

# CompTIA

## Exam Questions DA0-001

CompTIA Data+ Certification Exam



**NEW QUESTION 1**

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

**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.

**NEW QUESTION 2**

A financial institution is reporting on sales performance to a company at the account level. Due to the sensitive nature of the government the does il with, some account information is not shown. Which of the following fields should be masked?

- A. Sales volume
- B. Start date
- C. Product name
- D. Customer name

**Answer: D**

**Explanation:**

Customer name is the field that should be masked, because it contains sensitive information that could identify the government accounts that the financial institution deals with. Masking is a technique that replaces or obscures sensitive data with dummy or random data, such as asterisks or hashes. Masking can help protect the privacy and security of the data, while still allowing for some analysis and reporting. Therefore, the correct answer is D. References: [Data Masking | Definition, Techniques & Examples - Talend], [Data masking - Wikipedia]

**NEW QUESTION 3**

A user receives a large custom report to track company sales across various date ranges. The user then completes a series of manual calculations for each date range. Which of the following should an analyst suggest so the user has a dynamic, seamless experience?

- A. Create multiple reports, one for each needed date range.
- B. Build calculations into the report so they are done automatically.
- C. Add macros to the report to speed up the filtering and calculations process.
- D. Create a dashboard with a date range picker and calculations built in.

**Answer: D**

**Explanation:**

Create a dashboard with a date range picker and calculations built in. This is because a dashboard is a type of visualization that displays multiple charts or graphs on a single page, usually to provide an overview or summary of some data or information. A dashboard can be used to track company sales across various date ranges by showing different metrics and indicators related to sales, such as revenue, volume, or growth. By creating a dashboard with a date range picker and calculations built in, the analyst can suggest a way for the user to have a dynamic, seamless experience, which means that the user can interact with and customize the dashboard according to their needs or preferences, as well as avoid any manual work or errors. For example, a date range picker is a type of feature or function that allows users to select or adjust the time period for which they want to see the data on the dashboard, such as daily, weekly, monthly, or quarterly. A date range picker can make the dashboard dynamic, as it can automatically update or refresh the dashboard with new data based on the selected time period. Calculations are mathematical operations or expressions that can be performed on the data on the dashboard, such as addition, subtraction, multiplication, division, average, sum, etc. Calculations can make the dashboard seamless, as they can eliminate the need for manual calculations for each date range, as well as ensure accuracy and consistency of the results. The other ways are not the best ways to provide a dynamic, seamless experience for the user. Here is why:

? Creating multiple reports, one for each needed date range would not provide a dynamic, seamless experience for the user, but rather create a static, cumbersome experience, which means that the user cannot interact with or customize the reports according to their needs or preferences, as well as have to deal with multiple files or pages. For example, creating multiple reports would make it difficult for the user to compare or contrast the sales across different date ranges, as well as increase the workload and complexity of managing and maintaining the reports.

? Building calculations into the report so they are done automatically would not provide a dynamic, seamless experience for the user, but rather provide a partial, limited experience, which means that the user can only benefit from one aspect or feature of the report, but not from others. For example, building calculations into the report would help with avoiding manual work or errors, but it would not help with interacting with or customizing the report according to different date ranges.

? Adding macros to the report to speed up the filtering and calculations process would not provide a dynamic, seamless experience for the user, but rather provide an advanced, complex experience, which means that the user would need to have some technical skills or knowledge to use or apply the macros, as well as face some potential risks or challenges. For example, adding macros to the report would require the user to know how to write or run the macros, which are a type of code or script that automates certain tasks or actions on the report, such as

filtering or calculating the data. Adding macros to the report could also expose the user to some security or compatibility issues, such as viruses, malware, or errors.

#### NEW QUESTION 4

Which of the following concepts should be applied if a data set with 40 fields needs to be pared down to 20 fields and contains similar data across multiple fields?

- A. Duplication
- B. Consolidation
- C. Compliance
- D. Standardization

**Answer: B**

#### Explanation:

Consolidation is the process of combining multiple elements into a single, more effective or coherent whole. In the context of data analytics, consolidation would involve merging similar fields to reduce the overall number of fields in a dataset. This is particularly useful when a dataset contains redundant or similar data across multiple fields, as it helps to simplify the data structure and improve efficiency. Techniques such as dimensionality reduction are often applied to achieve this, where the goal is to retain the most informative and representative features of the data while reducing the number of total features. References:

? Applied Dimensionality Reduction — 3 Techniques using Python1.

? Seven Techniques for Data Dimensionality Reduction2.

? Best practices when working with datasets3.

? Effectively Handling Large Datasets4.

#### NEW QUESTION 5

An analysts building a monthly report for production and wants to ensure the audience is aware of its once-a-month cadence. Which of the following is the MOST important to convey that information?

- A. The date of the dashboard build
- B. The data refresh date
- C. A report summary
- D. Frequently asked questions

**Answer: A**

#### Explanation:

This is because the date of the dashboard build is the most important component to convey that information, which is the once-a-month cadence of the monthly report for production. The date of the dashboard build can convey that information by indicating when the dashboard was created or updated, as well as showing the frequency or interval of the dashboard creation or update. For example, the date of the dashboard build can convey that information by displaying a date format that includes the month and year, such as January 2020, February 2020, etc., or by displaying a text format that includes the word ??monthly??, such as Monthly Report for Production - January 2020, Monthly Report for Production - February 2020, etc. The other components are not the most important components to convey that information. Here is why:

? The data refresh date is a component that indicates when the data on the dashboard was refreshed or retrieved from the source or system, such as a database, a cloud service, or a web application. The data refresh date does not convey that information, but rather conveys how current or up-to-date the data on the dashboard is.

? A report summary is a component that provides an overview or a highlight of the main findings or insights from the dashboard, such as key metrics, indicators, or trends. A report summary does not convey that information, but rather conveys what the dashboard is about or what it shows.

? Frequently asked questions is a component that provides answers or explanations to common or expected questions from the audience or users of the dashboard, such as how to use or interpret the dashboard, what are the assumptions or limitations of the dashboard, etc. Frequently asked questions does not convey that information, but rather conveys how to understand or interact with the dashboard.

#### NEW QUESTION 6

A sales team wants visibility of current sales numbers, pipeline, and team performance. The team would also like to see calculations of individuals?? earned commissions and projected commissions based on sales, but they want that information to be kept confidential. Which of the following would be the BEST way to provide this visibility?

- A. Create a dashboard displaying a data refresh date so users know the current sales numbers and configure permissions to control access.
- B. Create a dashboard for sales numbers, pipeline, and team and individual performance for the management team.
- C. Create a dashboard with filters for the overall team, individuals, and managemen
- D. Users can filter to see the data they want.
- E. Create a dashboard with views for team, individuals, and managemen
- F. Configure permissions to control access.

**Answer: D**

#### Explanation:

Create a dashboard with views for team, individuals, and management. Configure permissions to control access. This is because a dashboard is a type of visualization that displays multiple charts or graphs on a single page, usually to provide an overview or summary of some data or information. A dashboard can be used to provide visibility of current sales numbers, pipeline, and team performance by showing different metrics and indicators related to these aspects. By creating a dashboard with views for team, individuals, and management, the analyst can customize the content and layout of the dashboard for different audiences and purposes. By configuring permissions to control access, the analyst can ensure that the confidential information, such as individuals?? earned commissions and projected commissions based on sales, is only visible to the authorized users. The other ways are not the best way to provide this visibility. Here is why: Creating a dashboard displaying a data refresh date so users know the current sales numbers and configuring permissions to control access would not be sufficient to provide visibility of pipeline and team performance, as well as individuals?? earned commissions and projected commissions based on sales. The dashboard would only show the current sales numbers and the date when the data was updated, which would not give a comprehensive or detailed view of the sales situation.

Creating a dashboard for sales numbers, pipeline, and team and individual performance for the management team would not be appropriate to provide visibility for the sales team, as they would not have access to the dashboard or the information they need. The dashboard would only be available for the management team, which would limit the transparency and collaboration among the sales team members.

Creating a dashboard with filters for the overall team, individuals, and management would not be secure to provide visibility of confidential information, such as individuals?? earned commissions and projected commissions based on sales. The dashboard would allow users to filter and see the data they want, which could

expose sensitive or personal information to unauthorized users.

#### NEW QUESTION 7

A site reliability team wants to monitor the stability of their website. so they can proactively diagnose issues when they occur Which of the following deliverables would best suit their needs?

- A. A self-serve dashboard of website performance that updates in real time
- B. A weekly log report of site visits and user actions
- C. A portal that is refreshed daily and reports errors classified by type
- D. A daily summary email indicating website outages for the previous day

**Answer:** A

#### Explanation:

The best deliverable that would suit the site reliability team's needs is A. A self-serve dashboard of website performance that updates in real time.

A self-serve dashboard is a visual display of the most important information needed to achieve one or more objectives, consolidated and arranged on a single screen so the information can be monitored at a glance. A self-serve dashboard of website performance that updates in real time would allow the site reliability team to easily and quickly access the information they need about the stability of their website, such as uptime, response time, error rate, traffic volume, etc. A self-serve dashboard would also enable the team to proactively diagnose issues when they occur, by providing alerts, notifications, or drill-down options. A self-serve dashboard would also be more interactive and engaging than a report or an email.

A weekly log report of site visits and user actions would not be a good deliverable for the site reliability team's needs, because it would not provide timely or relevant information about the stability of their website. A weekly log report would be too infrequent and delayed to monitor and diagnose issues when they occur.

A weekly log report would also focus on the behavior and actions of the users, rather than the performance and functionality of the website.

A portal that is refreshed daily and reports errors classified by type would not be a good deliverable for the site reliability team's needs, because it would not provide real-time or comprehensive information about the stability of their website. A portal that is refreshed daily would be too slow and outdated to monitor and diagnose issues when they occur. A portal that reports errors classified by type would be too narrow and limited to capture the full picture of the website performance.

A daily summary email indicating website outages for the previous day would not be a good deliverable for the site reliability team's needs, because it would not provide real-time or actionable information about the stability of their website. A daily summary email would be too late and retrospective to monitor and diagnose issues when they occur. A daily summary email indicating website outages would also be too passive and generic to help the team resolve or prevent issues in the future.

#### NEW QUESTION 8

Which of the following programming languages are best suited for analysis and machine- learning applications? (Select two).

- A. Ruby
- B. Rust
- C. PHP
- D. Python
- E. Kotlin
- F. R

**Answer:** DF

#### NEW QUESTION 9

Which of the following data cleansing issues will be fixed when a DISTINCT function is applied?

- A. Missing data
- B. Duplicate data
- C. Redundant data
- D. Invalid data

**Answer:** B

#### Explanation:

This is because duplicate data refers to data that is repeated or copied in a data set, which can affect the quality and validity of the analysis. A DISTINCT function is a type of function that removes duplicate values from a column or a table, leaving only unique values. For example, a DISTINCT function in SQL that can achieve this is:

```
SELECT DISTINCT column_name FROM table_name;
```

The other data cleansing issues will not be fixed by applying a DISTINCT function. Here is why:

Missing data refers to data that is absent or incomplete in a data set, which can affect the accuracy and reliability of the analysis. A DISTINCT function does not help with missing data, because it does not fill in or impute the missing values.

Redundant data refers to data that is unnecessary or irrelevant for the analysis, which can affect the efficiency and performance of the analysis. A DISTINCT function does not help with redundant data, because it does not remove or filter out the redundant values.

Invalid data refers to data that is incorrect or inaccurate in a data set, which can affect the validity and reliability of the analysis. A DISTINCT function does not help with invalid data, because it does not validate or correct the invalid values.

#### NEW QUESTION 10

A sales director has requested a report for individual team members within the division be developed. The director would like the report to be shared with all team members, but individual team members should not be identifiable within the report Which of the following access requirements would support the director's needs?

- A. Create an acceptable use policy for the sales data.
- B. Release the report as user-group-based access and include data masking.
- C. Get a data use agreement from the individual team members.



D. Provide the report based on role and include data encryption.

**Answer:** B

#### NEW QUESTION 10

A marketing analytics team received customer transaction data from two different sources. The data is complete and accurate; however, the field names appear to be inconsistent. Given the following tables:

Online transactions:

Customer_ID	Channel	Segment	Amount (\$)
001	Online	Existing	3,000
002	Online	Existing	4,000
003	Online	New	1,500

Store transactions:

Customer_ID	Source	Segment	Amount (\$)
001	In-store	New	1,000
004	In-store	Existing	4,000
005	In-store	New	3,500

Which of the following is considered best practice if the team wants to consolidate the files and conduct further analysis?

- A. Standardize the field names.
- B. Recode the data values.
- C. Overwrite the field names in one of the tables.
- D. Edit the field names in the data dictionary.

**Answer:** A

#### Explanation:

When consolidating data from different sources, it is crucial to standardize field names to ensure consistency across datasets. This process involves aligning the field names so that they are the same in both tables, which simplifies the merging of data and subsequent analysis. Standardizing field names helps in maintaining data integrity and avoids confusion that may arise from having different names for the same data point. Recode the data values (B) would not be necessary unless the data values themselves are inconsistent or in different formats. Overwriting the field names in one of the tables © could lead to loss of information or confusion. Editing the field names in the data dictionary (D) is helpful, but it does not address the immediate need to harmonize the field names in the actual datasets.

References:

? Best practices in data management.

? Principles of data integration and consolidation.

#### NEW QUESTION 11

A data analyst wants to create "Income Categories" that would be calculated based on the existing variable "Income". The "Income Categories" would be as follows:

Income category 1: less than \$1.

Income category 2: more than \$1 and less than \$20,000. Income category 3: more than \$20,001 and less than \$40,000. Income category 4: more than \$40,001.

Which of the following data manipulation techniques should the data analyst use to create "Income Categories"?

- A. Data merge
- B. Derived variables
- C. Data blending
- D. Data append

**Answer:** B

#### Explanation:

The correct answer is B: Derived variables Derived variables are variables that you create by calculating or categorizing variables that already exist in your data set.

Data merge is incorrect. Data merging is the process of combining two or more data sets into a single data set. Data blending is incorrect.

Data blending involves pulling data from different sources and creating a single, unique, dataset for visualization and analysis.

Data append is incorrect. A data append is a process that involves adding new data elements to an existing database.

#### NEW QUESTION 13

Which of the following data manipulation techniques is an example of a logical function?

- A. WHERE
- B. AGGREGATE
- C. BOOLEAN
- D. IF

**Answer:** D

#### Explanation:

This is because an IF function is a type of logical function that returns a value based on a condition or a set of conditions. An IF function can be used to

manipulate data by applying different actions or calculations depending on whether the condition is true or false. For example, an IF function in Excel that can achieve this is:

=IF (condition, value\_if\_true, value\_if\_false)

The other data manipulation techniques are not examples of logical functions. Here is why:

? WHERE is a type of clause that filters data based on a condition or a set of conditions. A WHERE clause can be used to manipulate data by selecting only the rows that satisfy the condition(s). For example, a WHERE clause in SQL that can achieve this is:

```
SELECT column_name FROM table_name WHERE condition;
```

? AGGREGATE is a type of function that performs a calculation on a group of values, such as sum, average, count, etc. An AGGREGATE function can be used to manipulate data by summarizing or aggregating the values in a column or a table. For example, an AGGREGATE function in SQL that can achieve this is:

```
SELECT AGGREGATE(column_name) FROM table_name;
```

? BOOLEAN is a type of data type that represents two possible values: true or false.

A BOOLEAN data type can be used to manipulate data by storing or returning logical values based on a condition or a set of conditions. For example, a BOOLEAN data type in Python that can achieve this is:

```
boolean_variable = condition
```

#### NEW QUESTION 15

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 text, symbols, numbers, or punctuation marks. Strings are enclosed in quotation marks, such as ??Hello??. ??123??. or ??!@#??. Strings can be manipulated, concatenated, sliced, indexed, formatted, and searched using various methods and functions. A string is different from other data types, such as boolean, integer, or float, which represent logical values (true or false), whole numbers, or decimal numbers respectively. Therefore, the correct answer is B. References: What is a String? | Definition and Examples, Python String Methods

#### NEW QUESTION 16

Given the table below:

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

Which of the following boxes indicates that a Type II error has occurred?

- A. 1
- B. 2
- C. 3
- D. 4

**Answer:** C

#### Explanation:

A Type II error is a false negative conclusion, which means failing to reject a null hypothesis that is actually false. In the table, box 3 indicates that a Type II error has occurred, because it shows that the null hypothesis is accepted when it is false in reality.

This means that the statistical test failed to detect a significant difference or relationship that actually exists. References: Type I & Type II Errors | Differences, Examples, Visualizations - Scribbr, Type I and type II errors - Wikipedia

#### NEW QUESTION 18

During data cleansing, an analyst conducts measures of central tendency on a data set. Which of the following data is the analyst attempting to identify?

- A. Duplicate
- B. Missing
- C. Outlying

D. Invalid

Answer: C

NEW QUESTION 21

Analytics reports should follow corporate style guidelines.

- A. True.
- B. False.

Answer: A

NEW QUESTION 23

A JSON file is an example of:

- A. structured data.
- B. web data.
- C. machine data.
- D. processed data.

Answer: A

Explanation:

A JSON (JavaScript Object Notation) file is a text-based format for representing structured data based on JavaScript object syntax. It is commonly used for transmitting data in web applications (e.g., sending some data from the server to the client, so it can be displayed on a web page, or vice versa). JSON files are human-readable and can be interpreted by various programming languages, making them ideal for data interchange<sup>123</sup>. JSON files typically contain an array of objects, with each object representing a record with a series of name-value pairs. This structured format is both easy to understand and write by humans and easy for machines to parse and generate<sup>4</sup>.

References:

- ? JSON??s official definition and syntax rules<sup>1</sup>.
- ? A beginner??s guide to JSON and its data types<sup>2</sup>.
- ? Understanding the JSON file format<sup>3</sup>.
- ? Detailed explanation of JSON as a structured data format<sup>4</sup>.

NEW QUESTION 26

A data analyst received the information in the table below from a recently completed marketing campaign:

Channels	Clicks	Orders
Display	580	55
PPC	800	100
Social	1,200	220
Mobile	300	60
SEO	620	85

Which of the following is the total order conversion rate?

- A. 13.2%
- B. 14.8%
- C. 22.3%
- D. 85.2%

Answer: B

Explanation:

The correct answer is A. 13.2%.  
The total order conversion rate is the ratio of the total number of orders to the total number of clicks, expressed as a percentage. To calculate the total order conversion rate, we need to sum up the clicks and orders from all the channels, and then divide the orders by the clicks and multiply by 100.  
Using the data from the table, we can do the following:  
? Total clicks = 580 + 800 + 1,200 + 300 + 620 = 3,500  
? Total orders = 55 + 100 + 220 + 60 + 85 = 520  
? Total order conversion rate = (520 / 3,500) x 100 = 14.857%  
? Rounding to one decimal place, we get 14.9% Therefore, the total order conversion rate is 14.9%.

NEW QUESTION 31



Which of the following will MOST likely be streamed live?

- A. Machine data
- B. Key-value pairs
- C. Delimited rows
- D. Flat files

**Answer:** A

**Explanation:**

Machine data is the most likely type of data to be streamed live, as it refers to data generated by machines or devices, such as sensors, web servers, network devices, etc. Machine data is often produced continuously and in large volumes, requiring real-time processing and analysis. Other types of data, such as key-value pairs, delimited rows, and flat files, are more likely to be stored in databases or files and processed in batches.

**NEW QUESTION 35**

An analyst is reviewing the following data: Car IDSpeed

123155  
 566436  
 564418  
 650567  
 546436  
 645638

Which of the following should the analyst include in the measures of central tendency for speed?

- A. Mode = 38 Range = 31 Mean = 42.5
- B. Range = 49 Max = 67 Min = 18
- C. Mode = 36 Max = 67 Min = 18
- D. Mode = 36 Median = 37 Mean = 41.5

**Answer:** D

**Explanation:**

The measures of central tendency include the mode, median, and mean. The mode is the value that appears most frequently in a data set. In this case, the speed of 36 appears twice, making it the mode. The median is the middle value when a data set is ordered from least to greatest; for these speeds, when ordered (18, 36, 36, 38, 55, 67), the median is the average of the two middle numbers, which is  $(\frac{36 + 38}{2} = 37)$ . The mean is the average of all values, calculated as  $(\frac{18 + 36 + 36 + 38 + 55 + 67}{6} = 41.7)$ . References:

? The calculation of the mode, median, and mean is based on standard statistical formulas and definitions.

The measures of central tendency for speed include the mode, median, and mean. To calculate these, we first need to organize the data:

? Speeds in ascending order: 18, 36, 36, 38, 55, 67

? Mode is the value that appears most frequently, which is 36, as it appears twice.

? Median is the middle value when the data is ordered. Since we have an even number of observations, we take the average of the two middle values (36 and 38), resulting in 37.

? Mean is the sum of all values divided by the number of values.  $(18+36+36+38+55+67)/6=41.5$

Thus, the correct option is D, which includes Mode = 36, Median = 37, and Mean = 41.5. The range, maximum, and minimum values, although useful in understanding data dispersion, are not measures of central tendency and are therefore not relevant to this specific question.

**NEW QUESTION 38**

A data analyst has a set with more than 40.000 rows in the sample schema below:

Name	Birth date - sales system	Birth date - marketing system	Birth date - accounting system
Tom	1/4/1989		
Frank		7/5/1994	
Carrie		8/3/1973	
Joe			3/2/2001

The analyst would like to create one column that contains the customers?? birth dates. Which of the following data quality dimensions would BEST explain the reason for compilation?

- A. Data accuracy
- B. Data completeness
- C. Data duplication
- D. Data integrity

**Answer:** D

**Explanation:**

Data integrity is the dimension that measures the consistency and validity of data across different data sources. In this case, the data analyst wants to create one column that contains the customers?? birth dates, but the data is stored in different formats and locations in the sample schema. For example, some customers have their birth dates in the customer table, while others have their birth years in the sales table. To compile the data into one column, the data analyst needs to ensure that the data is consistent and valid across the tables. Therefore, data integrity is the best explanation for the reason for compilation. References: Data Quality Dimensions - DATAVERSITY, The 6 Data Quality Dimensions with Examples | Collibra



#### NEW QUESTION 40

Given the customer table below:

Customer_ID	Active_flag	Segment	Store_ID	Spend
004	N	Nursery	004C	\$7,000
009	Y	Prime	004A	\$2,000
008	N	Prime	004D	\$6,000
003	Y	Nursery	004U	\$1,000
002	Y	Prime	004S	\$2,000
001	N	Prime	004A	\$1,500
007	Y	Prime	004D	\$2,000

Which of the following chart types is the most appropriate to represent the average spending of active customers vs. inactive customers?

- A. Pie chart
- B. Heat graph
- C. Scatter plot
- D. Line chart

**Answer:** A

#### Explanation:

A Pie chart is the most suitable for representing the average spending of active customers versus inactive customers. Pie charts are effective for comparing parts of a whole, which makes them ideal for visually displaying the proportion of spend between two distinct groups. They are widely used to depict percentage distributions and are straightforward, allowing immediate analysis of the active vs. inactive customer spending distribution at a glance.

#### NEW QUESTION 42

Given the following table:

Code	New_Measure	Old_Measure
A	10	12
B	14	12
C	5	12
D	9	12

Which of the following methods is the best way to describe the changes in the values in the table?

- A. Average
- B. Range
- C. Standard deviation
- D. Median

**Answer:** B

#### NEW QUESTION 44

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?

- A. `<span>Upload Your Short Videos</span><button>upload now</button>`
- B. `<p>Upload Your Short Videos</p><p>upload now</p>`
- C. `<h1>Upload Your Short Videos</h1><button>upload now</button>`
- D. `<h1>Upload Your Short Videos</h1><h1>upload now</h1>`

**Answer:** C

#### Explanation:

The HTML commands that would help the developer to complete the task successfully are

`<h1>Upload Your Short Videos</h1>` and `<button>upload now</button>`. The `<h1>` tag defines a heading level 1, which is the largest and most important heading on a webpage. The `<button>` tag defines a clickable button that can perform some action when clicked. The other options are not suitable for the task, as they either use the wrong tags or do not create a clickable button. The `<span>` tag defines a section of text with no specific meaning or formatting. The `<p>` tag defines a paragraph of text. The `<h1>` tag does not exist in HTML. Reference: HTML Tags - W3Schools

#### NEW QUESTION 48

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

- A. A stratified phone survey of 100 people that is conducted between 2:00 p.
- B. and 3:00 p.m.
- C. A systematic survey that is sent to 100 single-family homes in the county
- D. Surveys sent to ten randomly selected homes within 5mi (8km) of the county??s office
- E. Surveys sent to 100 randomly selected homes that are reflective of the population

**Answer: D**

**Explanation:**

Surveys sent to 100 randomly selected homes that are reflective of the population. This is because a random sample is a type of sample that is selected by using a random method, such as a lottery or a computer-generated number, which ensures that every element in the population has an equal chance of being selected. A random sample can result in a representative sample, which means that the sample reflects the characteristics and diversity of the population. By sending surveys to 100 randomly selected homes that are reflective of the population, the analyst can ensure that the sample is representative of the county??s households and their income levels. The other sampling methods are not likely to result in a representative sample. Here is why:  
A stratified phone survey of 100 people that is conducted between 2:00 p.m. and 3:00 p.m. would result in a biased sample, which means that the sample favors or excludes certain groups or elements in the population. By conducting the survey only between 2:00 p.m. and 3:00 p.m., the analyst would miss out on people who are not available or reachable at that time, such as those who are working or sleeping. This could affect the representativeness and generalizability of the sample.  
A systematic survey that is sent to 100 single-family homes in the county would result in an unrepresentative sample, which means that the sample does not reflect the characteristics and diversity of the population. By sending surveys only to single-family homes, the analyst would ignore other types of households, such as apartments, condos, or mobile homes. This could affect the accuracy and reliability of the sample.  
Surveys sent to ten randomly selected homes within 5mi (8km) of the county??s office would result in a small sample, which means that the sample size is too low to capture the variability and diversity of the population. By sending surveys only to ten homes within a limited area, the analyst would miss out on many households that are located in different parts of the county. This could affect the precision and confidence of the sample.

**NEW QUESTION 50**

Which of the following techniques is used to quantify data?

- A. Decoding
- B. Enumeration
- C. Coding
- D. Structure

**Answer: C**

**Explanation:**

Answer C. Coding  
Coding is a technique that is used to quantify data, especially qualitative data that are not expressed numerically. Coding involves assigning codes, such as numbers, letters, symbols, or colors, to different categories or themes that emerge from the data. For example, if you have a set of survey responses that ask about the satisfaction level of customers, you can code them as follows:  
? Very satisfied = 5  
? Satisfied = 4  
? Neutral = 3  
? Dissatisfied = 2  
? Very dissatisfied = 1  
By coding the data, you can convert them into quantitative data that can be analyzed using statistical methods, such as calculating the mean, median, mode, frequency, or percentage of each category<sup>12</sup>.  
Option A is incorrect, as decoding is not a technique that is used to quantify data, but rather a process of interpreting or translating data from one form to another. For example, decoding can involve converting binary codes into text or images, or decrypting ciphertext into plaintext<sup>3</sup>.  
Option B is incorrect, as enumeration is not a technique that is used to quantify data, but rather a process of listing or naming data in a specific order. For example, enumeration can involve listing the names of the states in alphabetical order, or naming the planets in order of their distance from the sun<sup>4</sup>.  
Option D is incorrect, as structure is not a technique that is used to quantify data, but rather a property or characteristic of data that describes how they are organized or arranged. For example, structure can refer to the format, type, or schema of data, such as structured, semi-structured, or unstructured data.

**NEW QUESTION 52**

Which one of the following programming languages is specifically designed for use in analytics applications?

- A. Python.
- B. R
- C. C++
- D. Java.

**Answer: B**

**NEW QUESTION 54**

Which of the following is a non-parametric test?

- A. One-sample t-test
- B. Two-way ANOVA
- C. Correlation coefficient
- D. Spearman's rank correlation

**Answer: D**

**Explanation:**

The correct answer is D. Spearman??s rank correlation.  
Spearman??s rank correlation is a non-parametric test that measures the strength and direction of the relationship between two variables that are ranked (ordinal) or continuous. Spearman??s rank correlation does not assume that the data follows a normal distribution or that the variables are linearly related. Spearman??s

rank correlation is based on the ranks of the data rather than the actual values<sup>12</sup>

- \* A. One-sample t-test is not correct, because it is a parametric test that compares the mean of a sample to a specified value. One-sample t-test assumes that the data follows a normal distribution and has a known population standard deviation<sup>34</sup>
- \* B. Two-way ANOVA is not correct, because it is a parametric test that compares the means of two or more groups that are influenced by two independent factors. Two-way ANOVA assumes that the data follows a normal distribution, has homogeneous variances, and has independent observations.
- \* C. Correlation coefficient is not correct, because it is a parametric test that measures the strength and direction of the linear relationship between two continuous variables. Correlation coefficient assumes that the data follows a bivariate normal distribution and has no outliers.

#### NEW QUESTION 58

Which of the following best describes the law of large numbers?

- A. As a sample size decreases, its standard deviation gets closer to the average of the whole population.
- B. As a sample size grows, its mean gets closer to the average of the whole population
- C. As a sample size decreases, its mean gets closer to the average of the whole population.
- D. When a sample size double
- E. the sample is indicative of the whole population.

**Answer: B**

#### Explanation:

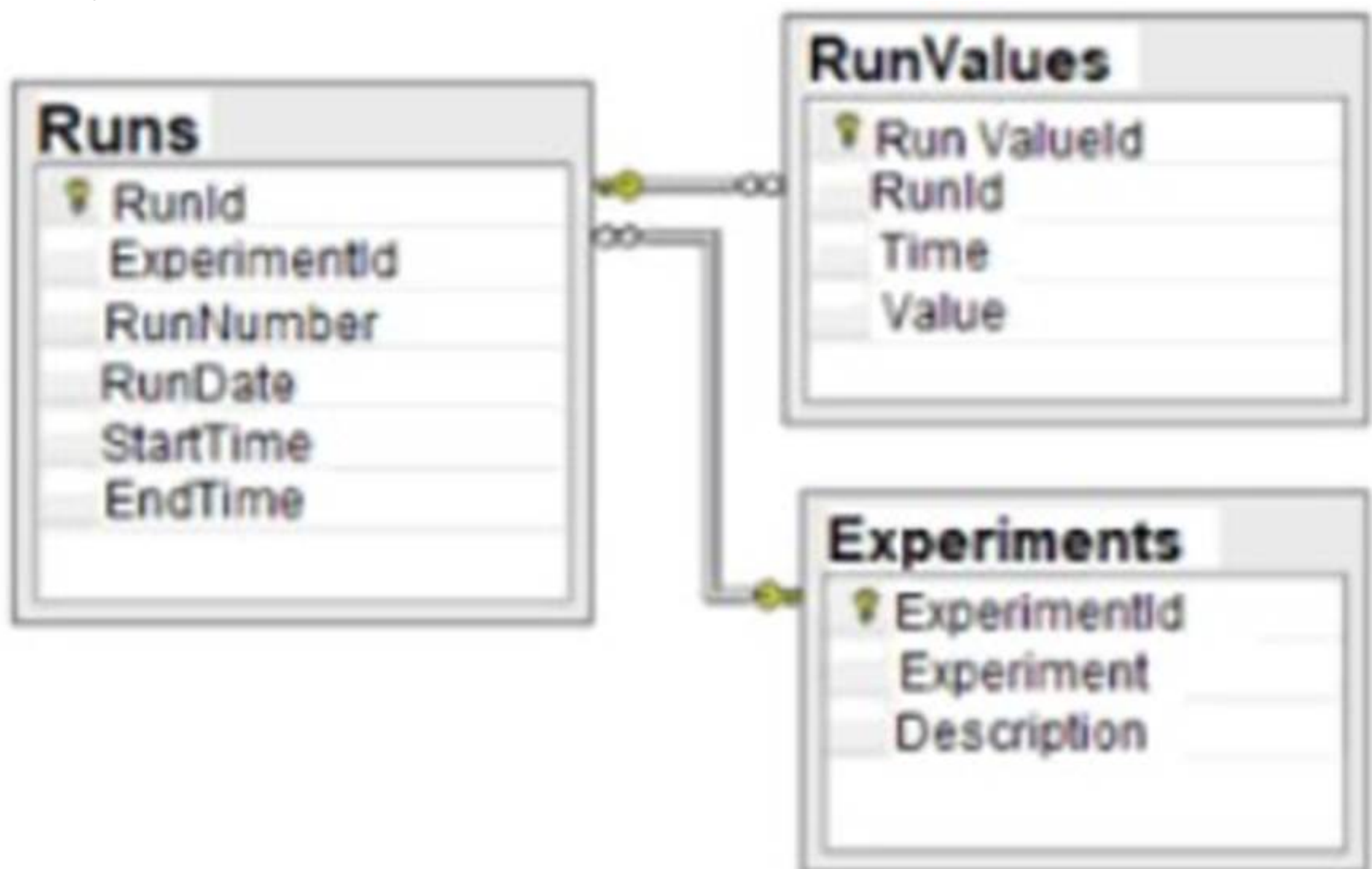
The best answer is B. As a sample size grows, its mean gets closer to the average of the whole population.

The law of large numbers, in probability and statistics, states that as a sample size grows, its mean gets closer to the average of the whole population. This is due to the sample being more representative of the population as it increases in size. The law of large numbers guarantees stable long-term results for the averages of some random events<sup>1</sup>

- \* A. As a sample size decreases, its standard deviation gets closer to the average of the whole population is not correct, because it confuses the concepts of standard deviation and mean. Standard deviation is a measure of how much the values in a data set vary from the mean, not how close the mean is to the population average. Also, as a sample size decreases, its standard deviation tends to increase, not decrease, because the sample becomes less representative of the population.
- \* C. As a sample size decreases, its mean gets closer to the average of the whole population is not correct, because it contradicts the law of large numbers. As a sample size decreases, its mean tends to deviate from the average of the whole population, because the sample becomes less representative of the population.
- \* D. When a sample size doubles, the sample is indicative of the whole population is not correct, because it does not specify how close the sample mean is to the population average. Doubling the sample size does not necessarily make the sample indicative of the whole population, unless the sample size is large enough to begin with. The law of large numbers does not state a specific number or proportion of samples that are indicative of the whole population, but rather describes how the sample mean approaches the population average as the sample size increases indefinitely.

#### NEW QUESTION 61

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:**

A relational database is a type of database that organizes data into tables, where each table has a fixed number of columns and a variable number of rows. Each row in a table represents a record or an entity, and each column represents an attribute or a property of that entity. The tables are linked by common fields, called keys, which enable the database to establish relationships between the data. A relational database schema is a diagram that shows the structure and organization of the tables, columns, keys, and constraints in a relational database. The diagram given in the question is an example of a relational database schema, as it shows two tables: ??Runs?? and ??Experiments??, with their respective columns, data types, and primary keys. The ??Runs?? table also has a foreign key that references the ??ExperimentId?? column in the ??Experiments?? table, indicating a relationship between the two tables. Therefore, the correct answer is D. References: What is a database schema? | IBM, Database Schema - Javatpoint

**NEW QUESTION 64**

Which of the following should be accomplished NEXT after understanding a business requirement for a data analysis report?

- A. Rephrase the business requirement.
- B. Determine the data necessary for the analysis
- C. Build a mock dashboard/presentation layout.
- D. Perform exploratory data analysis.

**Answer: B**

**Explanation:**

The next step after understanding a business requirement for a data analysis report is to determine the data necessary for the analysis. This step involves identifying the data sources, variables, metrics, and dimensions that are relevant and sufficient to answer the business question or problem. This step also involves assessing the availability, quality, and accessibility of the data, and planning how to collect, clean, and prepare the data for analysis. The other options are not the next steps after understanding a business requirement, but rather subsequent steps in the data analysis process. Rephrasing the business requirement is a step that can help clarify and refine the business question or problem before determining the data necessary for the analysis. Building a mock dashboard/presentation layout is a step that can help design and visualize the report before performing the data analysis. Performing exploratory data analysis is a step that can help explore and summarize the data before drawing conclusions and recommendations from the data. Reference: Data Analysis Process - DataCamp

**NEW QUESTION 65**

Which of the following types of analyses should be used to evaluate the connections and anomalies in a data set when either known patterns are being violated or new patterns are emerging?

- A. Correlation
- B. Descriptive
- C. Graph
- D. Regression

**Answer: C**

**NEW QUESTION 67**

A data scientist wants to see which products make the most money and which products attract the most customer purchasing interest in their company. Which of the following data manipulation techniques would he use to obtain this information?

- A. Data append
- B. Data blending
- C. Normalize data
- D. Data merge

**Answer: B**

**Explanation:**

The correct answer is B: Data blending.

Data blending is combining multiple data sources to create a single, new dataset, which can be presented visually in a dashboard or other visualization and can then be processed or analyzed. Enterprises get their data from a variety of sources, and users may want to temporarily bring together different datasets to compare data relationships or answer a specific question. Data append is incorrect. Data append is a process that involves adding new data elements to an existing database. An example of a common data append would be the enhancement of a company's customer files. A data append takes the information they have, matches it against a larger database of business data, allowing the desired missing data fields to be added. Normalize data is incorrect.

Data normalization is the process of structuring your relational customer database, following a series of normal forms. This improves the accuracy and integrity of your data while ensuring that your database is easier to navigate. Data merge is incorrect. Data merging is the process of combining two or more data sets into a single data set.

**NEW QUESTION 71**

Alex wants to use data from his corporate sale, CRM, and shipping systems to try and predict future sales. Which of the following systems is the most appropriate? Choose the best answer.

- A. Data mart.
- B. OLAP.
- C. Data Warehouse.
- D. OLTP.

**Answer: C**

**Explanation:**

Correct Answer: C. Data Warehouse.

Data warehouse bring together data from multiple systems used by an organization. A data mart is too narrow, as Alex needs data from across multiple divisions. OLAP is a broad term of analytical processing, and OLTP systems are transactional and not ideal for this task.



**NEW QUESTION 73**

Which of the following best describes the process of examining data for statistics and information about the data?

- A. Cleansing
- B. search
- C. Profiling
- D. Governance

**Answer:** C

**Explanation:**

Data profiling is the process of examining data for statistics and information about the data, such as the structure, format, quality, and content of the data. Data profiling can help to understand the characteristics, patterns, relationships, and anomalies of the data, as well as to identify and resolve any errors, inconsistencies, or missing values in the data. Data profiling can be done using various tools and methods, such as spreadsheets, databases, or programming languages<sup>12</sup>.

**NEW QUESTION 78**

Which of the following variable name formats would be problematic if used in the majority of data software programs?

- A. First\_Name\_
- B. FirstName
- C. First\_Name
- D. First Name

**Answer:** D

**Explanation:**

This is because First Name is a variable name format that would be problematic if used in most of the data software programs, such as Excel, SQL, or Python. This is because First Name contains a space between two words, which could cause confusion or errors in the data software programs, as they might interpret the space as a separator or a delimiter between two different variables or values, rather than as part of a single variable name. For example, in SQL, a space is used to separate keywords, clauses, or expressions in a statement, such as SELECT, FROM, WHERE, etc. Therefore, using First Name as a variable name in SQL could result in a syntax error or an unexpected result. The other variable name formats would not be problematic if used in most of the data software programs. Here is why:  
? First\_Name\_ is a variable name format that uses an underscore (\_) to separate two words, which is a common and acceptable practice in most of the data software programs, as it helps to improve the readability and clarity of the variable name. For example, in Python, an underscore is used to follow the PEP 8 style guide for naming variables, which recommends using lowercase letters and underscores for multi-word variable names.  
? FirstName is a variable name format that uses camel case to separate two words, which is another common and acceptable practice in most of the data software programs, as it helps to reduce the length and complexity of the variable name. For example, in Excel, camel case is used to follow the VBA naming conventions for naming variables, which recommends using mixed case letters for multi-word variable names.  
? First\_Name is a variable name format that also uses an underscore (\_) to separate two words, which is also a common and acceptable practice in most of the data software programs, as it helps to improve the readability and clarity of the variable name. For example, in SQL, an underscore is used to follow the ANSI SQL naming standards for naming variables, which recommends using lowercase letters and underscores for multi-word variable names.

**NEW QUESTION 83**

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

- A. Replace missing data.
- B. Remove duplicate data.
- C. Replace redundant data.
- D. Remove invalid data.

**Answer:** A

**Explanation:**

This is because missing data is a type of data quality issue that occurs when data is absent or incomplete in a data set, which can affect the accuracy and reliability of the analysis or process. Missing data can be caused by various factors, such as human error, system error, or non-response. Missing data can be addressed by using various methods, such as replacing missing data, which means filling in or imputing the missing values with some reasonable estimates, such as mean, median, mode, or regression. The other methods are not used to address missing data. Here is why:  
? Remove duplicate data is a type of method that eliminates or reduces duplicate data, which is a type of data quality issue that occurs when data is repeated or copied in a data set. Removing duplicate data does not address missing data, but rather affects the quantity and validity of the data.  
? Replace redundant data is a type of method that eliminates or reduces redundant data, which is a type of data quality issue that occurs when data is unnecessary or irrelevant for the analysis or purpose. Replacing redundant data does not address missing data, but rather affects the efficiency and performance of the analysis or process.  
? Remove invalid data is a type of method that eliminates or reduces invalid data, which is a type of data quality issue that occurs when data is incorrect or inaccurate in a data set. Removing invalid data does not address missing data, but rather affects the validity and reliability of the analysis or process.

**NEW QUESTION 88**

An analyst is required to run a text analysis of data that is found in articles from a digital news outlet. Which of the following would be the BEST technique for the analyst to apply to acquire the data?

- A. Web scraping
- B. Sampling
- C. Data wrangling
- D. ETL

**Answer:** A

**Explanation:**

This is because web scraping is a technique that allows the analyst to extract data from web pages, such as articles from a digital news outlet. Web scraping can be done using various tools and methods, such as Python libraries, browser extensions, or online services. The other techniques are not suitable for acquiring data from web pages. Here is why:

Sampling is a technique that involves selecting a subset of data from a larger population, usually for statistical analysis or testing purposes. Sampling does not help the analyst to acquire data from web pages, but rather to reduce the amount of data to be analyzed. Data wrangling is a technique that involves transforming and cleaning data to make it suitable for analysis or visualization. Data wrangling does not help the analyst to acquire data from web pages, but rather to improve the quality and usability of the data.

ETL stands for Extract, Transform, and Load, which is a process that involves moving data from one or more sources to a destination, such as a data warehouse or a database. ETL does not help the analyst to acquire data from web pages, but rather to store and organize the data.

#### NEW QUESTION 93

Which of the following is the BEST reason to use database views instead of tables?

- A. Views reduce the need for repetitive, complex data joins.
- B. Views allow for the storage of temporary data.
- C. whereas tables do not.
- D. Views allow for the joining of multiple data sources, whereas tables do not.
- E. Views can be used to restrict sensitive information.

**Answer:** A

#### Explanation:

Views are virtual tables that are created by querying one or more base tables or other views. Views do not store any data, but only show the result of a query. One of the main advantages of using views is that they can reduce the need for repetitive, complex data joins. For example, if a query involves joining multiple tables with many conditions, creating a view can simplify the query and make it easier to reuse. Therefore, the correct answer is A. References: [What is a Database View? | Definition & Examples - Vertabelo], [Database Views - GeeksforGeeks]

#### NEW QUESTION 94

Which of the following query statements would be used when filtering data in a relational database management system? (Select two).

- A. ORDER BY
- B. HAVING
- C. WHERE
- D. SELECT
- E. INSERT
- F. GROUP BY

**Answer:** BC

#### NEW QUESTION 95

An analyst has written the following code: `SELECT *  
FROM Cust_table`

`WHERE age > 60 AND City = "New York"`

Which of the following criteria is the analyst retrieving?

- A. All customers older than age 60 in New York state
- B. All customers aged 60 and older in New York state
- C. All customers older than age 60 in New York City
- D. All customers younger than age 60 in New York City

**Answer:** C

#### Explanation:

The SQL query provided is selecting all records from the `Cust_table` where the `age` column has values greater than 60 and the `City` column matches `??New York??`. The `>` operator selects values that are strictly greater than the comparison value, so it does not include customers aged exactly 60. The term `??New York??` in the context of a city database typically refers to New York City, not the state of New York. Therefore, the correct answer is that the analyst is retrieving data for all customers older than age 60 in New York City.

References:

? The use of the `>` operator in SQL is to select values greater than the specified value1.

? Understanding the `WHERE` clause in SQL and its use in filtering records based on specified conditions2.

? Clarification on the distinction between city and state names in database records3.

#### NEW QUESTION 96

Joseph is interpreting a left skewed distribution of test scores. Joe scored at the mean, Alfonso scored at the median, and Gaby scored at the end of the tail. Who had the highest score?

- A. Joseph
- B. Joe
- C. Alfonso
- D. Gaby

**Answer:** C

#### Explanation:

Alfonso had the highest score. A left skewed distribution is a distribution where the tail is longer on the left side than on the right side, meaning that most of the values are clustered on the right side and there are some outliers on the left side. In a left skewed distribution, the mean is less than the median, which is less than the mode. Therefore, Joseph, who scored at the mean, had the lowest score, Gaby, who scored at the end of the tail, had the second lowest score, and Alfonso, who scored at the median, had the highest score. Reference: Skewness - Statistics How To

**NEW QUESTION 97**

Which one of the following is NOT a common data integration tool?

- A. XSS
- B. ELT
- C. ETL
- D. APIs

**Answer:** A

**Explanation:**

Cross-site Scripting (XSS) is a security vulnerability usually found in websites and/or web applications that accept user input. XSS is a client-side vulnerability that targets other application users, while SQL injection is a server-side vulnerability that targets the application's database. How do I prevent XSS in PHP? Filter your inputs with a whitelist of allowed characters and use type hints or type casting.

**NEW QUESTION 100**

An analyst needs to summarize the number of people in Chicago in 2022 using the following set of data:

Name	City	Year	Grade
Chloe	Chicago	2022	A
Blake	Chicago	2023	B
Carter	Chicago	2022	A
Kim	Detroit	2021	C

Which of the following steps should the analyst use to provide results? (Select two).

- A. Aggregation
- B. Sorting
- C. Filtering
- D. Indexing
- E. Cleaning
- F. Replacing

**Answer:** AC

**NEW QUESTION 104**

A data analyst is compiling a report that a Chief Executive Officer needs for an impromptu meeting. The report should include information on the previous day's performance. Which of the following reports should the analyst provide?

- A. Tactical
- B. Ad hoc
- C. Dynamic
- D. Recurring

**Answer:** B

**NEW QUESTION 107**

A data analyst needs to perform a full outer join of a customer's orders using the tables below:

Sales\_table

Cust_id	Order_id	Order_qty
Tc - 5858	Od - 9800	50
Tc - 5833	Od - 9801	68
Tc - 5890	Od - 9802	103

Order\_table

Order_id	Order_qty
Od - 9803	102
Od - 9800	50
Od - 9802	103
Od - 9805	80
Od - 9804	70

Which of the following is the mean of the order quantity?

- A. 73.5
- B. 76.5
- C. 78.8
- D. 81.5

Answer: D

Explanation:

The correct answer is D. OUTER JOIN, seven rows.  
 An OUTER JOIN is a type of SQL join that returns all the rows from both tables, regardless of whether there is a match or not. If there is no match, the missing side will have null values. An OUTER JOIN can be either a LEFT JOIN, a RIGHT JOIN, or a FULL JOIN, depending on which table's rows are preserved.  
 Using the example tables, a FULL OUTER JOIN query would look like this:  
 SELECT Cust\_id, Order\_id, Order\_qty FROM Sales\_table FULL OUTER JOIN Order\_table ON Sales\_table.Order\_id = Order\_table.Order\_id;  
 The result of this query would be:  
 Cust\_id | Order\_id | Order\_qty  
 -----  
 1 | 1 | 100  
 2 | 2 | 50  
 3 | 3 | 25  
 4 | 4 | 75  
 NULL | 5 | 10  
 NULL | 6 | 20  
 NULL | 7 | 15  
 As you can see, the query returns seven rows, one for each order in either table. The orders that are not in the Sales\_table have null values for the Cust\_id column.  
 To find the mean of the order quantity, we need to sum up the order quantities and divide by the number of rows. In this case, the mean is (100 + 50 + 25 + 75 + 10 + 20 + 15) / 7 = 42.14. Rounding to one decimal place, we get 42.1 as the mean of the order quantity.

NEW QUESTION 112

An analyst is reporting on the average income for a county and is reviewing the following data:

Name	Address	Yearly income
Jessica Jones	145 Stonebridge Avenue	\$634,900
Spencer James	1567 Watercress	\$135,000
Olivia Baker	456 Harvard Road	\$95,000
Layla Harding	5674 Yarding Street	\$37,000

Which of the following is the reason the analyst would need to cleanse the data in this data set?

- A. Data completeness
- B. Data outliers
- C. Duplicate data
- D. Missing values



Answer: B

NEW QUESTION 117

A data analyst is creating a dashboard and trying to identify the type of information that should be included. Which of the following should the analyst consider first?

- A. Data refresh rate
- B. Consumer types
- C. Access permissions
- D. Data sources and attributes

Answer: D

Explanation:

The answer is D. Data sources and attributes.  
Short Explanation: The data analyst should consider the data sources and attributes first when creating a dashboard, because they determine what kind of information can be included and how it can be displayed. The data sources and attributes define the origin, quality, format, and structure of the data that will be used for the dashboard. They also affect the data refresh rate, the consumer types, and the access permissions of the dashboard<sup>12</sup>  
\* A. Data refresh rate is not the first thing to consider, because it depends on the data sources and attributes. The data refresh rate is how often the data in the dashboard is updated or refreshed to reflect the latest changes. The data refresh rate can vary depending on the type, frequency, and availability of the data sources<sup>1</sup>  
\* B. Consumer types are not the first thing to consider, because they depend on the data sources and attributes. The consumer types are the intended audiences or users of the dashboard, who may have different needs, preferences, and expectations for the dashboard. The consumer types can influence the design, layout, and functionality of the dashboard. However, the consumer types cannot be determined without knowing what kind of data is available and relevant for them<sup>1</sup>  
\* C. Access permissions are not the first thing to consider, because they depend on the data sources and attributes. The access permissions are the rules or policies that govern who can view, edit, or share the dashboard. The access permissions can protect the confidentiality, integrity, and availability of the data in the dashboard. However, the access permissions cannot be set without knowing what kind of data is involved and who needs to access it<sup>1</sup>

NEW QUESTION 118

Given the information in the following tables:

Online transactions:

Customer ID	Channel	Segment	Amount (\$)
001	Online	Existing	3,000
002	Online	Existing	4,000
003	Online	New	1,500

In-store transactions:

Customer ID	Channel	Segment	Amount (\$)
001	In-store	New	1,000
004	In-store	Existing	4,000
005	In-store	New	3,500

Which of the following describes merging these tables to create a master file that includes all transactions for both online and in-store sales?

- A. Data audit
- B. Data completeness
- C. Data validation
- D. Data consolidation

Answer: D

Explanation:

Merging tables to create a master file that includes all transactions for both online and in- store sales is best described as data consolidation. This process involves combining data from various sources into a single, unified dataset. Data consolidation is essential for providing a comprehensive view of all transactions, which can be used for analysis, reporting, and decision-making purposes.  
References: The answer is based on standard data management practices and the definition of data consolidation. No specific external documents were referenced for this response.

NEW QUESTION 120

A stakeholder wants to see daily sales targets organized in a dashboard by country, state, city, and ZIP Code. Which of the following delivery considerations must a data analyst take into account when creating the dashboard?

- A. Variable formatting
- B. Drill-down capability
- C. Saved searches
- D. Access permissions

**Answer:** B

#### NEW QUESTION 122

Which of the following is a relational database?

- A. SQL
- B. Excel
- C. JSON
- D. NoSQL

**Answer:** A

#### NEW QUESTION 123

A data analyst is using a two-tailed, independent t-test to determine whether the type of stretching, dynamic or static, has any influence on a dancer's flexibility. Which of the following is the alternative hypothesis?

- A. A dancer's flexibility is improved through static stretching.
- B. The change in a dancer's flexibility is not equal to zero.
- C. There is a difference in a dancer's flexibility between static and dynamic stretching.
- D. The means of the static and dynamic stretching groups do not differ from each other.

**Answer:** C

#### NEW QUESTION 124

Under which of the following circumstances should the null hypothesis be accepted when  $\alpha = 0.05$ ?

- A. When p is 0.00003
- B. When p is 0.001
- C. When p is 0.04
- D. When p is 0.06

**Answer:** C

#### 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 error<sup>12</sup>.

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.

#### NEW QUESTION 126

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?

- A. Indexing.
- B. Subset of records.
- C. Temporary table in the query set.
- D. Parametrization.

**Answer:** D

#### Explanation:

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.

#### NEW QUESTION 131

Which of the following roles is responsible for ensuring an organization's data quality, security, privacy, and regulatory compliance?

- A. Data owner.
- B. Data steward.
- C. Data custodian.
- D. Data processor.

Answer: B

Explanation:

Correct answer B. Data steward.  
A data steward is responsible for leading an organization's data governance activities, which include data quality, security, privacy, and regulatory compliance.

NEW QUESTION 134

Which of the following is a KPI metric for tracking sales performance?

- A. Order status percentage
- B. Customer acquisition percentage
- C. Gross profit percentage
- D. Click-through rate percentage

Answer: C

Explanation:

Gross profit percentage is a key performance indicator (KPI) that measures the profitability of a company's sales by showing the percentage of revenue that exceeds the cost of goods sold (COGS). It is a critical metric for tracking sales performance because it directly reflects the efficiency of a company in managing its production costs and the profitability of its products. This KPI is essential for understanding the financial health of a business and making informed decisions about pricing, cost control, and sales strategies.

References:

- ? Sales KPIs are essential for measuring the effectiveness of sales activities and the profitability of those efforts1.
- ? Gross profit percentage is highlighted as a crucial metric for assessing the financial success of sales initiatives2.
- ? Understanding the difference between sales metrics and KPIs, and the importance of gross profit percentage as a KPI1.
- ? The significance of gross profit percentage in evaluating sales team performance and guiding business decisions3.

NEW QUESTION 137

A data analyst has been asked to create one table that has each employee's first name, last name, sales, and address. The sales and addresses are listed in the tables below:

Table 1

First name	Last name	Sales
John	Knox	\$30
John	Johnson	\$10
John	Sinclair	\$70
Bob	Sinclair	\$100

Table 2

First name	Last name	Address
John	Knox	2851 N. Southport
John	Johnson	457 Bridle Ridge
John	Sinclair	1067 Windwood Lane
Bob	Sinclair	71 S. Wacker Drive

Which of the following steps should the analyst take to create the table?

- A. Transpose the first name and last name in both table
- B. Use lookup to pull the address field from Table 2 into Table 1.
- C. Use lookup with the first name or first name to pull the address field from Table 2 into Table 1.
- D. Use the append formula in both tables for the first name and last nam
- E. Use lookup topull the address field from Table 2 into Table 1.
- F. Create a column that concatenates the first name and last name in each tabl

G. Use concatenate and lookup to bring the address field into Table 1.

**Answer:** D

#### NEW QUESTION 142

A gambler thinks that a coin is fair and is equally likely to turn up heads or tails when the coin is flipped. Which of the following tests should the gambler use to test this hypothesis?

- A. t-test
- B. Chi-squared test
- C. Rank sum test
- D. Ratio test

**Answer:** B

#### NEW QUESTION 146

A data analyst needs to create a dashboard using the company's yearly revenue data sets. Which of the following would be the best way to plot the information to show the top- performing region?

- A. A line chart
- B. A waterfall chart
- C. A heat map
- D. A stacked bar chart

**Answer:** D

#### NEW QUESTION 148

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:

Getting stakeholder approval is the next step the analyst should take in the dashboard creation process, after confirming the data sources and creating a wireframe. Stakeholder approval means getting feedback and validation from the intended users or clients of the dashboard, to ensure that it meets their expectations and requirements. This step helps to avoid rework and ensure customer satisfaction. References: CompTIA Data+ Certification Exam Objectives, page 14

#### NEW QUESTION 152

An analyst wants to combine two data sets into a single spreadsheet. Column names from the first spreadsheet are listed in rows in the second spreadsheet. Which of the following is the first step the analyst should take to combine the data sets?

- A. Blend
- B. Merge
- C. Concatenate
- D. Transpose

**Answer:** C

#### NEW QUESTION 153

Five dogs have the following heights in millimeters: 300, 430, 170, 470, 600  
Which of the following is the mean height for the five dogs?

- A. 394mm
- B. 405mm
- C. 493mm
- D. 504mm

**Answer:** B

#### Explanation:

The mean height for the five dogs is 405mm. The mean, or average, is a measure of central tendency that represents the sum of all values divided by the number of values. To calculate the mean height for the five dogs, we can use the following formula:  $\text{Mean} = (300 + 430 + 170 + 470 + 600) / 5 = 2020 / 5 = 404$   
We can round up the result to the nearest millimeter, which is 405mm. The other options are not correct, as they are either too high or too low than the actual mean. Reference: [Mean - Math is Fun]

#### NEW QUESTION 158

An employer needs to maintain adequate office staffing during the winter and wants to track storm data. Which of the following data collection methods should the employer use?

- A. Web scraping



- B. Public databases
- C. Observations
- D. Weather surveys

Answer: B

Explanation:

For an employer looking to maintain adequate office staffing during winter while tracking storm data, the most effective method would be to use public databases. These databases often contain comprehensive records of weather patterns and storm data collected and verified by reputable meteorological organizations. Utilizing public databases allows for access to historical and real-time data that is crucial for making informed decisions about staffing during adverse weather conditions.

Web scraping (A) is not the most reliable method, as it may involve extracting data from various websites that might not always provide verified or consistent information. Observations © can be subjective and may not cover a wide enough area to be effective for decision-making on a larger scale. Weather surveys (D) could provide insights, but they are not as immediate or comprehensive as the data available in public databases. References:

? The systematic review on Big Data Analytics in Weather Forecasting suggests that big data techniques and technologies can manage and analyze the huge volume of weather data from different resources, which supports the use of public databases1.

? NOAA??s approach to detecting severe weather events using instruments and receiving information from storm spotters indicates the importance of reliable, collected data, which is typically stored in public databases2.

? The National Weather Service??s use of observational data collected by various instruments, which are then fed into forecast models, further emphasizes the value of established data collection methods over individual observations or surveys3.

NEW QUESTION 163

An analyst runs a report on a daily basis, and the number of datapoints must be validated before the data can be analyzed. The number of datapoints increases each day by approximately 20% of the total number from the day before. On a given day, the number of datapoints was 8,798. Which of the following should be the total number of datapoints on the next day?

- A. 7,038
- B. 9,600
- C. 10,600
- D. 10,800

Answer: C

Explanation:

This is because the number of datapoints increases each day by approximately 20% of the total number from the day before. Therefore, to find the number of datapoints on the next day, we can use the formula:

Next day = Current day \* (1 + 20%)

Plugging in the given values, we get:

Next day = 8,798 \* (1 + 0.2)

Next day = 8,798 \* 1.2

Next day = 10,557.6

Since we are dealing with whole numbers, we can round up the result to the nearest integer, which is 10,600.

NEW QUESTION 166

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

**Answer:** C

#### NEW QUESTION 171

A data analyst has been asked to create a sales report that calculates the rolling 12-month average for sales. If the report will be published on November 1, 2020, which of the following months should the report cover?

- A. October 1, 2019 to October 31, 2020
- B. October 31, 2020 to November 1, 2021
- C. November 1, 2019 to October 31, 2020
- D. October 31, 2019 to October 31, 2020

**Answer:** A

#### Explanation:

The report should cover the months from October 1, 2019 to October 31, 2020. A rolling 12-month average is a type of moving average that calculates the average of the last 12 months of data for each month. It is useful for smoothing out seasonal fluctuations and identifying long-term trends in the data. To calculate the rolling 12-month average for sales for November 1, 2020, the analyst needs to use the sales data from the previous 12 months, starting from November 1, 2019 and ending on October 31, 2020. The other options are either too short or too long to cover the required period.

#### NEW QUESTION 174

Which of the following types of analysis is used when comparing last week's sales to the previous week's sales?

- A. Trend analysis
- B. Exploratory analysis
- C. Prescriptive analysis
- D. Link analysis

**Answer:** A

#### NEW QUESTION 177

A company wants to know how its customers interact with an e-commerce website based on clicks over items. Which of the following is the primary requirement for this report?

- A. Data content
- B. Frequency
- C. Filtering
- D. Views

**Answer:** B

#### NEW QUESTION 181

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.

**Answer:** C

#### Explanation:

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.

#### NEW QUESTION 185

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 means arranging the data in a specific order, such as ascending or descending, based on one or more criteria. Sorting can be applied to any column in the table, such as sales or state. References: CompTIA Data+ Certification Exam Objectives, page 11

**NEW QUESTION 187**

The number of phone calls that the 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:**

Discrete data is a type of data that can only take certain values, usually whole numbers or integers. Discrete data can be counted, but not measured. For example, the number of students in a class, the number of books in a library, or the number of phone calls that a call center receives in a day are all examples of discrete data. Discrete data is different from continuous data, which can take any value within a range, and can be measured with precision. For example, the height of a person, the weight of a fruit, or the temperature of a room are all examples of continuous data. Therefore, the correct answer is D. References: [Discrete vs Continuous Data: Definition and Examples - Statistics How To], [Discrete Data - Definition and Examples | Math Goodies]

**NEW QUESTION 191**

Which of the following is the most likely reason for a data analyst to optimize a query using parameterization?

- A. To return a subset of records
- B. To insert a temporary table
- C. To prevent SQL injections
- D. To increase the query speed

**Answer:** C

**Explanation:**

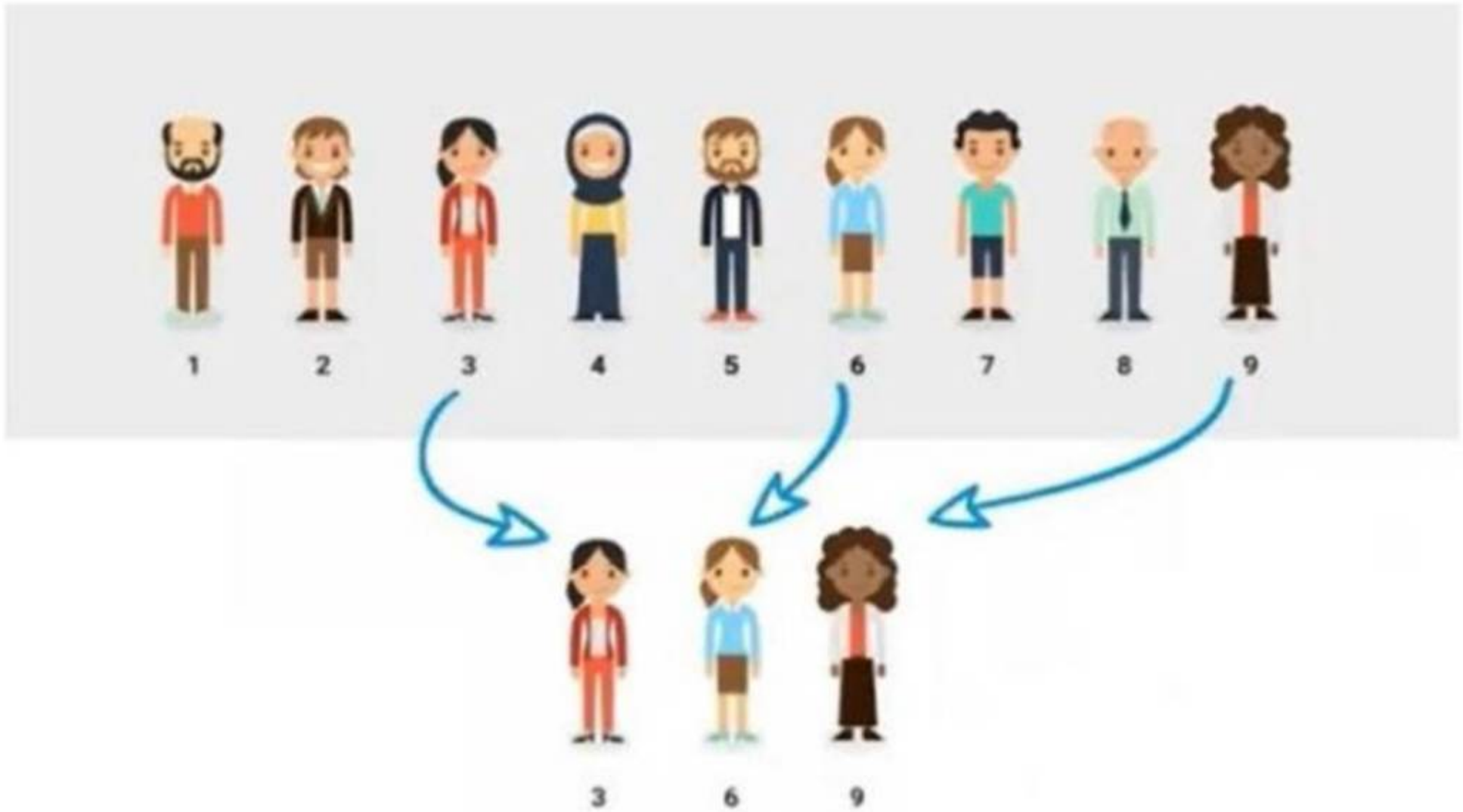
Parameterization in SQL queries is a technique used to prevent SQL injection, which is a common security vulnerability that allows an attacker to interfere with the queries that an application makes to its database. By using parameterized queries, the database can distinguish between code and data, regardless of the input received. This method ensures that an attacker cannot change the intent of a query, even if SQL commands are inserted by the attacker. While parameterization can also affect performance by enabling consistent query execution plans, its primary purpose is to enhance security.

References:

- ? Medium article on SQL Query Optimization<sup>1</sup>.
- ? MSSQLTips on SQL Query Performance<sup>2</sup>.
- ? Blog post on SQL Performance Optimization<sup>3</sup>.
- ? SQL Easy guide on improving SQL Query Performance<sup>4</sup>.
- ? LearnSQL.com on SQL for Data Analysis<sup>5</sup>.

**NEW QUESTION 194**

Given the diagram below:



Which of the following types of sampling is depicted in the image?

- A. Stratified
- B. Random
- C. Cluster
- D. Systematic

**Answer:** D

**Explanation:**

Systematic sampling is a type of sampling where the sample is selected by following a fixed interval. For example, every 10th person in a list is chosen for the sample. In the image, the sample is selected by choosing every 3rd person in the line, starting from person number 1. This is an example of systematic sampling. References: Types of Sampling Techniques in Data Analytics You Should Know, Sampling Methods | Types, Techniques & Examples - Scribbr

**NEW QUESTION 196**

Which of the following data governance concepts fits into the security requirements category?

- A. Data transmission
- B. Data deletion
- C. Data use agreements
- D. Personally identifiable information

**Answer:** D

**NEW QUESTION 201**

Which of the following is an example of PII?

- A. Age
- B. Name
- C. Ethnicity
- D. Gender

**Answer:** B

**Explanation:**

A name is an example of personally identifiable information (PII), which is any data that can be used to identify someone, either on its own or with other relevant data. A name is a direct identifier, which means that it can uniquely identify a person without the need for any additional information. For example, a full name, such as John Smith, can be used to distinguish or trace an individual's identity<sup>1</sup>. Other examples of direct identifiers include:

- ? Social Security Number
- ? Passport number
- ? Driver's license number
- ? Email address
- ? Phone number

**NEW QUESTION 202**

Andy is a pricing analyst for a retailer. Using a hypothesis test, he wants to assess whether people who receive electronic coupons spend more on average. What should Andy's null hypothesis be?



- A. People who receive electronic coupons spend more on average.
- B. People who receive electronic coupons spend less on average.
- C. People who receive electronic coupons do not spend more on average.
- D. People who do not receive electronic coupons spend more on average.

**Answer:** C

**Explanation:**

The null hypothesis presumes the status quo. Andy is testing whether or not people who receive an electronic coupon spend more on average, so, the null hypothesis states that people who receive the coupon do spend more on average.

#### NEW QUESTION 204

A data analyst is designing a dashboard that will provide a story of sales and determine which site is providing the highest sales volume per customer. The analyst must choose an appropriate chart to include in the dashboard. The following data is available:

Site	Customers	Sales volume	Average sales per customer
A1	2236	\$3,415,372.00	\$1,527.45
A2	885	\$1,405,437.00	\$1,588.06
A3	333	\$952,723.00	\$2,861.03
B1	483	\$4,871,380.00	\$10,085.67
B2	2969	\$780,381.00	\$262.84
B4	2357	\$4,917,436.00	\$2,086.31
C1	1524	\$1,135,204.00	\$744.88
C2	878	\$614,964.00	\$700.41
C3	1925	\$4,035,100.00	\$2,096.16

Which of the following types of charts should be considered?

- A. Include a line chart using the site and average sales per customer.
- B. Include a pie chart using the site and sales to average sales per customer.
- C. Include a scatter chart using sales volume and average sales per customer.
- D. Include a column chart using the site and sales to average sales per customer.

**Answer:** C

**Explanation:**

A scatter chart using sales volume and average sales per customer is the best type of chart to include in the dashboard. A scatter chart is a type of chart that displays the relationship between two numerical variables using dots or markers. A scatter chart can show how one variable affects another, how strong the correlation is between them, and how the data points are distributed. In this case, a scatter chart can show the story of sales and determine which site is providing the highest sales volume per customer by plotting the sales volume on the x-axis and the average sales per customer on the y-axis. Each dot on the chart will represent a site, and the analyst can easily compare the sites based on their position on the chart. A site with a high sales volume and a high average sales per customer will be in the upper right quadrant, indicating a high performance. A site with a low sales volume and a low average sales per customer will be in the lower left quadrant, indicating a low performance. A site with a high sales volume and a low average sales per customer will be in the lower right quadrant, indicating a high volume but low value. A site with a low sales volume and a high average sales per customer will be in the upper left quadrant, indicating a low volume but high value. A scatter chart can also show if there is a positive or negative correlation between the two variables, or if there is no correlation at all. A positive correlation means that as one variable increases, so does the other. A negative correlation means that as one variable increases, the other decreases. No correlation means that there is no relationship between the two variables.

The other types of charts are not as suitable for this purpose. A line chart is a type of chart that displays the change of one or more variables over time using lines. A line chart can show trends, patterns, and fluctuations in the data. However, in this case, there is no time variable involved, so a line chart would not be appropriate. A pie chart is a type of chart that displays the proportion of each category in a whole using slices of a circle. A pie chart can show how each category contributes to the total and compare the relative sizes of each category. However, in this case, there are two numerical variables involved, so a pie chart would not be able to show their relationship. A column chart is a type of chart that displays the comparison of one or more variables across categories using vertical bars. A column chart can show how each category differs from each other and rank them by size. However, in this case, a column chart would not be able to show the relationship between sales volume and average sales per customer, as it would only show one variable for each site.

#### NEW QUESTION 205

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

- A. Normalize the data.
- B. Standardize the data.
- C. Rescale the data.
- D. Aggregate the data.

**Answer:** C

**Explanation:**

The analyst should rescale the data to fix this issue. Rescaling is a process of transforming data from one scale to another, such as changing the units of measurement. In this case, the analyst needs to rescale the temperatures from Fahrenheit to Celsius, which are two different scales for measuring temperature. To do this, the analyst can use the following formula:

$\text{Celsius} = (\text{Fahrenheit} - 32) * 5/9$

This formula converts each temperature value from Fahrenheit to Celsius by subtracting 32

and multiplying by 5/9. For example, if the temperature is 68°F, the rescaled value in Celsius is:

$\text{Celsius} = (68 - 32) * 5/9$  Celsius = 20°C

Rescaling the data can help the analyst to report the temperatures in a consistent and accurate way, and to avoid any confusion or errors that may arise from using different scales. Rescaling can also make the data more comparable and compatible with other data sources or standards that use the same scale<sup>12</sup>.

#### NEW QUESTION 207

A user imports a data file into the accounts payable system each day. On a regular basis, the field input is not what the system is expecting, so it results in an error for the row and a broken import process. To resolve the issue, the user opens the file, finds the error in the row, and manually corrects it before attempting the import again. The import sometimes breaks on subsequent attempts, though. Which of the following changes should be made to this process to reduce the number of errors?

- A. Delete all incorrect inputs and upload the corrected file.
- B. Have the user manually review the file for data completeness before loading it
- C. Create a data field to data type validator to run the file through prior to import.
- D. Spot-check the file prior to import to catch and correct field errors.

**Answer:** C

#### Explanation:

A data field to data type validator is a tool or a process that checks if the data in each field of a file matches the expected data type, such as text, number, date, etc. A data field to data type validator can help to identify and correct any errors or inconsistencies in the data before importing it into the accounts payable system. This would reduce the number of errors and broken imports, as well as save time and effort for the user.

#### NEW QUESTION 210

An organization would like to add a secondary email field to its customer database in order to enrich the customer profiles. Which of the following data manipulation techniques should the analyst use to add this information?

- A. Blend
- B. Merge
- C. Append
- D. Aggregate

**Answer:** C

#### NEW QUESTION 212

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

References: [What are Dynamic Reports? | Sisense], Static vs Dynamic Reports - What's The Difference? | datapine

#### NEW QUESTION 214

A data analyst is performing a data merge within a spreadsheet using the tables below:

<https://www.bing.com/images/blob?bcid=S1XCF9p02M4GjpbGxHj0Irlaj9sw.....4c>

Table 1

Last name	Sales
Knox	\$30
Johnson	\$10
Sinclair	\$70

Table 2

Last name	Address
Knox	2851 N. Southport
Johnson	467 Bridle Ridge
Sinclair	1067 Windwood Lane

The analyst is attempting to pull the addresses from Table 2 into Table 1 using the last names and is receiving an error message. Which of the following steps can the analyst perform to fix the error?

- A. Use concatenate to combine the tables.
- B. Ensure the formula is pulling from right to left.
- C. Sort the data by the last name field.
- D. Review the spelling and data type.

**Answer:** D

**Explanation:**

The error in merging data from Table 2 into Table 1 using last names could be due to discrepancies in spelling or data type between the two tables. It is essential to ensure that the last names are spelled consistently and that the data types are compatible for a successful merge. Option D suggests reviewing these aspects, which can potentially resolve the error, ensuring that each last name in Table 1 accurately corresponds to the same last name in Table 2, allowing for a successful data pull of addresses.

References: This answer is based on general data analytics practices and does not reference a specific document.

**NEW QUESTION 216**

An analyst needs to determine the appropriate data type for the following sample data: sample data collected:  
Which of the following data types should be used for this data?

- A. Text
- B. Float
- C. Alphanumeric
- D. Numeric

**Answer:** B

**NEW QUESTION 217**

An analyst is working with a data set that lists individuals' first and last names in separate columns. Which of the following processes should the analyst use to combine the first and last names into a single spreadsheet cell?

- A. Transpose
- B. Blend
- C. Concatenate
- D. Merges

**Answer:** C

**NEW QUESTION 221**

Given the table below:

Name	Gender	Level	Code	Region
James	Male	College	P	ON
Paul	Female	Elementary	A	BC
Sean	College	College	S	QC
Dad	Male	High school	D	AT
Nathan	Female	College	E	QC
Ahmed	Female	University	L	ON

Which of the following variables can be considered inconsistent, and how many distinct values should the variable have?

- A. Name, one
- B. Gender, two
- C. Level, three
- D. Code, four
- E. Region, five

**Answer:** B

**Explanation:**

The table provided shows an inconsistency in the ??Gender?? column, which lists three distinct values: Male, Female, and College. This is inconsistent because ??College?? is not a gender category. The ??Gender?? column should only have two distinct values, typically ??Male?? and ??Female??, to accurately represent gender data. This error could be due to a data entry mistake or a misclassification during data collection.

In data analysis, it??s crucial to ensure that categorical variables like gender are consistent and correctly classified, as this can significantly impact the analysis results. Data cleaning processes often involve identifying and correcting such inconsistencies to maintain the integrity of the data set.

References:

- ? Data quality management principles emphasize the importance of consistency in data values, especially for categorical variables like gender<sup>1</sup>.
- ? Best practices in data cleaning include checking for and rectifying inconsistencies or misclassifications in data sets<sup>2</sup>.
- ? The importance of accurate data classification is highlighted in data analysis literature, as it directly affects the validity of the analysis results<sup>3</sup>.

**NEW QUESTION 225**

Which of the following query optimization techniques involves examining only the data that is needed for a particular task?

- A. Making a temporary table
- B. Creating a flat file
- C. Indexing documents
- D. Creating an execution plan

**Answer:** C

**Explanation:**

The correct answer is C. Indexing documents.

Indexing documents is a query optimization technique that involves creating a data structure that allows faster access to the data in the documents. Indexing documents can reduce the amount of data that needs to be scanned for a particular query, thus improving the performance and efficiency of the query. Indexing documents can also help with searching, sorting, filtering, and aggregating the data in the documents<sup>12</sup>

**NEW QUESTION 229**

A data analyst has received a data set that contains actual and projected sales for the fourth quarter of 2019. Which of the following statistical methods should the analyst use to find the measure of dispersion?

- A. Mean
- B. Variance
- C. Correlation
- D. Confidence interval

**Answer:** B

**Explanation:**

The measure of dispersion is used to describe the spread of data around a central value. In the context of a data set containing actual and projected sales, the measure of dispersion will help to understand the variability or consistency of sales figures. The variance is the most appropriate statistical method for finding the measure of dispersion because it calculates the average of the squared differences from the Mean, providing a clear picture of data spread. It is especially useful in comparing the spread between different data sets and understanding the distribution of data points.

? Mean is a measure of central tendency, not dispersion.

? Correlation measures the relationship between two variables, not the spread of a single variable.

? Confidence intervals are used to estimate the range within which a population parameter will fall, but they do not measure dispersion within the data set itself.

References:

- ? Measures of Dispersion in Statistics<sup>1</sup>
- ? Measures of Dispersion - Definition, Formulas, Examples<sup>2</sup>
- ? Statistical dispersion - Wikipedia<sup>3</sup>



**NEW QUESTION 230**

A collections manager has a team calling customers who are past due on their accounts in an attempt to collect payments. The manager receives the call list in the form of a printed report that is generated by the accounting department at the beginning of each week. Consequently, the collections team calls some customers who have made payments in the time since the report was last printed. Which of the following reporting enhancements could the accounting department implement to best reduce the number of calls on current accounts?

- A. Modify the date range on the report
- B. Include a time stamp on the report.
- C. Increase the frequency of report generation.
- D. Add a report run date to the report.

**Answer:** C

**Explanation:**

The best reporting enhancement that the accounting department could implement to reduce the number of calls on current accounts is C. Increase the frequency of report generation.

By increasing the frequency of report generation, the accounting department could provide the collections manager with more up-to-date information on the customers who are past due on their accounts. This would help to avoid calling customers who have made payments in the time since the last report was printed, and thus reduce the number of calls on current accounts. Increasing the frequency of report generation would also improve the accuracy and timeliness of the data, and enhance the efficiency and effectiveness of the collections process.

Modifying the date range on the report, including a time stamp on the report, or adding a report run date to the report would not be sufficient to reduce the number of calls on current accounts. These enhancements would only provide information on when the report was generated or what period it covers, but they would not change the fact that the report could be outdated by the time it reaches the collections manager. Therefore, these enhancements would not solve the problem of calling customers who have already paid their accounts.

**NEW QUESTION 234**

An analyst in a consumer bank department wants to showcase the concentration of accounts opened in the United States by ZIP Code to describe the effectiveness of the bank's marketing campaigns. Which of the following would be the best way to visualize the data?

- A. A stacked chart
- B. A tree map
- C. A waterfall chart
- D. A geographic map

**Answer:** D

**NEW QUESTION 237**

A data analyst is attempting to understand how ice cream consumption is affected by different attributes. such as cost, temperature. and income level. Which of the following regression analyses should the data analyst perform to understand this relationship?

- A. Logistic
- B. Ordinary least squares
- C. Cox
- D. Polynomial

**Answer:** B

**Explanation:**

Answer: B. Ordinary least squares

Ordinary least squares (OLS) is a type of linear regression that is used to fit a regression model that describes the relationship between one or more predictor variables and a numeric response variable. Use when: The relationship between the predictor variable(s) and the response variable is reasonably linear. The response variable is a continuous numeric variable<sup>1</sup>.

In this case, the data analyst is interested in understanding how ice cream consumption (the response variable) is affected by different attributes, such as cost, temperature, and income level (the predictor variables). Assuming that these variables have a linear relationship, OLS can be used to estimate the coefficients of the regression equation that best fits the data. OLS can also provide measures of goodness-of-fit, such as R-squared and adjusted R-squared, and test the significance of the coefficients using t-tests and F- tests<sup>2</sup>.

Option A is incorrect, as logistic regression is used to fit a regression model that describes the relationship between one or more predictor variables and a binary response variable. Use when: The response variable is binary – it can only take on two values<sup>1</sup>. Ice cream consumption is not a binary variable, but rather a continuous numeric variable.

Option C is incorrect, as Cox regression is used to fit a regression model that describes the relationship between one or more predictor variables and a survival time response variable. Use when: The response variable is the time until an event of interest occurs, such as death, failure, or recovery<sup>3</sup>. Ice cream consumption is not a survival time variable, but rather a continuous numeric variable.

Option D is incorrect, as polynomial regression is used to fit a regression model that describes the relationship between one or more predictor variables and a numeric response variable. Use when: The relationship between the predictor variable(s) and the response variable is non-linear<sup>1</sup>. If there is no evidence of non-linearity in the data, polynomial regression may not be appropriate, as it may overfit the data and produce unreliable estimates.

**NEW QUESTION 240**

Which of the following is an object associated with a table that sorts and stores table row data in a key-value pair?

- A. Foreign key
- B. Function
- C. Stored procedure
- D. Clustered index

**Answer:** D

**NEW QUESTION 242**

You are working with a dataset and need to swap the values in rows with those in columns. What action do you need to perform?

- A. Recording
- B. Filtering.
- C. Aggregation.
- D. Transposition.

**Answer:** D

**Explanation:**

Transpose creates a new data file in which the rows and columns in the original data file are transposed so that cases (rows) become variables and variables (columns) become cases. Transpose automatically creates new variable names and displays a list of the new variable names. Transposing data is useful for data analysis. At times, we have to pull data from various files with different formats for analysis and preparing reports. In such circumstances, we may have to transpose some data from one file to the other. In excel, we can transpose data in multiple ways.

#### NEW QUESTION 247

A reporting analyst is creating a dashboard that shows the year-over-year performance for a sales organization. Which of the following is the best visual for the analyst use to illustrate the organization's performance?

- A. Pie chart
- B. Scatter plot
- C. Heat map
- D. Line chart

**Answer:** D

#### NEW QUESTION 249

Which one of the following is a measure of dispersion?

- A. Variance.
- B. Mode.
- C. Median.
- D. Mean.

**Answer:** A

#### NEW QUESTION 250

Different people manually type a series of handwritten surveys into an online database. Which of the following issues will MOST likely arise with this data? (Choose two.)

- A. Data accuracy
- B. Data constraints
- C. Data attribute limitations
- D. Data bias
- E. Data consistency
- F. Data manipulation

**Answer:** AE

**Explanation:**

? Data accuracy refers to the extent to which the data is correct, reliable, and free of errors. When different people manually type a series of handwritten surveys into an online database, there is a high chance of human error, such as typos, misinterpretations, omissions, or duplications. These errors can affect the quality and validity of the data and lead to incorrect or misleading analysis and decisions.

? Data consistency refers to the extent to which the data is uniform and compatible across different sources, formats, and systems. When different people manually type a series of handwritten surveys into an online database, there is a high chance of inconsistency, such as different spellings, abbreviations, formats, or standards. These inconsistencies can affect the integration and comparison of the data and lead to confusion or conflicts.

Therefore, to ensure data quality, it is important to have clear and consistent rules and procedures for data entry, validation, and verification. It is also advisable to use automated tools or methods to reduce human error and inconsistency.

#### NEW QUESTION 255

An analyst is currently working on a ticket for revamping a company-wide dashboard that has been in use for five years. Which of the following should be the first step in the development process?

- A. Talk to the group that made the request to determine the desired goal.
- B. Make changes to a frequently used report that is already in production.
- C. Build an additional dashboard with fewer views that are tailored toward each specific team.
- D. Develop a more streamMined dashboard to roll out by the next delivery date.

**Answer:** A

**Explanation:**

The first step in the development process of revamping a company-wide dashboard should be to talk to the group that made the request to determine the desired goal. This would help to understand the needs, expectations, and preferences of the stakeholders, as well as the scope, purpose, and objectives of the project. Talking to the group that made the request would also help to establish a clear communication channel, build rapport and trust, and solicit feedback and suggestions.

#### NEW QUESTION 260

Which of the following should an analyst do to best summarize the data on a data set?

- A. Filtering
- B. Aggregation
- C. Sorting
- D. Concatenation

**Answer:** B

**NEW QUESTION 262**

A salesperson who is prospecting potential clients collected the following data:

ID	Name	LName	Phone	Email
1	Jacob	Smith	(303)445-2323	jsmith@abc.com
2	Hans	Williams	(302)546-4588	hws@emc.com
3	Martha	Dion	(304)254-6575	dion@mail.com
4	Jules	Martin	(300)563-3435	jmartinxyz.com
5	Sabrina	Huggins	(323)655-3475	shug@emc.com

Which of the following is an issue with this data?

- A. Duplicate data
- B. Invalid data
- C. Missing value
- D. Redundant data

**Answer:** C

**NEW QUESTION 267**

A military commander would like to see the health scorecards of the troops daily and filter them based on gender and rank. Considering this data is PHI, which of the following would be the best way for the commander to view the information?

- A. An emailed report
- B. A password-protected dashboard
- C. A daily printout of a report
- D. A cloud-hosted spreadsheet

**Answer:** B

**Explanation:**

A password-protected dashboard is a type of web-based application that can display the health scorecards of the troops in a secure and interactive way. A password-protected dashboard can provide the following benefits for the commander:

- ? It can protect the PHI data from unauthorized access or disclosure by requiring a valid username and password to log in. This can ensure that only the commander and other authorized personnel can view the information<sup>12</sup>
- ? It can allow the commander to filter the data based on gender and rank by using drop-down menus, sliders, checkboxes, or other controls. This can enable the commander to customize the view and focus on the relevant data<sup>13</sup>
- ? It can update the data daily by connecting to a data source that refreshes automatically or on demand. This can ensure that the commander always sees the latest and most accurate information<sup>14</sup>
- ? It can present the data in a visual and intuitive way by using charts, graphs, tables, or other elements. This can help the commander to understand and analyze the data more easily and effectively<sup>1</sup>

**NEW QUESTION 268**

Randy scored 76 on a math test, Katie scored 86 on a science test, Ralph scored 80 on a history test, and Jean scored 80 on an English test. The table below contains the mean and standard deviation of the scores for each of the courses:

Course	Mean	Standard deviation
Math	70	2
Science	80	3
History	75	2
English	90	1

Using this information, which of the following students had the BEST score?

- A. Randy
- B. Katie
- C. Ralph
- D. Jean

**Answer:** B

**Explanation:**

To compare the students' scores, we need to standardize them by using the z-score formula, which is:

$$z = \frac{(x - \mu)}{\sigma}$$

where  $x$  is the raw score,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. The z-score tells us how many standard deviations a score is above or below the mean. A higher z-score means a better score relative to the average.

Using the table, we can calculate the z-scores for each student as follows:

Randy:  $z = (76 - 70) / 2 = 3$  Katie:  $z = (86 - 80) / 3 = 2$  Ralph:  $z = (80 - 75) / 2 = 2.5$  Jean:  $z = (80 - 90) / 1 = -10$

The student with the highest z-score is Randy, with a z-score of 3. This means that Randy scored 3 standard deviations above the mean in math, which is the best performance among the four students. Therefore, the correct answer is A.

References: Comparing with z-scores (video) | Z-scores | Khan Academy, 17 Important Data Visualization Techniques | HBS Online

**NEW QUESTION 273**

A company notifies its employees that emails will be automatically moved to a cloud-based server in 180 days. Which of the following describes this concept?

- A. Data deletion
- B. Data processing
- C. Data retention
- D. Data constraints

**Answer: C**

**NEW QUESTION 277**

An analyst is creating a resource to improve users' experience when they select specific records based on particular dates. Which of the following should the analyst use to create a resource that best meets user needs?

- A. Drop-down menu
- B. Date range
- C. Text field
- D. Frequency

**Answer: B**

**Explanation:**

A drop-down menu is a graphical user interface element that allows users to select one option from a list of options that are hidden until the user clicks on the menu. A drop-down menu can be used to create a resource that best meets user needs when they select specific records based on particular dates, because:

? A drop-down menu can provide a predefined list of dates or date ranges that are relevant and valid for the records, such as today, yesterday, last week, last month, custom range, etc. This can help users to avoid typing errors or invalid dates in a text field, and to save time and effort in entering the dates.

? A drop-down menu can also provide a calendar or a date picker that allows users

to select a specific date or a range of dates from a graphical representation of a calendar. This can help users to visualize and compare the dates, and to easily adjust or modify their selection.

? A drop-down menu can improve the user experience by making the interface more

compact and organized, as it only shows one option at a time and hides the rest of the options until the user clicks on the menu. This can help users to focus on their selection and to avoid clutter and distraction.

**NEW QUESTION 281**

Which of the following is most likely to be used as a data-mining ETL tool?

- A. SSIS
- B. Stata
- C. SPSS
- D. Cognos

**Answer: A**

**NEW QUESTION 283**

Given the data below:

First,Last,Company,Phone_number
John,Smith,Lee Shoes,(617) 310-5525
Charles,Wilson,Space Missiles Inc.,(203) 528-4466
Margaret,Lee,Lion Electronics,(515) 713-4817
Jennifer,Gonzalez,Private Financial Ltd.,(901) 207-1311

In which of the following file formats is the data presented?

- A. Xs
- B. CSV



- C. RIF
- D. XML

**Answer:** B

**Explanation:**

The data is presented in a CSV (comma-separated values) file format, which is a plain text format that stores tabular data. Each line of the file is a data record, and each record consists of one or more fields separated by commas. The first line of the file usually contains the names of the fields, also known as the header. In this case, the data has four fields: Name, Age, Gender, and Occupation. Therefore, the correct answer is B. References: CSV File (What It Is & How to Open One), Comma-separated values - Wikipedia

**NEW QUESTION 287**

What R package makes it easy to work with dates?

- A. Lubridate.
- B. Datemath.
- C. Stringr.
- D. ggplot.

**Answer:** A

**Explanation:**

Lubridate is an R package that makes it easier to work with dates and times.

**NEW QUESTION 288**

An analyst is designing a dashboard that will provide a story of the sales and sales customer ratio. The following data is available:

Site	Customers	New customers	Percentage of new customers	Sales volume	Average sales per customer
A1	2236	277	12%	\$3,415,372.00	\$1,527.45
A2	885	300	34%	\$1,405,437.00	\$1,588.06
A3	333	200	60%	\$952,723.00	\$2,861.03
B1	483	167	35%	\$4,871,380.00	\$10,085.67
B2	2969	235	8%	\$780,381.00	\$262.84
B3	2357	153	6%	\$4,917,436.00	\$2,086.31
C1	1524	180	12%	\$1,135,204.00	\$744.88
C2	878	150	17%	\$614,964.00	\$700.41
C2	1925	142	7%	\$4,035,100.00	\$2,096.16

Which of the following charts should the analyst consider including in the dashboard?

- A. A column chart with site and sales
- B. A line chart with site and sales
- C. A pie chart with site and sales
- D. A scatter chart with site and sales

**Answer:** A

**Explanation:**

For a dashboard that aims to tell a story about sales and the sales customer ratio, a column chart is an effective choice. Column charts are particularly useful for showing data changes over a period of time or for illustrating comparisons among items. In this case, a column chart can clearly display the sales figures for each site, allowing for easy comparison across different sites. Additionally, it can be used to represent the sales customer ratio by showing the proportion of sales per customer, which can provide insights into customer behavior and sales effectiveness.

? Line charts are best suited for displaying data trends over time, rather than for comparing individual categories.

? Pie charts could show the proportion of sales for each site, but they are not as effective as column charts for comparing multiple categories.

? Scatter charts are used to show the relationship between two variables, which is not the focus in this scenario.

References:

? Effective Use of Column Charts<sup>1</sup>

? Choosing the Right Chart for Your Data<sup>2</sup>

? Sales Dashboards: Examples & Templates<sup>3</sup>

**NEW QUESTION 291**

An analyst needs to join two tables of data together for analysis. All the names and cities in the first table should be joined with the corresponding ages in the second table, if applicable.

**Table 1**

Name	City
Jane Smith	Detroit
John Smith	Dallas
Candace Johnson	Atlanta
Kyle Jacobs	Chicago

**Table 2**

Name	Age
John Smith	34
John Smith	56
Candace Johnson	45
Kyle Jacobs	39

Which of the following is the correct join the analyst should complete. and how many total rows will be in one table?

- A. INNER JOIN, two rows
- B. LEFT JOIN, four rows
- C. four rows
- D. RIGHT JOIN, five rows
- E. five rows
- F. OUTER JOIN, seven rows

**Answer:** B

**Explanation:**

The correct join the analyst should complete is B. LEFT JOIN, four rows.

A LEFT JOIN is a type of SQL join that returns all the rows from the left table, and the matched rows from the right table. If there is no match, the right table will have null values. A LEFT JOIN is useful when we want to preserve the data from the left table, even if there is no corresponding data in the right table.

Using the example tables, a LEFT JOIN query would look like this:

```
SELECT t1.Name, t1.City, t2.Age FROM Table1 t1 LEFT JOIN Table2 t2 ON t1.Name = t2.Name;
```

The result of this query would be:

Name City Age Jane Smith Detroit NULL John Smith Dallas 34 Candace Johnson Atlanta 45 Kyle Jacobs Chicago 39

As you can see, the query returns four rows, one for each name in Table1. The name John Smith appears twice in Table2, but only one of them is matched with the name in Table1. The name Jane Smith does not appear in Table2, so the age column has a null value for that row.

**NEW QUESTION 292**

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. Data removal

**Answer:** DE

**Explanation:**

Data encryption and data masking are two actions that can be taken when transmitting data to mitigate the chance of a data leak occurring. Data encryption means transforming data into an unreadable format that can only be decrypted with a key. Data masking means hiding or replacing sensitive data with fictitious or anonymized data. Both methods protect the confidentiality and integrity of the data in transit. References: CompTIA Data+ Certification Exam Objectives, page 13

**NEW QUESTION 294**

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:**

A static report is a type of report that shows a snapshot of data at a specific point in time. A static report does not change or update automatically, unless the data source is refreshed or the report is regenerated. A static report is suitable for situations where the data does not change frequently or where historical data is needed for comparison or analysis. In this case, the data analyst is asked to create a sales report for the second-quarter 2020 board meeting, which will include a review of the business's performance through the second quarter. The board meeting will be held on July 15, 2020, after the numbers are finalized. This means that the data analyst does not need to show real-time or dynamic data, but rather a fixed and accurate view of the sales data for the second quarter. Therefore, a static report would be the best way to meet this stakeholder requirement. Therefore, the correct answer is A. References: What are Static Reports? | Sisense, Static vs Dynamic Reports - What's The Difference? | datapine

**NEW QUESTION 299**

Which of the following is an example of a discrete variable?

- A. The temperature of a hot tub
- B. The height of a horse
- C. The time to complete a task
- D. The number of people in an office

**Answer:** D

**Explanation:**

A discrete variable is a variable that can only take on a finite number of values, such as integers or categories. The number of people in an office is an example of a discrete variable, as it can only be a whole number. The temperature of a hot tub, the height of a horse, and the time to complete a task are examples of continuous variables, as they can take on any value within a range. Reference: CompTIA Data+ (DAO-001) Practice Certification Exams | Udemy

**NEW QUESTION 304**

Which of the following is a best practice when updating a legacy data source?

- A. Placing old data in new fields
- B. Keeping only the most recent data
- C. Creating a codebook to document field changes
- D. Removing the data source from production

**Answer:** C

**Explanation:**

When updating a legacy data source, it is a best practice to create a codebook to document field changes. A codebook serves as a detailed guide and record of the data structure, definitions, and any transformations or modifications made to the data fields. This documentation is crucial for maintaining data integrity, ensuring consistency, and facilitating future data use and understanding. It provides a reference that can be invaluable for data analysts, developers, and any stakeholders who need to work with the data.

Creating a codebook is preferred over placing old data in new fields, which can lead to confusion and data integrity issues. Keeping only the most recent data may result in the loss of valuable historical information. Removing the data source from production is not a practice related to updating data but rather to retiring a data source<sup>1234</sup>.

References:

- ? Legacy Data Migration: A Comprehensive Guide | OpenGeeksLab
- ? How to Successfully Complete Legacy Database Migration
- ? Methods for Saving and Integrating Legacy Data - DATAVERSITY
- ? Legacy Data Digitization - Learn The Best Practices

**NEW QUESTION 307**

A data analyst needs to create a dashboard to help identify trends in the data sets. Which of the following is an appropriate consideration for dashboard development?

- A. Data sources and attributes
- B. Frequently asked questions
- C. A report from the data source
- D. A comparison of data sets

**Answer:** A

**Explanation:**

When creating a dashboard to identify trends in data sets, the most appropriate consideration is the data sources and attributes. This is because the quality, reliability, and structure of the data sources directly influence the dashboard's ability to accurately reflect trends. Attributes, such as the type of data and the time frame it covers, are crucial for trend analysis. A well-designed dashboard should provide a clear and intuitive representation of the data, allowing for easy identification of trends and patterns. Frequently asked questions (B) can inform the design of the dashboard but are not a direct consideration for the development process itself. A report from the data source © might be an output of the dashboard but does not guide its development. A comparison of data sets (D) could be a feature of the dashboard, but the underlying data sources and attributes must be considered first to ensure accurate and meaningful comparisons. References:

- ? Best practices in dashboard design emphasize the importance of understanding and consolidating different data sources and creating a mix of useful metrics, which aligns with the choice of data sources and attributes<sup>1</sup>.
- ? Fundamental dashboard design principles include the clear and efficient display of information, which is dependent on the proper selection and use of data

sources and attributes2.

? Effective dashboard communication is achieved by using colors, shapes, sizes, labels, and legends meaningfully, all of which rely on the underlying data sources and attributes3.

#### **NEW QUESTION 312**

.....



## 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

### DA0-001 Practice Exam Features:

- \* DA0-001 Questions and Answers Updated Frequently
- \* DA0-001 Practice Questions Verified by Expert Senior Certified Staff
- \* DA0-001 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- \* DA0-001 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

**100% Actual & Verified — Instant Download, Please Click**  
**[Order The DA0-001 Practice Test Here](#)**