

# DSA-C03 Test Torrent, New DSA-C03 Dumps Questions



P.S. Free & New DSA-C03 dumps are available on Google Drive shared by ExamDiscuss: <https://drive.google.com/open?id=1BUkKY0c14Oeo5L4x1XZABbrqN4XnV3m8>

We provide free demo for you to have a try before buying DSA-C03 exam braindumps. Free demo will help you have a better understanding of what you are going to buy, and we also recommend you try the free demo before buying. Moreover, DSA-C03 exam braindumps of us will offer you free update for one year, and you can get the latest version of the exam dumps if you choose us. And the update version for DSA-C03 Exam Dumps will be sent to your email automatically, and you just need to receive them.

All points of questions are correlated with the newest and essential knowledge. The second one of DSA-C03 test guide is emphasis on difficult and hard-to-understand points. Experts left notes for your reference, and we believe with their notes things will be easier. In addition, the new supplementary will be sent to your mailbox if you place order this time with beneficial discounts at intervals. So our DSA-C03 Exam Questions mean more intellectual choice than other practice materials.

>> DSA-C03 Test Torrent <<

## Marvelous DSA-C03 Test Torrent, New DSA-C03 Dumps Questions

Will you feel nervous in the exam? If you do, just choose us, our DSA-C03 Soft test engine can stimulate the real exam environment, which will help you know the procedure of the exam, and will strengthen your confidence. Moreover DSA-C03 exam dumps are high-quality, and we have professional experts to compile them, and they can help you pass the exam just one time. We offer you free demo to have a try for DSA-C03 Exam Dumps, and free update for one year. If you indeed have questions, just contact with us.

## Snowflake SnowPro Advanced: Data Scientist Certification Exam Sample Questions (Q288-Q293):

### NEW QUESTION # 288

A data scientist is exploring customer purchase data in Snowflake to identify high-value customer segments. They have a table named 'CUSTOMER\_TRANSACTIONS' with columns 'CUSTOMER\_ID', 'TRANSACTION\_DATE', and 'PURCHASE\_AMOUNT'. They want to calculate the interquartile range (IQR) of 'PURCHASE\_AMOUNT' for each customer. Which SQL query using Snowsight is the most efficient and accurate way to calculate and display the IQR for each 'CUSTOMER

ID?

□

- A. Option E
- B. Option D
- C. Option C
- D. Option A
- E. Option B

**Answer: A**

Explanation:

Option E, using 'QUANTILE, is the most accurate way to calculate the IQR. 4)' returns an array representing the quartiles (0%, 25%, 50%, 75%, 100%). Subtracting the 25th percentile (index 1) from the 75th percentile (index 3) gives the IQR. Other options either approximate the percentiles (APPROX\_PERCENTILE), calculate the range (MAX-MIN), or calculate standard deviation, none of which directly give the IQR. Option B while syntactically valid is less performant and returns the IQR on entire table not grouped by customer.

### NEW QUESTION # 289

You're working with a large dataset of user transactions in Snowflake. You need to identify potential outliers in transaction amounts (TRANSACTION AMOUNT) for each user (CUSER ID). Your goal is to flag transactions that are more than 3 standard deviations away from the mean transaction amount for that specific user. Which of the following approaches, utilizing Snowflake's statistical functions and window functions, would be MOST efficient and accurate for achieving this?

- A. Using a correlated subquery to calculate the mean and standard deviation for each user and then filtering the transactions.
- B. Calculating the overall mean and standard deviation for all transactions and filtering transactions based on those global statistics.
- C. Using window functions to calculate the mean and standard deviation for each user within the same query, and then comparing each transaction amount to the calculated range.
- D. Exporting the data to a Python environment, performing the calculations using Pandas, and then re-importing the results to Snowflake.
- E. Creating a stored procedure that iterates through each user and calculates the mean and standard deviation individually.

**Answer: C**

Explanation:

Using window functions (option C) is the most efficient and accurate approach. It allows you to calculate the mean and standard deviation for each user within the same query, avoiding the overhead of correlated subqueries (option A) or the inaccuracy of global statistics (option B). Options D and E are less efficient due to data transfer and procedural logic overhead. Correlated subquery will lead to performance issue and is not advisable for bigger datasets.

### NEW QUESTION # 290

You are evaluating a binary classification model's performance using the Area Under the ROC Curve (AUC). You have the following predictions and actual values. What steps can you take to reliably calculate this in Snowflake, and which snippet represents a crucial part of that calculation? (Assume tables 'predictions' with columns 'predicted\_probability' (FLOAT) and 'actual\_value' (BOOLEAN); TRUE indicates positive class, FALSE indicates negative class). Which of the below code snippet should be used to calculate the 'True positive Rate' and 'False positive Rate' for different thresholds

- A. Export the 'predicted\_probability' and 'actual\_value' columns to a local Python environment and calculate the AUC using scikit-learn.
- B. The AUC cannot be reliably calculated within Snowflake due to limitations in SQL functionality for statistical analysis.
- C. Calculate AUC directly within a Snowpark Python UDF using scikit-learn's function. This avoids data transfer overhead, making it highly efficient for large datasets. No further SQL is needed beyond querying the predictions data.
- 
- D. The best way to calculate AUC is to randomly guess the probabilities and see how it performs.
- E. Using only SQL, Create a temporary table with calculated True Positive Rate (TPR) and False Positive Rate (FPR) at different probability thresholds. Then, approximate the AUC using the trapezoidal rule.
- 

**Answer: C,E**

Explanation:

Options A and C are correct. Option A demonstrates calculating AUC directly within Snowflake using a Snowpark Python UDF and scikit-learn's . This is efficient for large datasets as it avoids data transfer. Option C correctly outlines the process of calculating TPR and FPR using SQL and approximating AUC using the trapezoidal rule, another viable approach within Snowflake. Option B is incorrect; AUC can be calculated reliably within Snowflake. Option D is inefficient due to data transfer. Option E is blatantly incorrect.

#### NEW QUESTION # 291

You are developing a churn prediction model using Snowpark Python and Scikit-learn. After initial model training, you observe significant overfitting. Which of the following hyperparameter tuning strategies and code snippets, when implemented within a Snowflake Python UDF, would be MOST effective to address overfitting in a Ridge Regression model and how can you implement a reproducible model with minimal code?

- A. Option C
- B. Option E
- C. Option A
- D. Option D
- E. Option B

**Answer: D,E**

Explanation:

Options B and D are correct because they employ techniques to mitigate overfitting. Option B uses 'RandomizedSearchCV' with cross-validation and a fixed 'random\_state', making the search reproducible and preventing overfitting by evaluating performance on multiple validation sets. Option D leverages 'BayesianSearchCV', which uses a probabilistic model to efficiently explore the hyperparameter space, also with cross-validation and a fixed random state making search reproducible. Both methods aim to find a balance between model complexity and generalization ability. Option A is incorrect because it does not use cross-validation, which is crucial for preventing overfitting. Option C is incorrect because manual tuning without a systematic search and cross-validation is prone to bias and overfitting. Finally, option E is incorrect because while using a modern algorithm, it lacks a random state, making it difficult to reproduce the outcome.

#### NEW QUESTION # 292

You are building a machine learning model using Snowflake data to predict customer churn. Your dataset includes a 'CUSTOMER TYPE' column with the following possible values: 'New', 'Returning', and 'VIP'. You need to perform one-hot encoding on this column. Which of the following Snowflake SQL queries correctly implements one-hot encoding for the 'CUSTOMER TYPE' column, creating separate binary columns for each customer type ('IS NEW', 'IS RETURNING', 'IS VIP')?

- A. Option D
- B. Option A
- C. Option E
- D. Option C
- E. Option B

**Answer: B,D,E**

Explanation:

Options A, B, and C are all valid ways to perform one-hot encoding in Snowflake. Option A uses the standard 'CASE statement, Option B leverages the 'IFF function (inline IF), and Option C uses 'DECODE', all achieving the same result of creating binary indicators for each category. Option D is incorrect because it uses GET DDL, which retrieves DDL statements, not for comparison. Option E is incorrect because it does not represent three separate columns of binary columns for each customer type. Therefore, options A, B, and C are the correct approaches to generate separate binary columns for one-hot encoding.

#### NEW QUESTION # 293

.....

The ExamDiscuss is a leading platform that has been offering top-rated and real SnowPro Advanced: Data Scientist Certification Exam (DSA-C03) exam questions for quick SnowPro Advanced: Data Scientist Certification Exam Certification Exam. The DSA-

What's more, part of that ExamDiscuss DSA-C03 dumps now are free: <https://drive.google.com/open?id=1BUkKY0c14Oeo5L4x1XZABbrqN4XnV3m8>