DBCLOUD
EXAMS

# QUESTIONS & ANSWERS

verified



**Technology Provider:** Databricks

**Exam:** Certified Machine Learning Associate

**Number of questions in the database:** 140

**Exam Version:** NOV.2026

Question #1Topic 1

A machine learning engineer has created a Feature Table new_table using Feature Store Client fs. When creating the table, they specified a metadata description with key information about the Feature Table. They now want to retrieve that metadata programmatically.

Which of the following lines of code will return the metadata description?

A. There is no way to return the metadata description programmatically.

B. fs.create_training_set("new_table")

C. fs.get_table("new_table").description

D. fs.get_table("new_table").load_df()

E. fs.get_table("new_table")

**Correct Answer: C**

Question #2Topic 1

A data scientist has a Spark DataFrame spark_df. They want to create a new Spark DataFrame that contains only the rows from spark_df where the value in column price is greater than 0.

Which of the following code blocks will accomplish this task?

A. spark_df[spark_df["price"] > 0]

B. spark_df.filter(col("price") > 0)

C. SELECT * FROM spark_df WHERE price > 0

D. spark_df.loc[spark_df["price"] > 0,:]

E. spark_df.loc[:,spark_df["price"] > 0]

**Correct Answer: B**

https://dbcloud.academy/

Question #3Topic 1

A health organization is developing a classification model to determine whether or not a patient currently has a specific type of infection. The organization's leaders want to maximize the number of positive cases identified by the model.

Which of the following classification metrics should be used to evaluate the model?

A. RMSE

B. Precision

C. Area under the residual operating curve

D. Accuracy

E. Recall

**Correct Answer: E**

Question #4Topic 1

In which of the following situations is it preferable to impute missing feature values with their median value over the mean value?

A. When the features are of the categorical type

B. When the features are of the boolean type

C. When the features contain a lot of extreme outliers

D. When the features contain no outliers

E. When the features contain no missing values

**Correct Answer: C**

https://dbcloud.academy/

Question #5Topic 1

A data scientist has replaced missing values in their feature set with each respective feature variable's median value. A colleague suggests that the data scientist is throwing away valuable information by doing this.

Which of the following approaches can they take to include as much information as possible in the feature set?

A. Impute the missing values using each respective feature variable's mean value instead of the median value

B. Refrain from imputing the missing values in favor of letting the machine learning algorithm determine how to handle them

C. Remove all feature variables that originally contained missing values from the feature set

D. Create a binary feature variable for each feature that contained missing values indicating whether each row's value has been imputed

E. Create a constant feature variable for each feature that contained missing values indicating the percentage of rows from the feature that was originally missing

**Correct Answer: D**

Question #6Topic 1

A data scientist is wanting to explore summary statistics for Spark DataFrame spark_df. The data scientist wants to see the count, mean, standard deviation, minimum, maximum, and interquartile range (IQR) for each numerical feature.

Which of the following lines of code can the data scientist run to accomplish the task?

A. spark_df.summary ()

B. spark_df.stats()

C. spark_df.describe().head()

D. spark_df.printSchema()

E. spark_df.toPandas()

**Correct Answer: A**

https://dbcloud.academy/