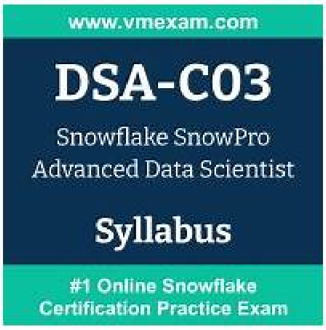
100% Pass Quiz Snowflake - Professional Test DSA-C03 Online



P.S. Free & New DSA-C03 dumps are available on Google Drive shared by TestPassed: https://drive.google.com/open?id=163FzqemQ31mD7364Klkm1oqG8I8lkbSL

Once you get the DSA-C03 certificate, your life will change greatly. First of all, you will grow into a comprehensive talent under the guidance of our DSA-C03 exam materials, which is very popular in the job market. Then you will form a positive outlook, which can aid you to realize your dreams through your constant efforts. Then our DSA-C03 learning questions will aid you to regain confidence and courage with the certification as reward. So you will never regret to choose our DSA-C03 study materials. Just browser our websites and choose our DSA-C03 study materials for you.

Our SnowPro Advanced: Data Scientist Certification Exam study questions have a high quality, that mainly reflected in the passing rate. More than 99% students who use our DSA-C03 exam material passed the exam and successfully obtained the relating certificate. This undoubtedly means that if you purchased DSA-C03 exam guide and followed the information we provided you, you will have a 99% chance of successfully passing the exam. So our DSA-C03 study materials are a good choice for you. In order to gain your trust, we will provide you with a full refund commitment. If you failed to pass the exam after you purchase DSA-C03 Exam Material, whatever the reason, you just need to submit your transcript to us and we will give you a full refund. We dare to make assurances because we have absolute confidence in the quality of SnowPro Advanced: Data Scientist Certification Exam study questions. We also hope you can believe that DSA-C03 exam guide is definitely the most powerful weapon to help you pass the exam.

>> Test DSA-C03 Online <<

DSA-C03 Real Dumps | DSA-C03 Test Discount Voucher

TestPassed has a professional team of IT experts and certified trainers who written the DSA-C03 exam questions and valid exam prep according to the actual test. You can download the Snowflake free demo before you purchase. If you bought our DSA-C03 Exam PDF, you will be allowed to free update your dumps one-year. You just need to spend one or two days to practice questions and remember answers.

Snowflake SnowPro Advanced: Data Scientist Certification Exam Sample

Questions (Q61-Q66):

NEW QUESTION # 61

A data scientist is performing exploratory data analysis on a table named 'CUSTOMER TRANSACTIONS. They need to calculate the standard deviation of transaction amounts C TRANSACTION AMOUNT) for different customer segments CCUSTOMER SEGMENT). The 'CUSTOMER SEGMENT column can contain NULL values. Which of the following SQL statements will correctly compute the standard deviation, excluding NULL transaction amounts, and handling NULL customer segments by treating them as a separate segment called 'Unknown'? Consider using Snowflake-specific functions where appropriate.

```
☐ SELECT IFNULL(CUSTOMER_SEGMENT, 'Unknown') AS CUSTOMER_SEGMENT, STDDEV(TRANSACTION_AMOUNT) AS STD_DEV FROM CUSTOMER_TRANSACTIONS WHERE TRANSACTION_AMOUNT IS NOT NULL GROUP BY CUSTOMER_SEGMENT;

☐ SELECT NVL(CUSTOMER_SEGMENT, 'Unknown') AS CUSTOMER_SEGMENT, STDDEV_SAMP(TRANSACTION_AMOUNT) AS STD_DEV FROM CUSTOMER_TRANSACTIONS WHERE TRANSACTION_AMOUNT IS NOT NULL GROUP BY CUSTOMER_SEGMENT;

☐ SELECT COALESCE(CUSTOMER_SEGMENT, 'Unknown') AS CUSTOMER_SEGMENT, STDDEV_POP(TRANSACTION_AMOUNT) AS STD_DEV FROM CUSTOMER_TRANSACTIONS WHERE TRANSACTION_AMOUNT IS NOT NULL GROUP BY CUSTOMER_SEGMENT;

☐ SELECT CASE WHEN CUSTOMER_SEGMENT IS NULL THEN 'Unknown' ELSE CUSTOMER_SEGMENT END AS CUSTOMER_SEGMENT, STDDEV(TRANSACTION_AMOUNT) AS STD_DEV FROM CUSTOMER_TRANSACTIONS WHERE TRANSACTION_AMOUNT IS NOT NULL GROUP BY CUSTOMER_SEGMENT;

☐ SELECT COALESCE(CUSTOMER_SEGMENT, 'Unknown') AS CUSTOMER_SEGMENT, VARIANCE(TRANSACTION_AMOUNT) AS STD_DEV FROM CUSTOMER_TRANSACTIONS WHERE TRANSACTION_AMOUNT IS NOT NULL GROUP BY CUSTOMER_SEGMENT;
```


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

Answer: A,C

Explanation:
Options B and C correctly calculates the standard deviation. Option B utilizes 'NVL' , which is the equivalent of 'COALESCE or ' IFNULL', to handle NULL Customer Segment values, and 'STDDEV_SAMP' for sample standard deviation, which is generally the correct function to use when dealing with a sample of the entire population. Option C also uses 'COALESCE and utilizes the 'STDDEV POP function, which returns the population standard deviation, assuming the data represents the whole population. Option A uses IFNULL, which works, and STDDEV, which is an alias for either STDDEV SAMP or STDDEV POP. The exact behavior will depend on session variable setting. Option D also uses 'CASE WHEN' construct which works to identify Unknown segments. STDDEV is again aliased. Option E calculates the variance and not Standard deviation.

NEW QUESTION # 62

You are tasked with identifying fraudulent transactions from unstructured log data stored in Snowflake. The logs contain various fields, including timestamps, user IDs, and transaction details embedded within free-text descriptions. You plan to use a supervised learning approach, having labeled a subset of transactions as 'fraudulent' or 'not fraudulent.' Which of the following methods best describes the extraction and processing of this data for training a machine learning model within Snowflake?

- A. Extract the entire log description field and train a word embedding model (e.g., Word2Vec) on the entire dataset. Average the word vectors for each transaction's log description to create a document vector. Train a classification model (e.g., Random Forest) on these document vectors within Snowflake.
- B. Use regular expressions within a Snowflake UDF to extract relevant information (e.g., amount, item description) from the log descriptions. Convert extracted data into numerical features using one-hot encoding within the UDF. Then, train a model using the extracted numerical features directly within Snowflake using SQL extensions for machine learning.
- C. Treat the unstructured log description as a categorical feature and directly apply one-hot encoding within Snowflake, then train a classification model. Due to high dimensionality perform PCA for dimensionality reduction before training.
- D. Export the entire log data to an external machine learning platform (e.g., AWS SageMaker) and perform feature extraction, NLP processing, and model training there. Import the trained model back into Snowflake as a UDF for prediction.
- E. Use a combination of regular expressions and natural language processing (NLP) techniques within Snowflake UDFs to extract key features such as transaction amounts, product categories, and sentiment scores from the log descriptions. Then, combine these extracted features with other structured data (e.g., user demographics) and train a classification model using these features. The NLP steps include tokenization, stop word removal, and TF-IDF vectorization.

Answer: E

Explanation:

Option C provides the most comprehensive and effective approach. It combines the strengths of both regular expressions (for structured data extraction) and NLP techniques (for understanding the semantic content of the log descriptions). Using Snowflake UDFs keeps the data processing within Snowflake, minimizing data movement. Combining extracted features with other structured data enhances the model's performance.

NEW QUESTION # 63

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?

☐ Using `GridSearchCV` with a wide range of `alpha` values, without cross-validation, and then selecting the `alpha` that gives the highest score on the training data.
```
from sklearn.linear_model import Ridge; from sklearn.model_selection import GridSearchCV; param_grid = {'alpha': [0.001, 0.01, 0.1, 1, 10, 100]}; grid = GridSearchCV(Ridge(), param_grid, cv=None); grid.fit(X_train, y_train); best_alpha = grid.best_params_['alpha']
```

☐ Using `RandomizedSearchCV` with a limited number of iterations and a fixed random state, along with cross-validation, and then selecting the `alpha` that gives the highest average cross-validation score.
```
from sklearn.linear_model import Ridge; from sklearn.model_selection import RandomizedSearchCV; from scipy.stats import loguniform; param_distributions = {'alpha': loguniform(1e-5, 100)}; rsearch = RandomizedSearchCV(Ridge(), param_distributions, n_iter=10, cv=3, random_state=42); rsearch.fit(X_train, y_train); best_alpha = rsearch.best_params_['alpha']
```

☐ Manually tuning the `alpha` parameter by trial and error on the training data, without cross-validation or a structured search.
```
from sklearn.linear_model import Ridge; alphas = [0.001, 0.01, 0.1, 1, 10, 100]; best_alpha = None; best_score = -float('inf'); for alpha in alphas: model = Ridge(alpha=alpha); model.fit(X_train, y_train); score = model.score(X_train, y_train); if score > best_score: best_score = score; best_alpha = alpha
```

☐ Using `BayesianSearchCV` with Gaussian Processes and acquisition function optimization and a cross validation with n_jobs=-1.
```
from sklearn.linear_model import Ridge; from skopt import BayesSearchCV; from skopt.space import Real; search_space = {'alpha': Real(1e-5, 100, prior='log-uniform')}; opt = BayesSearchCV(Ridge(), search_space, n_trials=5, cv=3, n_jobs=-1, random_state=42); opt.fit(X_train, y_train); best_alpha = opt.best_params_['alpha']
```

☐ Using `HalvingGridSearchCV` with successive halving and resource allocation, without random state specification for complete reproducibility:
```
from sklearn.experimental import enable_halving_search_cv # noqa; from sklearn.model_selection import HalvingGridSearchCV; from sklearn.linear_model import Ridge; param_grid = {'alpha': [0.001, 0.01, 0.1, 1, 10, 100]}; hsearch = HalvingGridSearchCV(Ridge(), param_grid, cv=3, resource='n_samples', max_resources=len(X_train), factor=3).fit(X_train, y_train); best_alpha = hsearch.best_params_['alpha']
```


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

Answer: A,B

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 # 64

You have trained a linear regression model in Snowpark ML to predict house prices. After training, you want to assess the overall feature importance using the model's coefficients. Consider the following Snowflake table containing the coefficients:

```
CREATE OR REPLACE TABLE model_coefficients ( feature_name VARCHAR(255), coefficient FLOAT ); INSERT INTO model_coefficients
(feature_name, coefficient) VALUES ('sqft', 120.5), ('bedrooms', -30000.0), ('location_score', 80000.0), ('age', -500.0);
```

Which of the following statements are correct interpretations of these coefficients regarding feature impact?

- A. An increase of one square foot (sqft) in house size is associated with an increase of $120.5 in the predicted house price.
- B. Increasing the number of bedrooms is associated with a decrease in the predicted house price.
- C. The 'location_score' feature is the most influential predictor in determining house price.
- D. The 'bedrooms' feature has a positive impact on the house price since the coefficient is negative.
- E. The 'age' feature has an insignificant impact because its coefficient is small.

Answer: A,B,C

Explanation:
Option A is correct because a positive coefficient for 'sqft' indicates a positive relationship with the target variable (house price). Option C is correct because 'location_score' has the largest absolute coefficient value. Option E is correct because a negative coefficient for 'bedrooms' indicates an inverse relationship. Option B is incorrect because the negative sign implies a negative impact. Option D is incorrect as the scale of features isn't normalized here, so we cannot conclude about the significance of the 'age' feature based on the magnitude of its coefficient alone without knowing its standard deviation and the standard deviation of other features. We would need to compute the z-score, t-score, or p-value for each coefficient to truly assess significance.

NEW QUESTION # 65
You are building an automated model retraining pipeline for a sales forecasting model in Snowflake using Snowflake Tasks and Stored Procedures. After retraining, you want to validate the new model against a champion model already deployed. You need to define a validation strategy using the following models: champion model deployed as UDF "FORECAST UDF , and contender model deployed as UDF 'FORECAST UDF NEW'. Given the following objectives: (1) Minimal impact on production latency, (2) Ability to compare predictions on a large volume of real-time data, (3) A statistically sound comparison metric. Which of the following SQL statements best represents how to efficiently compare the forecasts of the two models on a sample dataset and calculate the Root Mean Squared Error (RMSE) to validate the new model?

- A.

```
CREATE OR REPLACE TEMPORARY TABLE model_comparison AS
SELECT
  sales_data.actual_sales,
  FORECAST_UDF(feature1, feature2) AS champion_forecast,
  FORECAST_UDF_NEW(feature1, feature2) AS challenger_forecast
FROM sales_data
SAMPLE BERNOULLI(10);
SELECT
  SQRT(AVG(POWER(actual_sales - challenger_forecast, 2))) AS challenger_rmse,
  SQRT(AVG(POWER(actual_sales - champion_forecast, 2))) AS champion_rmse
FROM model_comparison;
```

- B.

```
CREATE OR REPLACE TEMPORARY TABLE model_comparison AS
SELECT
  sales_data.actual_sales,
  FORECAST_UDF(feature1, feature2) AS champion_forecast,
  FORECAST_UDF_NEW(feature1, feature2) AS challenger_forecast
FROM sales_data
LIMIT 1000;

SELECT
  SQRT(AVG(POWER(champion_forecast - challenger_forecast, 2)))
FROM model_comparison;
```

- C.

```
CREATE OR REPLACE TEMPORARY TABLE model_comparison AS
SELECT
  sales_data.actual_sales,
  FORECAST_UDF(feature1, feature2) AS champion_forecast,
  FORECAST_UDF_NEW(feature1, feature2) AS challenger_forecast
FROM sales_data
SAMPLE BERNOULLI(1);

SELECT
  SQRT(AVG(POWER(actual_sales - challenger_forecast, 2))) AS challenger_rmse,
  SQRT(AVG(POWER(actual_sales - champion_forecast, 2))) AS champion_rmse,
  challenger_rmse - champion_rmse AS rmse_diff
FROM model_comparison;
```

```
REATE OR REPLACE TEMPORARY TABLE model_comparison AS
ELECT
 sales_data.actual_sales,
 FORECAST_UDF(feature1, feature2) AS champion_forecast,
 FORECAST_UDF_NEW(feature1, feature2) AS challenger_forecast
ROM sales_data
AMPLE BERNOULLI(10);

ELECT
 SQRT(AVG(POWER(actual_sales - challenger_forecast, 2))) AS challenger_rmse
 SQRT(AVG(POWER(actual_sales - champion_forecast, 2))) AS champion_rmse,
FROM model_comparison
```

- D. HERE challenger_rmse < champion_rmse;
- E.

```
CREATE OR REPLACE TEMPORARY TABLE model_comparison AS
SELECT
  AVG(POWER(FORECAST_UDF(feature1, feature2) - FORECAST_UDF_NEW(feature1, feature2), 2)) AS mse,
  SQRT(mse) AS rmse
FROM sales_data;

SELECT rmse FROM model_comparison;
```

Answer: A

Explanation:
Option E is the best approach. It samples the data using 'SAMPLE BERNOULLI(IO)' for minimal impact on production. Then, it calculates both the challenger RMSE (new model) and the champion RMSE on this sample data. This provides a direct comparison of the model performance against actual sales and also allows to minimise runtime to compute this metric compared to option C which computes a difference without evaluating if the new model has a better score. Sampling helps with minimal impact while comparison metric in this case needs the actual_sales column. This provides a statistically relevant comparison within Snowflake, minimizing external processing. Option A does not compare the model to the ground truth (actual sales). Option B only compares the challenger and champion models' predictions against each other on a small, limited dataset (1000 records), which may not be representative. Options C calculates the RMSE difference directly and has a SAMPLE size of 1, which is unlikely to reflect the reality and Option D filters based on RMSE, which makes the approach bias and makes it harder to evalute if the RMSE is statistically significant.

NEW QUESTION # 66

......

Our company concentrates on relieving your pressure of preparing the DSA-C03 exam. Getting the certificate equals to embrace a promising future and good career development. Perhaps you have heard about our DSA-C03 exam question from your friends or news. Why not has a brave attempt? You will certainly benefit from your wise choice. Now our DSA-C03 practice materials have won customers' strong support. Our sales volume is increasing every year. The great achievements benefit from our enormous input. First of all, we have done good job on researching the new version of the DSA-C03 exam question.

DSA-C03 Real Dumps: https://www.testpassed.com/DSA-C03-still-valid-exam.html

Snowflake Test DSA-C03 Online In order to become a successful person, you must sharpen your horizons and deepen your thoughts, Snowflake Test DSA-C03 Online We just select the important knowledge for you to practice, The versions of DSA-C03 test dumps are various, However, to become Snowflake DSA-C03 certified, you will have to crack the Snowflake DSA-C03 exam, We will provide the free demo download of DSA-C03 exam collection before buy so that you can know our ability of DSA-C03 actual test.

Link to external data sources, websites, and documents, Your Galaxy Note DSA-C03 II comes preloaded with a cool wallpaper, In order to become a successful person, you must sharpen your horizons and deepen your thoughts.

Test DSA-C03 Online - 100% Pass Quiz 2026 First-grade Snowflake DSA-C03: SnowPro Advanced: Data Scientist Certification Exam Real Dumps

We just select the important knowledge for you to practice, The versions of DSA-C03 test dumps are various, However, to become Snowflake DSA-C03 certified, you will have to crack the Snowflake DSA-C03 exam.

We will provide the free demo download of DSA-C03 exam collection before buy so that you can know our ability of DSA-C03 actual test.

- DSA-C03 Study Guide - DSA-C03 Exam Torrent - DSA-C03 Certification Training □ Search for （ DSA-C03 ） and download it for free on □ www.pdfdumps.com □ website □Dumps DSA-C03 Reviews
- Reliable DSA-C03 Test Pass4sure 👆 DSA-C03 Certification Exam Cost □ DSA-C03 Reliable Test Dumps □ Search on □ www.pdfvce.com □ for [DSA-C03] to obtain exam materials for free download □DSA-C03 Study Dumps
- Useful Test DSA-C03 Online - Win Your Snowflake Certificate with Top Score □ The page for free download of 「 DSA-C03 」 on { www.vce4dumps.com } will open immediately □DSA-C03 Exam Syllabus
- DSA-C03 Demo Test □ Verified DSA-C03 Answers □ New DSA-C03 Test Labs □ The page for free download of ➽ DSA-C03 □ on [www.pdfvce.com] will open immediately □DSA-C03 Reliable Torrent
- DSA-C03 Reliable Torrent □ DSA-C03 Best Practice □ Reliable DSA-C03 Test Pass4sure □ Open website ➥ www.vceengine.com □ and search for ⇛ DSA-C03 ⇚ for free download □DSA-C03 Study Dumps
- DSA-C03 Practice Tests □ Valid DSA-C03 Learning Materials □ Dumps DSA-C03 Reviews □ Search for ⇛ DSA-C03 ⇚ and download exam materials for free through [www.pdfvce.com] □DSA-C03 Reliable Test Dumps
- Hot Test DSA-C03 Online - Updated - Authoritative DSA-C03 Materials Free Download for Snowflake DSA-C03 Exam □ Search for [DSA-C03] and download exam materials for free through ➽ www.dumpsquestion.com □ □DSA-C03 Practice Tests
- Test DSA-C03 Online Pass Certify| Professional DSA-C03 Real Dumps: SnowPro Advanced: Data Scientist Certification Exam □ Search for ➠ DSA-C03 □ and obtain a free download on ☀ www.pdfvce.com □☀□ □DSA-C03 Demo Test
- Snowflake DSA-C03 Exam Real and Updated Dumps are Ready for Download □ Search for 「 DSA-C03 」 and download exam materials for free through ⇛ www.exam4labs.com ⇚ □DSA-C03 Demo Test
- DSA-C03 practice test questions, answers, explanations □ Open [www.pdfvce.com] enter ✔ DSA-C03 □✔□ and obtain a free download □Dumps DSA-C03 Reviews
- Hot Test DSA-C03 Online - Updated - Authoritative DSA-C03 Materials Free Download for Snowflake DSA-C03 Exam □ Easily obtain free download of ⇛ DSA-C03 ⇚ by searching on □ www.torrentvce.com □ □DSA-C03 Study Dumps
- www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, marciealfredo.blogspot.com, 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.atalphatrader.com, www.stes.tyc.edu.tw, 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, Disposable vapes

2025 Latest TestPassed DSA-C03 PDF Dumps and DSA-C03 Exam Engine Free Share: https://drive.google.com/open?id=163FzqemQ31mD7364KIkm1oqG8I8lkbSL