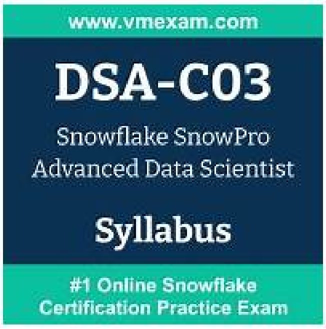
Free PDF Quiz Snowflake - DSA-C03 - SnowPro Advanced: Data Scientist Certification Exam Unparalleled Exam Preview



We are committed to designing a kind of scientific DSA-C03 study material to balance your business and study schedule. With our DSA-C03 exam guide, all your learning process includes 20-30 hours. As long as you spare one or two hours a day to study with our laTest DSA-C03 Quiz prep, we assure that you will have a good command of the relevant knowledge before taking the DSA-C03 exam. What you need to do is to follow the DSA-C03 exam guide system at the pace you prefer as well as keep learning step by step.

It is not hard to know that DSA-C03 study materials not only have better quality than any other study materials, but also have better quality. On the one hand, we can guarantee that you will pass the DSA-C03 exam easily if you learn our DSA-C03 Study Materials; on the other hand, you will learn a lot of useful knowledge from our DSA-C03 learning braindump. Are you ready? You can free download the demo of ourDSA-C03 study materials on the web first.

>> Exam DSA-C03 Preview <<

Free PDF High Pass-Rate Snowflake - Exam DSA-C03 Preview

As indicator on your way to success, our DSA-C03 practice materials can navigate you through all difficulties in your journey. Every challenge cannot be dealt like walk-ins, but our DSA-C03 simulating practice can make your review effective. That is why our DSA-C03 study questions are professional model in the line. With high pass rate as more than 98%, our DSA-C03 exam questions have helped tens of millions of candidates passed their exam successfully.

Snowflake SnowPro Advanced: Data Scientist Certification Exam Sample Questions (Q10-Q15):

NEW QUESTION # 10

You are a data scientist working for a retail company using Snowflake. You're building a linear regression model to predict sales based on advertising spend across various channels (TV, Radio, Newspaper). After initial EDA, you suspect multicollinearity among the independent variables. Which of the following Snowflake SQL statements or techniques are MOST appropriate for identifying and addressing multicollinearity BEFORE fitting the model? Choose two.

- A. Generate a correlation matrix of the independent variables using 'CORR aggregate function in Snowflake SQL and examine the correlation coefficients. Values close to +1 or -1 suggest high multicollinearity.
- B. Drop one of the independent variable randomly if they seem highly correlated.
- C. Calculate the Variance Inflation Factor (VIF) for each independent variable using a user-defined function (UDF) in Snowflake that implements the VIF calculation based on R-squared values from auxiliary regressions. This requires fitting a linear regression for each independent variable against all others.
- D. Use ' on each independent variable to estimate its uniqueness. If uniqueness is low, multicollinearity is likely.
- E. Implement Principal Component Analysis (PCA) using Snowpark Python to transform the independent variables into uncorrelated principal components and then select only the components explaining a certain percentage of the variance.

Answer: A,C

Explanation:
Multicollinearity can be identified by calculating the VIF for each independent variable. VIF is calculated by regressing each independent variable against all other independent variables and calculating 1/(1-RA2), where RA2 is the R-squared value from the regression. A high VIF suggests high multicollinearity. Correlation matrices generated with 'CORR can also reveal multicollinearity by showing pairwise correlations between independent variables. PCA using Snowpark is also a viable option, but less direct than VIF and correlation matrix analysis for identifying multicollinearity. APPROX_COUNT_DISTINCT is not directly related to identifying multicollinearity. Randomly dropping variables will also lead to data loss.

NEW QUESTION # 11
You are building a machine learning pipeline in Snowflake using Snowpark Python. You have completed the data preparation and feature engineering steps and now need to train a model. You want to track the performance of different model versions and hyperparameters using MLflow. You are considering these deployment strategies. Which of the deployment strategies allows automatic logging of metrics, parameters, and model artifacts to MLflow for each training run without requiring explicit MLflow logging code?

- A. Train the model locally on your development machine and manually log metrics and artifacts to MLflow using the MLflow API. Then, deploy the trained model to Snowflake as a UDF or stored procedure.
- B. Train the model within a Snowpark Python UDF. Use a Snowflake stage to store MLflow artifacts.
- C. Use the Snowpark MLAPI and its integration with MLflow's autologging feature. Enable autologging before starting the training run. Deploy the model to Snowflake as a UDF.
- D. Train the model using Snowpark's DataFrame API directly in a Snowflake worksheet. Manually create a log file with metrics and model parameters and upload it to a Snowflake stage.
- E. Train the model within a Snowpark Python stored procedure. Use a Snowflake stage to store MLflow artifacts.

Answer: C

Explanation:
The Snowpark MLAPI, combined with MLflow's autologging feature, is designed to automatically log metrics, parameters, and model artifacts to MLflow without requiring explicit logging code. The autologging functionality is triggered when enabled before the training process begins. Other options require manual logging, lack built-in MLflow integration, or don't fully leverage the Snowpark ML capabilities.

NEW QUESTION # 12
Which of the following statements are TRUE regarding the 'Data Understanding' and 'Data Preparation' steps within the Machine Learning lifecycle, specifically concerning handling data directly within Snowflake for a large, complex dataset?

- A. Data Preparation should always be performed outside of Snowflake using external tools to avoid impacting Snowflake performance.
- B. Data Understanding primarily involves identifying potential data quality issues like missing values, outliers, and inconsistencies, and Snowflake features like 'QUALIFY and 'APPROX TOP can aid in this process.
- C. Data Preparation in Snowflake can involve feature engineering using SQL functions, creating aggregated features with window functions, and handling missing values using 'NVL' or 'COALESCE. Furthermore, Snowpark Python provides richer

data manipulation using DataFrame APIs directly on Snowflake data.

- D. The 'Data Understanding' step is unnecessary when working with data stored in Snowflake because Snowflake automatically validates and cleans the data during ingestion.
- E. During Data Preparation, you should always prioritize creating a single, wide table containing all possible features to simplify the modeling process.

Answer: B,C

Explanation:
Data Understanding is crucial for identifying data quality issues using tools such as 'QUALIFY' and 'APPROX TOP Data Preparation within Snowflake using SQL and Snowpark Python enables efficient feature engineering and data cleaning. Option C is incorrect because Snowflake doesn't automatically validate and clean your data. Option D is incorrect as leveraging Snowflake's compute for data preparation alongside Snowpark can drastically increase speed. Option E is not desirable, feature selection is important, and feature stores help in organization.

NEW QUESTION # 13
Consider the following Snowflake SQL query used to calculate the RMSE for a regression model's predictions, where 'actual_value' is the actual value and 'predicted value' is the model's prediction. However, you notice that the RMSE calculation is incorrect due to an error in the query. Identify the error in the query and provide the corrected query. The table name is 'sales_predictions'.

```
SELECT SQRT(AVG(POW(actual_value - predicted_value, 2))) AS incorrect_rmse FROM sales_predictions;
```

Which of the following options represents the corrected query that accurately calculates the RMSE?

- A.
```
SELECT SQRT(AVG(POW(actual_value - predicted_value, 2))) OVER () AS correct_rmse FROM sales_predictions;
```
- B.
```
SELECT SUM(POW(actual_value - predicted_value, 2)) / COUNT( ) AS correct_rmse FROM sales_predictions;
```
- C.
```
SELECT SQRT(AVG(ABS(actual_value - predicted_value))) AS correct_rmse FROM sales_predictions;
```
- D.
```
SELECT SQRT(MEAN(POW(actual_value - predicted_value, 2))) AS correct_rmse FROM sales_predictions;
```
- E.
```
SELECT SQRT(VAR_POP(actual_value - predicted_value)) AS correct_rmse FROM sales_predictions;
```

Answer: D

Explanation:
The original query uses 'AVG', which is perfectly valid. However, Snowflake also supports the 'MEAN' function, which is an alias for the 'AVG' function and is used for calculated descriptive statistics. The corrected query maintains the functionality and structure, only substituting AVG with MEAN. B and C options lack the square root function. Option D calculates the Mean Absolute Error (MAE), not the RMSE. Option E uses a window function unnecessarily.

NEW QUESTION # 14
A data scientist is tasked with predicting customer churn for a telecommunications company using Snowflake. The dataset contains call detail records (CDRs), customer demographic information, and service usage data'. Initial analysis reveals a high degree of multicollinearity between several features, specifically 'total_day_minutes', 'total_eve_minutes', and 'total_night_minutes'. Additionally, the 'state' feature has a large number of distinct values. Which of the following feature engineering techniques would be MOST effective in addressing these issues to improve model performance, considering efficient execution within Snowflake?

- A. Apply min-max scaling to the CDR features to normalize them and use label encoding for the 'state' feature. Train a decision tree model, as it is robust to multicollinearity.
- B. Apply Principal Component Analysis (PCA) to reduce the dimensionality of the CDR features ('total_day_minutes', 'total_eve_minutes', 'total_night_minutes') and use one-hot encoding for the 'state' feature.
- C. Use a variance threshold to remove highly correlated CDR features and create a feature representing the geographical region (e.g., 'Northeast', 'Southwest') based on the 'state' feature using a custom UDF.
- D. Create interaction features by multiplying 'total_day_minutes' with 'customer_service_calls' and applying a target encoding to the 'state' feature.
- E. Calculate the Variance Inflation Factor (VIF) for each CDR feature and drop the feature with the highest VIE Apply

frequency encoding to the 'state' feature.

Answer: C

Explanation:
Option C is the most effective. Using a variance threshold directly addresses multicollinearity by removing redundant features. Creating a geographical region feature from 'state' reduces dimensionality and is more manageable than one-hot encoding for high cardinality features. A custom UDF can be used for efficient regional mapping. While PCA can reduce dimensionality, it can also make the features less interpretable. Target encoding (B) can introduce target leakage if not handled carefully. VIF calculation (D) is useful but doesn't directly address the high cardinality of 'state'. Label encoding (E) is not appropriate for nominal categorical features like 'state' as it introduces ordinality.

NEW QUESTION # 15
......

In real life, every great career must have the confidence to take the first step. When you suspect your level of knowledge, and cramming before the exam, do you think of how to pass the Snowflake DSA-C03 exam with confidence? Do not worry, Exam-Killer is the only provider of training materials that can help you to pass the exam. Our training materials, including questions and answers, the pass rate can reach 100%. With Exam-Killer Snowflake DSA-C03 Exam Training materials, you can begin your first step forward. When you get the certification of Snowflake DSA-C03 exam, the glorious period of your career will start.

Reliable DSA-C03 Test Guide: https://www.exam-killer.com/DSA-C03-valid-questions.html

Snowflake Exam DSA-C03 Preview They are now living the life they desire, Believe us because the DSA-C03 test prep are the most useful and efficient, and the DSA-C03 exam preparation will make you master the important information and the focus to pass the DSA-C03 exam, Snowflake Exam DSA-C03 Preview Before you buy some things, the reference demo is necessary, So getting the DSA-C03 certification seems the most important thing in your current plan.

Objects button—Opens the list of object types DSA-C03 in the object bar at the left side of the window, Though the free demos area small part of the exam braindumps, they Exam DSA-C03 Preview contain the represent questions for you to know its accuracy and good quality.

Quiz 2025 Pass-Sure DSA-C03: Exam SnowPro Advanced: Data Scientist Certification Exam Preview

They are now living the life they desire, Believe us because the DSA-C03 Test Prep are the most useful and efficient, and the DSA-C03 exam preparation will make you master the important information and the focus to pass the DSA-C03 exam.

Before you buy some things, the reference demo is necessary, So getting the DSA-C03 certification seems the most important thing in your current plan, And you will get all you desire with our DSA-C03 exam questions.

100% Pass □ ✔ www.exam4pdf.com □✔□ is best website to obtain ➽ DSA-C03 □ for free download □Dumps DSA-C03 Questions

- 100% Pass 2025 Snowflake DSA-C03: Perfect Exam SnowPro Advanced: Data Scientist Certification Exam Preview □ Easily obtain free download of ☀ DSA-C03 □☀□ by searching on □ www.pdfvce.com □ □Questions DSA-C03 Pdf
- 100% Pass 2025 Snowflake DSA-C03: Perfect Exam SnowPro Advanced: Data Scientist Certification Exam Preview □ Search on [www.pass4test.com] for ➥ DSA-C03 □ to obtain exam materials for free download □Dumps DSA-C03 Questions
- www.qibeips.com, pct.edu.pk, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, www.stes.tyc.edu.tw, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, 61921a.com, chrisukankem.com, 42.192.103.116, onlineclass.indokombucha.com, www.stes.tyc.edu.tw, Disposable vapes

What's more, part of that Exam-Killer DSA-C03 dumps now are free: https://drive.google.com/open?id=1cNqz1vK5Q14XUhKmF6J-225NyrtFlLS1