VCE & PDF
Geekcert.com

DSA-C02^{Q&As}

SnowPro Advanced: Data Scientist Certification

Pass Snowflake DSA-C02 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.geekcert.com/dsa-c02.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Snowflake
Official Exam Center

⚙ **Instant Download** After Purchase

⚙ **100% Money Back** Guarantee

⚙ **365 Days** Free Update

⚙ **800,000+** Satisfied Customers

SATISFACTION GUARANTEED
100%
SATISFACTION GUARANTEED

QUESTION 1

Data Scientist can query, process, and transform data in a which of the following ways using Snowpark Python. Choose 2.

A. Query and process data with a DataFrame object.

B. Write a user-defined tabular function (UDTF) that processes data and returns data in a set of rows with one or more columns.

C. SnowPark currently do not support writing UDTF.

D. Transform Data using DataIKY tool with SnowPark API.

Correct Answer: AC

Explanation:

Query and process data with a DataFrame object. Refer to Working with DataFrames in Snowpark Python.

Convert custom lambdas and functions to user-defined functions(UDFs) that you can call to process data.

Write a user-defined tabular function (UDTF) that processes data and returns data in a set of rows with one or more columns.

Write a stored procedure that you can call to process data, or automate with a task to build a data pipeline.

QUESTION 2

Consider a data frame df with 10 rows and index [\\'r1\\', \\'r2\\', \\'r3\\', \\'row4\\', \\'row5\\', \\'row6\\', \\'r7\\', \\'r8\\', \\'r9\\', \\'row10\\']. What does the expression g = df.groupby(df.index.str.len()) do?

A. Groups df based on index values

B. Groups df based on length of each index value

C. Groups df based on index strings

D. Data frames cannot be grouped by index values. Hence it results in Error.

Correct Answer: D

Explanation: Data frames cannot be grouped by index values. Hence it results in Error.

QUESTION 3

Which of the following metrics are used to evaluate classification models?

A. Area under the ROC curve

B. F1 score

C. Confusion matrix

D. All of the above

Correct Answer: D

Explanation:

Evaluation metrics are tied to machine learning tasks. There are different metrics for the tasks of classification and regression. Some metrics, like precision-recall, are useful for multiple tasks. Classification and regression are examples of

supervised learning, which constitutes a majority of machine learning applications. Using different metrics for performance evaluation, we should be able to im-prove our model\'s overall predictive power before we roll it out for production on

unseen data. Without doing a proper evaluation of the Machine Learning model by using different evaluation metrics, and only depending on accuracy, can lead to a problemwhen the respective model is deployed on unseen data and may

end in poor predictions.

Classification metrics are evaluation measures used to assess the performance of a classification model. Common metrics include accuracy (proportion of correct predictions), precision (true positives over total predicted positives), recall (true

positives over total actual positives), F1 score (har-monic mean of precision and recall), and area under the receiver operating characteristic curve (AUC-ROC).

Confusion Matrix

Confusion Matrix is a performance measurement for the machine learning classification problems where the output can be two or more classes. It is a table with combinations of predicted and actual values.

It is extremely useful for measuring the Recall, Precision, Accuracy, and AUC-ROC curves.

The four commonly used metrics for evaluating classifier performance are:

1.

 Accuracy: The proportion of correct predictions out of the total predictions.

2.

 Precision: The proportion of true positive predictions out of the total positive predictions (precision = true positives / (true positives + false positives)).

3.

 Recall (Sensitivity or True Positive Rate): The proportion of true positive predictions out of the total actual positive instances (recall = true positives / (true positives + false negatives)).

4.

 F1 Score: The harmonic mean of precision and recall, providing a balance between the two metrics (F1 score = 2 * ((precision * recall) / (precision + recall))). These metrics help assess the classifier\'s effectiveness in correctly classifying

instances of different classes.

Understanding how well a machine learning model will perform on unseen data is the main purpose behind working with these evaluation metrics. Metrics like accuracy, precision, recall are good ways to evaluate classification models for

balanced datasets, but if the data is imbalanced then other methods like ROC/AUC perform better in evaluating the model performance.

ROC curve isn\\'t just a single number but it\\'s a whole curve that provides nuanced details about the behavior of the classifier. It is also hard to quickly compare many ROC curves to each other.

QUESTION 4

Which tools helps data scientist to manage ML lifecycle and Model versioning? Choose 2.

A. MLFlow

B. Pachyderm

C. Albert

D. CRUX

Correct Answer: AB

Explanation:

Model versioning in a way involves tracking the changes made to an ML model that has been previously built. Put differently, it is the process of making changes to the configurations of an ML Model. From another perspective, we can see

model versioning as a feature that helps Machine Learning Engineers, Data Scientists, and related personnel create and keep multiple versions of the same model. Think of it as a way of taking notes of the changes you make to the model

through tweaking hyperparameters, retraining the model with more data, and so on. In model versioning, a number of things need to be versioned, to help us keep track of important changes. I\\'ll list and explain them below:

Implementation code: From the early days of model building to optimization stages, code or in this case source code of the model plays an important role. This code experiences significant changes during optimization stages which can easily

be lost if not tracked properly. Because of this, code is one of the things that are taken into consideration during the model versioning process.

Data: In some cases, training data does improve significantly from its initial state during model op-timization phases. This can be as a result of engineering new features from existing ones to train our model on. Also there is metadata (data

about your training data and model) to consider versioning. Metadata can change different times over without the training data actually changing. We need to be able to track these changes through versioning

Model: The model is a product of the two previous entities and as stated in their explanations, an ML model changes at different points of the optimization phases through hyperparameter setting, model artifacts and learning coefficients.

Versioning helps take record of the different versions of a Machine Learning model. MLFlow and Pachyderm are the tools used to manage ML lifecycle and Model versioning.

QUESTION 5

Which object records data manipulation language (DML) changes made to tables, including inserts, updates, and deletes, as well as metadata about each change, so that actions can be taken using the changed data of Data Science Pipelines?

A. Task

B. Dynamic tables

C. Stream

D. Tags

E. Delta

F. OFFSET

Correct Answer: C

Explanation: A stream object records data manipulation language (DML) changes made to tables, including inserts, updates, and deletes, as well as metadata about each change, so that actions can be taken using the changed data. This process is referred to as change data capture (CDC). An individual table stream tracks the changes made to rows in a source table. A table stream (also referred to as simply a "stream") makes a "change table" available of what changed, at therow level, between two transactional points of time in a table. This allows querying and consuming a sequence of change records in a transactional fashion. Streams can be created to query change data on the following objects: Standard tables, including shared tables. Views, including secure views Directory tables Event tables

[DSA-C02 VCE Dumps](link)   [DSA-C02 Practice Test](link)   [DSA-C02 Exam Questions](link)