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.leads4pass.com/dsa-c02.html

100% Passing Guarantee
100% Money Back Assurance

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

QUESTION 1

Skewness of Normal distribution is _____

A. Negative

B. Positive

C. 0

D. Undefined

Correct Answer: C

Explanation:

Since the normal curve is symmetric about its mean, its skewness is zero. This is a theoretical explanation for mathematical proofs, you can refer to books or websites that speak on the same in detail.

QUESTION 2

Which of the following Snowflake parameter can be used to Automatically Suspend Tasks which are running Data science pipelines after specified Failed Runs?

A. SUSPEND_TASK

B. SUSPEND_TASK_AUTO_NUM_FAILURES

C. SUSPEND_TASK_AFTER_NUM_FAILURES

D. There is none as such available.

Correct Answer: C

Explanation:

Automatically Suspend Tasks After Failed Runs

Optionally suspend tasks automatically after a specified number of consecutive runs that either fail or time out. This feature can reduce costs by suspending tasks that consume Snowflake credits but fail to run to completion. Failed task runs

include runs in which the SQL code in the task body either produces a user error or times out. Task runs that are skipped, canceled, or that fail due to a sys-tem error are considered indeterminate and are not included in the count of failed

task runs.

Set the SUSPEND_TASK_AFTER_NUM_FAILURES = num parameter on a standalone task or the root task in a DAG. When the parameter is set to a value greater than 0, the following behavior applies to runs of the standalone task or DAG:

Standalone tasks are automatically suspended after the specified number of consecutive task runs either fail or time

out.

The root task is automatically suspended after the run of any single task in a DAG fails or times out the specified number of times in consecutive runs. The parameter can be set when creating a task (using CREATE TASK) or later (using

ALTER TASK). The setting applies to tasks that rely on either Snowflake-managed compute resources (i.e. serverless compute model) or user-managed compute resources (i.e. a virtual warehouse).

The SUSPEND_TASK_AFTER_NUM_FAILURES parameter can also be set at the account, database, or schema level. The setting applies to all standalone or root tasks contained in the modified object. Note that explicitly setting the

parameter at a lower (i.e. more granular) level overrides the parameter value set at a higher level.

QUESTION 3

In a simple linear regression model (One independent variable), If we change the input variable by 1 unit. How much output variable will change?

A. by 1

B. no change

C. by intercept

D. by its slope

Correct Answer: D

Explanation:

What is linear regression?

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable\\'s value is

called the independent variable.

Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatoryvariable, and the other is considered to be a dependent variable. For

example, a modeler might want to relate the weights of individuals to their heights using a linear regression model. A linear regression line has an equation of the form Y = a + bX, where X is the explanatory variable and Y is the dependent

variable. The slope of the line is b, and a is the intercept (the value of y when x = 0).

For linear regression Y=a+bx+error.

If neglect error then Y=a+bx. If x increases by 1, then Y = a+b(x+1) which implies Y=a+bx+b. So Y increases by its slope.

For linear regression Y=a+bx+error. If neglect error then Y=a+bx. If x increases by 1, then Y = a+b(x+1) which implies Y=a+bx+b. So Y increases by its slope.

QUESTION 4

Which of the following cross validation versions is suitable quicker cross-validation for very large datasets with hundreds of thousands of samples?

A. k-fold cross-validation

B. Leave-one-out cross-validation

C. Holdout method

D. All of the above

Correct Answer: C

Explanation:

Holdout cross-validation method is suitable for very large dataset because it is the simplest and quicker to compute version of cross-validation.

Holdout method

In this method, the dataset is divided into two sets namely the training and the test set with the basic property that the training set is bigger than the test set. Later, the model is trained on the training dataset and evaluated using the test

dataset.

QUESTION 5

All Snowpark ML modeling and preprocessing classes are in the _____ namespace?

A. snowpark.ml.modeling

B. snowflake.sklearn.modeling

C. snowflake.scikit.modeling

D. snowflake.ml.modeling

Correct Answer: D

Explanation:

All Snowpark ML modeling and preprocessing classes are in the snowflake.ml.modeling namespace. The Snowpark ML modules have the same name as the corresponding module from the sklearn namespace. For example, the Snowpark

ML module corresponding to sklearn.calibration is snow-flake.ml.modeling.calibration. The xgboost and lightgbm modules correspond to snowflake.ml.modeling.xgboost and snow-flake.ml.modeling.lightgbm, respectively.

Not all of the classes from scikit-learn are supported in Snowpark ML.

QUESTION 6

Which Python method can be used to Remove duplicates by Data scientist?

A. remove_duplicates()

B. duplicates()

C. drop_duplicates()

D. clean_duplicates()

Correct Answer: D

Explanation:

The drop_duplicates() method removes duplicate rows. dataframe.drop_duplicates(subset, keep, inplace, ignore_index)
Remove duplicate rows from the DataFrame:

1.import pandas as pd

2.data = {

3."name": ["Peter", "Mary", "John", "Mary"],

4."age": [50, 40, 30, 40],

5."qualified": [True, False, False, False]

6.}

7.

8.df = pd.DataFrame(data)

9.newdf = df.drop_duplicates()

QUESTION 7

Select the Data Science Tools which are known to provide native connectivity to Snowflake?

A. Denodo

B. DvSUM

C. DiYotta

D. HEX

Correct Answer: D

Explanation:

Hex -- collaborative data science and analytics platform Denodo -- data virtualization and federation platform DvSum -- data catalog and data intelligence platform Diyotta -- data integration and migration

QUESTION 8

Which ones are the known limitations of using External function? Choose all apply.

A. Currently, external functions cannot be shared with data consumers via Secure Data Sharing.

B. Currently, external functions must be scalar functions. A scalar external function re-turns a single value for each input row.

C. External functions have more overhead than internal functions (both built-in functions and internal UDFs) and usually execute more slowly

D. An external function accessed through an AWS API Gateway private endpoint can be accessed only from a Snowflake VPC (Virtual Private Cloud) on AWS and in the same AWS region.

Correct Answer: ABCD

QUESTION 9

What Can Snowflake Data Scientist do in the Snowflake Marketplace as Consumer? Choose all apply.

A. Discover and test third-party data sources.

B. Receive frictionless access to raw data products from vendors.

C. Combine new datasets with your existing data in Snowflake to derive new business in- sights.

D. Use the business intelligence (BI)/ML/Deep learning tools of her choice.

Correct Answer: ABCD

Explanation:

As a consumer, you can do the following:

Discover and test third-party data sources.

Receive frictionless access to raw data products from vendors.

Combine new datasets with your existing data in Snowflake to derive new business insights.

Have datasets available instantly and updated continually for users.

Eliminate the costs of building and maintaining various APIs and data pipelines to load and up-date data.

Use the business intelligence (BI) tools of your choice.

QUESTION 10

Which of the learning methodology applies conditional probability of all the variables with respec-tive the dependent variable?

A. Reinforcement learning

B. Unsupervised learning

C. Artificial learning

D. Supervised learning

Correct Answer: A

Explanation:

Supervised learning methodology applies conditional probability of all the variables with respective the dependent variable and generally conditional probability of variables is nothing but a basic method of estimating the statistics for few

random experiments. Conditional probability is thus the likelihood of an event or outcome occurring based on the occurrence of some other event or prior outcome. Two events are said tobe independent if one event occurring does not affect

the probability that the other event will occur.

QUESTION 11

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

QUESTION 12

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 13

Data providers add Snowflake objects (databases, schemas, tables, secure views, etc.) to a share us-ing. Which of the following options? Choose 2.

A. Grant privileges on objects to a share via Account role.

B. Grant privileges on objects directly to a share.

C. Grant privileges on objects to a share via a database role.

D. Grant privileges on objects to a share via a third-party role.

Correct Answer: BC

Explanation:What is a Share?

Shares are named Snowflake objects that encapsulate all of the information required to share a database.

Data providers add Snowflake objects (databases, schemas, tables, secure views, etc.) to a share using either or both of the following options:

Option 1: Grant privileges on objects to a share via a database role. Option 2: Grant privileges on objects directly to a share. You choose which accounts can consume data from the share by adding the accounts to the share.

After a database is created (in a consumer account) from a share, all the shared objects are accessible to users in the consumer account. Shares are secure, configurable, and controlled completely by the provider account:

New objects added to a share become immediately available to all consumers, providing real-time access to shared data.

Access to a share (or any of the objects in a share) can be revoked at any time.

QUESTION 14

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 15

Mark the correct steps for saving the contents of a DataFrame to aSnowflake table as part of Moving Data from Spark to Snowflake?

A. Step 1.Use the PUT() method of the DataFrame to construct a DataFrameWriter. Step 2.Specify SNOWFLAKE_SOURCE_NAME using the NAME() method. Step 3.Use the dbtable option to specify the table to which data is written. Step 4.Specify the connector options using either the option() or options() method. Step 5.Use the save() method to specify the save mode for the content.

B. Step 1.Use the PUT() method of the DataFrame to construct a DataFrameWriter. Step 2.Specify SNOWFLAKE_SOURCE_NAME using the format() method. Step 3.Specify the connector options using either the option() or options() method. Step 4.Use the dbtable option to specify the table to which data is written. Step 5.Use the save() method to specify the save mode for the content.

C. Step 1.Use the write() method of the DataFrame to construct a DataFrameWriter. Step 2.Specify SNOWFLAKE_SOURCE_NAME using the format() method. Step 3.Specify the connector options using either the option() or options() method. Step 4.Use the dbtable option to specify the table to which data is written. Step 5.Use the mode() method to specify the save mode for the content.

D. Step 1.Use the writer() method of the DataFrame to construct a DataFrameWriter. Step 2.Specify SNOWFLAKE_SOURCE_NAME using the format() method. Step 3.Use the dbtable option to specify the table to which data is written. Step 4.Specify the connector options using either the option() or options() method. Step 5.Use the save() method to specify the save mode for the content.

Correct Answer: C

Explanation:

Moving Data from Spark to Snowflake

The steps for saving the contents of a DataFrame to a Snowflake table are similar to writing from Snowflake to Spark:

1.

Use the write() method of the DataFrame to construct a DataFrameWriter.

2.

Specify SNOWFLAKE_SOURCE_NAME using the format() method.

3.

Specify the connector options using either the option() or options() method.

4.

Use the dbtable option to specify the table to which data is written.

5.

Use the mode() method to specify the save mode for the content.

Examples

1.df.write

2..format(SNOWFLAKE_SOURCE_NAME)

3..options(sfOptions)

4..option("dbtable", "t2")

5..mode(SaveMode.Overwrite)

6..save()