

Snowflake DSA-C03 Actual Exam Dumps Materials are the best simulate product - TestkingPass



BTW, DOWNLOAD part of TestkingPass DSA-C03 dumps from Cloud Storage: <https://drive.google.com/open?id=1fzJPHj3JNlPvstH1EjBc64a5KPlsR5O0>

No matter how good the product is users will encounter some difficult problems in the process of use, and how to deal with these problems quickly becomes a standard to test the level of product service. Our DSA-C03 study materials are not exceptional also, in order to enjoy the best product experience, as long as the user is in use process found any problem, can timely feedback to us, for the first time you check our DSA-C03 Study Materials performance, professional maintenance staff to help users solve problems.

Everybody knows that Snowflake is an influential company with high-end products and best-quality service. It will be a long and tough way to pass DSA-C03 exam test, especially for people who have no time to prepare the DSA-C03 Questions and answers. So choosing right DSA-C03 dumps torrent is very necessary and important for people who want to pass test at first attempt.

>> **Reliable DSA-C03 Dumps Ppt** <<

Pass Guaranteed Snowflake - DSA-C03 –Reliable Reliable Dumps Ppt

Snowflake Certified professionals are often more sought after than their non-certified counterparts and are more likely to earn higher

salaries and promotions. Moreover, cracking the SnowPro Advanced: Data Scientist Certification Exam (DSA-C03) exam helps to ensure that you stay up to date with the latest trends and developments in the industry, making you more valuable assets to your organization.

Snowflake SnowPro Advanced: Data Scientist Certification Exam Sample Questions (Q68-Q73):

NEW QUESTION # 68

A data scientist is using association rule mining with the Apriori algorithm on customer purchase data in Snowflake to identify product bundles. After generating the rules, they obtain the following metrics for a specific rule: Support = 0.05, Confidence = 0.7, Lift = 1.2. Consider that the overall purchase probability of the consequent (right-hand side) of the rule is 0.4. Which of the following statements are CORRECT interpretations of these metrics in the context of business recommendations for product bundling?

- A. Customers who purchase the items in the antecedent are 70% more likely to also purchase the items in the consequent, compared to the overall purchase probability of the consequent.
- B. The lift value of 1.2 indicates that customers are 20% more likely to purchase the consequent items when they have also purchased the antecedent items, compared to the baseline purchase probability of the consequent items.
- C. The rule applies to 5% of all transactions in the dataset, meaning 5% of the transactions contain both the antecedent and the consequent.
- D. The lift value of 1.2 suggests a strong negative correlation between the antecedent and consequent, indicating that purchasing the antecedent items decreases the likelihood of purchasing the consequent items.
- E. The confidence of 0.7 indicates that 70% of transactions containing the antecedent also contain the consequent.

Answer: B,C,E

Explanation:

Option A is correct because support represents the proportion of transactions that contain both the antecedent and the consequent. Option D is correct because confidence represents the proportion of transactions containing the antecedent that also contain the consequent. Option E is correct because $\text{lift} = \text{confidence} / (\text{probability of consequent})$. Therefore, lift of 1.2 means confidence is 1.2 times the probability of the consequent. Hence 20% more likely than the baseline. Option B is incorrect because lift, not confidence, captures the relative likelihood compared to the baseline. Option C is incorrect because a $\text{lift} > 1$ suggests a positive correlation, not a negative one.

NEW QUESTION # 69

You have trained a complex machine learning model using Snowpark for Python and are now preparing it for production deployment using Snowpark Container Services. You have containerized the model and pushed it to a Snowflake-managed registry. However, you need to ensure that only authorized users can access and deploy this model. Which of the following actions MUST you take to secure your model in the Snowflake Model Registry, ensuring appropriate access control, and minimizing the risk of unauthorized deployment or modification?

- A. Grant the 'USAGE' privilege on the stage where the model files are stored to all users who need to deploy the model.
- B. Store the model outside of Snowflake managed registry and use external authentication to control access.
- C. Grant the 'READ' privilege on the container registry to all users who need to deploy the model. Create a custom role with the 'APPLY MASKING POLICY' privilege and grant this role to the deployment team.
- D. Grant the 'USAGE' privilege on the database and schema containing the model registry, grant the 'READ' privilege on the registry itself, and grant the 'EXECUTE TASK' privilege to the deployment team for the deployment task.
- E. Create a custom role, grant the 'USAGE' privilege on the database and schema containing the model registry, grant the 'READ' privilege on the registry, and then grant this custom role to only those users authorized to deploy the model. Consider masking sensitive model parameters using masking policies.

Answer: E

Explanation:

Option D is the correct answer because it provides the most secure and granular access control. 'USAGE' on the database and schema allows access to the container registry. 'READ' on the registry allows viewing of model metadata without modification. Creating a custom role and granting it to specific users limits access to only authorized personnel. Utilizing masking policies further secures sensitive parameters. Option A is incorrect because it does not control access to the registry itself. 'USAGE' privilege on a stage alone is insufficient for managing model registry access. Option B is incorrect because 'APPLY MASKING POLICY' is not relevant for controlling access to the model registry. Option C is partially correct, but 'EXECUTE TASK' grants unnecessary privileges related to task execution, which is beyond the scope of registry access. It also lacks fine-grained control over who can

deploy. Option E is incorrect because while it offers security, it bypasses the advantages of using Snowflake's managed registry.

NEW QUESTION # 70

You are preparing a dataset in Snowflake for a K-means clustering algorithm. The dataset includes features like 'age', 'income' (in USD), and 'number of transactions'. 'Income' has significantly larger values than 'age' and 'number of transactions'. To ensure that all features contribute equally to the distance calculations in K-means, which of the following scaling approaches should you consider, and why? Select all that apply:

- A. Do not scale the data, as K-means is robust to differences in feature scales.
- B. Apply PowerTransformer to transform income and StandardScaler to other features to handle skewness.
- C. Apply RobustScaler to handle outliers and then StandardScaler or MinMaxScaler to further scale the features.
- D. Apply MinMaxScaler to all three features to scale them to a range between 0 and 1 .
- E. Apply StandardScaler to all three features ('age', 'income', 'number_of_transactions') to center the data around zero and scale it to unit variance.

Answer: C,D,E

Explanation:

K-means clustering is sensitive to the scale of the features because it relies on distance calculations. Features with larger values will have a disproportionate influence on the clustering results. StandardScaler centers the data around zero and scales it to unit variance, which ensures that all features have a similar range and variance. MinMaxScaler scales the features to a range between 0 and 1, which also addresses the issue of different scales. RobustScaler handles outliers which will then use the other two scaling techniques. Therefore A, B and D are the appropriate scaling techniques. C is not correct as K-means relies on distance calculations and not scaling the data could give some feature a larger weight which isn't the desired outcome. Option E: Using PowerTransformer on 'income' to reduce skewness and StandardScaler on the other features can be a valid approach, but it depends on the distribution of 'income' and the presence of outliers. If 'income' is highly skewed and/or contains outliers, this combination might be more effective than using StandardScaler or MinMaxScaler alone.

NEW QUESTION # 71

You have trained a complex Random Forest model in Snowflake to predict loan default risk. You wish to understand the individual and combined effects of 'credit_score' and 'debt_to_income_ratio' on the predicted probability of default. Which approach is MOST suitable for visualizing and interpreting these relationships?

- A. Create a two-way Partial Dependence Plot (PDP) showing the interaction between 'credit_score' and 'debt_to_income_ratio'.
- B. Fit a simpler linear model (e.g., Logistic Regression) to the data and interpret its coefficients.
- C. Calculate feature importance using SNOWFLAKE.ML.FEATURE_IMPORTANCE and focus on the features with the highest scores.
- D. Generate individual Partial Dependence Plots (PDPs) for 'credit_score' and 'debt_to_income_ratio'.
- E. Examine the model's overall accuracy (e.g., AUC) and assume the relationships are well-represented.

Answer: A

Explanation:

The correct answer is C. While individual PDPs (option B) provide insights into the individual effects of each feature, a two-way PDP specifically visualizes and helps interpret the interaction between 'credit_score' and 'debt_to_income_ratio'. This is crucial for understanding how the combined effect of these features influences the predicted probability of default. Feature importance (option A) indicates feature relevance but doesn't show the nature of the relationship. Simplifying the model (option D) sacrifices the complexity captured by the Random Forest. Overall accuracy (option E) doesn't provide specific insights into feature relationships.

NEW QUESTION # 72

You are building a fraud detection model for an e-commerce platform. One of the features is 'purchase_amount', which ranges from \$1 to \$10,000. The data has a skewed distribution with many small purchases and a few very large ones. You need to normalize this feature for your model, which uses gradient descent. Which normalization technique(s) would be most suitable in Snowflake, considering the data characteristics and the need to handle potential future outliers?

- A. Z-score standardization using the following SQL:

```
(purchase_amount - AVG(purchase_amount) OVER ()) / STDEV(purchase_amount) OVER (
```

- B. Unit Vector normalization (L2 Normalization) using SQL:

```
purchase_amount / SQRT(SUM(purchase_amount * purchase_amount) OVER (
```

- C. Robust scaling using interquartile range (IQR) in a stored procedure with Python:

```
:RETURNS STRING
:LANGUAGE PYTHON
:RUNTIME_VERSION = '3.8'
:PACKAGES = ('snowflake-snowpark-python')
:HANDLER = 'main'
:AS $$

import snowflake.snowpark.functions as f
from snowflake.snowpark import Session

def main(session: Session):
    df = session.table(table_name)
    q1 = df.approx_quantile(column_name, 0.25)
    q3 = df.approx_quantile(column_name, 0.75)
    iqr = q3 - q1
    df = df.with_column(column_name + '_scaled', (f.col(column_name) - q1) / iqr)
    df.write.mode('overwrite').save_as_table(table_name + '_scaled')
    return 'Robust scaling complete'
:$
```

- D. Power Transformer (e.g., Yeo-Johnson) implemented with Snowpark Python:

```
CREATE OR REPLACE PROCEDURE power_transform(table_name STRING, column_name STRING)
RETURNS STRING
LANGUAGE PYTHON
RUNTIME_VERSION = '3.8'
PACKAGES = ('snowflake-snowpark-python', 'scikit-learn')
HANDLER = 'main'
AS $$

import snowflake.snowpark.functions as f
from snowflake.snowpark import Session
from sklearn.preprocessing import PowerTransformer
import pandas as pd

def main(session: Session):
    df = session.table(table_name)
    # Fetch data to Pandas DataFrame for transformation
    pdf = df.select(column_name).to_pandas()

    pt = PowerTransformer(method='yeo-johnson', standardize=False)
    transformed_data = pt.fit_transform(pdf)
    # Create a new Snowpark DataFrame from the transformed data
    transformed_df = session.create_dataframe(transformed_data, schema=[column_name + '_transformed'])

    # Combine transformed data back with the original data
    combined_df = df.select(f.col('')).cross_join(transformed_df)

    # Write the combined dataframe back to a new table
    combined_df.write.mode('overwrite').save_as_table(table_name + '_transformed')

    return f'Power Transformation complete for {column_name}'
;$
```

- E. Min-Max scaling using the following SQL:

```
(purchase_amount - MIN(purchase_amount) OVER ()) \ (MAX(purchase_amount) OVER () - MIN(purchase_amount) OVER (
```

Answer: C,D

Explanation:

Options C and D are the most suitable. Robust scaling (C) is effective because it uses the IQR, making it less sensitive to outliers compared to Min-Max scaling (A) or Z-score standardization (B). The Snowflake UDF handles potential outliers by not being

myportal.utt.edu.tt, myportal.utt.edu.tt, epsf-eg.com, www.stes.tyc.edu.tw, lms.ait.edu.za, pct.edu.pk, 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, Disposable vapes

DOWNLOAD the newest TestkingPass DSA-C03 PDF dumps from Cloud Storage for free: <https://drive.google.com/open?id=1fzJPHj3JNlpVstH1EjBc64a5KPlsR5O0>