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.pass2lead.com/da0-001.html

100% Passing Guarantee
100% Money Back Assurance

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

⚙ **Instant Download** After Purchase

⚙ **100% Money Back** Guarantee

⚙ **365 Days** Free Update

⚙ **800,000+** Satisfied Customers

QUESTION 1

Which one of the following values will appear first if they are sorted in descending order?

A. Aaron.

B. Molly.

C. Xavier.

D. Adam.

Correct Answer: C

Explanation: The value that will appear first if they are sorted in descending order is Xavier. Descending order means arranging values from the largest to the smallest, or from the last to the first in alphabetical order. In this case, Xavier is the last name in alphabetical order, so it will appear first when sorted in descending order. The other names will appear in the following order: Molly, Adam, Aaron. Reference: Sorting Data - W3Schools

QUESTION 2

Which of the following is an example of a flat file?

A. CSV file

B. PDF file

C. JSON file

D. JPEG file

Correct Answer: A

A CSV file is a type of flat file that stores data as plain text in a table-like structure with rows and columns. Each row represents a single record, while columns represent fields or attributes of the data. A CSV file uses commas or other delimiters to separate the values in each row. A CSV file can be easily imported or exported by various applications and programs12

QUESTION 3

Which of the ing is the correct ion for a tab-delimited spre file?

A. tap

B. tar

C. sv

D. az

Correct Answer: C

Explanation: A tab-delimited spreadsheet file is a type of flat text file that uses tabs as delimiters to separate data values in a table. The file extension for a tab-delimited spreadsheet file is usually .tsv, which stands for tab-separated values. Therefore, the correct answer is C. References: [Tab-separated values - Wikipedia], [What is a TSV File? | How to Open, Edit and Convert TSV Files]

QUESTION 4

Which of the following techniques is used to quantify data?

A. Decoding

B. Enumeration

C. Coding

D. Structure

Correct Answer: C

Answer: C. Coding Coding is a technique that is used to quantify data, especially qualitative data that are not expressed numerically. Coding involves assigning codes, such as numbers, letters, symbols, or colors, to different categories or themes that emerge from the data. For example, if you have a set of survey responses that ask about the satisfaction level of customers, you can code them as follows: Very satisfied = 5 Satisfied = 4 Neutral = 3 Dissatisfied = 2 Very dissatisfied = 1 By coding the data, you can convert them into quantitative data that can be analyzed using statistical methods, such as calculating the mean, median, mode, frequency, or percentage of each category12. Option A is incorrect, as decoding is not a technique that is used to quantify data, but rather a process of interpreting or translating data from one form to another. For example, decoding can involve converting binary codes into text or images, or decrypting ciphertext into plaintext3. Option B is incorrect, as enumeration is not a technique that is used to quantify data, but rather a process of listing or naming data in a specific order. For example, enumeration can involve listing the names of the states in alphabetical order, or naming the planets in order of their distance from the sun4. Option D is incorrect, as structure is not a technique that is used to quantify data, but rather a property or characteristic of data that describes how they are organized or arranged. For example, structure can refer to the format, type, or schema of data, such as structured, semi-structured, or unstructured data.

QUESTION 5

An analyst is working with the income data of suburban families in the United States. The data set has a lot of outliers, and the analyst needs to provide a measure that represents the typical income. Which of the following would BEST fulfill the analyst\\'s goal?

A. Median

B. Mean

C. Mode

D. Standard deviation

Correct Answer: A

his is because median is a type of statistical measure that represents the typical value or central tendency of a data set, which means that it divides the data set into two equal halves, such that half of the values are above it and half are below it. Median can be used to provide a measure that represents the typical income of suburban families in the United

States, especially when the data set has a lot of outliers, which means that it has values that are unusually high or low compared to the rest of the data set. Median can provide a measure that represents the typical income of suburban families in the United States, because it is not affected or skewed by the outliers, as it only depends on the middle value or the middle two values of the data set, regardless of how extreme or distant the outliers are. For example, median can provide a measure that represents the typical income of suburban families in the United States, by finding the income value that splits the data set into two equal groups of families, such that 50% of the families have higher incomes and 50% have lower incomes. The other statistical measures are not the best measures to represent the typical income of suburban families in the United States. Here is why:

Mean is a type of statistical measure that represents the average value or central tendency of a data set, which means that it is the sum of all the values divided by the number of values. Mean is not a good measure to represent the typical income of suburban families in the United States, especially when the data set has a lot of outliers, because it is affected or skewed by the outliers, as it takes into account all the values in the data set, regardless of how extreme or distant they are. For example, mean can provide a measure that does not represent the typical income of suburban families in the United States, by finding the income value that is influenced by a few very high or very low incomes, which could make it higher or lower than most of the incomes in the data set. Mode is a type of statistical measure that represents the most frequent value or mode of a data set, which means that it is the value that occurs most often in the data set. Mode is not a good measure to represent the typical income of suburban families in the United States, especially when the data set has a lot of outliers, because it is not representative or indicative of the central tendency or distribution of the data set, as it only depends on the count or occurrence of a single value or a few values in the data set, regardless of how common or rare they are. For example, mode can provide a measure that does not represent the typical income of suburban families in the United States, by finding the income value that is repeated more often than others, which could be an outlier or an anomaly in the data set. Standard deviation is a type of statistical measure that represents the amount of dispersion or variation of a data set, which means that it quantifies how much the values in a data set vary or deviate from the mean or average of the data set. Standard deviation is not a measure that represents the typical income of suburban families in the United States, but rather a measure that describes the spread or distribution of their incomes, as well as identifies any outliers or extreme values in their incomes. For example, standard deviation can provide a measure that describes how diverse or homogeneous their incomes are, as well as how far their incomes are from their average income.

QUESTION 6

While reviewing survey data, a research analyst notices data is missing from all the responses to a single question. Which of the following methods would BEST address this issue?

A. Replace missing data.

B. Remove duplicate data.

C. Replace redundant data.

D. Remove invalid data.

Correct Answer: A

Explanation: This is because missing data is a type of data quality issue that occurs when data is absent or incomplete in a data set, which can affect the accuracy and reliability of the analysis or process. Missing data can be caused by various factors, such as human error, system error, or non-response. Missing data can be addressed by using various methods, such as replacing missing data, which means filling in or imputing the missing values with some reasonable estimates, such as mean, median, mode, or regression. The other methods are not used to address missing data. Here is why:

Remove duplicate data is a type of method that eliminates or reduces duplicate data, which is a type of data quality issue that occurs when data is repeated or copied in a data set. Removing duplicate data does not address missing data, but rather affects the quantity and validity of the data. Replace redundant data is a type of method that eliminates

![Pass2Lead](https://Pass2Lead.com)
or reduces redundant data, which is a type of data quality issue that occurs when data is unnecessary or irrelevant for the analysis or purpose. Replacing redundant data does not address missing data, but rather affects the efficiency and performance of the analysis or process. Remove invalid data is a type of method that eliminates or reduces invalid data, which is a type of data quality issue that occurs when data is incorrect or inaccurate in a data set. Removing invalid data does not address missing data, but rather affects the validity and reliability of the analysis or process.

QUESTION 7

Amanda needs to create a dashboard that will draw information from many other data sources and present it to business leaders.

Which one of the following tools is least likely to meet her needs?

A. QuickSight.

B. Tableau.

C. Power BI.

D. SPSS Modeler.

Correct Answer: D

SPSS Modeler.

QuickSight, Tableau, and Power BI are all powerful analytics and reporting tools that can pull data from a variety of sources. SPSS Modeler is a powerful predictive analytics platform that is designed to bring predictive intelligence to decisions

made by individuals, groups, systems and your enterprise.

QUESTION 8

An analyst is preparing a report that contains weather data. The temperatures are shown in Fahrenheit. but they must be reported in Celsius. Which of the following should the analyst do to fix this issue?

A. Normalize the data.

B. Standardize the data.

C. Rescale the data.

D. Aggregate the data.

Correct Answer: C

The analyst should rescale the data to fix this issue. Rescaling is a process of transforming data from one scale to another, such as changing the units of measurement. In this case, the analyst needs to rescale the temperatures from Fahrenheit to Celsius, which are two different scales for measuring temperature. To do this, the analyst can use the following formula: Celsius = (Fahrenheit - 32) * 5/9 This formula converts each temperature value from Fahrenheit to Celsius by subtracting 32 and multiplying by 5/9. For example, if the temperature is 68

QUESTION 9

Which of the following are reasons to conduct data cleansing? (Select two).

A. To perform web scraping

B. To track KPIs

C. To improve accuracy

D. To review data sets

E. To increase the sample size

F. To calculate trends

Correct Answer: CF

Two reasons to conduct data cleansing are: To improve accuracy: Data cleansing helps to ensure that the data is correct, consistent, and reliable. This can improve the quality and validity of the analysis, as well as the decision-making and outcomes based on the data12 To calculate trends: Data cleansing helps to remove or resolve any errors, outliers, or missing values that could distort or skew the data. This can help to identify and measure the patterns, changes, or relationships in the data over time13

QUESTION 10

Given the following report:

Quarterly Customer Service Report

Table 1. Frequency of Ticket Statuses

Status	Count
Reported	11
In-Progress	323 •
Closed	554

Table 2. Occurrence of Target Phrases

Target Phrases	Count
Have a great day!	1200
It is my pleasure to assist you.	70
Can you please hold?	7352

Most tickets are being addressed soon after being reported. Asking customers to hold is the most commonly used target phrase.

Which of the following components need to be added to ensure the report is point-in-time and static? (Choose two.)

A. A control group for the phrases

B. A summary of the KPIs

C. Filter buttons for the status

D. The date when the report was last accessed

E. The time period the report covers

F. The date on which the report was run

Correct Answer: E

The date on which the report was run. This is because the time period the report covers and the date on which the report was run are two components that need to be added to ensure the report is point-in-time and static, which means

that the report shows the data as it was at a specific moment or interval in time, and does not change or update with new data. By adding the time period the report covers and the date on which the report was run, the analyst can indicate when and for how long the data was collected and analyzed, as well as avoid any confusion or ambiguity about the currency or validity of the data. The other components do not need to be added to ensure the report is point-in-time and static. Here is why:

A control group for the phrases is a type of group that serves as a baseline or a reference for comparison with another group that is exposed to some treatment or intervention, such as a target phrase in this case. A control group for the phrases does not need to be added to ensure the report is point-in-time and static, because it does not affect the time frame or the stability of the data. However, a control group for the phrases could be useful for evaluating the effectiveness or impact of the target phrases on customer satisfaction or retention. A summary of the KPIs is a type of document that provides an overview or a highlight of the key performance indicators (KPIs), which are measurable values that indicate how well an organization or a process is achieving its goals or objectives. A summary of the KPIs does not need to be added to ensure the report is point-in-time and static, because it does not affect the time frame or the stability of the data. However, a summary of the KPIs could be useful for communicating or presenting the main findings or insights from the report. Filter buttons for the status are a type of feature or function that allows users to select or deselect certain values or categories in a column or a table, such as ticket statuses in this case. Filter buttons for the status do not need to be added to ensure the report is point-in- time and static, because they do not affect the time frame or the stability of the data. However, filter buttons for the status could be useful for exploring or analyzing different aspects or segments of the data.

QUESTION 11

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 12

A county in Illinois is conducting a survey to determine the mean annual income per household. The county is 427sq mi (2.65q km). Which of the following sampling methods would MOST likely result in a representative sample?

A. A stratified phone survey of 100 people that is conducted between 2:00 p.m. and 3:00 p.m

B. A systematic survey that is sent to 100 single-family homes in the county

C. Surveys sent to ten randomly selected homes within 5mi (8km) of the county\\'s office

D. Surveys sent to 100 randomly selected homes that are reflective of the population

Correct Answer: D

Explanation: Surveys sent to 100 randomly selected homes that are reflective of the population. This is because a random sample is a type of sample that is selected by using a random method, such as a lottery or a computer-generated number, which ensures that every element in the population has an equal chance of being selected. A random sample can result in a representative sample, which means that the sample reflects the characteristics and diversity of the population. By sending surveys to 100 randomly selected homes that are reflective of the population, the analyst can ensure that the sample is representative of the county\\'s households and their income levels. The other sampling methods are not likely to result in a representative sample. Here is why: A stratified phone survey of 100 people that is conducted between 2:00 p.m. and 3:00 p.m. would result in a biased sample, which means that the sample favors or excludes certain groups or elements in the population. By conducting the survey only between 2:00 p.m. and 3:00 p.m., the analyst would miss out on people who are not available or reachable at that time, such as those who are working or sleeping. This could affect the representativeness and generalizability of the sample. A systematic survey that is sent to 100 single-family homes in the county would result in an unrepresentative sample, which means that the sample does not reflect the characteristics and diversity of the population. By sending surveys only to single-family homes, the analyst would ignore other types of households, such as apartments, condos, or mobile homes. This

![Pass2Lead](https://Pass2Lead.com)
could affect the accuracy and reliability of the sample. Surveys sent to ten randomly selected homes within 5mi (8km) of the county\\'s office would result in a small sample, which means that the sample size is too low to capture the variability and diversity of the population. By sending surveys only to ten homes within a limited area, the analyst would miss out on many households that are located in different parts of the county. This could affect the precision and confidence of the sample.

QUESTION 13

A data analyst needs to create a weekly recurring report on sales performance and distribute it to all sales managers. Which of the following would be the BEST method to automate and ensure successful delivery for this task?

A. Use scheduled report delivery.

B. Implement subscription access delivery.

C. Print out a copy.

D. Upload the report to the server.

Correct Answer: A

Explanation: Scheduled report delivery is a feature that allows a data analyst to automate the generation and distribution of a report at a specified time and frequency. This would be the best method to ensure that the sales managers receive the weekly report on sales performance without manual intervention. Subscription access delivery is a feature that allows users to subscribe to a report and access it on demand, but it does not automate the delivery. Printing out a copy or uploading the report to the server are manual methods that require more time and effort from the data analyst. Reference: CertMaster Practice for Data+ Exam Prep - CompTIA

QUESTION 14

Which one of the following would not normally be considered a summary statistic?

A. z-score.

B. Mean.

C. Variance.

D. Standard deviation.

Correct Answer: A

Simply put, a z-score (also called a standard score) gives you an idea of how far from the mean a data point is. But more technically it\\'s a measure of how many standard deviations below or above the population mean a raw score is. A z-score can be placed on a normal distribution curve.

QUESTION 15

A data analyst is asked to create a sales report for the second-quarter 2020 board meeting, which will include a review of the business\\'s performance through the second quarter. The board meeting will be held on July 15, 2020, after the numbers are finalized. Which of the following report types should the data analyst create?

A. Static

B. Real-time

C. Self-service

D. Dynamic

Correct Answer: A

Explanation: A dynamic report is a type of report that shows data that changes or updates automatically based on certain criteria or parameters. A dynamic report can allow users to interact with the data, filter it, drill down into it, or visualize it in different ways. A dynamic report is suitable for situations where the data changes frequently or where real-time or near-real-time data is needed for decision making or analysis. In this case, the data analyst is asked to create a sales report for the second-quarter 2020 board meeting, which will include a review of the business\'s performance through the second quarter. The board meeting will be held on July 15, 2020, after the numbers are finalized. This means that the data analyst does not need to show real-time or dynamic data, but rather a fixed and accurate view of the sales data for the second quarter. Therefore, a static report would be the best way to meet this stakeholder requirement. Therefore, the correct answer is A. References: [What are Dynamic Reports? | Sisense], Static vs Dynamic Reports - What\'s The Difference? | datapine

[Latest DA0-001 Dumps](#)   [DA0-001 Study Guide](#)   [DA0-001 Braindumps](#)