

CompTIA

Exam Questions DA0-002

CompTIA Data+ Exam (2025)



NEW QUESTION 1

A data analyst encounters an issue with new software and a code that they are using. The analyst includes print statements in the code to try to identify the issue, without success. An informal peer review of the code also produces the same result. The analyst confirms that the software is updated to the latest version and compatible with the code. Which of the following troubleshooting steps should the analyst take next?

- A. Use the old software and preexisting code, since both were functional.
- B. Contact the IT department and inform them that the software has a bug.
- C. Escalate to the department manager and ask for assistance.
- D. Research the issue online and see if a solution is available.

Answer: D

Explanation:

This question pertains to the Data Governance domain, focusing on troubleshooting and maintaining data quality in software processes. The analyst has already tried basic debugging and confirmed compatibility, so the next step involves seeking external resources.

? Use the old software and preexisting code, since both were functional (Option A):

Reverting to old software avoids solving the issue and may introduce other risks (e.g., security vulnerabilities).

? Contact the IT department and inform them that the software has a bug (Option B):

Assuming a bug without further investigation is premature, especially since compatibility was confirmed.

? Escalate to the department manager and ask for assistance (Option C): Escalation

is a later step after exhausting technical troubleshooting options.

? Research the issue online and see if a solution is available (Option D): Researching online (e.g., forums, documentation) is a logical next step to find solutions or identify known issues, especially after local debugging fails.

The DA0-002 Data Governance domain includes "data quality control concepts," and researching online is a standard troubleshooting step to maintain data process integrity. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance.

=====

NEW QUESTION 2

A sales manager wants a dashboard that shows sales aggregated by region and identifies high-volume sales by salesperson per region. Which of the following communication techniques best displays this information?

- A. Defined parameters
- B. Filter options
- C. Level of detail
- D. User persona

Answer: B

Explanation:

This question pertains to the Visualization and Reporting domain, focusing on dashboard features for displaying data. The dashboard needs to show aggregated sales by region and allow identification of high-volume sales by salesperson within each region.

? Defined parameters (Option A): Parameters set specific values (e.g., a date range), but they don't directly enable interaction with aggregated data.

? Filter options (Option B): Filter options allow the user to select a region and then view salespersons within that region, enabling the identification of high-volume sales per region interactively.

? Level of detail (Option C): Level of detail determines the granularity of data shown but doesn't facilitate interactive exploration.

? User persona (Option D): User personas guide dashboard design but aren't a communication technique for displaying data.

The DA0-002 Visualization and Reporting domain emphasizes "translating business requirements to form the appropriate visualization," and filter options best enable the interactive analysis required.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 4.0 Visualization and Reporting.

=====

NEW QUESTION 3

A data analyst needs to get an accurate idea of how data components are automated. Which of the following types of documentation should the analyst review first?

- A. Data flow diagram
- B. Data explainability report
- C. Data dictionary
- D. Data lineage

Answer: A

Explanation:

This question pertains to the Data Concepts and Environments domain, focusing on documentation for understanding data processes. The analyst needs to understand automation of data components, which involves data movement and processes.

? Data flow diagram (Option A): A data flow diagram (DFD) visualizes how data moves through systems, including automated processes, making it the best starting point.

? Data explainability report (Option B): This is related to AI/ML model transparency, not data automation.

? Data dictionary (Option C): A data dictionary defines data elements, not how they're automated.

? Data lineage (Option D): Data lineage tracks data origin and transformations but doesn't focus on automation processes.

The DA0-002 Data Concepts and Environments domain includes understanding "data schemas and dimensions," and a data flow diagram is key for visualizing automation. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.

=====

NEW QUESTION 4

Which of the following best represents a type of infrastructure that requires a company to purchase and maintain all of its own servers?

- A. Private
- B. Cloud
- C. Hybrid
- D. Public

Answer: A

Explanation:

This question pertains to the Data Concepts and Environments domain, focusing on types of server infrastructure. The task is to identify an infrastructure where a company owns and maintains all servers.

? Private (Option A): A private infrastructure (often on-premises) means the company owns and maintains its own servers, typically in a private data center, which matches the requirement.

? Cloud (Option B): Cloud infrastructure is managed by third-party providers, not owned by the company.

? Hybrid (Option C): Hybrid combines on-premises and cloud, so not all servers are owned by the company.

? Public (Option D): Public infrastructure is a cloud model shared across multiple organizations, not owned by the company.

The DA0-002 Data Concepts and Environments domain includes understanding "data environments," and a private infrastructure requires the company to purchase and maintain its own servers.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.

=====

NEW QUESTION 5

A data analyst is gathering data from multiple tables in a database. The analyst needs certain columns from each table. Which of the following is the best method to accomplish this task?

- A. Aggregate
- B. Union
- C. Nest
- D. Join

Answer: D

Explanation:

This question falls under the Data Acquisition and Preparation domain, focusing on combining data from multiple tables. The analyst needs specific columns from each table, suggesting a method to combine data horizontally based on relationships.

? Aggregate (Option A): Aggregation (e.g., SUM, COUNT) summarizes data, not suitable for combining columns from tables.

? Union (Option B): Union stacks tables vertically, requiring identical structures, but the analyst needs specific columns, likely based on relationships, not a vertical stack.

? Nest (Option C): Nesting is used for hierarchical data (e.g., JSON), not for combining relational tables.

? Join (Option D): A join (e.g., INNER JOIN) combines tables horizontally based on a common key, allowing the analyst to select specific columns from each table, which fits the task.

The DA0-002 Data Acquisition and Preparation domain includes "executing data manipulation," and joining tables is the best method for combining specific columns from multiple tables.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 2.0 Data Acquisition and Preparation.

=====

NEW QUESTION 6

A data analyst needs to provide a weekly sales report for the Chief Financial Officer. Which of the following delivery methods is the most appropriate?

- A. A granular daily report in a dashboard
- B. A detailed text document
- C. A spreadsheet with raw data
- D. A high-level email

Answer: D

Explanation:

This question pertains to the Visualization and Reporting domain, focusing on report delivery methods for a specific audience. The Chief Financial Officer (CFO) needs a weekly sales report, suggesting a concise, executive-level summary.

? A granular daily report in a dashboard (Option A): Daily granularity is too frequent for a weekly report, and a dashboard might be too interactive for a CFO's needs.

? A detailed text document (Option B): A detailed document is too lengthy for a CFO, who typically needs a summary.

? A spreadsheet with raw data (Option C): Raw data requires further analysis, which isn't appropriate for an executive-level report.

? A high-level email (Option D): A high-level email provides a concise summary of weekly sales, tailored for an executive like a CFO, making it the most appropriate delivery method.

The DA0-002 Visualization and Reporting domain emphasizes "translating business requirements to form the appropriate visualization," and a high-level email is best for delivering a weekly summary to a CFO.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 4.0 Visualization and Reporting.

NEW QUESTION 7

The human resources department wants to know the number of employees who earn \$125,000 or more. However, the department is concerned about duplicates in the dataset. Given the following table:

Employee_ID	Level	Salary
001	1	10000
002	2	

20000
003
2
256000
004
2
125000
001
1
10000
002
2
20000
Which of the following SQL statements resolves this issue?

- A. SELECT DISTINCT Employee_ID FROM Employee WHERE Salary >= 125000
- B. SELECT COUNT(DISTINCT Employee_ID) FROM Employee WHERE Salary >= 125000
- C. SELECT DISTINCT Employee_ID FROM Employee WHERE Salary > 125000
- D. SELECT COUNT(Employee_ID) FROM Employee WHERE Salary >= 125000

Answer: B

Explanation:
This question falls under the Data Analysis domain, focusing on SQL queries to handle duplicates while counting employees. The task is to count unique employees with a salary of \$125,000 or more, addressing duplicates in the dataset.
? Option A: SELECT DISTINCT Employee_ID FROM Employee WHERE Salary >= 125000 This lists unique Employee_IDs but doesn't provide a count, which the department needs.
? Option B: SELECT COUNT(DISTINCT Employee_ID) FROM Employee WHERE Salary >= 125000 This counts unique Employee_IDs (using DISTINCT) with a salary of \$125,000 or more, correctly addressing duplicates and providing the required count (2 employees: 003 and 004).
? Option C: SELECT DISTINCT Employee_ID FROM Employee WHERE Salary > 125000 This lists unique Employee_IDs with a salary strictly greater than \$125,000 (missing 004), and doesn't provide a count.
? Option D: SELECT COUNT(Employee_ID) FROM Employee WHERE Salary >= 125000 This counts all rows without addressing duplicates, resulting in an incorrect count (2 rows, but only 2 unique employees).
The DAO-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods using SQL queries," and COUNT(DISTINCT) is the correct method to count unique employees while handling duplicates.
Reference: CompTIA Data+ DAO-002 Draft Exam Objectives, Domain 3.0 Data Analysis.
=====

NEW QUESTION 8

A data analyst calculated the average score per student without making any changes to the following table:
Student Subject Score
123 Math 100
123 Biology 80
234 Math 96
123 Biology 80
345 Biology 88
234 Math 96
Which of the following exploration techniques should the analyst have considered before calculating the average?

- A. Duplication
- B. Redundancy
- C. Binning
- D. Grouping

Answer: A

Explanation:
This question pertains to the Data Governance domain, focusing on data quality issues that affect analysis. The table contains duplicate rows, which would skew the average score calculation if not addressed.
? Student 123: Math (100), Biology (80), Biology (80)– Duplicate Biology score.
? Student 234: Math (96), Math (96)– Duplicate Math score.
? Student 345: Biology (88)– No duplicates.
? Duplication (Option A): The table has duplicate rows (e.g., Student 123's Biology score of 80 appears twice), which would inflate the average if not removed. The analyst should have checked for duplicates before calculating the average.
? Redundancy (Option B): Redundancy refers to unnecessary fields (e.g., storing the same data in multiple columns), not duplicate rows.
? Binning (Option C): Binning groups data into categories, not relevant for addressing duplicates in averaging.
? Grouping (Option D): Grouping (e.g., GROUP BY in SQL) might be part of the solution, but the issue to identify is duplication.
The DAO-002 Data Governance domain includes "data quality control concepts," and checking for duplication is critical to ensure accurate calculations like averages. Reference: CompTIA Data+ DAO-002 Draft Exam Objectives, Domain 5.0 Data Governance.
=====

NEW QUESTION 9

Which of the following data sources makes online data consumption easier?

- A. Data mart
- B. Web scraping
- C. Database
- D. Application programming interface

Answer: D

Explanation:

This question pertains to the Data Concepts and Environments domain, focusing on data sources that facilitate online data access. The task is to identify a source that simplifies online data consumption.

? Data mart (Option A): A data mart stores structured data for specific business areas, typically accessed internally, not designed for online consumption.

? Web scraping (Option B): Web scraping extracts data from websites but requires parsing and cleaning, which isn't necessarily "easier."

? Database (Option C): Databases store data but aren't inherently designed for online consumption without an interface.

? Application programming interface (Option D): An API provides a structured way to access data online, often in formats like JSON, making data consumption easier for applications and users.

The DA0-002 Data Concepts and Environments domain includes understanding "data sources," and APIs are specifically designed to simplify online data access.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.

=====

NEW QUESTION 10

An analyst is building a reporting deck. The deck must include tracking and visualizing metrics and row-level security. Which of the following actions should the analyst take after meeting all of the requirements?

- A. Show a mock-up to the team.
- B. Explain the desired level of reporting detail.
- C. Present an analysis of the data.
- D. Find out the project due date.

Answer: A

Explanation:

This question pertains to the Visualization and Reporting domain, focusing on the process of building a reporting deck. After meeting requirements (tracking metrics, visualizing data, and row-level security), the next step involves validation with stakeholders.

? Show a mock-up to the team (Option A): Creating a mock-up and presenting it to the team ensures alignment on design and functionality before finalizing, which is a standard next step in report development.

? Explain the desired level of reporting detail (Option B): This should have been done earlier during requirements gathering, not after meeting them.

? Present an analysis of the data (Option C): Data analysis might inform the deck, but the task focuses on building the deck, not presenting analysis.

? Find out the project due date (Option D): The due date should have been established earlier in the project planning phase.

The DA0-002 Visualization and Reporting domain includes "translating business requirements to form the appropriate visualization," and showing a mock-up ensures the reporting deck meets stakeholder expectations.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 4.0 Visualization and Reporting.

=====

NEW QUESTION 10

Which of the following explains the purpose of UAT?

- A. To begin the software application development process to enhance user experience
- B. To ensure all parts of the software application work together after each sprint
- C. To review software application crashes, create patches, and deploy to users
- D. To validate and verify that a software application meets the needs and requirements of users

Answer: D

Explanation:

This question is related to the Data Governance domain of DA0-002, which includes understanding processes like User Acceptance Testing (UAT) to ensure data-related applications meet governance and quality standards. UAT is a critical step in ensuring software aligns with user needs and organizational requirements.

? To begin the software application development process to enhance user experience (Option A): UAT occurs near the end of development, not at the beginning.

? To ensure all parts of the software application work together after each sprint (Option B): This describes integration testing, not UAT, which focuses on user validation.

? To review software application crashes, create patches, and deploy to users (Option C): This refers to post-deployment maintenance, not UAT.

? To validate and verify that a software application meets the needs and requirements of users (Option D): UAT is specifically designed to ensure the software meets user requirements and functions as intended in a real-world scenario, aligning with governance standards for quality.

The DA0-002 Data Governance domain emphasizes "data quality control concepts" (similar to DA0-001, web ID: 1), which include ensuring applications meet user needs through processes like UAT.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance

=====

NEW QUESTION 13

A data company needs a visualization that shows the availability zones from the last ten years and any future availability zones that the company will be using in the next five years. Which of the following is the most appropriate visualization to display this information?

- A. Bar chart
- B. Mosaic plot
- C. Map
- D. Pie chart

Answer: C

Explanation:

This question falls under the Visualization and Reporting domain of CompTIA Data+ DA0-002, focusing on selecting the appropriate visualization for a specific dataset. The task is to show availability zones over a 15-year period (past and future), which involves a geographical element since availability zones are typically location-based.

? Bar chart (Option A): Bar charts are good for comparing categorical data but don't effectively show geographical locations or time-based trends across zones.

? Mosaic plot (Option B): Mosaic plots display relationships between categorical variables, not suitable for geographical or time-series data.

? Map (Option C): A map can display availability zones geographically, with annotations or layers to show changes over time (past 10 years and future 5 years), making it the most appropriate visualization.

? Pie chart (Option D): Pie charts show proportions of a whole, not suitable for geographical or time-based data.

The DA0-002 Visualization and Reporting domain emphasizes "translating business requirements to form the appropriate visualization," and a map is the best choice for displaying geographical availability zones over time.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 4.0 Visualization and Reporting.

=====

NEW QUESTION 15

The sales department wants to include the composition of total sales amounts across all three sales channels in a report. Given the following sample sales table:

Sales channel	Month	Sales (million \$)
Digital	January	135
Store	February	145
Online	March	165
Store	April	200
Store	May	125
Online	June	155
Digital	July	120
Online	August	145
Digital	September	160

Which of the following visualizations is the most appropriate?

- A. Pivot table
- B. Pie chart
- C. KPI card
- D. Box plot

Answer: B

Explanation:

This question pertains to the Visualization and Reporting domain, focusing on selecting the appropriate visualization for a specific requirement. The task is to show the composition of total sales across three channels, which involves showing proportions.

? Pivot table (Option A): A pivot table summarizes data but isn't a visualization; it's more for data exploration.

? Pie chart (Option B): A pie chart shows the proportion of total sales for each channel (Digital, Store, Online), which is ideal for displaying composition.

? KPI card (Option C): A KPI card displays a single metric, not suitable for showing composition across multiple channels.

? Box plot (Option D): A box plot shows data distribution (e.g., quartiles), not proportions.

The DA0-002 Visualization and Reporting domain emphasizes "translating business requirements to form the appropriate visualization," and a pie chart is best for showing the composition of totals.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 4.0 Visualization and Reporting.

=====

NEW QUESTION 17

The following SQL code returns an error in the program console: SELECT firstName, lastName, SUM(income) FROM companyRoster SORT BY lastName, income

Which of the following changes allows this SQL code to run?

- A. SELECT firstName, lastName, SUM(income) FROM companyRoster HAVING SUM(income) > 10000000
- B. SELECT firstName, lastName, SUM(income) FROM companyRoster GROUP BY firstName, lastName
- C. SELECT firstName, lastName, SUM(income) FROM companyRoster ORDER BY firstName, income
- D. SELECT firstName, lastName, SUM(income) FROM companyRoster

Answer: B

Explanation:

This question falls under the Data Analysis domain, focusing on SQL query correction. The query uses an aggregate function (SUM) but has two issues: it uses "SORT BY" (incorrect syntax) and lacks a GROUP BY clause for non-aggregated columns.

? The query selects firstName, lastName, and SUM(income), but firstName and lastName are not aggregated, requiring a GROUP BY clause.

? "SORT BY" is incorrect; the correct syntax is "ORDER BY."

? Option A: SELECT firstName, lastName, SUM(income) FROM companyRoster HAVING SUM(income) > 10000000 This adds a HAVING clause but doesn't fix the GROUP BY issue, so it's still invalid.

? Option B: SELECT firstName, lastName, SUM(income) FROM companyRoster GROUP BY firstName, lastName This adds the required GROUP BY clause for firstName and lastName, fixing the aggregation error. While it removes the ORDER BY, the query will run without it, addressing the primary error.

? Option C: SELECT firstName, lastName, SUM(income) FROM companyRoster

ORDER BY firstName, income This fixes "SORT BY" to "ORDER BY" but doesn't address the missing GROUP BY, so the query remains invalid.

? Option D: SELECT firstName, lastName, SUM(income) FROM

companyRoster This removes the ORDER BY but still lacks the GROUP BY clause, making it invalid.

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods using SQL queries," and adding GROUP BY fixes the aggregation error, allowing the query to run.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.
=====

NEW QUESTION 19

Which of the following elements is the most important to include in a dashboard for internal technical audiences?

- A. Methodology section
- B. Dynamic features
- C. Key performance indicators
- D. Company branding

Answer: C

Explanation:

This question pertains to the Visualization and Reporting domain, focusing on dashboard design for specific audiences. Internal technical audiences (e.g., data analysts, IT staff) need actionable, data-driven insights.

? Methodology section (Option A): Methodology is important for research reports, not dashboards, especially for technical audiences who prioritize data.

? Dynamic features (Option B): Dynamic features (e.g., interactivity) are useful but not the most critical element for technical audiences.

? Key performance indicators (Option C): KPIs provide critical metrics (e.g., system uptime, error rates) that technical audiences need to monitor and act on, making this the most important element.

? Company branding (Option D): Branding is more relevant for external audiences, not internal technical ones.

The DA0-002 Visualization and Reporting domain emphasizes "translating business requirements to form the appropriate visualization," and KPIs are essential for technical dashboards.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 4.0 Visualization and Reporting.
=====

NEW QUESTION 24

Before distributing a report, a marketing analyst notices that the total distinct promotional email messages is less than the combined total of emails sent. Which of the following is the most likely reason for this difference?

- A. The aggregation did not include all emails.
- B. Some emails were not delivered.
- C. The report failed to run properly.
- D. A recipient received duplicate emails.

Answer: D

Explanation:

This question falls under the Data Analysis domain, focusing on analyzing discrepancies in data reports. The total distinct messages are fewer than the total emails sent, indicating a specific issue.

? The aggregation did not include all emails (Option A): If the aggregation missed emails, the total sent would be lower, not the distinct count.

? Some emails were not delivered (Option B): Undelivered emails would reduce the total sent, but the scenario implies the total sent is accurate.

? The report failed to run properly (Option C): A report failure would likely cause broader issues, not a specific discrepancy between distinct and total counts.

? A recipient received duplicate emails (Option D): If recipients received duplicates, the total emails sent would be higher than the distinct messages (unique email content), explaining the difference.

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods," and identifying duplicates is a common analysis task to explain such discrepancies.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.
=====

NEW QUESTION 29

A data analyst is creating a forecast for a product line introduced early last year. Which of the following should the analyst use to create projected sales and customer satisfaction for next year?

- A. Standard deviation and constraints
- B. Mean and median
- C. Boolean data and an array
- D. Numerical and ordinal attributes

Answer: D

Explanation:

This question pertains to the Data Analysis domain, focusing on data types and methods for forecasting. The task involves projecting sales (numerical) and customer satisfaction (likely ordinal, e.g., ratings), requiring appropriate data attributes.

? Standard deviation and constraints (Option A): Standard deviation measures data spread, and constraints are conditions, neither of which directly supports forecasting.

? Mean and median (Option B): Mean and median are descriptive statistics, not sufficient for forecasting future values.

? Boolean data and an array (Option C): Boolean data (true/false) and arrays (data structures) are not relevant for forecasting sales and satisfaction.

? Numerical and ordinal attributes (Option D): Sales are numerical (e.g., units sold), and customer satisfaction is often ordinal (e.g., 1-5 ratings). These attributes are suitable for forecasting models (e.g., time-series analysis for sales, regression for satisfaction).

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods," and numerical and ordinal attributes are key for forecasting sales and satisfaction.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.
=====

NEW QUESTION 32

Which of the following data repositories stores unaltered data?

- A. Data lake
- B. Data warehouse
- C. Data table
- D. Data factory

Answer: A

Explanation:

This question falls under the Data Concepts and Environments domain, focusing on data repositories. The task is to identify a repository that stores data in its original, unaltered form.

? Data lake (Option A): A data lake stores raw, unaltered data in its native format (structured, semi-structured, or unstructured), making it the correct choice.

? Data warehouse (Option B): A data warehouse stores processed, structured data, often transformed for analysis, not unaltered.

? Data table (Option C): A data table is a structure within a database, not a repository, and may contain altered data.

? Data factory (Option D): A data factory (e.g., Azure Data Factory) is a data integration service, not a repository for storing data.

The DA0-002 Data Concepts and Environments domain includes understanding "different types of databases and data repositories," and a data lake is designed to store unaltered data.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.

=====

NEW QUESTION 37

A data analyst must combine service calls into low-, medium-, and high-priority levels in order to analyze organizational responses. Which of the following techniques should the analyst use for this task?

- A. Augmentation
- B. Imputation
- C. Scaling
- D. Binning

Answer: D

Explanation:

This question pertains to the Data Analysis domain, focusing on techniques for categorizing data. The task involves grouping service calls into priority levels (low, medium, high), which requires segmenting numerical or ordinal data into discrete categories.

? Augmentation (Option A): Augmentation involves adding data (e.g., in machine learning), not categorizing existing data.

? Imputation (Option B): Imputation fills in missing values, not relevant for categorizing priority levels.

? Scaling (Option C): Scaling adjusts numerical data to a common range (e.g., normalization), not suitable for creating priority categories.

? Binning (Option D): Binning groups continuous or ordinal data into discrete categories (e.g., assigning calls to low, medium, or high priority based on a metric like response time), which fits the task.

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods," and binning is a standard technique for categorizing data for analysis. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.

=====

NEW QUESTION 42

A data analyst needs to create a combined report that includes information from the following two tables:

Managers table

ID

First_name Last_name Job_title 1001

John Doe Manager 1002

Jane Roe Director

Non-managers table

ID First_name Last_name Job_title 1003

Robert Roe

Business Analyst 1004

Jane Doe

Sales Representative 1005

John Roe

Operations Analyst

Which of the following query methods should the analyst use for this task?

- A. Group
- B. Join
- C. Union
- D. Nested

Answer: C

Explanation:

This question pertains to the Data Acquisition and Preparation domain, focusing on combining data from two tables. Both tables have the same structure (ID, First_name, Last_name, Job_title) and contain employee data, suggesting the task is to create a single list of all employees.

? Group (Option A): Grouping (e.g., GROUP BY in SQL) is for aggregation (e.g., counting employees by job title), not combining tables into a single report.

? Join (Option B): Joining tables (e.g., INNER JOIN) requires a common key and combines tables horizontally, but there's no indication of a relationship between the tables (e.g., no shared key beyond ID, which isn't linked).

? Union (Option C): UNION combines the rows of two tables with the same structure into a single result set, removing duplicates, which is ideal for creating a combined report of all employees from both tables.

? Nested (Option D): Nested queries (e.g., subqueries) are used for complex filtering, not for combining tables into a single list.

The DA0-002 Data Acquisition and Preparation domain includes "executing data manipulation," and UNION is the best method for combining two tables with identical structures into a single report.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 2.0 Data Acquisition and Preparation.

NEW QUESTION 43

A data analyst is creating a new dataset that involves bringing together the following datasets:

Name ID
Date of birth
Frank 23525
3/19
Martha 11290
6/13
Ellen 12141
11/4
ID
Address City State 23525
1234 Harding Chicago
IL 11040
935 Terrace Hills Chino
CA 11290
2 Speedway Miami
FL

Which of the following would be the output if the data analyst does a FULL JOIN?

- A. Name IDDate of birth Address CityState Frank 235253/191234 Harding ChicagoILMartha 112906/13935 Terrace Hills ChinoCAEllen 1214111/42 Speedway MiamiFL
- B. Name IDDate of birth Address CityState Frank 235253/191234 Harding ChicagoILMartha112906/13935 Terrace Hills ChinoCAEllen 1214111/42 Speedway MiamiFL
- C. Name IDDate of birth Address CityState Frank 235253/191234 Harding ChicagoILMartha 112906/132 Speedway MiamiFLEllen 1214111/4935 Terrace Hills ChinoCA
- D. Name IDDate of birth Address CityState Frank 235253/191234 Harding ChicagoILMartha 112906/132 Speedway MiamiFLEllen 1214111/411040935 Terrace Hills ChinoCA

Answer: D

Explanation:

This question falls under theData Concepts and Environmentsdomain, focusing on database operations like joins. A FULL JOIN combines all rows from both tables, including matches and non-matches, filling in NULLs where there??s no corresponding data.

? The first table has IDs: 23525 (Frank), 11290 (Martha), 12141 (Ellen).

? The second table has IDs: 23525, 11040, 11290.

? A FULL JOIN includes all IDs: 23525, 11290, 12141, 11040.

? Option A: Incorrect; it includes a row for Ellen with "2 Speedway," but Ellen??s ID (12141) doesn??t match any address, and 11040 is missing.

? Option B: Identical to Option A, so incorrect for the same reasons.

? Option C: Incorrect; it mismatches addresses (e.g., Ellen with 935 Terrace Hills, which belongs to 11040).

? Option D: Correct; it includes all IDs, with NULLs for non-matching rows (Ellen has no address, and 11040 has no name).

The DA0-002 Data Concepts and Environments domain includes understanding "data schemas and dimensions," such as performing joins in relational databases.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.

=====

NEW QUESTION 48

Given the following dataset:

Day
Number of Guests
Monday 455
Tuesday 346
Wednesday 382
Thursday 563
Friday 887
Saturday 934
Sunday 346

Which of the following is the mode?

- A. 346
- B. 446
- C. 455
- D. 559

Answer: A

Explanation:

This question falls under theData Analysisdomain, focusing on statistical measures. The mode is the value that appears most frequently in a dataset.

? Monday: 455

? Tuesday: 346

? Wednesday: 382

? Thursday: 563

? Friday: 887

? Saturday: 934

? Sunday: 346

The value 346 appears twice (Tuesday and Sunday), while all other values (455, 382, 563, 887, 934) appear once. Thus, the mode is 346.

? Option A: 346– Correct, as it??s the most frequent value.

? Option B: 446– Incorrect, as 446 doesn??t appear in the dataset.

? Option C: 455– Incorrect, as 455 appears only once.

? Option D: 559– Incorrect, as 559 doesn??t appear in the dataset.

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods," and the mode is a fundamental measure of central tendency. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.

NEW QUESTION 50

A data analyst troubleshoots a dashboard every day for a week. Which of the following techniques best addresses how to validate the data moving forward?

- A. Inquiring about structure changes
- B. Setting up monitoring alerts
- C. Reaching out to users daily
- D. Rebuilding the dashboard

Answer: B

Explanation:

This question pertains to the Data Governance domain, focusing on ensuring data quality and reliability in dashboards over time. Daily troubleshooting indicates a recurring issue, and the task is to validate data moving forward.

? Inquiring about structure changes (Option A): This might identify past issues but doesn't provide ongoing validation.

? Setting up monitoring alerts (Option B): Monitoring alerts can automatically notify the analyst of data issues (e.g., missing updates, errors), providing a proactive way to validate data continuously.

? Reaching out to users daily (Option C): This is inefficient and reactive, not a sustainable validation method.

? Rebuilding the dashboard (Option D): Rebuilding might fix current issues but doesn't ensure future validation.

The DA0-002 Data Governance domain includes "data quality control concepts," such as implementing monitoring to ensure data reliability in dashboards.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance.

NEW QUESTION 55

A user needs a report that shows the main causes of customer churn rate in a three-year period. Which of the following methods provides this information?

- A. Inferential
- B. Descriptive
- C. Prescriptive
- D. Predictive

Answer: B

Explanation:

This question falls under the Data Analysis domain, focusing on analytical methods for reporting. The task is to identify the causes of customer churn over three years, which involves analyzing historical data.

? Inferential (Option A): Inferential statistics make predictions or generalizations about a population, not focused on identifying causes in historical data.

? Descriptive (Option B): Descriptive analytics summarizes historical data to identify patterns and causes (e.g., reasons for churn), which fits the task.

? Prescriptive (Option C): Prescriptive analytics provides recommendations, which goes beyond identifying causes.

? Predictive (Option D): Predictive analytics forecasts future outcomes, not focused on historical causes.

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods," and descriptive analytics is best for identifying causes in historical data.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.

NEW QUESTION 58

Which of the following best describes the semi-structured data that is gathered when web scraping?

- A. JSON
- B. CSV
- C. CSS
- D. HTML

Answer: A

Explanation:

This question pertains to the Data Acquisition and Preparation domain, which in DA0-002 includes understanding data acquisition concepts and the types of data gathered from various sources, such as web scraping. Web scraping involves extracting data from websites, and the data gathered is often semi-structured, meaning it has some organizational structure but isn't fully relational like a database table.

? JSON (Option A): JSON (JavaScript Object Notation) is a semi-structured data format commonly used in web applications. Web scraping often retrieves data in JSON format via APIs or embedded scripts, as it's lightweight and structured with key-value pairs, making it ideal for semi-structured data.

? CSV (Option B): CSV (Comma-Separated Values) is a structured format typically used for tabular data. It's not commonly the direct output of web scraping, though scraped data might be converted to CSV later.

? CSS (Option C): CSS (Cascading Style Sheets) is used for styling web pages and isn't a data format, making it irrelevant for describing scraped data.

? HTML (Option D): HTML (HyperText Markup Language) is the structure of web pages and is often the raw format scraped during web scraping. While HTML is semi-structured due to its tag-based hierarchy, it's primarily a markup language, not a data format, and the actual data extracted is often parsed into formats like JSON.

The DA0-002 Data Acquisition and Preparation domain aligns with the DA0-001 focus on "data acquisition concepts" (web ID: 14), which includes identifying formats like JSON as semi-structured data commonly acquired through web scraping. JSON is the best fit here due to its prevalence in web data exchange.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 2.0 Data Acquisition and Preparation

NEW QUESTION 60

Which of the following best describes the reason an analyst would reference a data dictionary versus a source's metadata?

- A. To gather information and resources about the data
- B. To find the content and specific attributes for a dataset
- C. To find a summary of basic information about the dataset
- D. To gather information about the availability of the data

Answer: B

Explanation:

This question is part of the Data Concepts and Environments domain, focusing on the purpose of data documentation tools like data dictionaries and metadata. The question compares their uses.

? To gather information and resources about the data (Option A): This is too vague and not specific to a data dictionary's purpose.

? To find the content and specific attributes for a dataset (Option B): A data dictionary provides detailed definitions of data elements (e.g., field names, types, descriptions), which is more specific than metadata, which often includes broader information like creation date or source.

? To find a summary of basic information about the dataset (Option C): This better describes metadata, which provides high-level summaries, not detailed attributes.

? To gather information about the availability of the data (Option D): Neither a data dictionary nor metadata typically focuses on data availability.

The DAO-002 Data Concepts and Environments domain includes understanding "data schemas and dimensions," and a data dictionary is specifically used to find detailed attributes of a dataset.

Reference: CompTIA Data+ DAO-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.

=====

NEW QUESTION 64

Due to new reporting requirements, a data analyst must add new classification codes to historical data. Which of the following is the best technique for this task?

- A. Append
- B. Binning
- C. Parsing
- D. Union

Answer: A

Explanation:

This question falls under the Data Acquisition and Preparation domain, focusing on modifying historical data. The task is to add new classification codes to existing data, which involves adding new rows or columns.

? Append (Option A): Appending adds new rows to a dataset, which is suitable if the classification codes are new records (e.g., a new table of codes to combine with historical data). If the codes are a new column, a join or update might be used, but append fits the context of adding new data.

? Binning (Option B): Binning groups data into categories, not suitable for adding classification codes.

? Parsing (Option C): Parsing breaks down data (e.g., splitting strings), not relevant for adding codes.

? Union (Option D): Union stacks tables with identical structures, but the task involves adding new data (codes) to historical data, not combining identical tables.

The DAO-002 Data Acquisition and Preparation domain includes "executing data manipulation," and appending is a common technique for adding new data to historical datasets.

Reference: CompTIA Data+ DAO-002 Draft Exam Objectives, Domain 2.0 Data Acquisition and Preparation.

=====

NEW QUESTION 65

A data analyst is preparing a survey for Paralympic Games athletes. Which of the following should the analyst consider when creating this survey?

- A. Idioms
- B. Color contrast
- C. Refresh speed
- D. Granularity

Answer: B

Explanation:

This question pertains to the Visualization and Reporting domain, focusing on survey design considerations, particularly for accessibility. The survey is for Paralympic athletes, who may have visual impairments, requiring specific design considerations.

? Idioms (Option A): Idioms (e.g., colloquial phrases) might confuse non-native speakers, but they're not a primary survey design concern for Paralympic athletes.

? Color contrast (Option B): High color contrast ensures readability for athletes with visual impairments (e.g., color blindness), a critical accessibility consideration for Paralympic surveys.

? Refresh speed (Option C): Refresh speed is relevant for dashboards, not static surveys.

? Granularity (Option D): Granularity refers to data detail levels, not a survey design consideration.

The DAO-002 Visualization and Reporting domain includes "translating business requirements to form the appropriate visualization," and color contrast is a key accessibility factor in survey design for diverse audiences.

Reference: CompTIA Data+ DAO-002 Draft Exam Objectives, Domain 4.0 Visualization and Reporting.

=====

NEW QUESTION 70

A data analyst receives a new data source that contains employee IDs, job titles, dates of birth, addresses, years of service, and employees' birth months. Which of the following inconsistencies should the analyst identify?

- A. Redundancy
- B. Equivalence

- C. Parallel
- D. Duplication

Answer: A

Explanation:

This question falls under the Data Governance domain, focusing on identifying data quality issues. The dataset includes dates of birth and birth months, which suggests a potential inconsistency.

? Redundancy (Option A): The dataset includes both dates of birth (e.g., 1990-05-

15) and birth months (e.g., May), which is redundant because the birth month can be derived from the date of birth, indicating a data quality issue.

? Equivalence (Option B): Equivalence isn't a standard data quality term in this

context; it might refer to data matching, which isn't the issue here.

? Parallel (Option C): Parallel isn't a recognized data quality term; it might relate to processing, not data inconsistencies.

? Duplication (Option D): Duplication refers to identical records, but the issue here is redundant fields, not duplicate rows.

The DA0-002 Data Governance domain includes "data quality control concepts," and redundancy is a key inconsistency when the same information is stored in multiple forms unnecessarily.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance.

=====

NEW QUESTION 73

Which of the following data repositories stores unstructured and structured data?

- A. Data store
- B. Data silo
- C. Data mart
- D. Data lake

Answer: D

Explanation:

This question falls under the Data Concepts and Environments domain of CompTIA Data+ DA0-002, which involves understanding different types of data repositories and their characteristics. The task is to identify a repository that can store both unstructured and structured data.

? Data store (Option A): A data store is a general term for any data repository, but

it's not specific enough to confirm it stores both unstructured and structured data.

? Data silo (Option B): A data silo is an isolated data repository, often structured, and not typically designed for unstructured data.

? Data mart (Option C): A data mart is a subset of a data warehouse, focused on structured data for specific business areas, not unstructured data.

? Data lake (Option D): A data lake is a centralized repository that stores raw data in its native format, including both structured (e.g., tables) and unstructured (e.g., text, images) data, making it the correct choice.

The DA0-002 Data Concepts and Environments domain includes understanding "different types of databases and data repositories," and a data lake is specifically designed to handle both unstructured and structured data.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.

=====

NEW QUESTION 78

Which of the following best explains the purpose of data lineage?

- A. To see the steps and path of data flow through different systems
- B. To better understand the granularity of data variable relationships
- C. To track data transformations from acquisition through reporting
- D. To look up data definitions, ensuring consistent use across business units

Answer: C

Explanation:

This question pertains to the Data Concepts and Environments domain, focusing on the purpose of data lineage. Data lineage involves tracking the lifecycle of data.

? To see the steps and path of data flow through different systems (Option A): This describes a data flow diagram, not data lineage, which focuses on transformations rather than just flow.

? To better understand the granularity of data variable relationships (Option B): This relates to data modeling, not the purpose of data lineage.

? To track data transformations from acquisition through reporting (Option C): Data lineage tracks the journey of data, including transformations (e.g., cleaning, aggregation) from its source to its final use in reporting, which is its primary purpose.

? To look up data definitions, ensuring consistent use across business units (Option D): This describes a data dictionary, not data lineage.

The DA0-002 Data Concepts and Environments domain includes understanding "data schemas and dimensions," and data lineage specifically tracks transformations across the data lifecycle.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.

NEW QUESTION 81

A data analyst pulls a table similar to the following one:

ID

Type ID Phone 1

Full Time Full Time 1 Mobile

2

Part Time Part Time 2 Work

3

Full Time Full Time 3 Mobile

Which of the following best explains the data issue with Type ID?

- A. Redundancy
- B. Outlier
- C. Missing data

D. Duplication

Answer: A

Explanation:

This question is part of the Data Concepts and Environments domain, focusing on identifying data quality issues. The table shows Type and TypeID columns, where TypeID seems to repeat information from Type with an additional identifier.

? Redundancy (Option A): The TypeID column (e.g., "Full Time 1") redundantly includes the Type value ("Full Time") with an extra identifier, which is unnecessary and could be simplified by using a numeric ID instead.

? Outlier (Option B): Outliers are data points that deviate significantly, which isn't applicable here.

? Missing data (Option C): There are no missing values in the table.

? Duplication (Option D): Duplication refers to identical rows, but the rows here are unique; the issue is with the column content.

The DA0-002 Data Concepts and Environments domain includes understanding "data schemas and dimensions," and redundancy is a common data quality issue in schema design.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.

=====

NEW QUESTION 84

A sales manager wants to understand how sales are trending year over year. Which of the following chart types is the most appropriate to display the information?

- A. Line
- B. Donut
- C. Bubble
- D. Hierarchy

Answer: A

Explanation:

This question falls under the Visualization and Reporting domain, focusing on selecting the appropriate visualization for a specific data trend. The task is to show sales trends over time (year over year).

? Line (Option A): Line charts are ideal for displaying trends over time, such as year-over-year sales, as they clearly show changes and patterns across a continuous time axis.

? Donut (Option B): Donut charts show proportions or percentages of a whole, not suitable for time-based trends.

? Bubble (Option C): Bubble charts display three dimensions of data (e.g., size, x-axis, y-axis), not ideal for simple time trends.

? Hierarchy (Option D): Hierarchy charts (e.g., treemaps) show nested relationships, not time-based trends.

The DA0-002 Visualization and Reporting domain emphasizes "translating business requirements to form the appropriate visualization," and a line chart is best for time-series trends.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 4.0 Visualization and Reporting.

=====

NEW QUESTION 87

Which of the following is the most efficient to use when programming repeatable tasks?

- A. LLM
- B. Deep learning
- C. NLP
- D. RPA

Answer: D

Explanation:

This question pertains to the Data Acquisition and Preparation domain, focusing on tools for automating repeatable tasks in data processes. The task is to identify the most efficient programming method.

? LLM (Option A): Large Language Models (e.g., GPT) are for text generation, not efficient for repeatable data tasks.

? Deep learning (Option B): Deep learning is for complex pattern recognition (e.g., image classification), not efficient for simple repeatable tasks.

? NLP (Option C): Natural Language Processing is for text analysis, not general repeatable tasks.

? RPA (Option D): Robotic Process Automation (RPA) automates repetitive, rule-based tasks (e.g., data entry, file processing), making it the most efficient for programming repeatable tasks.

The DA0-002 Data Acquisition and Preparation domain includes "executing data manipulation," and RPA is a widely used method for automating repeatable data tasks. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 2.0 Data Acquisition and Preparation.

=====

NEW QUESTION 90

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

DA0-002 Practice Exam Features:

- * DA0-002 Questions and Answers Updated Frequently
- * DA0-002 Practice Questions Verified by Expert Senior Certified Staff
- * DA0-002 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- * DA0-002 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The DA0-002 Practice Test Here](#)