SnowFlake

DSA-C02

SnowPro Advanced: Data Scientist

Question: 145

Which ones are the type of visualization used for Data exploration in Data Science?

A. Heat Maps

B. Newton AI

C. Feature Distribution by Class

D. 2D-Density Plots

E. Sand Visualization

- wrong

Answer: A

Explanation:

For data exploration in Data Science, visualizations are used to understand the underlying patterns, relationships, anomalies, and distributions in the data. Among the given options:

A. Heat Maps - Correct. Heat maps are used to represent data values using colors. They can visualize the distribution of values across two dimensions, making them useful for understanding correlations or patterns.

C. Feature Distribution by Class - Correct. This visualization displays the distribution of a feature (or multiple features) for different classes. It can be helpful in classification problems to understand how features vary across classes.

D. 2D-Density Plots - Correct. These plots are a way to represent the distribution of two continuous variables, usually

with contour lines or color-coded regions. It helps in visualizing where data points are concentrated.

B. Newton AI - Incorrect. Newton AI is not a type of visualization. It might refer to an AI platform or product, but it's not a visualization method used in data exploration.

E. Sand Visualization - Incorrect. This isn't a recognized standard type of visualization in Data Science.

So, the correct answers are A, C, and D.

Question: 146

Which one is not the feature engineering techniques used in ML data science world?

A. Imputation
B. Binning
C. One hot encoding
D. Statistical

Answer: D

Explanation:

Feature engineering is the pre-processing step of machine learning, which is used to transform raw data into features that can be used for creating a predictive model using Machine learning or statistical Modelling.

What is a feature?

Generally, all machine learning algorithms take input data to generate the output. The input data re-mains in a tabular form consisting of rows (instances or observations) and columns (variable or at-tributes), and these attributes are often known as features. For example, an image is an instance in computer vision, but a line in the image could be the feature. Similarly, in NLP, a document can be an observation, and the word count could be the feature. So, we can say a feature is an attribute that impacts a problem or is useful for the problem.

What is Feature Engineering?

Feature engineering is the pre-processing step of machine learning, which extracts features from raw data. It helps to represent an underlying problem to predictive models in a better way, which as a result, improve the accuracy of the model for unseen data. The predictive model contains predictor variables and an outcome variable, and while the feature engineering process selects the most useful predictor variables for the model.

Some of the popular feature engineering techniques include:

Question: 147

Skewness of Normal distribution is _____

A. Negative
B. Positive
C. 0
D. Undefined

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: 148

What is the formula for measuring skewness in a dataset?

A. MEAN - MEDIAN
B. MODE - MEDIAN
C. (3(MEAN - MEDIAN))/ STANDARD DEVIATION
D. (MEAN - MODE)/ STANDARD DEVIATION

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: 149

What Can Snowflake Data Scientist do in the Snowflake Marketplace as Provider?

A. Publish listings for free-to-use datasets to generate interest and new opportunities among the Snowflake customer base.
B. Publish listings for datasets that can be customized for the consumer.
C. Share live datasets securely and in real-time without creating copies of the data or im-posing data integration tasks on the consumer.
D. Eliminate the costs of building and maintaining APIs and data pipelines to deliver data to customers.

Answer: A,B,C,D

Explanation:

All are correct!

About the Snowflake Marketplace

You can use the Snowflake Marketplace to discover and access third-party data and services, as well as market your own data products across the Snowflake Data Cloud.

As a data provider, you can use listings on the Snowflake Marketplace to share curated data offerings with many consumers simultaneously, rather than maintain sharing relationships with each individual consumer. With Paid Listings, you can also charge for your data products.

As a consumer, you might use the data provided on the Snowflake Marketplace to explore and access the following:

Historical data for research, forecasting, and machine learning.

Up-to-date streaming data, such as current weather and traffic conditions.

Specialized identity data for understanding subscribers and audience targets.

New insights from unexpected sources of data.

The Snowflake Marketplace is available globally to all non-VPS Snowflake accounts hosted on Amazon Web Services, Google Cloud Platform, and Microsoft Azure, with the exception of Microsoft Azure Government. Support for Microsoft Azure Government is planned.

Question: 150

What Can Snowflake Data Scientist do in the Snowflake Marketplace as Consumer?

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.

Answer: A,B,C,D

Explanation:

As a consumer, you can do the following:

ã» Discover and test third-party data sources.

ã» Receive frictionless access to raw data products from vendors.

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

ã» Have datasets available instantly and updated continually for users.

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

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

Question: 151

Which one is the incorrect option to share data in Snowflake?

A. a Listing, in which you offer a share and additional metadata as a data product to one or more accounts.
B. a Direct Marketplace, in which you directly share specific database objects (a share) to another account in your region using Snowflake Marketplace.
C. a Direct Share, in which you directly share specific database objects (a share) to another account in your region.
D. a Data Exchange, in which you set up and manage a group of accounts and offer a share to that group.

Answer: B

Explanation:

Options for Sharing in Snowflake

You can share data in Snowflake using one of the following options:

ã» a Listing, in which you offer a share and additional metadata as a data product to one or more ac-counts,

ã» a Direct Share, in which you directly share specific database objects (a share) to another account in your region,

ã» a Data Exchange, in which you set up and manage a group of accounts and offer a share to that group.

Question: 152

Data providers add Snowflake objects (databases, schemas, tables, secure views, etc.) to a share using.

Which of the following options?

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.

Answer: A,B,C

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:

ã» 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: 153

Secure Data Sharing do not let you share which of the following selected objects in a database in your account with other Snowflake accounts?

A. Sequences
B. Tables
C. External tables
D. Secure UDFs

Answer: A

Explanation:

Secure Data Sharing lets you share selected objects in a database in your account with other Snow-flake accounts.

You can share the following Snowflake database objects:

Tables

External tables

Secure views

Secure materialized views

Secure UDFs

Snowflake enables the sharing of databases through shares, which are created by data providers and âimportedâ by data consumers.

Question: 154

Which one is incorrect understanding about Providers of Direct share?

A. A data provider is any Snowflake account that creates shares and makes them available to other Snowflake accounts to consume.
B. As a data provider, you share a database with one or more Snowflake accounts.
C. You can create as many shares as you want, and add as many accounts to a share as you want.
D. If you want to provide a share to many accounts, you can do the same via Direct Share.

Answer: D

Explanation:

If you want to provide a share to many accounts, you might want to use a listing or a data ex-change.

Question: 155

As Data Scientist looking out to use Reader account, Which ones are the correct considerations about Reader Accounts for Third-Party Access?

A. Reader accounts (formerly known as âread-only accountsâ) provide a quick, easy, and cost-effective way to share data without requiring the consumer to become a Snowflake customer.

B. Each reader account belongs to the provider account that created it.

C. Users in a reader account can query data that has been shared with the reader account, but cannot perform any of the DML tasks that are allowed in a full account, such as data loading, insert, update, and similar data manipulation operations.

D. Data sharing is only possible between Snowflake accounts.

- wrong

Answer: D

Explanation:

Let's evaluate each of the statements regarding Reader Accounts for Third-Party Access in Snowflake:

A. Correct. Reader accounts (formerly known as â read-only accounts â) are specifically designed for sharing data with third-party consumers without requiring them to be Snowflake customers. They allow consumers to query shared data sets.

B. Correct. Each reader account is indeed linked to the provider account that created it. The provider manages and incurs the costs of the reader account.

C. Correct. Users in a reader account can only query the shared data. They don't have permissions to perform DML operations (like insert, update, delete). Their primary purpose is to read and analyze the shared data.

D. Incorrect. While it's true that typical data sharing in Snowflake occurs between Snowflake accounts, the concept of the reader account was introduced specifically to share data with third parties that do not have a Snowflake account.

So, the correct considerations about Reader Accounts for Third-Party Access are options A, B, and C.

Question: 156

SHOW GRANTS OF SHARE product_s;

A. GRANT USAGE ON DATABASE product_db TO SHARE product_s;
B. CREATE DIRECT SHARE product_s;
C. GRANT SELECT ON TABLE sales_db. product_agg.Item_agg TO SHARE product_s;
D. ALTER SHARE product_s ADD ACCOUNTS=xy12345, yz23456;

Answer: C

Explanation:

CREATE SHARE product_s is the correct Snowsql command to create Share object.

Rest are correct ones.

Question: 157

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

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 âstreamâ) makes a âchange tableâ available of what changed, at the row 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:

ã» Standard tables, including shared tables.

ã» Views, including secure views

ã» Directory tables

ã» Event tables

Question: 158

Which are the following additional Metadata columns Stream contains that could be used for creating Efficient Data science Pipelines & helps in transforming only the New/Modified data only?

A. METADATA$ACTION
B. METADATA$FILE_ID
C. METADATA$ISUPDATE
D. METADATA$DELETE
E. METADATA$ROW_ID

Answer: A,C,E

Explanation:

A stream stores an offset for the source object and not any actual table columns or data. When que-ried, a stream accesses and returns the historic data in the same shape as the source object (i.e. the same column names and ordering) with the following additional columns: METADATA$ACTION

Indicates the DML operation (INSERT, DELETE) recorded.

METADATA$ISUPDATE

Indicates whether the operation was part of an UPDATE statement. Updates to rows in the source object are represented as a pair of DELETE and INSERT records in the stream with a metadata column METADATA$ISUPDATE values set to TRUE.

Note that streams record the differences between two offsets. If a row is added and then updated in the current offset, the delta change is a new row. The METADATA$ISUPDATE row records a FALSE

value.

METADATA$ROW_ID

Specifies the unique and immutable ID for the row, which can be used to track changes to specific rows over time.