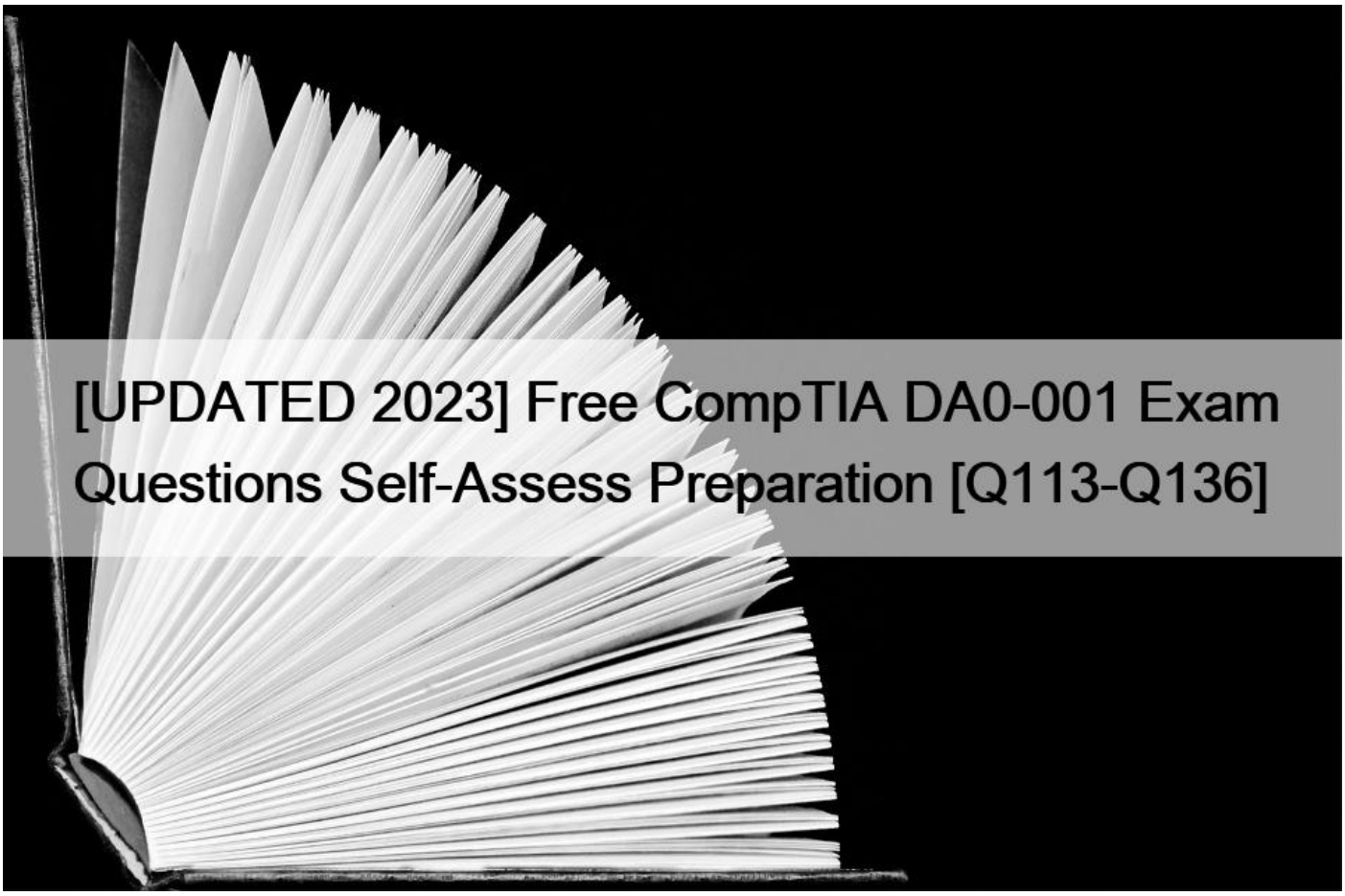
[UPDATED 2023 Free CompTIA DA0-001 Exam Questions Self-Assess Preparation [Q113-Q136



[UPDATED 2023] Free CompTIA DA0-001 Exam Questions Self-Assess Preparation
DA0-001 Free Sample Questions to Practice One Year Update

Objectives of CompTIA DA0-001 Exam

In this section, we will list the objectives you should know in order to pass your CompTIA Data+ certification exam.
The first objective is to understand how to become a successful IT professional with data skills.
The second objective is to learn about different types of data, including structured and unstructured data.
The third objective is to be able to demonstrate your knowledge of at least two types of data storage: online and offline.
The fourth objective is to be able to describe how a client can access information from various sources through technology such as social media or the web. **CompTIA DA0-001 exam dumps** he best way to prepare for the exams.
The fifth objective is for you to be able to apply concepts related to using software applications such as Microsoft Office suite, Google Docs, and Adobe Creative Cloud.

Things You Should Know about the CompTIA DA0-001 Exam

CompTIA Data+ is a certification that validates your understanding of the fundamentals of computer data storage and management. It's a foundational certification that can be used to secure entry-level positions in various fields, such as IT, business intelligence, and application development.

CompTIA Data+ is not the only certification available for this subject matter. The CompTIA A+ and Microsoft Certified Solutions Associate exams also cover similar material. However, unlike those exams, CompTIA D+ does not have a pre-requisite requirement for passing it. If you already have a working knowledge of networking, operating systems (OS), and some basic computer concepts, you can prepare for this exam without any prior experience in these areas. **CompTIA DA0-001 exam dumps** will help you to clear the concepts.
CompTIA Data+ is designed to ensure that candidates have a thorough understanding of how data storage works and the basic concepts involved in managing it. This includes an understanding of how information is stored electronically on hard drives, solid-state drives (SSDs), magnetic tapes, optical discs and other storage devices; along with how data can be retrieved from these devices using appropriate software tools.

Q113. An analyst modified a data set that had a number of issues. Given the original and modified versions:

Original data:

Var001	Var002	Var003	Var004
1	0	0	0
0	1	0	1
1	1	1	2
0	0	0	1

Modified data:

Var001	Var002	Var003	Var004
Yes	Absent	No payment	No
No	Present	No payment	Yes
Yes	Present	Payment	Maybe
No	Absent	No payment	Yes

Which of the following data manipulation techniques did the analyst use?
* Imputation
* Recoding
* Parsing
* Deriving

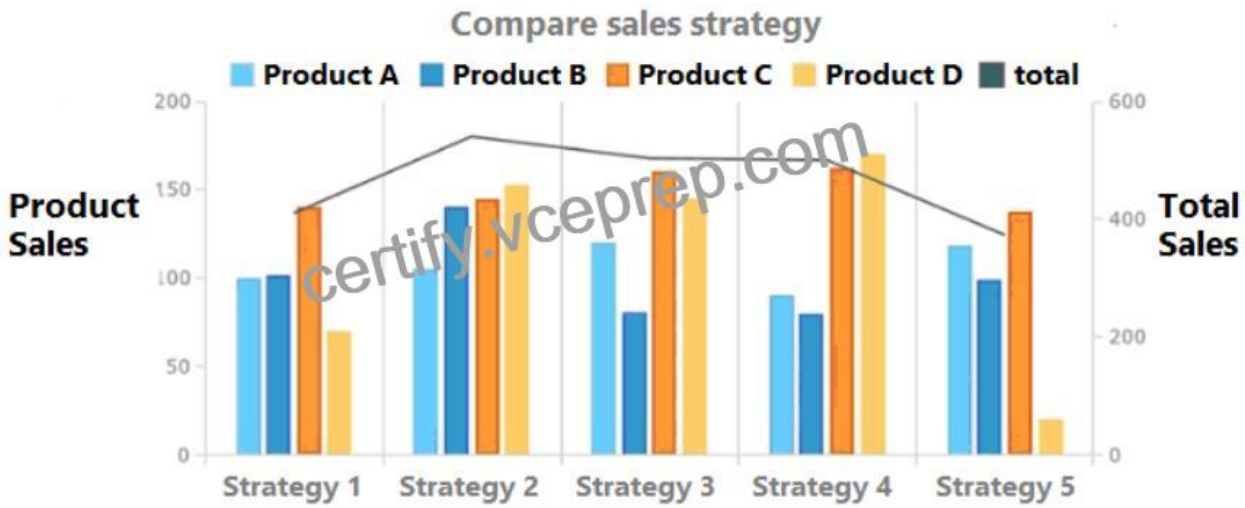
Explanation

The correct answer is B. Recoding.

Recoding is a data manipulation technique that involves changing the values or categories of a variable to make it more suitable for analysis. Recoding can be used to simplify or group the data, to correct errors or inconsistencies, or to create new variables from existing ones12 In the example, the analyst used recoding to change the values of Var001, Var002, Var003, and Var004 from numerical to textual form. The analyst also used recoding to assign meaningful labels to the values, such as

“Absent” for 0, “Present” for 1, “Low” for 2, “Medium” for 3, and “High” for 4. This makes the data more understandable and easier to analyze.

Q114. Given the following graph:



Which of the following summary statements upholds integrity in data reporting?
* Sales are approximately equal for Product A and Product B across all strategies.
* Strategy 4 provides the best sales in comparison to other strategies.
* While Strategy 2 does not result in the highest sales of Product D, over all products it appears to be the most effective.
* Product D should be promoted more than the other products in all strategies.

Q115. Jenny wants to study the academic performance of undergraduate sophomores and wants to determine the average grade point average at different points during an academic year.

What best describes the data set she needs?
* Sample.
* Observation.
* Variable.
* Population.

Explanation

Correct answer A. Sample.

Jenny does not have data for the entire population of all undergraduate sophomores. While a specific grade point average is an observation of variable, jenny needs sample data.

Q116. A cereal manufacturer wants to determine whether the sugar content of its cereal has increased over the years.

Which of the following is the appropriate descriptive statistic to use?
* Frequency
* Percent change
* Variance
* Mean

Explanation

This is because percent change is a type of descriptive statistic that measures the relative change or difference of a variable over

time, such as the sugar content of cereal over years in this case. Percent change can be used to determine whether the sugar content of cereal has increased over years by comparing the initial and final values of the sugar content, as well as calculating the ratio or proportion of the change. For example, percent change can be used to determine whether the sugar content of cereal has increased over years by finding out how much more (or less) sugar there is in cereal now than before, as well as expressing it as a fraction or a percentage of the original sugar content. The other descriptive statistics are not appropriate to use to determine whether the sugar content of cereal has increased over years. Here is why:

Frequency is a type of descriptive statistic that measures how often or how likely a value or an event occurs in a data set, such as how many times a certain sugar content appears in cereal in this case.

Frequency does not measure the relative change or difference of a variable over time, but rather measures the occurrence or chance of a variable at a given time.

Variance is a type of descriptive statistic that measures how much the values in a data set vary or deviate from the mean or average of the data set, such as how much variation there is in sugar content among different cereals in this case. Variance does not measure the relative change or difference of a variable over time, but rather measures the dispersion or spread of a variable at a given time.

Mean is a type of descriptive statistic that measures the average value or central tendency of a data set, such as what is the typical sugar content of cereal in this case. Mean does not measure the relative change or difference of a variable over time, but rather measures the summary or representation of a variable at a given time.

Q117. You are a database administrator and you create a precompiled query that is stored on the database server.

What have you created?
* Query execution plan.
* Stored procedure.
* Parameterized query.
* Function as a service.

A stored procedure is a set of Structured Query Language (SQL) statements with an assigned name, which are stored in a relational database management system (RDBMS) as a group, so it can be reused and shared by multiple programs.

SQL Server stored procedure is a batch of statements grouped as a logical unit and stored in the database. The stored procedure accepts the parameters and executes the T-SQL statements in the procedure, returns the result set if any.

Q118. After completing web scraping, which of the following file formats needs to be parsed?
* .html
* .txt
* .csv
* .tsv

Explanation

The correct answer is .html.

Short explanation: Web scraping is the process of extracting data from websites by parsing the HTML code of the web pages. HTML stands for HyperText Markup Language and it is the standard markup language for creating web pages and web applications. HTML files have the extension .html and they contain tags, elements, attributes, and content that define the structure and appearance of a web page. Web scraping tools need to parse the HTML files to extract the relevant data from the web pages12

Q119. Which of the following best describes a business analytics tool with interactive visualization and business capabilities and an interface that is simple enough for end users to create their own reports and dashboards?

Python
* R
* Microsoft Power Bl
* SAS

Explanation

The best answer is C. Microsoft Power BI.

Microsoft Power BI is a business analytics and business intelligence service by Microsoft. It aims to provide interactive visualizations and business intelligence capabilities with an interface simple enough for end users to create their own reports and dashboards. Power BI can connect to multiple data sources, clean and transform data, create custom calculations, and visualize data through charts, graphs, and tables. Power BI can be accessed through a web browser, mobile device, or desktop application and integrated with other Microsoft tools like Excel and SharePoint12 Python is not correct, because Python is a general-purpose programming language that can be used for various applications, including data analysis and visualization. However, Python is not a dedicated business analytics tool, and it requires coding or programming skills to create reports and dashboards.

R is not correct, because R is a programming language and software environment for statistical computing and graphics. R can be used for data analysis and visualization, but it is not a specialized business analytics tool, and it requires coding or programming skills to create reports and dashboards.

SAS is not correct, because SAS is a software suite for advanced analytics, business intelligence, data management, and predictive analytics. SAS can provide interactive visualizations and business capabilities, but it does not have an interface that is simple enough for end users to create their own reports and dashboards.

SAS also requires coding or programming skills to use its features.

Q120. Which of the following describes the method of sampling in which elements of data are selected randomly from each of the small subgroups within a population?
* Simple random
* Cluster
* Systematic
* Stratified

Explanation

This is because stratified is a type of sampling in which elements of data are selected randomly from each of the small subgroups within a population, such as age groups, gender groups, or income groups. Stratified sampling can be used to ensure that the sample is representative and proportional of the population, as well as reduce the sampling error or bias. For example, stratified sampling can be used to select a sample of voters from different political parties based on their proportion in the population. The other types of sampling are not the types of sampling in which elements of data are selected randomly from each of the small subgroups within a population. Here is why:

Simple random is a type of sampling in which elements of data are selected randomly from the entire population, without dividing it into any subgroups. Simple random sampling can be used to ensure that every element in the population has an equal chance of being selected, as well as avoid any systematic error or bias. For example, simple random sampling can be used to select a sample of students from a school by using a lottery or a computer-generated number.

Cluster is a type of sampling in which elements of data are selected randomly from a few large subgroups within a population, such as regions, districts, or schools. Cluster sampling can be used to reduce the cost and complexity of sampling, as well as increase the feasibility and convenience of sampling. For example, cluster sampling can be used to select a sample of households from a few

neighborhoods by using a map or a list.

Systematic is a type of sampling in which elements of data are selected at regular intervals from an ordered list or sequence within a population, such as every nth element or every kth element. Systematic sampling can be used to simplify and speed up the sampling process, as well as ensure that the sample covers the entire range or scope of the population. For example, systematic sampling can be used to select a sample of books from a library by using an alphabetical order or a numerical order.

Q121. 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?
* $2,466.18
* $2,667.60
* $3,082.72
* $12,330.88

Q122. Which of the following statements would be used to append two tables that have the same number of columns?
* UNION ALL
* MERGE
* GROUP BY
* JOIN

Explanation

The correct answer is A. UNION ALL.

UNION ALL is a SQL statement that appends two tables that have the same number of columns and compatible data types. UNION ALL preserves all the rows from both tables, including any duplicates12 B: MERGE is not correct, because MERGE is a SQL statement that combines the data of two tables based on a common column. MERGE can perform insert, update, or delete operations on the target table depending on the matching or non-matching rows from the source table34 C: GROUP BY is not correct, because GROUP BY is a SQL clause that groups the rows of a table based on one or more columns. GROUP BY is often used with aggregate functions, such as SUM, AVG, COUNT, etc., to calculate summary statistics for each group56 D: JOIN is not correct, because JOIN is a SQL clause that combines the data of two tables based on a common column or condition. JOIN can produce different results depending on the type of join, such as INNER JOIN, LEFT JOIN, RIGHT JOIN, etc.

Q123. Data validation should occur only when data is initially brought into a organization.
* True.
* False.

Q124. You recently downloaded a file containing website visitor logs from your organization’s web server.

What term best describes these logs at this point in the process?

* Schema.
* Intelligence.
* Information.
* Data.

Q125. You are working with a professional statistician to perform an analysis and would like to use a statistics package.

Which one of the following would be the most appropriate?
* Rapid Miner.
* QLIK.
* Power BI.
* Minitab.

Explanation

Minitab is statistical analysis software. It can be used for learning about statistics as well as statistical research.

Statistical analysis computer applications have the advantage of being accurate, reliable, and generally faster than computing statistics and drawing graphs by hand.

Q126. Which of the following best describes an exploratory analysis?
* Involves the use of descriptive statistics to understand observations
* Involves analysis of exploring data sets for performance tracking
* Involves the testing of specific hypotheses
* Involves the use of arithmetic algebra to determine the distribution

Explanation

answer: A. Involves the use of descriptive statistics to understand observations.

Exploratory data analysis (EDA) is a method of analyzing and investigating data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods. EDA involves the use of descriptive statistics, such as mean, median, mode, standard deviation, frequency, or percentage, to understand the distribution, central tendency, variability, and relationship of the data. EDA helps to see what the data can reveal beyond the formal modeling or hypothesis testing, and provides a better understanding of data set variables and the interactions between them1.

Q127. A table in a hospital database has a column for patient height in inches and a column for patient height in centimeters. This is an example of:
* dependent data.
* duplicate data.
* invalid data
* redundant data

Q128. Which of the following contains alphanumeric values?
* $10.1E^2$
* 13.6
* 1347
* A3J7

Q129. A junior web developer is developing a new application where users can upload short videos. The first task is to create a homepage that shows the headline “Upload Your Short Videos” and a clickable button that says “upload now”.

Which of the following HTML commands would help the developer to complete the task successfully?
* < span >Upload Your Short Videos< /span >< button >upload now< /button >
* < p >Upload Your Short Videos< /p >< p >upload now< /p >
* < hl >Upload Your Short Videos< /h1 >< button >upload now< /button >
* < hl >Upload Your Short Videos< /h1 >< hl >upload now< /h1 >
The correct answer is: Upload Your Short Videos

upload now

The two tags are used to define HTML headings. defines the most important heading. defines the least important heading.

Note: Only use one per page – this should represent the main heading/subject for the whole page. The tag defines a clickable button.

Q130. A web developer wants to ensure that malicious users can’t type SQL statements when they asked for input, like their username/userid.

Which of the following query optimization techniques would effectively prevent SQL Injection attacks?
* Indexing.
* Subset of records.
* Temporary table in the query set.
* Parametrization.
The correct answer is D: Parametrization. Parameterized SQL queries allow you to place parameters in an SQL query instead of a constant value. A parameter takes a value only when the query is executed, allowing the query to be reused with different values and purposes. Parameterized SQL statements are available in some analysis clients, and are also available through the Historian SDK.

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

Q131. What type of regulation is the Payment Card Industry Data Security Standard (PCIDSS)?
* Industry regulation
* Civil law
* Administrative law
* Criminal law

Q132. Which of the following data manipulation techniques is an example of a logical function?
* WHERE
* AGGREGATE
* BOOLEAN
* IF

Q133.

Name	Gender_flag	Level	Code	Region
James	Male	College	P	ON
Paul	Female	Elementary	A	BC
Sean	Male	College	S	QC
Dan	Female	Elementary	A	BC
Sam	Male	Elementary	A	BC
Ahmed	Male	University	L	ON
Tom	Male	Elementary	A	BC
Kim	Male	Elementary	A	BC
Pat	Female	Elementary	A	BC
Ben	Male	Elementary	A	BC
Ken	Male	High school	D	AT

Which of the following logical statements results in Table B?

* `IF Name = "James" and Gender_flag = "College" then delete`

* `IF Name = "Sam" and Gender_flag = "M`

The logical statement that results in Table B is Option D. Option D is a logical statement that uses the AND operator to combine two conditions: Name = “Tom” and Region = “BC”. The AND operator returns true only if both conditions are true, otherwise it returns false. Therefore, Option D will select only the rows from Table A that satisfy both conditions, which are rows 4, 5, 6, and 7. These rows form Table B, as shown below:

Name | Gender flag | Level | College | Code | Region Tom | Male | Elementary | A | BC | BC Kim | Female | Elementary | A | BC | BC Pat | Female | Elementary | A | BC | BC Ben | Male | Elementary | A | BC | BC The other options are not correct, as they use different logical operators or conditions that do not result in Table B. Option A uses the OR operator, which returns true if either condition is true, or both. Option A will select all the rows from Table A except row 3, which does not match either condition. Option B uses the NOT operator, which returns the opposite of the condition. Option B will select all the rows from Table A except rows 4, 5, 6, and 7, which match the condition. Option C uses a different condition, Region = “ON”, which does not match any row in Table A. Option C will select no rows from Table A. Reference: [SQL Logical Operators – W3Schools]

Q134. Which of the following is a common data analytics tool that is also used as an interpreted, high-level, general-purpose programming language?
* SAS
* Microsoft Power B1
* IBM SPSS
* Python

Explanation

The option that is a common data analytics tool that is also used as an interpreted, high-level, general-purpose programming language is Python. Python is a popular and versatile programming language that can be used for various purposes, such as web development, software development, automation, machine learning, and data analysis. Python has many features and libraries that make it suitable for data analytics, such as its simple syntax, dynamic typing, multiple paradigms, built-in data structures, NumPy, pandas, matplotlib, scikit-learn, etc. The other options are not programming languages, but software applications or platforms that are used for data analytics or related tasks. SAS is a software suite that provides advanced analytics, business intelligence, data management, and predictive analytics capabilities. Microsoft Power BI is a business analytics service that provides interactive visualizations and business intelligence capabilities. IBM SPSS is a software package that offers statistical analysis, data mining, text analytics, and predictive analytics capabilities. Reference: Python For Data Analysis – DataCamp

Q135. Andrew conducts a study and wants to capture eye color.

What kind of data is eye color?

Choose the best response.
* Discrete.
* Categorical.
* Continuous.
* Alphanumeric.

Correct answer B. Categorical.

Eye color can only fall into a certain range of values; as such, it is categorical.

Q136. Under which of the following circumstances should the null hypothesis be accepted when a = 0.05?
* When p is 0.00003
* When p is 0.001
* When p is 0.04
* When p is 0.06

Explanation

The null hypothesis should be accepted when the p-value is greater than the alpha level, which is the significance level of the test. The p-value is the probability of obtaining a test statistic at least as extreme as the one observed in the sample, assuming that the null hypothesis is true. The alpha level is the probability of rejecting the null hypothesis when it is true, which is also known as a type I error12.

In this case, the alpha level is 0.05, which means that there is a 5% chance of rejecting the null hypothesis when it is true. Therefore, to reject the null hypothesis, the p-value must be less than or equal to 0.05, which indicates that the test statistic is very unlikely to occur by chance under the null hypothesis. Conversely, to accept the null hypothesis, the p-value must be greater than 0.05, which indicates that the test statistic is not very unlikely to occur by chance under the null hypothesis.

Among the four options, only option D has a p-value that is greater than 0.05 (p = 0.06). Therefore, option D is the correct answer. When p = 0.06, it means that there is a 6% chance of obtaining a test statistic at least as extreme as the one observed in the sample, assuming that the null hypothesis is true. This probability is not very low, and therefore does not provide enough evidence to reject the null hypothesis.