



Free Questions for DSA-C02 by ebraindumps

Shared by Irwin on 29-01-2024

For More Free Questions and Preparation Resources

Check the Links on Last Page

Question 1

Question Type: MultipleChoice

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

Options:

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.

(Correct)

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.

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()

Question 2

Question Type: MultipleChoice

Which method is used for detecting data outliers in Machine learning?

Options:

- 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 high-lighting 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-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.

Question 3

Question Type: MultipleChoice

Which is the visual depiction of data through the use of graphs, plots, and informational graphics?

Options:

A- Data Interpretation

B- Data Virtualization

C- Data visualization

D- Data Mining

Answer:

D

Explanation:

Data visualization is the visual depiction of data through the use of graphs, plots, and informational graphics. Its practitioners use statistics and data science to convey the meaning behind data in ethical and accurate ways.

Question 4

Question Type: MultipleChoice

Which of the following is a Python-based web application framework for visualizing data and analyzing results in a more efficient and flexible way?

Options:

- A- StreamBI
- B- Streamlit
- C- Streamsets
- D- Rapter

Answer:

B

Explanation:

Streamlit is a Python-based web application framework for visualizing data and analyzing results in a more efficient and flexible way. It is an open source library that assists data scientists and academics to develop Machine Learning (ML) visualization dashboards in a short period of time. We can build and deploy powerful data applications with just a few lines of code.

Why Streamlit?

Currently, real-world applications are in high demand and developers are developing new libraries and frameworks to make on-the-go dashboards easier to build and deploy. Streamlit is a library that reduces your dashboard development time from days to hours.

Following are some reasons to choose the Streamlit:

It is a free and open-source library.

Installing Streamlit is as simple as installing any other python package

It is easy to learn because you won't need any web development experience, only a basic understanding of Python is enough to build a data application.

It is compatible with almost all machine learning frameworks, including Tensorflow and Pytorch, Scikit-learn, and visualization libraries such as Seaborn, Altair, Plotly, and many others.

Question 5

Question Type: MultipleChoice

To return the contents of a DataFrame as a Pandas DataFrame, Which of the following method can be used in SnowPark API?

Options:

A- REPLACE_TO_PANDAS

B- SNOWPARK_TO_PANDAS

C- CONVERT_TO_PANDAS

D- TO_PANDAS

Answer:

D

Explanation:

To return the contents of a DataFrame as a Pandas DataFrame, use the `to_pandas` method.

For example:

1. `>>> python_df = session.create_dataframe(['a', 'b', 'c'])`

2. `>>> pandas_df = python_df.to_pandas()`

Question 6

Question Type: MultipleChoice

Select the Correct Statements regarding Normalization?

Options:

- 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.

Question 7

Question Type: MultipleChoice

Which one is not Types of Feature Scaling?

Options:

- A- Economy Scaling
- B- Min-Max Scaling
- C- Standard Scaling
- D- Robust Scaling

Answer:

B

Explanation:

Feature Scaling

Feature Scaling is the process of transforming the features so that they have a similar scale. This is important in machine learning because the scale of the features can affect the performance of the model.

Types of Feature Scaling:

Min-Max Scaling: Rescaling the features to a specific range, such as between 0 and 1, by subtracting the minimum value and dividing by the range.

Standard Scaling: Rescaling the features to have a mean of 0 and a standard deviation of 1 by subtracting the mean and dividing by the standard deviation.

Robust Scaling: Rescaling the features to be robust to outliers by dividing them by the interquartile range.

Benefits of Feature Scaling:

Improves Model Performance: By transforming the features to have a similar scale, the model can learn from all features equally and avoid being dominated by a few large features.

Increases Model Robustness: By transforming the features to be robust to outliers, the model can become more robust to anomalies.

Improves Computational Efficiency: Many machine learning algorithms, such as k-nearest neighbors, are sensitive to the scale of the features and perform better with scaled features.

Improves Model Interpretability: By transforming the features to have a similar scale, it can be easier to understand the model's predictions.

Question 8

Question Type: MultipleChoice

Which one is not the types of Feature Engineering Transformation?

Options:

A- Scaling

B- Encoding

C- Aggregation

D- Normalization

Answer:

C

Explanation:

What is Feature Engineering?

Feature engineering is the process of transforming raw data into features that are suitable for machine learning models. In other words, it is the process of selecting, extracting, and transforming the most relevant features from the available data to build more accurate and efficient machine learning models.

The success of machine learning models heavily depends on the quality of the features used to train them. Feature engineering involves a set of techniques that enable us to create new features by combining or transforming the existing ones. These techniques help to highlight the most important patterns and relationships in the data, which in turn helps the machine learning model to learn from the data more effectively.

What is a Feature?

In the context of machine learning, a feature (also known as a variable or attribute) is an individual measurable property or characteristic of a data point that is used as input for a machine learning algorithm. Features can be numerical, categorical, or text-based, and they represent different aspects of the data that are relevant to the problem at hand.

For example, in a dataset of housing prices, features could include the number of bedrooms, the square footage, the location, and the age of the property. In a dataset of customer demographics, features could include age, gender, income level, and occupation.

The choice and quality of features are critical in machine learning, as they can greatly impact the accuracy and performance of the model.

Why do we Engineer Features?

We engineer features to improve the performance of machine learning models by providing them with relevant and informative input data. Raw data may contain noise, irrelevant information, or missing values, which can lead to inaccurate or biased model predictions. By engineering features, we can extract meaningful information from the raw data, create new variables that capture important patterns and relationships, and transform the data into a more suitable format for machine learning algorithms.

Feature engineering can also help in addressing issues such as overfitting, underfitting, and high di-mensionality. For example, by reducing the number of features, we can prevent the model from be-coming too complex or overfitting to the training data. By selecting the most relevant features, we can improve the model's accuracy and interpretability.

In addition, feature engineering is a crucial step in preparing data for analysis and decision-making in various fields, such as finance, healthcare, marketing, and social sciences. It can help uncover hidden insights, identify trends and patterns, and support data-driven decision-making.

We engineer features for various reasons, and some of the main reasons include:

Improve User Experience: The primary reason we engineer features is to enhance the user experience of a product or service. By adding new features, we can make the product more intuitive, efficient, and user-friendly, which can increase user satisfaction and engagement.

Competitive Advantage: Another reason we engineer features is to gain a competitive advantage in the marketplace. By offering unique and innovative features, we can differentiate our product from competitors and attract more customers.

Meet Customer Needs: We engineer features to meet the evolving needs of customers. By analyzing user feedback, market trends, and customer behavior, we can identify areas where new features could enhance the product's value and meet customer needs.

Increase Revenue: Features can also be engineered to generate more revenue. For example, a new feature that streamlines the checkout process can increase sales, or a feature that provides additional functionality could lead to more upsells or cross-sells.

Future-Proofing: Engineering features can also be done to future-proof a product or service. By an-ticipating future trends and potential customer needs, we can develop features that ensure the product remains relevant and useful in the long term.

Processes Involved in Feature Engineering

Feature engineering in Machine learning consists of mainly 5 processes: Feature Creation, Feature Transformation, Feature Extraction, Feature Selection, and Feature Scaling. It is an iterative process that requires experimentation and testing to find the best combination of features for a given problem. The success of a machine learning model largely depends on the quality of the features used in the model.

Feature Transformation

Feature Transformation is the process of transforming the features into a more suitable representation for the machine learning model. This is done to ensure that the model can effectively learn from the data.

Types of Feature Transformation:

Normalization: Rescaling the features to have a similar range, such as between 0 and 1, to prevent some features from dominating others.

Scaling: Rescaling the features to have a similar scale, such as having a standard deviation of 1, to make sure the model considers all features equally.

Encoding: Transforming categorical features into a numerical representation. Examples are one-hot encoding and label encoding.

Transformation: Transforming the features using mathematical operations to change the distribution or scale of the features. Examples are logarithmic, square root, and reciprocal transformations.

To Get Premium Files for DSA-C02 Visit

<https://www.p2pexams.com/products/dsa-c02>

For More Free Questions Visit

<https://www.p2pexams.com/snowflake/pdf/dsa-c02>

