Free Questions for Databricks-Certified-Data-Analyst-Associate by go4braindumps

Shared by Maxwell on 19-01-2024

For More Free Questions and Preparation Resources

Check the Links on Last Page

Question 1

A data analyst is working with gold-layer tables to complete an ad-hoc project. A stakeholder has provided the analyst with an additional dataset that can be used to augment the gold-layer tables already in use.

Which of the following terms is used to describe this data augmentation?

Options:

A- Data testing

B- Ad-hoc improvements

C- Last-mile

D- Last-mile ETL

E- Data enhancement

Answer:

E

Explanation:

Data enhancement is the process of adding or enriching data with additional information to improve its quality, accuracy, and usefulness. Data enhancement can be used to augment existing data sources with new data sources, such as external datasets, synthetic data, or machine learning models. Data enhancement can help data analysts to gain deeper insights, discover new patterns, and solve complex problems.Data enhancement is one of the applications of generative AI, which can leverage machine learning to generate synthetic data for better models or safer data sharing1.

In the context of the question, the data analyst is working with gold-layer tables, which are curated business-level tables that are typically organized in consumption-ready project-specific databases234.The gold-layer tables are the final layer of data transformations and data quality rules in the medallion lakehouse architecture, which is a data design pattern used to logically organize data in a lakehouse2. The stakeholder has provided the analyst with an additional dataset that can be used to augment the gold-layer tables already in use. This means that the analyst can use the additional dataset to enhance the existing gold-layer tables with more information, such as new features, attributes, or metrics. This data augmentation can help the analyst to complete the ad-hoc project more effectively and efficiently.

What is the medallion lakehouse architecture? - Databricks

Data Warehousing Modeling Techniques and Their Implementation on the Databricks Lakehouse Platform | Databricks Blog

What is the medallion lakehouse architecture? - Azure Databricks

What is a Medallion Architecture? - Databricks

Synthetic Data for Better Machine Learning | Databricks Blog

Question 2

In which of the following situations will the mean value and median value of variable be meaningfully different?

Options:

A- When the variable contains no outliers

B- When the variable contains no missing values

C- When the variable is of the boolean type

D- When the variable is of the categorical type

E- When the variable contains a lot of extreme outliers

Answer:

E

Explanation:

The mean value of a variable is the average of all the values in a data set, calculated by dividing the sum of the values by the number of values. The median value of a variable is the middle value of the ordered data set, or the average of the middle two values if the data set

has an even number of values. The mean value is sensitive to outliers, which are values that are very different from the rest of the data. Outliers can skew the mean value and make it less representative of the central tendency of the data. The median value is more robust to outliers, as it only depends on the middle values of the data.Therefore, when the variable contains a lot of extreme outliers, the mean value and the median value will be meaningfully different, as the mean value will be pulled towards the outliers, while the median value will remain close to the majority of the data1.Reference:Difference Between Mean and Median in Statistics (With Example) - BYJU'S

Question 3

Question Type: MultipleChoice

Which of the following statements describes descriptive statistics?

Options:

A- A branch of statistics that uses summary statistics to quantitatively describe and summarize data.

B- A branch of statistics that uses a variety of data analysis techniques to infer properties of an underlying distribution of probability.

C- A branch of statistics that uses quantitative variables that must take on a finite or countably infinite set of values.

D- A branch of statistics that uses summary statistics to categorically describe and summarize data.

E- A branch of statistics that uses quantitative variables that must take on an uncountable set of values.

Answer:

A

Explanation:

Descriptive statistics is a branch of statistics that uses summary statistics, such as mean, median, mode, standard deviation, range, frequency, or correlation, to quantitatively describe and summarize data. Descriptive statistics can help data analysts understand the main features of a data set, such as its central tendency, variability, or distribution. Descriptive statistics can also help data analysts visualize data using charts, graphs, or tables. Descriptive statistics do not make any inferences or predictions about the data, unlike inferential statistics, which use data analysis techniques to infer properties of an underlying population or probability distribution from a sample of data.Reference:Databricks - Descriptive Statistics,Databricks - Data Analysis with Databricks SQL

Question 4

Question Type: **MultipleChoice**

A data organization has a team of engineers developing data pipelines following the medallion architecture using Delta Live Tables. While the data analysis team working on a project is using gold-layer tables from these pipelines, they need to perform some additional processing of these tables prior to performing their analysis.

Which of the following terms is used to describe this type of work?

Options:

A- Data blending

B- Last-mile

C- Data testing

D- Last-mile ETL

E- Data enhancement

Answer:

D

Explanation:

Last-mile ETL is the term used to describe the additional processing of data that is done by data analysts or data scientists after the data has been ingested, transformed, and stored in the lakehouse by data engineers. Last-mile ETL typically involves tasks such as data cleansing, data enrichment, data aggregation, data filtering, or data sampling that are specific to the analysis or machine learning use case. Last-mile ETL can be done using Databricks SQL, Databricks notebooks, or Databricks Machine Learning.Reference:Databricks - Last-mile ETL,Databricks - Data Analysis with Databricks SQL

Question 5

A data team has been given a series of projects by a consultant that need to be implemented in the Databricks Lakehouse Platform.

Which of the following projects should be completed in Databricks SQL?

Options:

A- Testing the quality of data as it is imported from a source

B- Tracking usage of feature variables for machine learning projects

C- Combining two data sources into a single, comprehensive dataset

D- Segmenting customers into like groups using a clustering algorithm

E- Automating complex notebook-based workflows with multiple tasks

Answer:

C

Explanation:

Databricks SQL is a service that allows users to query data in the lakehouse using SQL and create visualizations and dashboards1.One of the common use cases for Databricks SQL is to combine data from different sources and formats into a single, comprehensive dataset that can be used for further analysis or reporting2.For example, a data analyst can use Databricks SQL to join data from a CSV file and a Parquet file, or from a Delta table and a JDBC table, and create a new table or view that contains the combined data3. This can help simplify the data management and governance, as well as improve the data quality and consistency.Reference:

Databricks SQL overview

Databricks SQL use cases

Joining data sources

Question 6

Question Type: MultipleChoice

Which of the following statements about adding visual appeal to visualizations in the Visualization Editor is incorrect?

Options:

A- Visualization scale can be changed.

B- Data Labels can be formatted.

C- Colors can be changed.

D- Borders can be added.

E- Tooltips can be formatted.

Answer:

D

Explanation:

The Visualization Editor in Databricks SQL allows users to create and customize various types of charts and visualizations from the query results. Users can change the visualization type, select the data fields, adjust the colors, format the data labels, and modify the tooltips. However, there is no option to add borders to the visualizations in the Visualization Editor.Borders are not a supported feature of the new chart visualizations in Databricks1. Therefore, the statement that borders can be added is incorrect.Reference:

New chart visualizations in Databricks | Databricks on AWS

Question 7

Question Type: **MultipleChoice**

A data analyst has been asked to configure an alert for a query that returns the income in the accounts_receivable table for a date range. The date range is configurable using a Date query parameter.

The Alert does not work.

Which of the following describes why the Alert does not work?

Options:

A- Alerts don't work with queries that access tables.

B- Queries that return results based on dates cannot be used with Alerts.

C- The wrong query parameter is being used. Alerts only work with Date and Time query parameters.

D- Queries that use query parameters cannot be used with Alerts.

E- The wrong query parameter is being used. Alerts only work with drogdown list query parameters, not dates.

Answer:

D

Explanation:

According to the Databricks documentation1, queries that use query parameters cannot be used with Alerts. This is because Alerts do not support user input or dynamic values. Alerts leverage queries with parameters using the default value specified in the SQL editor for each parameter. Therefore, if the query uses a Date query parameter, the alert will always use the same date range as the default value, regardless of the actual date. This may cause the alert to not work as expected, or to not trigger at all.Reference:

Databricks SQL alerts: This is the official documentation for Databricks SQL alerts, where you can find information about how to create, configure, and monitor alerts, as well as the limitations and best practices for using alerts.

Question 8

Question Type: **MultipleChoice**

An analyst writes a query that contains a query parameter. They then add an area chart visualization to the query. While adding the area chart visualization to a dashboard, the analyst chooses "Dashboard Parameter" for the query parameter associated with the area chart.

Which of the following statements is true?

Options:

A- The area chart will use whatever is selected in the Dashboard Parameter while all or the other visualizations will remain changed regardless of their parameter use.

B- The area chart will use whatever is selected in the Dashboard Parameter along with all of the other visualizations in the dashboard that use the same parameter.

C- The area chart will use whatever value is chosen on the dashboard at the time the area chart is added to the dashboard.

D- The area chart will use whatever value is input by the analyst when the visualization is added to the dashboard. The parameter cannot be changed by the user afterwards.

E- The area chart will convert to a Dashboard Parameter.

Answer:

B

Explanation:

A Dashboard Parameter is a parameter that is configured for one or more visualizations within a dashboard and appears at the top of the dashboard.The parameter values specified for a Dashboard Parameter apply to all visualizations reusing that particular Dashboard Parameter1. Therefore, if the analyst chooses "Dashboard Parameter" for the query parameter associated with the area chart, the area chart will use whatever is selected in the Dashboard Parameter along with all of the other visualizations in the dashboard that use the same parameter.This allows the user to filter the data across multiple visualizations using a single parameter widget2.Reference:Databricks SQL dashboards,Query parameters

Question 9

A data analyst has been asked to produce a visualization that shows the flow of users through a website.

Which of the following is used for visualizing this type of flow?

Options:

A- Heatmap

B- IChoropleth

C- Word Cloud

D- Pivot Table

E- Sankey

Answer:

E

Explanation:

A Sankey diagram is a type of visualization that shows the flow of data between different nodes or categories. It is often used to represent the movement of users through a website, as it can show the paths they take, the sources they come from, the pages they visit, and the outcomes they achieve. A Sankey diagram consists of links and nodes, where the links represent the volume or weight of

the flow, and the nodes represent the stages or steps of the flow. The width of the links is proportional to the amount of flow, and the color of the links can indicate different attributes or segments of the flow. A Sankey diagram can help identify the most common or popular user journeys, the bottlenecks or drop-offs in the flow, and the opportunities for improvement or optimization.Reference: The answer can be verified from Databricks documentation which provides examples and instructions on how to create Sankey diagrams using Databricks SQL Analytics and Databricks Visualizations. Reference links: Databricks SQL Analytics - Sankey Diagram, Databricks Visualizations - Sankey Diagram

Question 10

Question Type: **MultipleChoice**

A data analyst has been asked to provide a list of options on how to share a dashboard with a client. It is a security requirement that the client does not gain access to any other information, resources, or artifacts in the database.

Which of the following approaches cannot be used to share the dashboard and meet the security requirement?

Options:

A- Download the Dashboard as a PDF and share it with the client.

B- Set a refresh schedule for the dashboard and enter the client's email address in the 'Subscribers' box.

C- Take a screenshot of the dashboard and share it with the client.

D- Generate a Personal Access Token that is good for 1 day and share it with the client.

E- Download a PNG file of the visualizations in the dashboard and share them with the client.

Answer:

D

Explanation:

The approach that cannot be used to share the dashboard and meet the security requirement is D. Generating a Personal Access Token that is good for 1 day and sharing it with the client.This approach would give the client access to the Databricks workspace using the token owner's identity and permissions, which could expose other information, resources, or artifacts in the database1. The other approaches can be used to share the dashboard and meet the security requirement because:

A)Downloading the Dashboard as a PDF and sharing it with the client would only provide a static snapshot of the dashboard without any interactive features or access to the underlying data2.

B)Setting a refresh schedule for the dashboard and entering the client's email address in the "Subscribers" box would send the client an email with the latest dashboard results as an attachment or a link to a secure web page3. The client would not be able to access the Databricks workspace or the dashboard itself.

C)Taking a screenshot of the dashboard and sharing it with the client would also only provide a static snapshot of the dashboard without any interactive features or access to the underlying data4.

E)Downloading a PNG file of the visualizations in the dashboard and sharing them with the client would also only provide a static snapshot of the visualizations without any interactive features or access to the underlying data5.Reference:

1: Personal access tokens

2:Download as PDF

3:Automatically refresh a dashboard

4:Take a screenshot

5:Download a PNG file

Question 11

Question Type: **MultipleChoice**

A data analyst has created a Query in Databricks SQL, and now they want to create two data visualizations from that Query and add both of those data visualizations to the same Databricks SQL Dashboard.

Which of the following steps will they need to take when creating and adding both data visualizations to the Databricks SQL Dashboard?

Options:

A- They will need to alter the Query to return two separate sets of results.

B- They will need to add two separate visualizations to the dashboard based on the same Query.

C- They will need to create two separate dashboards.

D- They will need to decide on a single data visualization to add to the dashboard.

E- They will need to copy the Query and create one data visualization per query.

Answer:

B

Explanation:

A data analyst can create multiple visualizations from the same query in Databricks SQL by clicking the + button next to the Results tab and selecting Visualization. Each visualization can have a different type, name, and configuration. To add a visualization to a dashboard, the data analyst can click the vertical ellipsis button beneath the visualization, select + Add to Dashboard, and choose an existing or new dashboard. The data analyst can repeat this process for each visualization they want to add to the same dashboard.Reference:Visualization in Databricks SQL,Visualize queries and create a dashboard in Databricks SQL