Vendor: CompTIA

Exam Code: DA0-001

Exam Name: CompTIA Data+ Certification Exam

Version: DEMO

QUESTION 1
A data analyst needs to calculate the mean for Q1 sales using the data set below:

Product	Q1 sales
Ground beef	$2,667.60
Crab meet	$1,768.41
Swiss cheese	$3,182.40
Broccoli	$1,509.60
Vegetable spread	$3.202.87

Which of the following is the mean?

A. $2,466.18
B. $2,667.60
C. $3,082.72
D. $12,330.88

Answer: A
Explanation:
The mean can be calculated by adding up all the Q1 sales values and then dividing by the number of items in the data set. In this case, the total of all the Q1 sales is $12,330.88, and there are 5 items in the data set.
So, the mean is calculated as follows:

$12,330.88 ÷ 5 = $2,466.18

Therefore, the mean is $2,466.18.

QUESTION 2
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

Answer: D
Explanation:
A dashboard with filters at the top that the user can toggle would be the MOST efficient way to deliver this report. A dashboard allows for a clear and concise presentation of information with the ability to filter and view data in different ways. This allows the user to quickly access the information they need without having to navigate through multiple tabs or pages. Additionally, dashboards can be interactive and allow for real-time updates, providing the most up-to-date information. With the ability to filter by region, product, and time period, the dashboard provides the user with a comprehensive and customizable view of the data, making it the most efficient way to deliver this report.

QUESTION 3
A customer list from a financial services company is shown below:

Name	Number of credit cards	Age	Income
Sean	0	27	$60,000
Angela	4	31	$50,000
Terry	3	40	$170,000
Paula	1	25	$70,000
Malcolm	3	28	$150,000

A data analyst wants to create a likely-to-buy score on a scale from 0 to 100, based on an average of the three numerical variables: number of credit cards, age, and income. Which of the following should the analyst do to the variables to ensure they all have the same weight in the score calculation?

A. Recode the variables.
B. Calculate the percentiles of the variables.
C. Calculate the standard deviations of the variables.
D. Normalize the variables.

Answer: D
Explanation:
Normalize the variables. Normalization involves transforming the values of the variables so that they have the same scale and range, typically from 0 to 1. This allows the analyst to ensure that each variable has the same weight in the calculation, since all the values are on the same scale. Normalizing the variables will also make it easier to interpret the results and compare the scores of different customers. This is important in creating a likely-to-buy score because it ensures that all three variables have an equal impact on the final score, rather than one variable having a larger weight due to its scale or range.

QUESTION 4
Which of the following actions should be taken when transmitting data to mitigate the chance of a data leak occurring? (Choose two.)

A. Data identification
B. Data processing
C. Data reporting
D. Data encryption
E. Data masking
F. Fata removal

Answer: DE
Explanation:
Data encryption involves transforming the data into an unreadable format so that it cannot be understood without the appropriate decryption key. This protects the data from being intercepted or viewed by unauthorized parties during transmission.
Data masking involves obscuring sensitive data elements by replacing them with non-sensitive values. This allows the data to be used for testing, development, and reporting purposes without compromising its confidentiality. Data masking also helps to mitigate the risk of data leaks, since

the sensitive information is not accessible to unauthorized parties.
By using both encryption and masking, the chance of a data leak occurring during transmission can be significantly reduced, as the data is protected both in transit and at rest.

QUESTION 5
A data analyst has been asked to organize the table below in the following ways:

```
- By sales from high to low
- By state in alphabetic order
```

First_name	Last_name	Address	City	State	Sales
Ed	Edens	2851 N. Southport	Chicago	IL	$125,689
Pat	Mudd	710 Bridle Ridge Road	Eagan	MN	$101,259
Katie	Hofstad	2851 S. Windwood Lane	Rosemount	NY	$105,779
Edward	Frank	281 S. Northport	Chicago	IL	$456,231
Rachel	Newman	305 Big Timber Trail	Wheaton	CO	$99,876
Kaylyn	Korth	332 Richfield Drive	Lakeview	MN	$166,874

Which of the following functions will allow the data analyst to organize the table in this manner?

A. Conditional formatting
B. Grouping
C. Filtering
D. Sorting

Answer: D
Explanation:
Sorting is the function that will allow the data analyst to organize the table in the desired manner. Sorting allows the data analyst to arrange the data in a specific order, either in ascending or descending order, based on one or multiple columns. In this case, the analyst can sort the table by Age from high to low by clicking on the Age column and selecting "Sort Descending". They can then sort the table by Income in alphabetic order by clicking on the Income column and selecting "Sort A to Z". Sorting is a simple and efficient way to organize large data sets, making it easier for the analyst to analyze and present the data.

QUESTION 6
Which of the following BEST describes the issue in which character values are mixed with integer values in a data set column?

A. Duplicate data
B. Missing data
C. Data outliers
D. Invalid data type

Answer: D
Explanation:
If a data set column contains character values mixed with integer values, it would be considered to have an invalid data type. A data type refers to the type of information that a column can contain, such as numbers, text, or dates. Mixing character values with integer values in the same column would be considered invalid because it violates the expected data type for the column.

The other options (duplicate data, missing data, and data outliers) are not related to the data type of the values in the column.

QUESTION 7
Which of the following is a process that is used during data integration to collect, blend, and load data?

A. MDM
B. ETL
C. OLTP
D. BI

Answer: B
Explanation:
ETL (Extract, Transform, Load) is a process that is used during data integration to collect, blend, and load data. The process involves extracting data from multiple sources, transforming the data into a format that is suitable for analysis, and then loading the data into a target system, such as a data warehouse. ETL is commonly used to integrate data from various sources, such as databases, flat files, and web services, and to prepare the data for reporting and analysis. The other options (MDM, OLTP, and BI) are not related to the ETL process.

QUESTION 8
An analyst has received the requirements for an internal user dashboard. The analyst confirms the data sources and then creates a wireframe. Which of the following is the NEXT step the analyst should take in the dashboard creation process?

A. Optimize the dashboard.
B. Create subscriptions.
C. Get stakeholder approval.
D. Deploy to production.

Answer: C
Explanation:
Get stakeholder approval is the NEXT step the analyst should take in the dashboard creation process after creating the wireframe. After confirming the data sources and creating a wireframe of the internal user dashboard, it is important to get approval from stakeholders who will be using the dashboard. This includes both internal and external stakeholders, as well as end users who will be accessing the dashboard.
Stakeholder approval is necessary to ensure that the dashboard meets the requirements of the stakeholders, and that it is aligned with the overall goals of the organization. During the approval process, stakeholders may provide feedback or make changes to the wireframe, and the analyst can use this feedback to make any necessary updates to the dashboard design.

QUESTION 9
A data analyst has been asked to derive a new variable labeled "Promotion_flag" based on the total quantity sold by each salesperson. Given the table below:

Store_ID	Item	Salesperson	Quantity_sold	Promotion_flag
104	Pax-2	James	1,000,300	
204	Pax-3	Paul	234,578	
304	Pax-1	Peter	2,000,432	
404	Pax-2	Esther	1,089,678	
204	Pax-3	May	126,578	
304	Pax-1	Park	200,432	
404	Pax-2	Mabel	1,089,000	

Which of the following functions would the analyst consider appropriate to flag "Yes" for every salesperson who has a number above 1,000,000 in the Quantity_sold column?

A. Date
B. Mathematical
C. Logical
D. Aggregate

Answer: C
Explanation:
Logical functions would be appropriate to flag "Yes" for every salesperson who has a number above 1,000,000 in the Quantity_sold column. Logical functions in data analysis are used to make decisions based on a set of conditions. In this case, the analyst can use a logical function such as an IF statement, where the condition is that the salesperson's Quantity_sold is above 1,000,000. If the condition is met, then the Promotion_flag is set to "Yes," otherwise it is set to "No."
For example, the formula could be something like:

```
IF(Quantity_sold > 1000000, "Yes", "No")
```

This formula would then be applied to the entire column of Quantity_sold values, and the resulting Promotion_flag column would indicate whether each salesperson is eligible for a promotion based on the number of items they have sold. Logical functions are an essential tool in data analysis, as they allow analysts to make decisions based on specific criteria and automate certain processes in the data analysis process.

QUESTION 10
Given the diagram below:
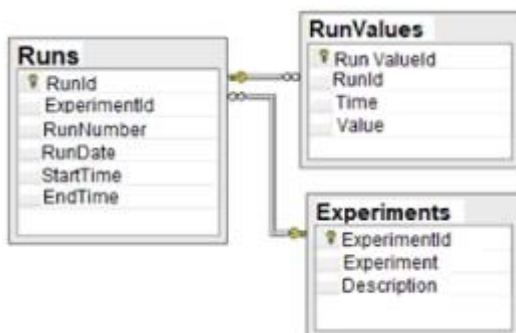
Which of the following data schemas shown?

A. Key-value pairs
B. Online transactional processing
C. Data lake
D. Relational database

Answer: D
Explanation:
The diagram shows a schema for a relational database, which is a type of database that stores data in tables that are related to each other through common columns or keys. In the diagram, the "Runs" table is connected to the "Athletes" and "Races" tables through the "Athlete_id" and "Race_id" columns, respectively. These columns are foreign keys that reference the primary keys in the "Athletes" and "Races" tables. This type of structure allows data to be stored and accessed in a structured and organized way, and enables relationships between different data entities to be captured and queried. The other options (key-value pairs, online transactional processing, and data lake) do not describe a relational database schema.

QUESTION 11
A company's marketing department wants to do a promotional campaign next month. A data analyst on the team has been asked to perform customer segmentation, looking at how recently a customer bought product, at what frequency, and at what value. Which of the following types of analysis would this practice be considered?

A. Prescriptive
B. Trend
C. Gap
D. Custer

Answer: D
Explanation:
Customer segmentation is the practice of dividing a customer base into smaller groups based on common characteristics such as demographics, buying behaviors, or psychographic traits. In this case, the analyst is looking at three characteristics: how recently a customer bought a product, how frequently they bought, and the value of their purchases. By grouping customers based on these characteristics, the company can tailor its promotional campaign to different segments, ensuring that the right message and offer is being communicated to the right people.

Customer segmentation is considered a descriptive analysis, as it involves looking at the data and grouping it into meaningful categories. This type of analysis is often used in marketing and sales to better understand customer behavior and develop targeted strategies for customer engagement and retention. By using customer segmentation, companies can more effectively allocate their marketing resources and create more personalized experiences for their customers.

QUESTION 12
A publishing group has requested a dashboard to track submissions before publication. A key requirement is that all changes are tracked, as multiple users will be checking out documents and editing them before submissions are considered final. Which of the following is the BEST way to meet this stakeholder requirement?

A. Display the version number next to each submission on the dashboard.

B. Present a data refresh date at the top of the dashboard.
C. Confirm the dashboard is adhering to the corporate style guide.
D. Use permissions to ensure users only see certain versions of the submissions.

Answer: A
Explanation:
Display the version number next to each submission on the dashboard. This option meets the stakeholder requirement by allowing them to track all changes by tracking the version number of each submission. The version number would increment with each change, allowing the user to easily see the latest version of each document and keep track of changes over time. This would be the best option to meet the requirement of tracking all changes.

QUESTION 13
The number of phone calls that call center receives in a day is an example of:

A. continuous data.
B. categorical data.
C. ordinal data.
D. discrete data.

Answer: D
Explanation:
The number of phone calls that a call center receives in a day is an example of discrete data. Discrete data refers to data that can only take on specific, distinct values within a given range. In this case, the number of phone calls can only be a whole number, and cannot take on any other value. For example, the call center might receive 100 calls one day, and 105 calls the next day, but it could not receive 100.5 calls. In contrast, continuous data refers to data that can take on any value within a given range, such as measurements or temperatures. Categorical data refers to data that can be divided into categories or groups, and ordinal data refers to data that has a specific order or ranking.

QUESTION 14
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

Answer: A
Explanation:
A static report is a report that is created using a snapshot of data at a specific point in time, and the data does not change unless the report is manually updated. In this case, the data analyst should create a static report for the second-quarter 2020 board meeting, which will be held on July 15, 2020, after the numbers are finalized. This means that the report will be based on data from the second quarter of 2020 that has already been collected and processed, and the data will not change unless the report is manually updated. The other options (real-time, self-service, and dynamic) are not applicable in this situation because the report is not being created for a live event or for real-time data, and it is not being created for self-service or dynamic use.

QUESTION 15
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

Answer: A
Explanation:
Non-personally identifiable information (NPI) is information that cannot be used to identify an individual. In this case, the cell phone device name would be considered NPI because it is not tied to a specific individual. The other options (customer's name, government ID number, and telephone number) are all personally identifiable information (PII) because they can be used to identify an individual. PII includes any information that can be used to distinguish or trace an individual's identity, either alone or when combined with other information that is linked or linkable to a specific individual.

QUESTION 16
Which of the following is the correct data type for text?

A. Boolean
B. String
C. Integer
D. Float

Answer: B
Explanation:
A string is a data type that represents a sequence of characters, such as letters, numbers, and symbols. Strings are often used to store and manipulate text data, such as names, addresses, and descriptions. The other options (Boolean, integer, and float) are not appropriate data types for text. A Boolean data type represents a true or false value, and an integer data type represents a whole number. A float data type represents a decimal number.

Thank You for Trying Our Product

Lead2pass Certification Exam Features:

★ More than **99,900** Satisfied Customers Worldwide.

★ Average **99.9%** Success Rate.

★ **Free Update** to match latest and real exam scenarios.

★ **Instant Download** Access! No Setup required.

★ Questions & Answers are downloadable in **PDF** format and **VCE** test engine format.

★ Multi-Platform capabilities - **Windows, Laptop, Mac, Android, iPhone, iPod, iPad**.

★ **100%** Guaranteed Success or **100%** Money Back Guarantee.

★ **Fast**, helpful support **24x7**.

View list of all certification exams: http://www.lead2pass.com/all-products.html

10% Discount Coupon Code: ASTR14