DA0-001 Q&As

CompTIA Data+

Pass CompTIA DA0-001 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.certbus.com/da0-001.html

100% Passing Guarantee
100% Money Back Assurance

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

QUESTION 1

Which one of the following programming languages is specifically designed for use in analytics applications?

A. Python.

B. R

C. C++

D. Java.

Correct Answer: B

QUESTION 2

A sales analyst needs to report how the sales team is performing to target. Which of the following files will be important in determining 2019 performance attainment?

A. 2018 goal data

B. 2018 actual revenue

C. 2019 goal data

D. 2019 commission plan

Correct Answer: C

Answer: C. 2019 goal data To report how the sales team is performing to target, the sales analyst needs to compare the actual sales revenue with the expected or planned sales revenue for the same period. The 2019 goal data is the file that contains the expected or planned sales revenue for the year 2019, which is the target that the sales team is aiming to achieve. By comparing the 2019 goal data with the 2019 actual revenue, the sales analyst can calculate the performance attainment, which is the percentage of the goal that was met by the sales team. Option A is incorrect, as 2018 goal data is not relevant for determining 2019 performance attainment. The 2018 goal data contains the expected or planned sales revenue for the year 2018, which is not the target that the sales team is aiming to achieve in 2019. Option B is incorrect, as 2018 actual revenue is not relevant for determining 2019 performance attainment. The 2018 actual revenue contains the actual sales revenue for the year 2018, which is not comparable with the 2019 goal data or the 2019 actual revenue. Option D is incorrect, as 2019 commission plan is not relevant for determining 2019 performance attainment. The 2019 commission plan contains the rules and rates for calculating and paying commissions to the sales team based on their performance attainment, but it does not contain the expected or planned sales revenue for the year 2019.

QUESTION 3

An analyst needs to provide a chart to identify the composition between the categories of the survey response data set:

Favorite color	Responses
Red	15
Blue	35
Green	25
Yellow	25
Total	100

Which of the following charts would be BEST to use?

A. Histogram

B. Pie

C. Line

D. Scatter pot

E. Waterfall

Correct Answer: B

Explanation: A pie chart is the best choice to show the composition between the categories of the survey response data set. A pie chart represents the whole with a circle, divided by slices into parts. Each slice shows the relative size of each category as a percentage of the total. A pie chart is useful when the categories are mutually exclusive and add up to 100%. The table shows the favorite color and the number of responses for each color, which can be easily converted into percentages. A pie chart can show how each color contributes to the total number of responses. Option A is incorrect because a histogram is used to show how data points are distributed along a numerical scale. The survey response data set is not numerical, but categorical. Option C is incorrect because a line chart is used to show trends or changes over time. The survey response data set does not have a time dimension. Option D is incorrect because a scatter plot is used to show the relationship between two numerical variables. The survey response data set does not have two numerical variables. Option E is incorrect because a waterfall chart is used to show how an initial value is increased or decreased by a series of intermediate values. The survey response data set does not have an initial value or intermediate values. References: How to Choose the Right Chart for Your Data - Infogram How to Choose the Right Data Visualization | Tutorial by Chartio Find the Best Visualizations for Your Metrics - The Data School How to choose the best chart or graph for your data

QUESTION 4

Joe. an analyst. tests the loading time on a dashboard he is preparing to go live and finds it is slower than he would like. Which of the following must occur to decrease the loading time?

A. Deploy the dashboard to production.

B. Change the field definitions.

C. Update the dashboard subscribers.

D. Optimize the dashboard.

Correct Answer: D

Optimizing the dashboard is the process of improving its performance and reducing its loading time by applying various techniques and best practices. Some of the common ways to optimize a dashboard are: Reducing the size and complexity of the data model, such as removing unnecessary columns, aggregating data at the source, or using data compression techniques12 Leveraging caching strategies, such as setting appropriate cache refresh intervals or utilizing Power BI\\'s built-in caching mechanisms, to minimize data retrieval delays2 Utilizing query folding, direct query, or live connection to enhance data processing efficiency and enable real-time data updates23 Optimizing DAX queries, such as avoiding nested calculations, using variables, or simplifying measures, to improve data calculation speed23 Reducing visualizations and calculations, such as using fewer or simpler charts, filters, or parameters, to speed up dashboard rendering12 Evaluating the impact of custom visuals on dashboard load time and avoiding or replacing those that are slow or inefficient2 Applying aggregation and summarization techniques, such as using extract filters, context filters, or level of detail expressions, to reduce the amount of data displayed on the dashboard1 Troubleshooting and resolving any issues that may cause slow dashboard load, such as network latency, server overload, or hardware limitations24

QUESTION 5

Consider the following dataset which contains information about houses that are for sale:

```
sonery=# select * from melb limit 5;
 houseid |      address      |      regionname       | type | rooms |    date    | distance |   price
---------+-------------------+-----------------------+------+-------+------------+----------+-----------
       1 | 85 Turner St      | Northern Metropolitan | h    |     2 | 2016-03-12 |      2.5 |  1.48e+06
       2 | 25 Bloomburg St   | Northern Metropolitan | h    |     2 | 2016-04-02 |      2.5 | 1.035e+06
       3 | 5 Charles St      | Northern Metropolitan | h    |     3 | 2017-04-03 |      2.5 | 1.465e+06
       4 | 40 Federation La  | Northern Metropolitan | h    |     3 | 2017-04-03 |      2.5 |    850000
       5 | 55a Park St       | Northern Metropolitan | h    |     4 | 2016-04-06 |      2.5 |   1.6e+06
(5 rows)
```

Which of the following string manipulation commands will combine the address and region name columns to create a full address?

full_address---------- 85 Turner St, Northern Metropolitan 25 Bloomburg St, Northern Metropolitan 5 Charles St, Northern Metropolitan 40 Federation La, Northern Metropolitan 55a Park St, Northern Metropolitan

A. SELECT CONCAT(address, \\' , \\' , regionname) AS full_address FROM melb LIMIT 5;

B. SELECT CONCAT(address, \\'-\\' , regionname) AS full_address FROM melb LIMIT 5;

C. SELECT CONCAT(regionname, \\' , \\' , address) AS full_address FROM melb LIMIT 5

D. SELECT CONCAT(regionname, \\'-\\' , address) AS full_address FROM melb LIMIT 5;

Correct Answer: A

The correct answer is A: SELECT CONCAT(address, \\' , \\' , regionname) AS full_address FROM melb LIMIT 5; String manipulation (or string handling) is the process of changing, parsing, splicing, pasting, or analyzing strings. SQL is used for managing data in a relational database. The CONCAT () function adds two or more strings together. Syntax CONCAT(stringl, string2,... string_n) Parameter Values Parameter Description stringl, string2, string_n Required. The strings to add together.

QUESTION 6

A web developer wants to ensure that malicious users can\\'t type SQL statements when they asked for input, like their username/userid.

Which of the following query optimization techniques would effectively prevent SQL Injection attacks?

A. Indexing.

B. Subset of records.

C. Temporary table in the query set.

D. Parametrization.

Correct Answer: D

The correct answer is D: Parametrization. Parameterized SQL queries allow you to place parameters in an SQL query instead of a constant value. A parameter takes a value only when the query is executed, allowing the query to be reused

with different values and purposes. Parameterized SQL statements are available in some analysis clients, and are also available through the Historian SDK.

For example, you could create the following conditional SQL query, which contains a parameter for the collector\\'s name: SELECT* FROM ExamsDigest WHERE coursename=? ORDER BY tagname SQL Injection is best prevented through

the use of parameterized queries.

QUESTION 7

Which of the following is a control measure for preventing a data breach?

A. Data transmission

B. Data attribution

C. Data retention

D. Data encryption

Correct Answer: D

Explanation: This is because data encryption is a type of control measure that prevents a data breach, which is an unauthorized or illegal access or use of data by an external or internal party. Data encryption can prevent a data breach by protecting and securing the data using a code or a key that scrambles or transforms the data into an unreadable or incomprehensible format, which can only be decoded or restored by authorized users who have the correct code or key. For example, data encryption can prevent a data breach by encrypting the data in transit or at rest, such as when the data is sent over a network or stored in a device. The other control measures are not used for preventing a data breach. Here is why:

Data transmission is a type of process that transfers and exchanges data between different sources or systems, such as databases, cloud services, or web applications. Data transmission does not prevent a data breach, but rather exposes the data to potential risks or threats during the transfer or exchange. However, data transmission can be made more secure and less vulnerable to a data breach by using encryption or other methods, such as authentication or authorization. Data attribution is a type of feature or function that assigns and tracks the ownership and origin of the

data, such as the creator, modifier, or source of the data. Data attribution does not prevent a data breach but rather provides information and evidence about the data provenance and history. However, data attribution can be useful for detecting and responding to a data breach by using audit logs or metadata to identify and trace any unauthorized or illegal access or use of the data. Data retention is a type of policy or standard that specifies and regulates the storage and preservation of the data, such as the duration, location, or format of the data. Data retention does not prevent a data breach, but rather affects the availability and accessibility of the data for future use or reference. However, data retention can be optimized and aligned with the legal and ethical requirements and standards of the industry or the organization to reduce the risk or impact of a data breach.

QUESTION 8

Given the data below: In which of the following file formats is the data presented?

A. Xs

B. CSV

C. RIF

D. XML

Correct Answer: B

Explanation: The data is presented in a CSV (comma-separated values) file format, which is a plain text format that stores tabular data. Each line of the file is a data record, and each record consists of one or more fields separated by commas. The first line of the file usually contains the names of the fields, also known as the header. In this case, the data has four fields: Name, Age, Gender, and Occupation. Therefore, the correct answer is B. References: CSV File (What It Is and How to Open One), Comma-separated values - Wikipedia

QUESTION 9

Samantha needs to share a list of her organization\\'s top 50 customers with the VP of sales.

She would like to include the name of the customer, the business they represent, their contact information, and their total sales over the past year. The VP does not have any specialized analytics skills or software but would like to make some personal notes on the dataset.

What would be the best tool for Samantha to use to share this information?

A. Power BI.

B. Microsoft Excel.

C. Minitab.

D. SAS.

Correct Answer: B

Microsoft Excel.

This scenario presents a very simple use case where the business leader needs a dataset in an easy-to-access form and will not be performing any detailed analysis. A simple spreadsheet, such as Microsoft Excel, would be the best tool for

this job. There is no need to use a statistical analysis package, such as SAS or Minitab, as this would likely confuse the VP without adding any value. The same is true of an integrated analytics suite, such as Power BI.

QUESTION 10

Which of the following would be considered non-personally identifiable information?

A. Cell phone device name

B. Customer\\'s name

C. Government ID number

D. Telephone number

Correct Answer: A

Explanation: Non-personally identifiable information (non-PII) is any data that cannot be used to identify, contact, or locate a specific individual, either alone or combined with other sources. Non-PII can include aggregated statistics, anonymous data, device identifiers, IP addresses, cookies, and other types of information that do not reveal the identity or location of a person. Cell phone device name is an example of non-PII, as it does not reveal any personal information about the owner or user of the device. Therefore, the correct answer is A. References: What is Non-Personally Identifiable Information (Non-PII)? | Definition and Examples, What is Personally Identifiable Information (PII)? | Definition and Examples

QUESTION 11

A data analyst is creating a report that will provide information about various regions, products, and time periods. Which of the following formats would be the most efficient way to deliver this report?

A. A workbook with multiple tabs for each region

B. A daily email with snapshots of regional summaries

C. A static report with a different page for every filtered view

D. A dashboard with filters at the top that the user can toggle

Correct Answer: D

Explanation: The best format to deliver this report is D. A dashboard with filters at the top that the user can toggle. A

dashboard is a visual display of the most important information needed to achieve one or more objectives, consolidated and arranged on a single screen so the information can be monitored at a glance1 A dashboard with filters at the top that the user can toggle would allow the user to easily and quickly access the information they need about various regions, products, and time periods, without having to navigate through multiple tabs, pages, or emails. A dashboard with filters would also enable the user to compare and contrast different views of the data and see how they change over time. A dashboard with filters would also be more interactive and engaging than a static or email report2 A workbook with multiple tabs for each region would not be an efficient way to deliver this report, because it would require the user to switch between different tabs to see the information they need. This would make it harder to compare and contrast different regions, products, and time periods, and also increase the risk of errors or confusion. A workbook with multiple tabs would also be less visually appealing and more cluttered than a dashboard3 A daily email with snapshots of regional summaries would not be an efficient way to deliver this report, because it would limit the user\\'s ability to explore the data in depth and customize their view. A daily email would also be dependent on the frequency and timing of the email delivery, which might not match the user\\'s needs or preferences. A daily email would also be more likely to be ignored or deleted than a dashboard that is always accessible. A static report with a different page for every filtered view would not be an efficient way to deliver this report, because it would create a very long and cumbersome report that would be difficult to read and understand. A static report would also not allow the user to change or update the filters as they wish, or see how the data changes over time. A static report would also be less interactive and engaging than a dashboard.

QUESTION 12

You are working with a professional statistician to perform an analysis and would like to use a statistics package.

Which one of the following would be the most appropriate?

A. Rapid Miner.

B. QLIK.

C. Power BI.

D. Minitab.

Correct Answer: D

Minitab is statistical analysis software. It can be used for learning about statistics as well as statistical research. Statistical analysis computer applications have the advantage of being accurate, reliable, and generally faster than computing statistics and drawing graphs by hand.

QUESTION 13

Which of the following can be used to translate data into another form so it can only be read by a user who has a key or a password?

A. Data encryption.

B. Data transmission.

C. Data protection.

D. Data masking.

Correct Answer: A

Data encryption can be used to translate data into another form so it can only be read by a user who has a key or a password. Data encryption is a process of transforming data using an algorithm or a cipher to make it unreadable to anyone except those who have the key or the password to decrypt it. Data encryption is a common method of protecting data from unauthorized access, modification, or theft. Reference: Guide to CompTIA Data+ and Practice Questions - Pass Your Cert

QUESTION 14

An analyst is designing a dashboard to determine which site has the highest percentage of new customers. The analyst must choose an appropriate chart to include in the dashboard. The following data is available:

Site	Customers	New customers	Percentage of new customers
A1	2236	277	12%
A2	885	300	34%
A3	333	200	60%
B1	483	167	35%
B2	2969	235	8%
B3	2357	153	6%
C1	1524	180	12%
C2	878	150	17%
C3	1925	142	7%

Which of the following types of charts should be considered to BEST display the data?

A. Include a bar chart using the site and the percentage of new customers data.

B. Include a line chart using the site and the percentage of new customers data.

C. Include a pie chat using the site and percentage of new customers data.

D. Include a scatter chart using the site and the percent of new customers data.

Correct Answer: A

Explanation: This is because a bar chart is a type of chart that shows the value or the amount of a single variable for different categories or groups, such as the percentage of new customers for different sites in this case. A bar chart can be used to display and analyze the comparison, ranking, or proportion among the categories or groups, as well as identify any differences, similarities, or outliers in the data. For example, a bar chart can show which site has the highest or lowest percentage of new customers, as well as show how much each site contributes to the total percentage of new customers. The other types of charts are not the best charts to display the data. Here is why:

A line chart is a type of chart that shows the change or the trend of a single variable over time, such as the percentage of new customers over months or years in this case. A line chart can be used to display and analyze the movement, cycle, or pattern of the variable, as well as identify any peaks, valleys, or fluctuations in the data. For example, a line

chart can show how the percentage of new customers increases or decreases over time, as well as show if there are any seasonal or periodic variations in the data. A pie chart is a type of chart that shows the proportion or the percentage of a single variable for different categories or groups, such as the percentage of new customers for different sites in this case. A pie chart can be used to display and analyze the composition, distribution, or share of the variable, as well as identify any segments, slices, or fractions in the data. For example, a pie chart can show how much each site represents of the total percentage of new customers, as well as show if there are any dominant or minor sites in the data. A scatter chart is a type of chart that shows the relationship between two variables for each observation or unit in a data set, such as the percentage of new customers and another variable for each site in this case. A scatter chart can be used to display and analyze the correlation, trend, or pattern among the variables, as well as identify any outliers or clusters in the data. For example, a scatter chart can show if there is a positive, negative, or no correlation between the percentage of new customers and another variable, such as sales revenue or customer satisfaction.

QUESTION 15

While reviewing survey data, an analyst notices respondents entered "Jan," "January," and "01" as responses for the month of January. Which of the following steps should be taken to ensure data consistency?

A. Delete any of the responses that do not have "January" written out.

B. Replace any of the responses that have "01".

C. Filter on any of the responses that do not say "January" and update them to "January".

D. Sort any of the responses that say "Jan" and update them to "01".

Correct Answer: C

Explanation: Filter on any of the responses that do not say "January" and update them to "January". This is because filtering and updating are data cleansing techniques that can be used to ensure data consistency, which means that the data is uniform and follows a standard format. By filtering on any of the responses that do not say "January" and updating them to "January", the analyst can make sure that all the responses for the month of January are written in the same way. The other steps are not appropriate for ensuring data consistency. Here is why:

Deleting any of the responses that do not have "January" written out would result in data loss, which means that some information would be missing from the data set. This could affect the accuracy and reliability of the analysis. Replacing any of the responses that have "01" would not solve the problem of data inconsistency, because there would still be two different ways of writing the month of January: "Jan" and "January". This could cause confusion and errors in the analysis. Sorting any of the responses that say "Jan" and updating them to "01" would also not solve the problem of data inconsistency, because there would still be two different ways of writing the month of January: "01" and "January". This could also cause confusion and errors in the analysis.