

# ITPassLeader

Pass Your Next Certification Exam Fast!

Select a vendor... Select an test... Your email address [Free Download Demo](#)



Instant Download



365 Days Free Updates



Money Back Guarantee



Security & Privacy

Choose the version that fits your needs

PDF Version

Desktop Test Engine

Online Test Engine

Latest and Up-to-Date exam dumps with real exam questions answers.



Get 12-Months free updates without any extra charges.



Experience same exam environment before appearing in the certification exam.



100% exam passing guarantee in the first attempt.



20% discount on more than one license and 30% discount on 5+ license purchases.



100% secure purchase on SSL.



Completely private purchase without sharing your personal info with anyone.



<http://www.itpassleader.com>

High-praise Exam Dumps Questions grant you success by high pass rate - ITPassLeader

**Exam** : **DSA-C02**

**Title** : SnowPro Advanced: Data  
Scientist Certification Exam

**Vendor** : Snowflake

**Version** : DEMO

**NO.1** Data Scientist can query, process, and transform data in a which of the following ways using Snowpark Python. [Select 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 DataKY tool with SnowPark API.

**Answer:** A C

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.

**NO.2** Which tools helps data scientist to manage ML lifecycle & Model versioning?

- A.** MLFlow
- B.** Pachyderm
- C.** Albert
- D.** CRUX

**Answer:** A B

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 optimization 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 & Pachyderm are the tools used to manage ML lifecycle & Model versioning.

**NO.3** 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

**Answer:** A D E

Explanation

Type of visualization used for exploration:

Correlation heatmap

Class distributions by feature

Two-Dimensional density plots.

All the visualizations are interactive, as is standard for Plotly.

For More details, please refer the below link:

<https://towardsdatascience.com/data-exploration-understanding-and-visualization-72657f5eac41>

**NO.4** Select the Correct Statements regarding Normalization?

- A. Normalization technique uses minimum and max values for scaling of model.
- B. Normalization technique uses mean and standard deviation for scaling of model.
- C. Scikit-Learn provides a transformer RecommendedScaler for Normalization.
- D. Normalization got affected by outliers.

**Answer:** A D

Explanation

Normalization is a scaling technique in Machine Learning applied during data preparation to change the values of numeric columns in the dataset to use a common scale. It is not necessary for all datasets in a model. It is required only when features of machine learning models have different ranges.

Scikit-Learn provides a transformer called MinMaxScaler for Normalization.

This technique uses minimum and max values for scaling of model. It is useful when feature distribution is unknown. It got affected by outliers.

**NO.5** Data Scientist used streams in ELT (extract, load, transform) processes where new data inserted in-to a staging table is tracked by a stream. A set of SQL statements transform and insert the stream contents into a set of production tables. Raw data is coming in the JSON format, but for analysis he needs to transform it into relational columns in the production tables. which of the following Data transformation SQL function he can used to achieve the same?

- A. He could not apply Transformation on Stream table data.
- B. lateral flatten()
- C. METADATA\$ACTION ()
- D. Transpose()

**Answer:** B

Explanation

To know about lateral flatten SQL Function, please refer:

<https://docs.snowflake.com/en/sql-reference/constructs/join-lateral#example-of-using-lateral-with-flatten>

**NO.6** Which method is used for detecting data outliers in Machine learning?

- A. Scaler
- B. Z-Score
- C. BOXI
- D. CMIYC

**Answer:** B

Explanation

What are outliers?

Outliers are the values that look different from the other values in the data. Below is a plot highlighting the outliers in 'red' and outliers can be seen in both the extremes of data.

Reasons for outliers in data

Errors during data entry or a faulty measuring device (a faulty sensor may result in extreme readings).

Natural occurrence (salaries of junior level employees vs C-level employees) Problems caused by outliers Outliers in the data may causes problems during model fitting (esp. linear models).

Outliers may inflate the error metrics which give higher weights to large errors (example, mean squared error, RMSE).

Z-score method is of the method for detecting outliers. This method is generally used when a variable' distribution looks close to Gaussian. Z-score is the number of standard deviations a value of a variable is away from the variable' mean.

$Z\text{-Score} = (X - \text{mean}) / \text{Standard deviation}$

IQR method , Box plots are some more example of methods used to detect data outliers in Data science.

**NO.7** You previously trained a model using a training dataset. You want to detect any data drift in the new data collected since the model was trained.

What should you do?

- A. Create a new dataset using the new data and a timestamp column and create a data drift monitor that uses the training dataset as a baseline and the new dataset as a target.
- B. Create a new version of the dataset using only the new data and retrain the model.
- C. Retrained your training dataset after correcting data outliers & no need to introduce new data.
- D. Add the new data to the existing dataset and enable Application Insights for the service where the model is deployed.

**Answer:** A

Explanation

To track changing data trends, create a data drift monitor that uses the training data as a baseline and the new data as a target.

Model drift and decay are concepts that describe the process during which the performance of a model deployed to production degrades on new, unseen data or the underlying assumptions about the data change.

These are important metrics to track once models are deployed to production. Models must be regularly re-trained on new data. This is referred to as refitting the model. This can be done either on a periodic basis, or, in an ideal scenario, retraining can be triggered when the performance of the model degrades below a certain pre-defined threshold.