Free PDF Quiz Snowflake - DSA-C03 - SnowPro Advanced: Data Scientist Certification Exam Pass-Sure Accurate Test

BONUS!!! Download part of ExamCost DSA-C03 dumps for free: https://drive.google.com/open?id=1A4epWHnbbLWKOWKZn1Q5wDrMMtcmCwIr

We have security and safety guarantee, which mean that you cannot be afraid of virus intrusion and information leakage since we have data protection acts, even though you end up studying DSA-C03 test guide of our company, we will absolutely delete your personal information and never against ethic code to sell your message to the third parties. Our DSA-C03 Exam Questions will spare no effort to perfect after-sales services. Thirdly countless demonstration and customer feedback suggest that our SnowPro Advanced: Data Scientist Certification Exam study question can help them get the certification as soon as possible, thus becoming the elite, getting a promotion and a raise and so forth.

If you want to be an excellent elites in this line, you need to get the DSA-C03 certification, thus it can be seen through the importance of qualification examination. Only through qualification examination, has obtained the corresponding qualification certificate, we will be able to engage in related work, so the DSA-C03 Test Torrent is to help people in a relatively short period of time a great important tool to pass the qualification test. Choose our DSA-C03 study tool, can help users quickly analysis in the difficult point, and pass the DSA-C03 exam successfully.

>> DSA-C03 Accurate Test <<

Exam DSA-C03 Simulator Free & Exam DSA-C03 Study Solutions

Our DSA-C03 exam cram has been revised for lots of times to ensure all candidates can easily understand all knowledge parts. In the meantime, the learning process is recorded clearly in the system, which helps you adjust your learning plan. On the one hand, our company has benefited a lot from renovation. Customers are more likely to choose our products. On the other hand, the money we have invested is meaningful, which helps to renovate new learning style of the DSA-C03 Exam. So, why not buy our DSA-C03 test guide?

Snowflake SnowPro Advanced: Data Scientist Certification Exam Sample Questions (Q174-Q179):

NEW QUESTION # 174
You are analyzing customer transaction data in Snowflake to identify fraudulent activities. The 'TRANSACTION AMOUNT' column exhibits a right-skewed distribution. Which of the following Snowflake queries is MOST effective in identifying outliers based on the Interquartile Range (IQR) method, specifically targeting unusually large transaction amounts? Assume IQR is already calculated as variable and QI as and Q3 as in snowflake session.

- A. SELECT TRANSACTION ID FROM TRANSACTIONS WHERE TRANSACTION_AMOUNT > (SELECT + 3 FROM TRANSACTIONS);
- B. SELECT TRANSACTION ID FROM TRANSACTIONS WHERE TRANSACTION_AMOUNT < qi - (1.5 iqr);

- C. SELECT TRANSACTION ID FROM TRANSACTIONS WHERE TRANSACTION AMOUNT > q3 + (1.5 iqr);
- D. SELECT TRANSACTION ID FROM TRANSACTIONS WHERE TRANSACTION_AMOUNT > (SELECT MEDIAN(TRANSACTION AMOUNT) FROM TRANSACTIONS);
- E. SELECT TRANSACTION ID FROM TRANSACTIONS WHERE TRANSACTION_AMOUNT > (SELECT WITHIN GROUP (ORDER BY TRANSACTION_AMOUNT) FROM TRANSACTIONS);

Answer: C

Explanation:
Option B correctly implements the IQR method for identifying outliers. The formula 'Q3 + (1.5 identifies values significantly higher than the third quartile, which is appropriate for detecting outliers in a right-skewed distribution. Options A uses standard deviation, which is less robust to outliers. Option C finds the 95th percentile but might not isolate extreme outliers. Option D looks for lower bound outliers, which is not the case. Option E uses the median which not effective to find out outlier values.

NEW QUESTION # 175
You're building a customer segmentation model and need to aggregate data from various tables. You have the following tables in Snowflake: 'customer demographics' (customer id, age, city, income) 'customer transactionS (transaction_id, customer id, transaction_date, amount) 'product_details' (product_id, category) 'transaction_products' (transaction_id, product_id) Your goal is to create a single Snowpark DataFrame containing customer demographics along with the total amount spent by each customer on products within the 'Electronics' category in the last year. However, ensure that only customers with income greater than 50000 are considered and handle cases where customers have no transaction records, assigning a value of 0 to the 'total_electronics_spending' column for those customers. How can we achieve this using snowpark? Choose the correct options

- A. Create a complex SQL query within Snowpark using 'session.sql()' to perform all the joins, filtering, and aggregation in a single step. This will be the most efficient approach.
- B. Use a series of INNER JOINs to connect the tables and filter data, followed by grouping and aggregation. This approach guarantees accurate results with good performance.
- C. Create a Python UDF that performs the joins and aggregations. This offers flexibility and good performance when dealing with complex data transformations.
- D. Use a combination of LEFT JOINs and filtering. Start with 'customer_demographics' (filtered for income > 50000) as the base table and LEFT JOIN to subsequent tables. Use the 'coalesce' function to handle customers without transaction data.
- E. Create a temporary view to store total electronics expenditure of each customer and left join with customer demographics table.

Answer: A,D,E

Explanation:
Option B, C and D are correct. Option B is correct because using LEFT JOINs starting with 'customer_demographics' (after filtering for income) ensures all eligible customers are included. 'coalesce' is crucial for handling customers with no transactions, assigning a 0 value. Option C is also correct as using a temporary view is a valid solution to have electronics expenditure for each customer. Option D is correct as pushing down all operations to SQL within Snowpark can be highly performant, as it allows Snowflake to optimize the query execution. However, query readability and maintainability should also be considered. Option A is incorrect because it states that INNER JOINs should be used, but inner joins would exclude customers with no transaction data which is opposite to what is stated in the question. Option E is incorrect as UDFs can introduce performance overhead compared to native Snowpark DataFrame operations or direct SQL queries, especially for large datasets. Avoid UDF when the same output can be achieved without it.

NEW QUESTION # 176
You are a data scientist working with a Snowflake table named 'CUSTOMER DATA' that contains a 'PHONE NUMBER' column stored as VARCHAR. The 'PHONE NUMBER' column sometimes contains non-numeric characters like hyphens and parentheses, and in some rows the data is missing. You need to create a new table 'CLEANED CUSTOMER DATA' with a column named 'CLEANED PHONE NUMBER that contains only the numeric part of the phone number (as VARCHAR) and replaces missing or invalid phone numbers with NULL. Which of the following Snowpark Python code snippets achieves this most efficiently, ensuring no errors occur during the data transformation, and considers Snowflake's performance best practices?

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

- E. Option C

Answer: C

Explanation:
Option E is the most efficient because it leverages Snowpark's built-in functions for string manipulation and conditional logic directly. It first removes all non-numeric characters using 'regexp_replace' and then uses 'iff (if and only if) to replace empty strings (resulting from cleaning) with NULL. This approach avoids using UDFs (User-Defined Functions), which can introduce overhead. Option B, although using 'regexp_replace' , requires an additional 'with_column' to handle empty strings after cleaning. Option A introduces UDF that decreases performance. Option C calls UDF with undefined 'call_udf function and 'snowflake-snowpark-python' library. Option D is missing dataframe and its transformation is not happening on top of Dataframe. Option E is preferrable over Option B, as it uses the single transformation.

NEW QUESTION # 177
You are a data scientist working with a Snowflake table named 'CUSTOMER TRANSACTIONS' that contains sensitive PII data, including customer names and email addresses. You need to create a representative sample of 1% of the data for model development, ensuring that the sample is anonymized and protects customer privacy. The sample must be reproducible for future model iterations.
Which of the following steps are most appropriate using Snowpark for Python and SQL?

- A. Create a new table using 'CREATE TABLE AS SELECT statement combined with 'SAMPLE clause and SHA256 hashing functions in SQL to create the sample and anonymize data. Manually seed the random number generator in Python before executing the SQL statement via Snowpark.
- B. Use the 'QUALIFY OVER (ORDER BY RANDOM()) (SELECT COUNT( ) 0.01 FROM CUSTOMER_TRANSACTIONS)' clause with SHA256 on sensitive columns directly within a CREATE TABLE AS statement to generate an anonymized sample. The function should return only 1 percentage of row.
- C. Employ stratified sampling based on a customer segment column, then anonymize data. Use the TABLESAMPLE BERNOULLI function in SQL with a 1 percent sample rate. Apply SHA256 hashing to the 'customer_name' and 'email_addresS columns using SQL functions.
- D. Use Snowpark DataFrame's 'sample' function with a fraction of 0.01 and a fixed random seed. Before sampling, create a view that masks 'customer_name' and 'email_address' columns, and then sample from the view.
- E. Use the 'SAMPLE clause in a SQL query to extract 1% of the rows, then apply SHA256 hashing to the 'customer_name' and 'email_addresS columns within Snowpark using a UDF. Seed the sampling for reproducibility.

Answer: C,E

Explanation:
Options A and D are correct because they address both sampling and anonymization requirements while leveraging Snowflake's capabilities. Option A utilizes SAMPLE clause within a SQL query in Snowflake and then leverages UDF for SHA256 hashing of sensitive information. This is a practical and common data sampling/anonymization pattern. Option D employs stratified sampling based on a customer segment, TABLESAMPLE BERNOULLI and SHA256 hashing in SQL, which provides a solid anonymization and sampling strategy.Option B: Creating a view is a good practice, but it doesn't automatically anonymize the data, and directly sampling from the view without anonymization doesn't meet the security requirements. Option C: Manually seeding in Python doesn't guarantee reproducibility when SQL is executed separately, as Snowflake has its own random number generator. Option E does not guarantee reproduciblity, and the query complexity might introduce performance issues and is less readable compared to the other options.

NEW QUESTION # 178
You are tasked with presenting a business case to stakeholders demonstrating the value of a new machine learning model that predicts customer churn. The model has been trained on data within Snowflake, and you have various metrics such as accuracy, precision, recall, and F I-score. You also have feature importance scores generated using a SHAP (SHapley Additive exPlanations) explainer. Which of the following visualization strategies, when combined, would MOST effectively communicate the model's performance and impact to a non-technical audience, while also providing sufficient detail for technical stakeholders?

- A. A ROC curve (Receiver Operating Characteristic) showing the trade-off between true positive rate and false positive rate, paired with a detailed table of all feature importance scores generated by the SHAP explainer. Present statistical summaries, such as mean and standard deviation, of the top 5 feature values, grouped by predicted churn probability.
- B. A distribution plot (e.g., histogram or KDE) of the predicted churn probabilities, segmented by actual churn status (churned vs. not churned), combined with a SHAP force plot visualizing the feature contributions for a single, randomly

selected customer who churned. Add a section on potential cost savings from churn reduction.

- C. A simple bar chart showing the overall accuracy score of the model alongside a table detailing the precision, recall, and F I-score. Include a word cloud of the most important features from the SHAP values.
- D. A confusion matrix visualizing the true positives, true negatives, false positives, and false negatives, along with a summary plot of the SHAP values showing the impact of each feature on the model's prediction for a representative sample of customers. A line chart showing cumulative churn rate across different customer segments.
- E. A scatter plot showing the relationship between two key features identified by SHAP, colored by the model's churn prediction, and a table summarizing the model's performance metrics (accuracy, precision, recall, F I-score). Additionally, include a waterfall plot for a specific customer, illustrating how each feature contributes to the final prediction.

Answer: D,E

Explanation:
Options B and D provide a balanced approach for both technical and non-technical audiences- A confusion matrix (Option B) is easily understandable and shows model performance across different prediction outcomes. A summary plot of SHAP values clearly illustrates feature importance and direction of impact. A line chart showing cumulative churn rate across different customer segments highlights the business value-Option D is also highly effective because scatter plots can be easily understood, especially when colored by churn prediction- The table of model metrics provides necessary details. The waterfall plot brings the explanation down to an individual customer level, making the model's behavior more tangible. Options A, C and E have deficits- Option A lacks detailed performance visualization. Option C is technical and might confuse non-technical stakeholders. Option E has too many summary plots.

NEW QUESTION # 179
......

C03 ⇚ and download exam materials for free through ✔ www.pdfvce.com □✔□ □Latest DSA-C03 Exam Dumps

- Pass Guaranteed Quiz DSA-C03 - SnowPro Advanced: Data Scientist Certification Exam Pass-Sure Accurate Test □ Search on ➽ www.dumpsmaterials.com □ for { DSA-C03 } to obtain exam materials for free download □DSA-C03 Latest Exam Test
- Detail DSA-C03 Explanation □ DSA-C03 Exam Topics □ Detail DSA-C03 Explanation □ Search for □ DSA-C03 □ and download it for free immediately on ➡ www.pdfvce.com □ □Standard DSA-C03 Answers
- Pass Guaranteed 2026 Snowflake DSA-C03 Authoritative Accurate Test □ Download 「 DSA-C03 」 for free by simply entering □ www.prep4sures.top □ website □DSA-C03 Latest Exam Test
- New DSA-C03 Test Tutorial □ DSA-C03 Cost Effective Dumps □ DSA-C03 Exam Quiz ↙ The page for free download of （ DSA-C03 ） on ⇛ www.pdfvce.com ⇚ will open immediately □DSA-C03 Exam Training
- Exam DSA-C03 Simulator □ DSA-C03 Valid Test Questions □ DSA-C03 Valid Test Questions □ Open website （ www.testkingpass.com ） and search for ➡ DSA-C03 □□□ for free download □Exam DSA-C03 Simulator
- DSA-C03 Test Questions Vce □ DSA-C03 Latest Mock Test □ Latest DSA-C03 Exam Papers □ Open " www.pdfvce.com " and search for [DSA-C03] to download exam materials for free □DSA-C03 Latest Exam Test
- Standard DSA-C03 Answers □ DSA-C03 Latest Mock Test □ DSA-C03 Test Questions Vce □ Simply search for ➠ DSA-C03 □ for free download on ✔ www.prepawayexam.com □✔□ □DSA-C03 Valid Test Questions
- www.stes.tyc.edu.tw, curiosiityclasses.com, tradestockspro.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, learnonline.pk, myportal.utt.edu.tt, www.stes.tyc.edu.tw, Disposable vapes

BONUS!!! Download part of ExamCost DSA-C03 dumps for free: https://drive.google.com/open?id=1A4epWHnbbLWKOWKZn1Q5wDrMMtcmCwIr