

B. Use Amazon Mechanical Turk directly without a labeling workflow when quality controls are required.

**C. Use SageMaker Ground Truth.**

D. Use AWS Glue DataBrew to assign image bounding boxes.

**Answer: C**
C is correct. SageMaker Ground Truth provides managed labeling workflows and quality-control mechanisms. The strongest distractor is A. Rekognition can analyze images, but it does not replace a managed human labeling workflow when curated annotations are required.

## Question 12 : Content Domain 1: Data Preparation for Machine Learning (ML)

**The platform team at an energy company is improving a workflow for fraud-review outcomes. Before training, the engineer needs to evaluate whether label distribution creates bias risk. Which service should be used?**

A. Use SageMaker Inference Recommender to benchmark the dataset.

B. Use SageMaker Model Monitor only after deployment.

C. Use SageMaker Neo to compile the untrained model.

**D. Use SageMaker Clarify to calculate pre-training bias metrics.**

**Answer: D**
D is correct. SageMaker Clarify supports pre-training bias analysis, including class imbalance metrics. The strongest distractor is C. Neo optimizes trained models for deployment targets; it does not calculate pre-training bias metrics.

## Question 13 : Content Domain 1: Data Preparation for Machine Learning (ML)

**The platform team at an education platform is improving a workflow for shipment status events. A pipeline must validate completeness, schema consistency, and business rules in an AWS Glue workflow. Which capability is appropriate?**

**A. Use AWS Glue Data Quality rules.**

B. Use SageMaker automatic model tuning.

C. Use AWS CloudTrail event selectors.

D. Use Amazon S3 Transfer Acceleration.

**Answer: A**
A is correct. Glue Data Quality supports rule-based checks for data quality in Glue-oriented workflows. The strongest distractor is B. Automatic model tuning optimizes hyperparameters; it does not validate source data rules.

## Question 14 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A regulated workload at an automotive supplier has the following constraint. A training dataset contains personally identifiable information. What should the team do before model training?**

A. Make the bucket public so all training workers can read it.

**B. Classify the sensitive fields and apply masking, anonymization, or tokenization as required.**

C. Store PII values in endpoint tags.

D. Copy raw identifiers into model artifacts for convenience.

**Answer: B**
B is correct. Sensitive data should be identified and protected before training according to compliance and data-minimization needs. The strongest distractor is D. Copying raw identifiers into artifacts increases exposure and is not a minimization strategy.

## Question 15 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A data science team at a telecom operator needs the least operationally complex managed option. A preprocessing job writes a curated S3 dataset that will be queried column-by-column. Which output format best reduces scan volume? The initial implementation will run in eu-west-1.**

A. Write one JSON object per row without compaction.

B. Store the analytical dataset as newline-delimited text only.

**C. Write the curated dataset in Apache Parquet format.**

D. Write the curated dataset as uncompressed CSV files.

**Answer: C**
C is correct. Parquet is columnar and supports efficient compression and column pruning for analytical reads. The strongest distractor is D. CSV is portable, but it usually scans more bytes for column-oriented analytics.

## Question 16 : Content Domain 1: Data Preparation for Machine Learning (ML)

**During a design review at a payment processor, the team identifies the following requirement. A data lake contains S3 objects partitioned by event date. The engineer needs a serverless SQL query to validate a sample before training. Which service should be used? The initial implementation will run in eu-central-1.**

A. Use Amazon OpenSearch Service as the default SQL engine for S3.

B. Use Amazon Redshift provisioned clusters for a one-time inspection.

C. Use AWS Glue DataBrew only to run SQL queries.

**D. Use Amazon Athena to query the data in Amazon S3.**

**Answer: D**
D is correct. Athena provides serverless SQL querying directly over S3 data. The strongest distractor is B. A provisioned Redshift cluster can work for warehouse workloads, but it adds unnecessary management for a serverless S3 inspection.

## Question 17 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A regulated workload at a payment processor has the following constraint. The ML team must catalog raw S3 data and run repeatable Spark-based transformations. Which AWS service should it select? The initial implementation will run in eu-west-1.**

**A. Use AWS Glue and the AWS Glue Data Catalog.**

B. Use Amazon Athena as the primary distributed ETL runtime.

C. Use AWS Lake Formation as the Spark execution engine.

D. Use Amazon Data Firehose as the metadata catalog.

**Answer: A**
A is correct. AWS Glue provides managed ETL and integrates with the Glue Data Catalog. The strongest distractor is B. Athena is useful for SQL queries, but it is not the primary distributed ETL runtime for this requirement.

## Question 18 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A model for credit risk is being operationalized by an industrial manufacturer. Several training instances require concurrent POSIX-style access to shared feature files. Which storage category is appropriate? The initial implementation will run in us-east-1.**

A. Use Amazon DynamoDB to emulate POSIX file locking.

**B. Use a shared file system such as Amazon EFS or Amazon FSx that matches the throughput requirements.**

C. Use a single Amazon EBS volume attached to one instance.

D. Use Amazon S3 Glacier Deep Archive as the active training mount.

**Answer: B**
B is correct. EFS or FSx can provide shared file-system access for training workloads. The strongest distractor is C. A single EBS volume is not the normal choice for concurrent shared file access across multiple training instances.

## Question 19 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A model for customer churn is being operationalized by a property-management platform. A source emits continuous inventory snapshots. Records must be transformed as they arrive and written to S3. Which service is most suitable for stateful stream processing? The initial implementation will run in eu-central-1.**

A. Use AWS Glue crawlers on a one-minute schedule.

B. Use AWS DataSync to calculate streaming windows.

**C. Use Amazon Managed Service for Apache Flink.**

D. Use SageMaker batch transform for each event.

**Answer: C**
C is correct. Managed Service for Apache Flink is designed for stateful streaming transformations and windowed aggregations. The strongest distractor is A. Glue crawlers discover metadata; they do not perform stateful event-stream processing.

## Question 20 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A customer-support SaaS provider is preparing an ML solution for ticket priority. The team wants a managed stream-delivery path into S3 and does not need complex stateful processing. What is the best fit? The initial implementation will run in us-east-1.**

A. Use Amazon EBS snapshots as the delivery mechanism.

B. Use Amazon Managed Service for Apache Flink for every simple delivery use case.

C. Use AWS DataSync for event-by-event stream ingestion.

**D. Use Amazon Data Firehose to deliver the stream to Amazon S3.**

**Answer: D**
D is correct. Amazon Data Firehose is suited to managed stream delivery to destinations such as S3 with minimal operational effort. The strongest distractor is B. Flink is powerful for stateful processing, but it is unnecessarily complex when the main need is managed delivery.

## Question 21 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A payment processor is preparing an ML solution for ticket priority. A low-cardinality feature contains non-ordinal values such as product category. Which encoding is appropriate for a linear model? The initial implementation will run in ap-southeast-2.**

**A. Apply one-hot encoding.**

B. Drop the feature automatically because it is categorical.

C. Apply label encoding and treat the integer values as ordered magnitudes.

D. Replace each category with a random hash and assume the hash magnitude is meaningful.

**Answer: A**
A is correct. One-hot encoding represents low-cardinality non-ordinal categories without introducing a false ranking. The strongest distractor is C. Direct label encoding can incorrectly imply an ordinal relationship.

## Question 22 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A cost and reliability review at a cybersecurity vendor raises the following question. A data scientist wants to profile a dataset and create reusable preprocessing steps visually inside a SageMaker-oriented workflow. What is the best fit? The initial implementation will run in ap-southeast-2.**

A. Use Amazon Athena only as the visual transformation interface.

**B. Use Amazon SageMaker Data Wrangler.**

C. Use AWS Config rules to impute missing values.

D. Use SageMaker Model Monitor to clean the training data interactively.

**Answer: B**
B is correct. SageMaker Data Wrangler supports exploration, profiling, and data transformation for ML preparation. The strongest distractor is D. Model Monitor evaluates production data and model behavior; it is not the interactive preparation tool for this requirement.

## Question 23 : Content Domain 1: Data Preparation for Machine Learning (ML)

**The platform team at a video-streaming platform is improving a workflow for daily transaction records. The operations team wants recipe-based visual data cleaning without writing Spark code. What should it use? The initial implementation will run in us-east-1.**

A. Use AWS Lake Formation to normalize feature values.

B. Use Amazon QuickSight to alter source records.

**C. Use AWS Glue DataBrew.**

D. Use AWS Glue Data Quality only as the visual cleaning tool.

**Answer: C**
C is correct. Glue DataBrew provides visual data preparation and cleaning recipes. The strongest distractor is D. Glue Data Quality validates data rules, but it is not the primary visual cleaning interface.

## Question 24 : Content Domain 1: Data Preparation for Machine Learning (ML)

**During a design review at a public-sector analytics team, the team identifies the following requirement. The same engineered customer features are required for online inference and offline training. Which SageMaker capability supports this pattern? The initial implementation will run in eu-central-1.**

A. Use AWS Secrets Manager as the offline feature repository.

B. Use Amazon Athena query history as the online feature store.

C. Store features only in notebook-local CSV files.

**D. Use SageMaker Feature Store with the appropriate online and offline stores.**

**Answer: D**
D is correct. SageMaker Feature Store supports managed feature reuse for online and offline access patterns. The strongest distractor is C. Notebook-local files do not provide a governed shared feature repository or production online access.

## Question 25 : Content Domain 1: Data Preparation for Machine Learning (ML)

**An automotive supplier is preparing an ML solution for delivery delay. The team must create a managed workforce-based labeling job for training images. Which AWS service is designed for this need? The initial implementation will run in ap-southeast-2.**

**A. Use SageMaker Ground Truth.**

B. Use Amazon Rekognition labels as a guaranteed human-reviewed dataset.

C. Use AWS Glue DataBrew to assign image bounding boxes.

D. Use Amazon Mechanical Turk directly without a labeling workflow when quality controls are required.

**Answer: A**
A is correct. SageMaker Ground Truth provides managed labeling workflows and quality-control mechanisms. The strongest distractor is B. Rekognition can analyze images, but it does not replace a managed human labeling workflow when curated annotations are required.

## Question 26 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A food-delivery platform is preparing an ML solution for customer churn. Before training, the engineer needs to evaluate whether label distribution creates bias risk. Which service should be used? The initial implementation will run in eu-central-1.**

A. Use SageMaker Inference Recommender to benchmark the dataset.

**B. Use SageMaker Clarify to calculate pre-training bias metrics.**

C. Use SageMaker Neo to compile the untrained model.

D. Use SageMaker Model Monitor only after deployment.

**Answer: B**
B is correct. SageMaker Clarify supports pre-training bias analysis, including class imbalance metrics. The strongest distractor is C. Neo optimizes trained models for deployment targets; it does not calculate pre-training bias metrics.

## Question 27 : Content Domain 1: Data Preparation for Machine Learning (ML)

**During a design review at a customer-support SaaS provider, the team identifies the following requirement. The data engineering team needs automated data-quality rules before a training dataset is approved. What should it add? The initial implementation will run in eu-central-1.**

A. Use SageMaker automatic model tuning.

B. Use Amazon S3 Transfer Acceleration.

**C. Use AWS Glue Data Quality rules.**

D. Use AWS CloudTrail event selectors.

**Answer: C**
C is correct. Glue Data Quality supports rule-based checks for data quality in Glue-oriented workflows. The strongest distractor is A. Automatic model tuning optimizes hyperparameters; it does not validate source data rules.

## Question 28 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A cost and reliability review at a retail marketplace raises the following question. A training dataset contains personally identifiable information. What should the team do before model training? The initial implementation will run in eu-central-1.**

A. Copy raw identifiers into model artifacts for convenience.

B. Make the bucket public so all training workers can read it.

C. Store PII values in endpoint tags.

**D. Classify the sensitive fields and apply masking, anonymization, or tokenization as required.**

**Answer: D**
D is correct. Sensitive data should be identified and protected before training according to compliance and data-minimization needs. The strongest distractor is A. Copying raw identifiers into artifacts increases exposure and is not a minimization strategy.

## Question 29 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A regulated workload at an education platform has the following constraint. A preprocessing job writes a curated S3 dataset that will be queried column-by-column. Which output format best reduces scan volume? The team expects sporadic requests.**

**A. Write the curated dataset in Apache Parquet format.**

B. Write one JSON object per row without compaction.

C. Write the curated dataset as uncompressed CSV files.

D. Store the analytical dataset as newline-delimited text only.

**Answer: A**
A is correct. Parquet is columnar and supports efficient compression and column pruning for analytical reads. The strongest distractor is C. CSV is portable, but it usually scans more bytes for column-oriented analytics.

## Question 30 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A model for credit risk is being operationalized by a healthcare analytics company. A data lake contains S3 objects partitioned by event date. The engineer needs a serverless SQL query to validate a sample before training. Which service should be used? The team expects bursty daytime traffic.**

A. Use Amazon Redshift provisioned clusters for a one-time inspection.

**B. Use Amazon Athena to query the data in Amazon S3.**

C. Use AWS Glue DataBrew only to run SQL queries.

D. Use Amazon OpenSearch Service as the default SQL engine for S3.

**Answer: B**
B is correct. Athena provides serverless SQL querying directly over S3 data. The strongest distractor is A. A provisioned Redshift cluster can work for warehouse workloads, but it adds unnecessary management for a serverless S3 inspection.

## Question 31 : Content Domain 1: Data Preparation for Machine Learning (ML)

**The next release at an education platform includes a change to the ML workflow. The ML team must catalog raw S3 data and run repeatable Spark-based transformations. Which AWS service should it select? The team expects steady high traffic.**

A. Use Amazon Athena as the primary distributed ETL runtime.

B. Use Amazon Data Firehose as the metadata catalog.

**C. Use AWS Glue and the AWS Glue Data Catalog.**

D. Use AWS Lake Formation as the Spark execution engine.

**Answer: C**
C is correct. AWS Glue provides managed ETL and integrates with the Glue Data Catalog. The strongest distractor is A. Athena is useful for SQL queries, but it is not the primary distributed ETL runtime for this requirement.

## Question 32 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A regulated workload at an industrial manufacturer has the following constraint. A training workload repeatedly reads the same shared files from multiple instances. The code expects a mounted file system. What should the engineer use? The team expects steady high traffic.**

A. Use Amazon S3 Glacier Deep Archive as the active training mount.

B. Use a single Amazon EBS volume attached to one instance.

C. Use Amazon DynamoDB to emulate POSIX file locking.

**D. Use a shared file system such as Amazon EFS or Amazon FSx that matches the throughput requirements.**

**Answer: D**
D is correct. EFS or FSx can provide shared file-system access for training workloads. The strongest distractor is B. A single EBS volume is not the normal choice for concurrent shared file access across multiple training instances.

## Question 33 : Content Domain 1: Data Preparation for Machine Learning (ML)

**The platform team at a public-sector analytics team is improving a workflow for shipment status events. The ML platform needs windowed aggregations over an event stream before persisting features. Which managed service should it choose? The team expects sporadic requests.**

**A. Use Amazon Managed Service for Apache Flink.**

B. Use AWS Glue crawlers on a one-minute schedule.

C. Use SageMaker batch transform for each event.

D. Use AWS DataSync to calculate streaming windows.

**Answer: A**
A is correct. Managed Service for Apache Flink is designed for stateful streaming transformations and windowed aggregations. The strongest distractor is B. Glue crawlers discover metadata; they do not perform stateful event-stream processing.

## Question 34 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A cost and reliability review at a sports analytics company raises the following question. The team wants a managed stream-delivery path into S3 and does not need complex stateful processing. What is the best fit? The team expects bursty daytime traffic.**

A. Use Amazon EBS snapshots as the delivery mechanism.

**B. Use Amazon Data Firehose to deliver the stream to Amazon S3.**

C. Use Amazon Managed Service for Apache Flink for every simple delivery use case.

D. Use AWS DataSync for event-by-event stream ingestion.

**Answer: B**
B is correct. Amazon Data Firehose is suited to managed stream delivery to destinations such as S3 with minimal operational effort. The strongest distractor is C. Flink is powerful for stateful processing, but it is unnecessarily complex when the main need is managed delivery.

## Question 35 : Content Domain 1: Data Preparation for Machine Learning (ML)

**A model for inventory shortage is being operationalized by a travel-booking platform. A low-cardinality feature contains non-ordinal values such as product category. Which encoding is appropriate for a linear model? The team expects sporadic requests.**

A. Replace each category with a random hash and assume the hash magnitude is meaningful.

B. Apply label encoding and treat the integer values as ordered magnitudes.

**C. Apply one-hot encoding.**

D. Drop the feature automatically because it is categorical.

**Answer: C**
C is correct. One-hot encoding represents low-cardinality non-ordinal categories without introducing a false ranking. The strongest distractor is B. Direct label encoding can incorrectly imply an ordinal relationship.