DA0-001^{Q&As}

CompTIA Data+

Pass CompTIA DA0-001 Exam with 100% Guarantee

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

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

100% Passing Guarantee
100% Money Back Assurance

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

- **Instant Download** After Purchase

- **100% Money Back** Guarantee

- **365 Days** Free Update

- **800,000+** Satisfied Customers

QUESTION 1

Daniel is using the structured Query language to work with data stored in relational database.

He would like to add several new rows to a database table.

What command should he use?

A. SELECT.

B. ALTER.

C. INSERT.

D. UPDATE.

Correct Answer: C

INSERT The INSERT command is used to add new records to a database table. The SELECT command is used to retrieve information from a database. It\\'s the most commonly used command in SQL because it is used to pose queries to the database and retrieve the data that you\\'re interested in working with. The UPDATE command is used to modify rows in the database. The CREATE command is used to create a new table within your database or a new database on your server.

QUESTION 2

Given the below:

		Conclusion from statistical analysis	
		Accept the null hypothesis	Reject the null hypothesis
The true state of nature	Null hypothesis is true	1	3
	Null hypothesis is false	2	4

Which of the following numbers represents a Type I error?

A. 1

B. 2

C. 3

D. 4

Correct Answer: C

QUESTION 3

The ACME Corporation hired an analyst to detect data quality issues in their Excel documents. Which of the following are the most common issues? (Select TWO)

A. Apostrophe.

B. Commas.

C. Symbols.

D. Duplicates.

E. Misspellings.

Correct Answer: DE

1.

 Duplicates

2.

 Misspellings

The most common data quality issues are difficult to resolve in Excel because of their rigidity. It forces analysts to do a ton of manual work, which results in a high probability of an error being introduced to the data set. Those common issues

include:

-Blanks

-Nulls

-Outliers

-Duplicates

-Extra spaces

-Misspellings

-Abbreviations and domain-specific variations

-Formula error codes

When introduced, these errors can skew or even invalidate the resulting analysis. A smart tool would minimize the possibility of error by automating the manual work. In Excel, you might look for data quality issues in one of two ways. First,

you might use auto filters on specific columns to scan for anomalies and blanks or you might use a pivot table to find gaps and discrepancies.

In either case, you\'re scanning for the anomalies yourself. Suffice it to say that\'s not a very efficient process. It also means accuracy is only as good as the analyst\'s eye, so the probability of error varies throughout the day.

QUESTION 4

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:

A. dependent data.

B. duplicate data.

C. invalid data

D. redundant data

Correct Answer: D

Explanation: This is because redundant data is a type of data that is unnecessary or irrelevant for the analysis or purpose, which can affect the efficiency and performance of the analysis or process. Redundant data can be caused by having multiple data fields that store the same or similar information, such as patient height in inches and patient height in centimeters in this case. Redundant data can be eliminated or reduced by using data cleansing techniques, such as removing or merging the redundant data fields. The other types of data are not examples of data that is unnecessary or irrelevant for the analysis or purpose. Here is what they mean in terms of data quality:

Dependent data is a type of data that relies on or is influenced by another data field or value, such as a formula or a calculation that uses other data fields or values as inputs or outputs. Dependent data can be useful or important for the analysis or purpose, as it can provide additional information or insights based on the existing data. Duplicate data is a type of data that is repeated or copied in a data set, which can affect the quality and validity of the analysis or process. Duplicate data can be caused by having multiple records or rows that have the same or similar values for one or more data fields or columns, such as customer ID or order ID. Duplicate data can be eliminated or reduced by using data cleansing techniques, such as removing or filtering out the duplicate records or rows. Invalid data is a type of data that is incorrect or inaccurate in a data set, which can affect the validity and reliability of the analysis or process. Invalid data can be caused by having values that do not match the expected format, type, range, or rule for a data field or column, such as an email address that does not have an @ symbol or a date that does not follow the YYYY-MM-DD format. Invalid data can be eliminated or reduced by using data cleansing techniques, such as validating or correcting the invalid values.

QUESTION 5

A data analyst is developing a data dictionary that aligns with a company\\'s data management processes and policies. Which of the following best describes what should be included in the data dictionary?

A. Information containing the links to business data

B. Information explaining the business methodologies

C. Information containing definitions of the business data

D. Information describing the data analysis phases

Correct Answer: C