Exam Questions DA0-001

CompTIA Data+ Certification Exam

https://www.2passeasy.com/dumps/DA0-001/

NEW QUESTION 1
Taylor wants to investigate how manufacturing, marketing, and sales expenditures impact overall profitability for her company.
Which of the following systems is the most appropriate?

A. OLTP.
B. OLAP.
C. Data warehouse.
D. Data mart.

Answer: C

Explanation:

A Data mart is too narrow, because Taylor needs data from across multiple divisions. OLAP is a broad term for analytical processing, and OLTP systems are transactional and not ideal for the task. Since Taylor is working with data across multiple different divisions, she will work with a Data warehouse.

NEW QUESTION 2
A data set was recorded using multimedia technology. Which of the following is a necessary step on the way to interpretation?

A. Structural equation modeling
B. Transcription
C. Sequential analysis
D. Sampling

Answer: B

Explanation:
The correct answer is B. Transcription.
Transcription is a necessary step on the way to interpretation when a data set was recorded using multimedia technology. Multimedia technology refers to the use of various forms of media, such as audio, video, images, and text, to capture and present information1 Transcription is the process of converting multimedia data into written or textual form, which can then be analyzed using various methods and tools2 Transcription can help to make the data more accessible, searchable, and manageable, as well as to preserve the data for future use.
Structural equation modeling is not correct, because it is a statistical technique that tests the causal relationships between multiple variables using observed and latent variables. Structural equation modeling is not a necessary step on the way to interpretation, but rather an optional method that can be applied to certain types of data.
Sequential analysis is not correct, because it is a method of analyzing the order and timing of events or behaviors in a data set. Sequential analysis is not a necessary step on the way to interpretation, but rather an optional method that can be applied to certain types of data. Sampling is not correct, because it is the process of selecting a subset of data from a larger population for analysis. Sampling is not a necessary step on the way to interpretation, but rather a preliminary step that can be done before collecting or analyzing the data.

NEW QUESTION 3
You are working with a professional statistician to perform an analysis and would like to use a statistics package.
Which one of the following would be the most appropriate?

A. Rapid Miner.
B. QLIK.
C. Power BI.
D. Minitab.

Answer: D

Explanation:
Minitab is statistical analysis software. It can be used for learning about statistics as well as statistical research. Statistical analysis computer applications have the advantage of being accurate, reliable, and generally faster than computing statistics and drawing graphs by hand.

NEW QUESTION 4
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 5
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 6
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 7
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 8
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 9
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 10
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 10
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 12
An analyst is building a new dashboard for a user. After an initial conversation with the user. the analyst created a mock-up of the dashboard. Which of the following best explains why the analyst created the mock-up?

A. To identify the dimensions and measures
B. To send to the client after deploying the dashboard to production
C. To confirm important details before dashboard development begins
D. To receive client approval for the final dashboard design

Answer: C

Explanation:
Answer C. To confirm important details before dashboard development begins.
A dashboard mockup is a prototype of a finished dashboard directly in the product. It is a way to visualize the layout, design, and functionality of the dashboard before it is built with real data and code. A dashboard mockup can help the analyst to confirm important details
with the user, such as the business objectives, the key performance indicators, the data sources, the filters, the charts, and the interactivity. By creating a dashboard mockup, the analyst can get immediate feedback and validation from the user, and avoid wasting time and resources on developing a dashboard that does not meet the user??s expectations or needs1.

NEW QUESTION 15
Which of following is a non-relational database?

A. Neo4j
B. SQLite
C. MySQL
D. PostgreSQL

Answer: A

Explanation:
Neo4j is a type of non-relational database that uses a graph model to store data. A graph database is a database that represents data as nodes and edges, where nodes are entities and edges are relationships between them. A graph database can store complex and diverse data that is not easily structured in tables. A graph database can also perform fast and efficient queries on the data by traversing the connections between the nodes

NEW QUESTION 20
An analyst collected data that includes primary account numbers, expiration dates, and service codes. Which of the following data governance classifications is used to describe this data?

A. PI I
B. PCI
C. PBI
D. PHI

Answer: B

NEW QUESTION 22
Which of the following is a difference between a primary key and a unique key?

A. A unique key cannot take null values, whereas a primary key can take null values.
B. There can be only one primary key in a data set, whereas there can be multiple unique keys.
C. A primary key can take a value more than once, whereas a unique key cannot take a value more than once.
D. A primary key cannot be a date variable, whereas a unique key can be.

Answer: B

Explanation:
The correct answer is B. There can be only one primary key in a data set, whereas there can be multiple unique keys.
A primary key is a column or a set of columns that uniquely identifies each row in a table. A table can have only one primary key, which also enforces the NOT NULL constraint on the column(s) involved. A primary key can also be referenced by a foreign key of another table to establish a relationship between the tables12
A unique key is a column or a set of columns that also uniquely identifies each row in a table, but it is not the primary key. A table can have more than one unique key, which also allows one NULL value for the column(s) involved. A unique key can also be referenced by a foreign key of another table to establish a relationship between the tables12
Some of the differences between a primary key and a unique key are:
? A primary key creates a clustered index on the column(s), whereas a unique key creates a non-clustered index on the column(s)3
? A primary key does not allow any NULL values, whereas a unique key allows one
NULL value for the column(s)123
? A primary key can be a unique key, but a unique key cannot be a primary key12

NEW QUESTION 26
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{55 + 36 + 18 + 67 + 36 + 38}{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(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 28
Which of the following data types would a telephone number formatted as XXX-XXX-XXXX be considered?

A. Numeric
B. Date
C. Float
D. Text

Answer: D

Explanation:
A telephone number formatted as XXX-XXX-XXXX would be considered a text data type, as it is composed of alphanumeric characters and symbols. A numeric data type is composed of only numbers, such as integers or decimals. A date data type is composed of values that represent dates or times, such as YYYY-MM-DD or HH:MM:SS. A float data type is composed of numbers with fractional parts, such as 3.14 or 0.5. Reference: Guide to CompTIA Data+ and Practice Questions - Pass Your Cert

NEW QUESTION 31
A database consists of one fact table that is composed of multiple dimensions. Each dimension is represented by a denormalized table. This structure is an example of a:

A. non-relational schema.
B. galaxy schema.
C. snowflake schema.
D. star schema.

Answer: D

Explanation:
A star schema is a type of database schema that consists of one fact table and multiple dimension tables. The fact table contains the measures or metrics of the business process, such as sales, orders, or transactions. The dimension tables contain the attributes or characteristics of the business entities, such as products, customers, or locations. The fact table is connected to the dimension tables by foreign keys that reference the primary keys of the dimension tables. The fact table is located at the center of the schema, while the dimension tables are located at the edges, forming a star-like shape1.
A star schema is an example of a denormalized schema, which means that the dimension tables are not normalized and may contain redundant or repeated data. This is done to improve the performance and simplicity of queries, as there are fewer joins and tables involved. A star schema is suitable for data warehouses and business intelligence applications that require fast and efficient data retrieval2.

NEW QUESTION 34
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 35
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 category12.

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 plaintext3.
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 sun4.
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 38
An analyst develops an IT document and needs to describe the technical terms used in the document. Which of the following is where the analyst should include descriptions of the technical terms?

A. Glossary
B. System diagram
C. User requirements
D. Index

Answer: A

Explanation:
In technical documentation, a glossary is the designated section where definitions for technical terms are provided. It serves as a reference point for readers to understand specialized or uncommon words used within the document. Including descriptions of technical terms in a glossary ensures that readers have a consistent resource to refer to, which can improve comprehension and reduce misunderstandings12.
A system diagram (Option B) is a visual representation of the system??s components and their interactions, not a place for defining terms. User requirements (Option C) outline what end-users expect from the system, and an index (Option D) is an alphabetical list of topics covered in the document, usually with page numbers, but not definitions.
References:
? Creating effective technical documentation1.
? Best practices when writing technical descriptions3.

NEW QUESTION 42
Which of the following data types must be used when working with variables that require classification into two or more groups before analysis?

A. Discrete
B. Numerical
C. Alphanumeric
D. Categorical

Answer: D

NEW QUESTION 44
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 values12
* 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 deviation34
* 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 47
A data analyst needs to create a master file that includes customer information from the tables below:

Table 1: Online Transactions

Order_ID	Customer_ID	Date	Amount	Quantity
002A	002	03/01/2020	$800	109
001B	001	02/01/2020	$400	14
001B	001	02/01/2020	$400	14
001B	001	02/01/2020	$400	14
004C	004	06/01/2020	$700	52
003D	003	05/01/2020	$900	20

Table 2: In-store Transactions

Order_ID	Customer_ID	Date	Amount	Quantity
006A	006	04/01/2020	$200	59
007B	007	03/01/2020	$500	54
008C	008	02/01/2020	$600	15
009D	009	05/01/2020	$800	18
001E	001	07/01/2020	$300	50
003F	003	08/01/2020	$200	55

Table 3: Customer Table

Customer_ID	Segment	Region
001	New	BC
002	Existing	ON
003	New	MB
004	New	ON
005	Existing	AT
006	Existing	MB
007	New	QC
008	New	QC
009	Existing	BC

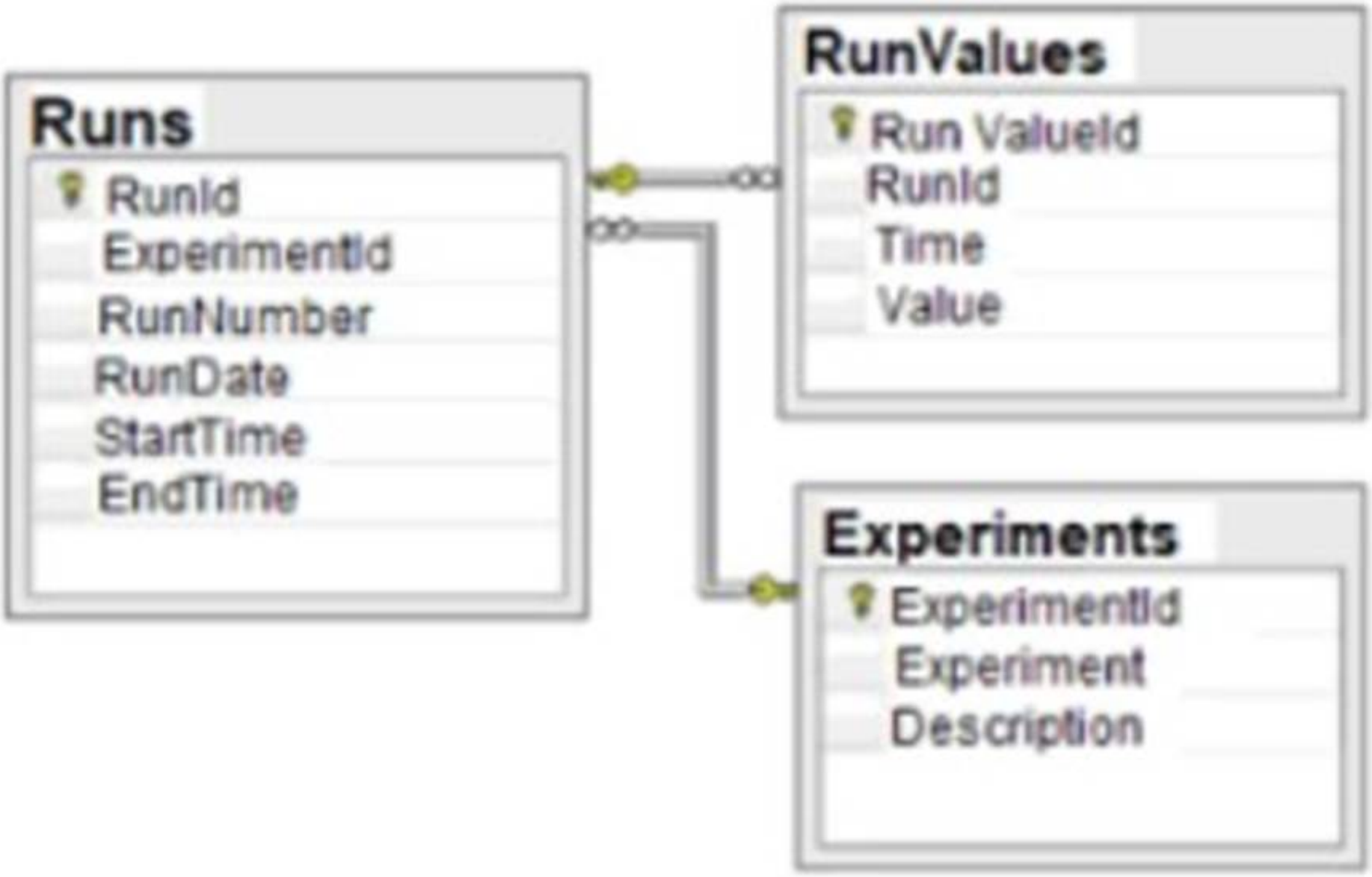
Given the three tables above, the analyst wants to filter down the information prior to joining it together. In which of the following orders should this data manipulation bo approached for the most efficient result?

A. Merge, append, deduplicate
B. Merge, deduplicate, append
C. Deduplicate, append, merge
D. Append, deduplicate, merge

Answer: B

Explanation:
For efficient data manipulation, the ideal order would be to first merge related tables to create a comprehensive set of records, then deduplicate to remove any redundant information. Lastly, appending additional data, such as from another source or table, ensures that all relevant data is included without redundancy before the final analysis. This order prevents unnecessary duplication of effort, such as deduplicating both before and after appending, which would be less efficient.
In the context of the tables provided, merging would likely involve combining customer information from the online and in-store transaction tables with the customer table. Deduplication would remove any redundant customer records that may exist across these tables. Finally, appending would involve adding any additional transaction records to the master file, ensuring a complete dataset for analysis.

NEW QUESTION 48
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 51
A customer list from a financial services company is shown below:

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

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

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

Answer: D

Explanation:
Normalizing the variables means scaling them to a common range, such as 0 to 1 or -1 to 1, so that they have the same weight in the score calculation. Recoding the variables means changing their values or categories, which would alter their meaning and distribution. Calculating the percentiles of the variables means ranking them relative to each other, which would not account for their actual magnitudes. Calculating the standard deviations of the variables means measuring their variability, which would not make them comparable. References: CompTIA Data+ Certification Exam Objectives, page 10

NEW QUESTION 56
Which of the following would be considered non-personally identifiable information?

A. Cell phone device name
B. Customer??s name
C. Government ID number
D. Telephone number

Answer: A

Explanation:
Non-personally identifiable information (non-PII) is any data that cannot be used to identify, contact, or locate a specific individual, either alone or combined with other sources. Non-PII can include aggregated statistics, anonymous data, device identifiers, IP addresses, cookies, and other types of information that do not reveal the identity or location of a person. Cell phone device name is an example of non-PII, as it does not reveal any personal information about the owner or user of the device. Therefore, the correct answer is A. References: What is Non-Personally Identifiable Information (Non-PII)? | Definition and Examples, What is Personally Identifiable Information (PII)? | Definition and Examples

NEW QUESTION 57
Kelly wants to get feedback on the final draft of a strategic report that has taken her six months to develop.
What can she do to get prevent confusion as see seeks feedback before publishing the report?
Choose the best answer.

A. Distribute the report to the appropriate stakeholders via email.
B. Use a watermark to identify the report as a draft.
C. Show the report to her immediate supervisor.
D. Publish the report on an internally facing website.

Answer: B

Explanation:

The best answer is to use a watermark to identify the report as a draft. A watermark is a faint image or text that appears behind the content of a document, indicating its status or ownership. By using a watermark, Kelly can clearly communicate that the report is not final and still subject to changes or feedback. This can prevent confusion among the readers and avoid any misuse or misinterpretation of the report. The other options are not as effective as using a watermark, as they either do not indicate the status of the report or do not reach the appropriate stakeholders. Distributing the report via email or publishing it on an internally facing website may not make it clear that the report is a draft and may cause confusion or errors. Showing the report to her immediate supervisor may not get enough feedback from other relevant stakeholders who may have different perspectives or insights. Reference: How to Add a Watermark in Microsoft Word - Lifewire

NEW QUESTION 61
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 languages12.

NEW QUESTION 62
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 66
A client has requested an analysis of all pet care items purchased by current customers and their social media connections in the past 12 months. Which of the following data analysis techniques would be the best choice given these requirements?

A. Trend analysis
B. Performance analysis
C. Link analysis
D. Exploratory data analysis

Answer: C

NEW QUESTION 69
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 74
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 78
??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 dat
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 79
Joseph is interpreting a left skewed distribution of test scores. Joe scored at the mean, Alfonso scored at the median, and gaby scored and 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 81
The current date is July 14, 2020. A data analyst has been asked to create a report that shows the company??s year-over-year Q2 2020 sales. Which of the following reports should
the analyst compare?

A. A Q2 2020 and Q4 2019
B. YTD 2020 and YTD 2019
C. Q2 2020 and Q2 2019
D. Q2 2020 and Q2 2021

Answer: C

Explanation:
To create a report that shows the company??s year-over-year Q2 2020 sales, the analyst should compare the sales data from Q2 2020 and Q2 2019. Year-over-year (YoY) analysis is a method of comparing the performance of a business or a financial instrument over the same period in different years. It helps to identify trends, growth patterns, and seasonal fluctuations. Q2 refers to the second quarter of a year, which is usually from April to June. Therefore, the correct answer is C. References: YoY - Year over Year Analysis - Definition, Explanation & Examples, What is an Annual Sales Report: Definition, metrics, and tips - Snov.io

NEW QUESTION 83
Emma is working in a data warehouse and finds a finance fact table links to an organization dimension, which in turn links to a currency dimension that not linked to the fact table.
What type of design pattern is the data warehouse using?

A. Star.
B. Sun.
C. Snowflake.
D. Comet.

Answer: C

Explanation:
Correct answer C. Snowflake.
Since the dimension links to a dimension that isn't connected to the fact table, it must be a Snowflake, with a Star, all dimensions link directly to the fact table, Sun

and Comet are not data warehouse design patterns.

NEW QUESTION 84
Which of the following descriptive statistical methods are measures of central tendency? (Choose two.)

A. Mean
B. Minimum
C. Mode
D. Variance
E. Correlation
F. Maximum

Answer: AC

Explanation:
Mean and mode are measures of central tendency, which describe the typical or most common value in a distribution of data. Mean is the arithmetic average of all the values in a dataset, calculated by adding up all the values and dividing by the number of values. Mode is the most frequently occurring value in a dataset. Other measures of central tendency include median, which is the middle value when the data is sorted in ascending or descending order.

NEW QUESTION 88
Which one of the following in 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 93
An analyst has generated a report that includes the number of months in the first two quarters of 2019 when sales exceeded $50,000:

Month	Sales	Sales_indicator
January 2019	$52,005	Exceeded $50,000
February 2019	$48,687	Not exceeded $50,000
March 2019	$50,255	Exceeded $50,000
April 2019	$38,924	Not exceeded $50,000
June 2019	$57,076	Exceeded $50,000
July 2019	$51,035	Exceeded $50,000

Which of the following functions did the analyst use to generate the data in the Sales_indicator column?

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

Answer: B

Explanation:
This is because a logical function is a type of function that returns a value based on a condition or a set of conditions. A logical function can be used to generate the data in the Sales_indicator column by comparing the values in the Sales column with a threshold of $50,000 and returning either ??Exceeded $50,000?? or ??Not exceeded $50,000?? accordingly. For example, a logical function in Excel that can achieve this is:

```
=IF(Sales>50000,"Exceeded $50,000","Not exceeded $50,000")
```

The other functions are not suitable for generating the data in the Sales_indicator column. Here is why:
Aggregate is a type of function that performs a calculation on a group of values, such as sum, average, count, etc. An aggregate function cannot generate the data in the Sales_indicator column because it does not compare the values in the Sales column with a threshold or return a text value based on a condition.
Date is a type of function that manipulates or extracts information from dates, such as year, month, day, etc. A date function cannot generate the data in the Sales_indicator column because it does not use the values in the Sales column or return a text value based on a condition.
Sort is a type of function that arranges the values in a column or a range in ascending or descending order. A sort function cannot generate the data in the

Sales_indicator column because it does not create a new column or return a text value based on a condition.

NEW QUESTION 97
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 preserved1
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 98
Angela is aggregating data from CRM system with data from an employee system.
While performing an initial quality check, she realizes that her employee ID is not associated with her identifier in the CRM system.
What kind of issues is Angela facing? Choose the best answer.

A. ETL process.
B. Record linkage.
C. ELT process.
D. System integration.

Answer: B

Explanation:

While this scenario describes a system integration challenge that can be solved with ETL or ELT, Angela is facing a Record linkage issue.

NEW QUESTION 103
Given the following customer and order tables:
Which of the following describes the number of rows and columns of data that would be present after performing an INNER JOIN of the tables?

A. Five rows, eight columns
B. Seven rows, eight columns
C. Eight rows, seven columns
D. Nine rows, five columns

Answer: B

Explanation:
This is because an INNER JOIN is a type of join that combines two tables based on a matching condition and returns only the rows that satisfy the condition. An INNER JOIN can be used to merge data from different tables that have a common column or a key, such as customer ID or order ID. To perform an INNER JOIN of the customer and order tables, we can use the following SQL statement:

```
SELECT * FROM customer INNER JOIN order ON customer.customer_id = order.customer_id;
```

This statement will select all the columns (*) from both tables and join them on the customer ID column, which is the common column between them. The result of this statement will be a new table that has seven rows and eight columns, as shown below:

customer_id	first_name	last_name	email	order_id	order_date	product	quantity
1	John	Smith	john.smith@email.com	1	2020-01-01	Book	2
2	Jane	Doe	jane.doe@email.com	2	2020-01-02	Pen	5
3	Bob	Lee	bob.lee@email.com	3	2020-01-03	Notebook	3
4	Mia	Chen	mia.chen@email.com	4	2020-01-04	Mug	4
5	Raj	Patel	raj.patel@email.com	null	null	null	null
null	null	null	null	null	null	null	null

The reason why there are seven rows and eight columns in the result table is because:
? There are seven rows because there are six customers and six orders in the original tables, but only five customers have matching orders based on the customer ID column. Therefore, only five rows will have data from both tables, while one row will have data only from the customer table (customer 5), and one row will have no data at all (null values).
? There are eight columns because there are four columns in each of the original tables, and all of them are selected and joined in the result table. Therefore, the result table will have four columns from the customer table (customer ID, first name, last name, and email) and four columns from the order table (order ID, order date, product, and quantity).

NEW QUESTION 105
While reviewing survey data, an analyst notices respondents entered ??Jan,?? ??January,?? and ??01?? as responses for the month of January. Which of the following steps should be taken to ensure data consistency?

A. Delete any of the responses that do not have ??January?? written out.
B. Replace any of the responses that have ??01??.
C. Filter on any of the responses that do not say ??January?? and update them to ??January??.
D. Sort any of the responses that say ??Jan?? and update them to ??01??.

Answer: C

Explanation:
Filter on any of the responses that do not say ??January?? and update them to ??January??. This is because filtering and updating are data cleansing techniques that can be used to ensure data consistency, which means that the data is uniform and follows a standard format. By filtering on any of the responses that do not say ??January?? and updating them to ??January??, the analyst can make sure that all the responses for the month of January are written in the same way. The other steps are not appropriate for ensuring data consistency. Here is why:
Deleting any of the responses that do not have ??January?? written out would result in data loss, which means that some information would be missing from the data set. This could affect the accuracy and reliability of the analysis.
Replacing any of the responses that have ??01?? would not solve the problem of data inconsistency, because there would still be two different ways of writing the month of January: ??Jan?? and ??January??. This could cause confusion and errors in the analysis. Sorting any of the responses that say ??Jan?? and updating them to ??01?? would also not solve the problem of data inconsistency, because there would still be two different ways of writing the month of January: ??01?? and ??January??. This could also cause confusion and errors in the analysis.

NEW QUESTION 110
A data analyst is developing a data dictionary that aligns with a company's data management processes and policies. Which of the following best describes what should be included in the data dictionary?

A. Information containing the links to business data

B. Information explaining the business methodologies
C. Information containing definitions of the business data
D. Information describing the data analysis phases

Answer: C

NEW QUESTION 111
Which of the following value is the measure of dispersion "range" between the scores of ten students in a test.
The scores of ten students in a test are 17, 23, 30, 36, 45, 51, 58, 66, 72, 77.

A. 90
B. 60
C. 70
D. 80

Answer: B

Explanation:

The correct answer is: 60
Range is the interval between the highest and the lowest score.
Range is a measure of variability or scatteredness of the varieties or observations among themselves and does not give an idea about the spread of the observations around some
central value. Symbolically R = Hs - Ls.
Where R = Range; Hs is the 'Highest score' and Ls is the Lowest Score.
The scores of ten students in a test are: 17, 23, 30, 36, 45, 51, 58, 66, 72, 77. The highest score is 77 and the lowest score is 17.
So the range is the difference between these two scores Range = 77 - 17 = 60

NEW QUESTION 114
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 115
Which of the following is the best variable formal to store a customer's age using the least possible amount of storage data?

A. Int
B. Float
C. Char
D. Double

Answer: A

NEW QUESTION 117
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 121
Which one the following is not considered an aggregate function?

A. SUM
B. MIN
C. SELECT
D. MAX

Answer: C

Explanation:
The option that is not considered an aggregate function is SELECT. An aggregate function is a function that performs a calculation on a set of values and returns a single value. Examples of aggregate functions are SUM, MIN, MAX, AVG, COUNT, etc. SELECT is not an aggregate function, but a SQL command that is used to select data from a table or a query. Reference: SQL Aggregate Functions - W3Schools

NEW QUESTION 124

Which one of the following is a common data warehouse schema?

A. Snowflake.
B. Square.
C. Spiral.
D. Sphere.

Answer: A

Explanation:

Snowflake enables data storage, processing, and analytic solutions that are faster, easier to use, and far more flexible than traditional offerings. The Snowflake data platform is not built on any existing database technology or ??big data?? software platforms such as Hadoop.

NEW QUESTION 125
Standardized tests are given to students in the middle of each month, and the results are ready by the end of the month. The superintendent needs a quick view of test performance. Which of the following would be the best recommendation to meet the superintendent's requirements?

A. A dashboard with a continuous data stream and saved searches
B. A report of test scores by classroom, emailed to the superintendent at the end of the month
C. A report of test scores with pie charts showing student performance
D. A dashboard with a scheduled delivery, the ability to filter scores by school, and bar charts for comparison

Answer: D

Explanation:
A dashboard with a scheduled delivery is an efficient way to provide a quick view of test performance. It allows for timely updates, which is crucial given that the superintendent needs the information promptly at the end of each month. The ability to filter scores by school enables the superintendent to easily segment and analyze the data as needed. Bar charts are effective for comparison and can visually communicate the performance across different schools or other categories, making it easier to identify trends and outliers at a glance.
References:
? Best practices in data visualization recommend using dashboards for real-time data monitoring and quick access to key metrics1.
? Guidelines for presenting performance data suggest that visual tools like bar charts are helpful in comparing and analyzing data effectively1.
? Educational performance data analysis often involves comparing scores across different schools or classrooms, which is facilitated by a well-designed dashboard2.

NEW QUESTION 128
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

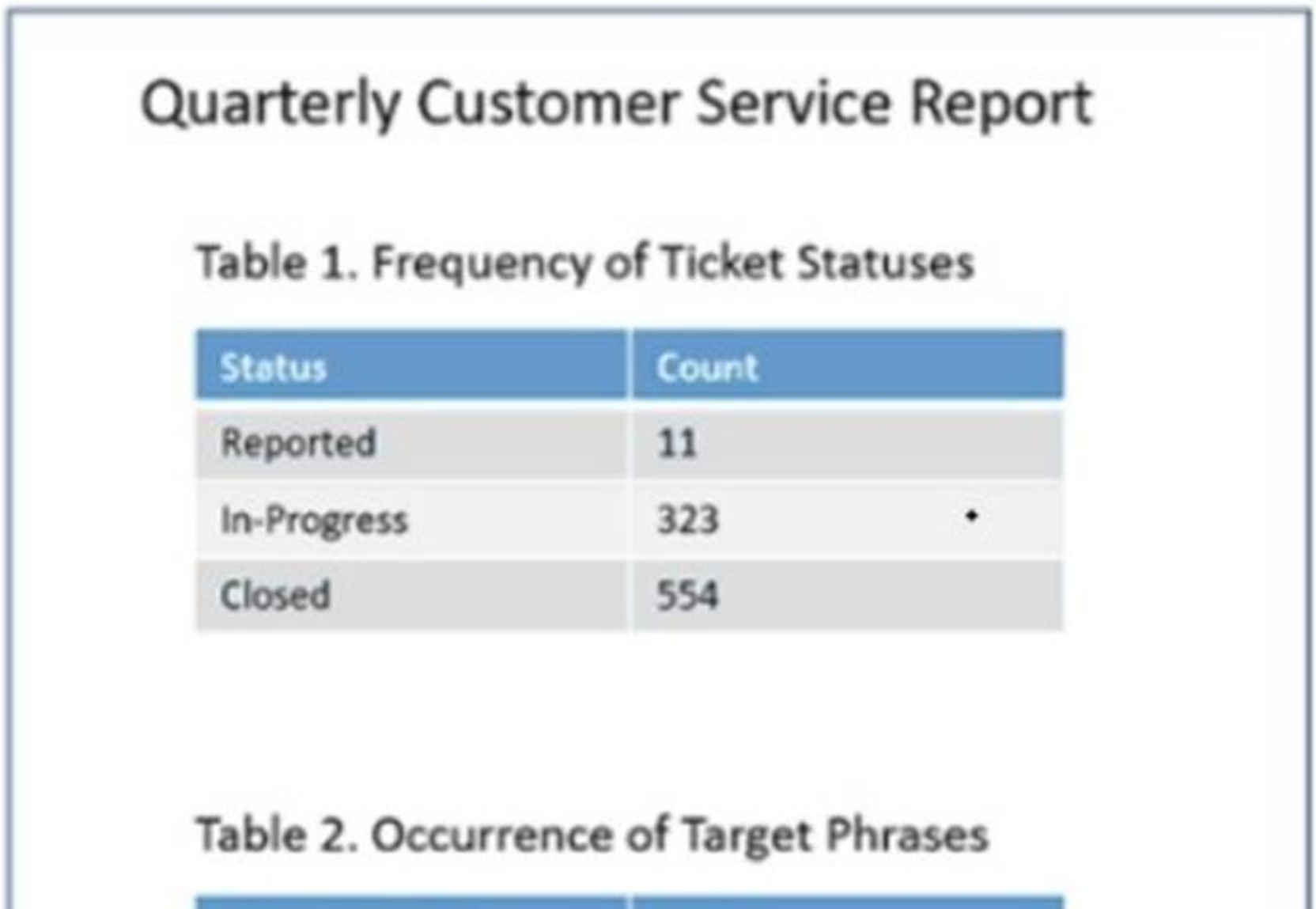
First name	Last name	Address
John	Knox	2851 N. Southport
John	Johnson	457 Bridle Ridge
John	Sinclair	1067 Windwood Lane
Bob	Sinclair	71 S. Wacker Drive

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 132
Given the following report:



Quarterly Customer Service Report

Table 1. Frequency of Ticket Statuses

Status	Count
Reported	11
In-Progress	323
Closed	554

Table 2. Occurrence of Target Phrases

Target Phrases	Count
Have a great day!	1200
It is my pleasure to assist you.	70
Can you please hold?	7352

Most tickets are being addressed soon after being reported. Asking customers to hold is the most commonly used target phrase.

Which of the following components need to be added to ensure the report is point-in-time and static? (Choose two.)

A. A control group for the phrases
B. A summary of the KPIs
C. Filter buttons for the status
D. The date when the report was last accessed
E. The time period the report covers
F. The date on which the report was run

Answer: E

Explanation:
The date on which the report was run. This is because the time period the report covers and the date on which the report was run are two components that need to be added to ensure the report is point-in-time and static, which means that the report shows the data as it was at a specific moment or interval in time, and does not change or update with new data. By adding the time period the report covers and the date on which the report was run, the analyst can indicate when and for how long the data was collected and analyzed, as well as avoid any confusion or ambiguity about the currency or validity of the data. The other components do not need to be added to ensure the report is point-in-time and static. Here is why:
A control group for the phrases is a type of group that serves as a baseline or a reference for comparison with another group that is exposed to some treatment or

intervention, such as a target phrase in this case. A control group for the phrases does not need to be added to ensure the report is point-in-time and static, because it does not affect the time frame or the stability of the data. However, a control group for the phrases could be useful for evaluating the effectiveness or impact of the target phrases on customer satisfaction or retention.
A summary of the KPIs is a type of document that provides an overview or a highlight of the key performance indicators (KPIs), which are measurable values that indicate how well an organization or a process is achieving its goals or objectives. A summary of the KPIs does not need to be added to ensure the report is point-in-time and static, because it does not affect the time frame or the stability of the data. However, a summary of the KPIs could be useful for communicating or presenting the main findings or insights from the report.
Filter buttons for the status are a type of feature or function that allows users to select or deselect certain values or categories in a column or a table, such as ticket statuses in this case. Filter buttons for the status do not need to be added to ensure the report is point-in- time and static, because they do not affect the time frame or the stability of the data. However, filter buttons for the status could be useful for exploring or analyzing different aspects or segments of the data.

NEW QUESTION 133
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 134
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 138
The total values in this month's revenue report are twice as much as last month's. Which of the following most likely occurred during the ETL process?

A. The data cleansing processes failed to execute.
B. The database connectivity failed.
C. The report included the previous month's data.
D. The data normalization processes failed.

Answer: C

NEW QUESTION 141
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 shouts 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 146
An analyst needs to provide a chart to identify the composition between the categories of the survey response data set:

Favorite color	Responses
Red	15
Blue	35
Green	25
Yellow	25
Total	100

Which of the following charts would be BEST to use?

A. Histogram
B. Pie
C. Line
D. Scatter pot
E. Waterfall

Answer: B

Explanation:
A pie chart is the best choice to show the composition between the categories of the survey response data set. A pie chart represents the whole with a circle, divided by slices into parts. Each slice shows the relative size of each category as a percentage of the total. A pie chart is useful when the categories are mutually exclusive and add up to 100%. The table shows the favorite color and the number of responses for each color, which can be easily converted into percentages. A pie chart can show how each color contributes to the total number of responses.
Option A is incorrect because a histogram is used to show how data points are distributed along a numerical scale. The survey response data set is not numerical, but categorical. Option C is incorrect because a line chart is used to show trends or changes over time. The survey response data set does not have a time dimension.
Option D is incorrect because a scatter plot is used to show the relationship between two numerical variables. The survey response data set does not have two numerical variables. Option E is incorrect because a waterfall chart is used to show how an initial value is increased or decreased by a series of intermediate values. The survey response data set does not have an initial value or intermediate values.
References:
? How to Choose the Right Chart for Your Data - Infogram
? How to Choose the Right Data Visualization | Tutorial by Chartio

? Find the Best Visualizations for Your Metrics - The Data School
? How to choose the best chart or graph for your data

NEW QUESTION 149
A data analyst is working with a team to create a dashboard for a client who requires on- demand access. Which of the following is the best delivery method to support the clients?? requirement?

A. Email
B. Scheduled
C. Subscription
D. Static

Answer: C

Explanation:
The best delivery method to support the client??s requirement is C. Subscription.
Short Explanation: A subscription is a delivery method that allows the client to access the dashboard on-demand, whenever they need it. A subscription can be set up by the data analyst or the client themselves, and it can be configured to send an email notification when the dashboard is updated or refreshed. A subscription also allows the client to view the dashboard online or download it as a file format of their choice12
* A. Email is not the best delivery method because it does not allow the client to access the dashboard on-demand. Email deliveries are sent at a fixed time or frequency, and they may not reflect the latest data or changes in the dashboard. Email deliveries also have limitations on the file size and format of the dashboard attachments1
* B. Scheduled is not the best delivery method because it does not allow the client to access the dashboard on-demand. Scheduled deliveries are similar to email deliveries, except that they are triggered by a specific event or condition, such as a data update or a threshold value. Scheduled deliveries also have the same limitations as email deliveries on the file size and format of the dashboard attachments1
* D. Static is not the best delivery method because it does not allow the client to access the dashboard on-demand. Static deliveries are one-time deliveries that are manually generated by the data analyst or the client. Static deliveries do not update or refresh automatically, and they may become outdated or irrelevant over time. Static deliveries also have limitations on the file size and format of the dashboard files3

NEW QUESTION 154
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 157
A company??s marketing department wants to do a promotional campaign next month. A data analyst on the team has been asked to perform customer segmentation, looking at how recently a customer bought the product, at what frequency, and at what value. Which of the following types of analysis would this practice be considered?

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

Answer: D

Explanation:
Customer segmentation is a type of cluster analysis, which is a method of grouping data points based on their similarities or differences. Cluster analysis can help identify patterns and trends in the data, as well as target specific groups of customers for marketing purposes. One common technique for customer segmentation is RFM analysis, which stands for recency, frequency, and monetary value. This technique assigns a score to each customer based on how recently they bought the product, how often they buy the product, and how much they spend on the product. These scores can then be used to create clusters of customers with different characteristics and preferences. Therefore, the correct answer is D. References: Cluster Analysis - Statistics Solutions, RFM Analysis: The Ultimate Guide for Customer Segmentation

NEW QUESTION 160
An analyst wants to extract data from a variety of sources and store the data in a cloud- based environment prior to cleaning. Which of the following integration techniques should the analyst use?

A. ETL
B. API
C. SQL
D. ELT

Answer: A

NEW QUESTION 161
A data analyst needs to write a SOL query measuring last month's website visits and distribute a summary report to the marketing team. Which of the following is the analyst creating?

A. Date range
B. Distribution list
C. Data content
D. Report view

Answer: D

NEW QUESTION 163
Exhibit.

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

Which of the following logical statements results in Table B?
A)

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

B)

```
IF Name = "Sam" and Gender_flag = "Male" then delete
```

C)

```
IF Name = "Pat" and Gender_flag = "Female" then delete
```

D)

```
IF Name - "Sean" and Gender_flag - "College" then delete
```

A. Option A
B. Option B
C. Option C
D. Option D

Answer: D

Explanation:
The logical statement that results in Table B is Option D. Option D is a logical statement that uses the AND operator to combine two conditions: Name = ??Tom?? and Region = ??BC??. The AND operator returns true only if both conditions are true, otherwise it returns false. Therefore, Option D will select only the rows from Table A that satisfy both conditions, which are rows 4, 5, 6, and 7. These rows form Table B, as shown below: Name | Gender flag | Level | College | Code | Region Tom | Male | Elementary | A | BC | BC Kim | Female | Elementary | A | BC | BC Pat | Female | Elementary | A | BC | BC Ben | Male | Elementary | A | BC | BC
The other options are not correct, as they use different logical operators or conditions that do not result in Table B. Option A uses the OR operator, which returns

true if either condition is true, or both. Option A will select all the rows from Table A except row 3, which does not match either condition. Option B uses the NOT operator, which returns the opposite of the condition. Option B will select all the rows from Table A except rows 4, 5, 6, and 7, which match the condition. Option C uses a different condition, Region = ??ON??, which does not match any row in Table A. Option C will select no rows from Table A. Reference: [SQL Logical Operators - W3Schools]

NEW QUESTION 164
A data engineer is creating a database field to capture whether a customer likes vanilla ice cream. Which of the following data types is the best to capture this information?

A. Integer
B. Boolean
C. Categorical
D. Numeric

Answer: B

NEW QUESTION 165
Which one of the following values will appear first if they are sorted in descending order?

A. Aaron.
B. Molly.
C. Xavier.
D. Adam.

Answer: C

Explanation:
The value that will appear first if they are sorted in descending order is Xavier. Descending order means arranging values from the largest to the smallest, or from the last to the first in alphabetical order. In this case, Xavier is the last name in alphabetical order, so it will appear first when sorted in descending order. The other names will appear in the following order: Molly, Adam, Aaron. Reference: Sorting Data - W3Schools

NEW QUESTION 168
The duration of a phone call in milliseconds is an example of:

A. ordinal data.
B. nominal data.
C. boolean data.
D. continuous data.

Answer: D

Explanation:
The correct answer is D. Continuous data.
Continuous data is a type of quantitative data that can take any value within a range and can be measured with infinite precision. Continuous data can be expressed as fractions, decimals, or percentages. Examples of continuous data are height, weight, temperature, time, speed, etc12
The duration of a phone call in milliseconds is an example of continuous data, because it can take any value within a range (from zero to infinity) and can be measured with infinite precision (up to milliseconds or even smaller units). The duration of a phone call in milliseconds can also be expressed as fractions, decimals, or percentages of a larger unit (such as seconds, minutes, or hours).
Ordinal data is not correct, because ordinal data is a type of qualitative or categorical data that can be ordered or ranked according to some criterion. Ordinal data can have a logical order, but the intervals between the values are not equal or meaningful. Examples of ordinal data are grades, ratings, ranks, etc12
Nominal data is not correct, because nominal data is a type of qualitative or categorical data that can be labeled or named without any order or ranking. Nominal data can have a finite number of categories or classes, but the categories have no intrinsic value or hierarchy. Examples of nominal data are gender, color, nationality, etc12
Boolean data is not correct, because boolean data is a type of binary data that can have only two possible values: true or false. Boolean data can be used to represent logical statements, conditions, or outcomes. Examples of boolean data are yes/no, on/off, 1/0, etc.

NEW QUESTION 172
Which of the following best describes a difference between JSON and XML?

A. JSON is quicker to read and write.
B. JSON has to use an end tag.
C. JSON strings are longer
D. JSON is much more difficult to parse.

Answer: A

Explanation:
The best answer is A. JSON is quicker to read and write.
JSON (JavaScript Object Notation) is a lightweight data-interchange format that is based on the JavaScript programming language and easy to understand and generate. JSON uses a simple syntax that consists of name-value pairs and arrays, and does not require any end tags or attributes. JSON is quicker to read and write than XML (Extensible Markup Language), which is a markup language that uses a tag structure to represent data items. XML has a more complex and verbose syntax that requires end tags, attributes, and namespaces123

NEW QUESTION 175
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 177
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 178
Which of the following data protection methods provides confidentiality for data in transit?

A. De-identification
B. Encryption
C. Masking
D. Anonymization

Answer: B

NEW QUESTION 179
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 documents12

NEW QUESTION 181
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 185
Which of the following are reasons to create and maintain a data dictionary? (Choose two.)

A. To improve data acquisition

B. To remember specifics about data fields
C. To specify user groups for databases
D. To provide continuity through personnel turnover
E. To confine breaches of PHI data
F. To reduce processing power requirements

Answer: BD

Explanation:
A data dictionary is a collection of metadata that describes the data elements in a database or dataset. It can help improve data acquisition by providing information about the data sources, formats, quality, and usage. It can also help remember specifics about data fields, such as their names, definitions, types, sizes, and relationships. Therefore, options B and D are correct.
Option A is incorrect because it is not a reason to create and maintain a data dictionary, but a benefit of doing so.
Option C is incorrect because specifying user groups for databases is not a function of a data dictionary, but a function of a database management system or a security policy.
Option E is incorrect because confining breaches of PHI data is not a function of a data dictionary, but a function of a data protection or encryption system.
Option F is incorrect because reducing processing power requirements is not a function of a data dictionary, but a function of a data compression or optimization system.

NEW QUESTION 186
Which one of the following would not normally be considered a summary statistic?

A. z-score.
B. Mean.
C. Variance.
D. Standard deviation.

Answer: A

Explanation:

Simply put, a z-score (also called a standard score) gives you an idea of how far from the mean a data point is. But more technically it's a measure of how many standard deviations below or above the population mean a raw score is. A z-score can be placed on a normal distribution curve.

NEW QUESTION 189
Given the following data table:

CandidateID	Status	Date	HireDate
01	Hired	05-23-87	05-23-87
02	Hired	11-30-96	11-30-96
03	Hired	13-05-99	13-05-99

Which of the following are appropriate reasons to undertake data cleansing? (Select two).

A. Non-parametric data
B. Missing data
C. Duplicate data
D. Invalid data
E. Redundant data
F. Normalized data

Answer: BD

Explanation:
Data cleansing is a critical process in data analytics to ensure the accuracy and quality of data. The reasons to undertake data cleansing include:
? Missing Data (B): Missing data can lead to incomplete analysis and biased
results. It is essential to identify and address gaps in the dataset to maintain the integrity of the analysis1.
? Invalid Data (D): Invalid data includes entries that are out of range, improperly
formatted, or illogical (e.g., a negative age). Such data can corrupt analysis and
lead to incorrect conclusions1.
Other options, such as non-parametric data (A), are not inherently errors but refer to a type of data that doesn??t assume a normal distribution. Duplicate data © and redundant data (E) could also be reasons for data cleansing, but they are not listed as options to select from in the provided image details. Normalized data (F) refers to data that has been processed to fit into a certain range or format and is typically not a reason for data cleansing. References:
? Understanding the importance of data quality and the impacts of missing and
invalid data on research outcomes1.
? Best practices in data cleansing2.
Data cleansing is required for various reasons, two of which are missing data (B) and invalid data (D). From the table provided, we can infer the necessity of cleansing in the context of ensuring data integrity and consistency. Missing data refers to the absence of data where it is expected, which can hinder analysis due to incomplete information. Invalid data refers to data that is incorrect, out of range, or in an inappropriate format, which can lead to inaccuracies in any analysis or report. Both these issues can significantly affect the outcomes of any data-related operations and thus need to be rectified through the data cleansing process.

NEW QUESTION 192

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 195
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 200
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 204
Which one of the following is a measure of dispersion?

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

Answer: A

NEW QUESTION 207
You are working with a dataset and want to change the names of categories that you used for different types of books.
What term best describes this action?

A. Recording.
B. Summarizing
C. Aggregating.
D. Filtering.

Answer: A

Explanation:
The term that best describes the action of changing the names of categories that you used for different types of books is recoding. Recoding is a process of transforming or modifying the values of a variable or a category to make them more meaningful, consistent, or accurate. For example, you can recode the names of book genres from ??Fiction??, ??Non-Fiction??, ??Biography??, etc. to ??FIC??, ??NF??, ??BIO??, etc. to make them shorter and easier to use. Reference: Recoding Data - SPSS Tutorials - LibGuides at
Kent State University

NEW QUESTION 208
Which of the following best describes an exploratory analysis?

A. Involves the use of descriptive statistics to understand observations
B. Involves analysis of exploring data sets for performance tracking
C. Involves the testing of specific hypotheses
D. Involves the use of arithmetic algebra to determine the distribution

Answer: A

Explanation:

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

NEW QUESTION 212
Given the following report:

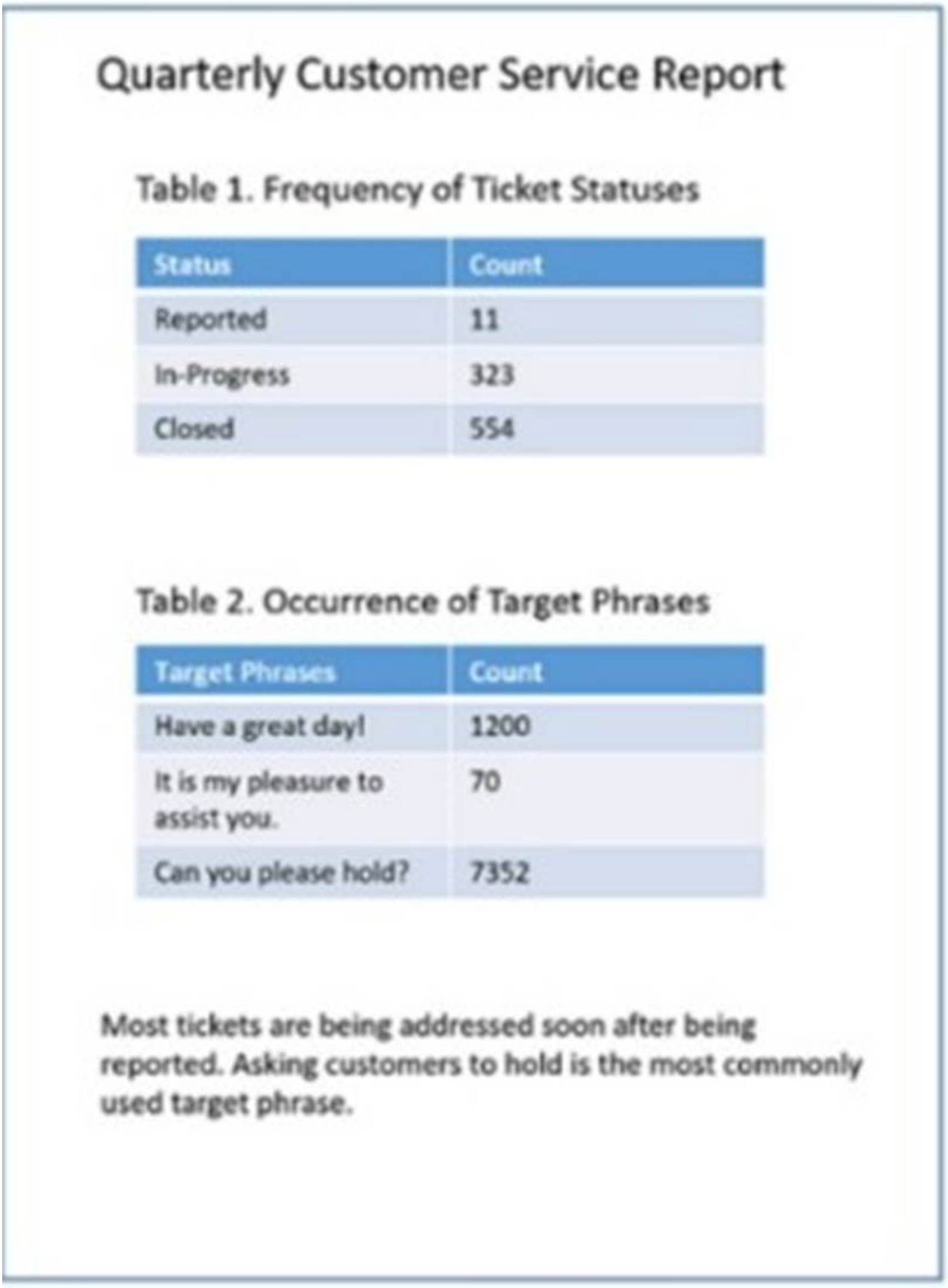

Quarterly Customer Service Report

Table 1. Frequency of Ticket Statuses

Status	Count
Reported	11
In-Progress	323
Closed	554

Table 2. Occurrence of Target Phrases

Target Phrases	Count
Have a great day!	1200
It is my pleasure to assist you.	70
Can you please hold?	7352

Most tickets are being addressed soon after being reported. Asking customers to hold is the most commonly used target phrase.

Which of the following components need to be added to ensure the report is point-in-time and static? (Select two).

A. A control group for the phrases
B. A summary of the KPIs
C. Filter buttons for the status

D. The date when the report was last accessed
E. The time period lhe report covers
F. The date on which the report was run

Answer: DF

Explanation:
To ensure that a report is point-in-time and static, it should include the date when the report was last accessed and the date on which the report was run. These components confirm the specific time frame the data represents, making the report a fixed reference that does not change with subsequent data updates or accesses. This is crucial for accurate historical analysis and for maintaining the integrity of the data as it was at the time of the report??s creation.
References:
? Best practices in business reporting.
? Importance of time-stamping in data analysis.
? Guidelines for creating static reports in data analytics.

NEW QUESTION 214
Which of the following BEST describes the issue in which character values are mixed with integer values in a data set column?

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

Answer: D

Explanation:
The invalid data type is the best description for the issue in which character values are mixed with integer values in a data set column. Invalid data type means that the data does not match the expected or required format or structure for a given variable or attribute. For example, if a column is supposed to store numerical values, but some rows contain text values, then those rows have an invalid data type. References: CompTIA Data+ Certification Exam Objectives, page 10

NEW QUESTION 216
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	(3OO)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 217
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 = (x -) /
where x is the raw score, is the mean, and 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 219
A database administrator needs to ensure only approved users can access specific database tables to perform financial functions. Which of the following is the best access control method for the administrator to use?

A. Role-based
B. Rule-based
C. Discretionary
D. Group-based

Answer: A

NEW QUESTION 222
A data analyst needs to present the results of an online marketing campaign to the marketing manager. The manager wants to see the most important KPIs and measure the return on marketing investment. Which of the following should the data analyst use to BEST communicate this information to the manager?

A. A real-time monitor that allows the manager to view performance the day the campaign was launched
B. A sell-service dashboard that allows the manager to look at the company??s annual budget performance
C. A spreadsheet of the raw data from all marketing campaigns and channels
D. A summary with statistics, conclusions, and recommendations from the data analyst

Answer: D

Explanation:
A summary with statistics, conclusions, and recommendations from the data analyst is the best way to communicate the results of an online marketing campaign to the marketing manager. A summary can provide a concise and clear overview of the most important KPIs and measure the return on marketing investment, as well as highlight the main findings and insights from the data analysis. A summary can also include actionable suggestions and best practices for improving the campaign performance and achieving the marketing objectives. A summary is different from other options, such as a real-time monitor, a self-service dashboard, or a spreadsheet of raw data, which may not provide enough context, interpretation, or guidance for the manager. Therefore, the correct answer is D. References: How to Write a Data Analysis Report: 6 Essential Tips, How to Write a Marketing Report (with Pictures) - wikiHow

NEW QUESTION 224
A data analyst has been asked to create an ad-hoc sales report for the Chief Executive Officer (CEO).
Which of the following should be included in the report?

A. The sales representatives' home addresses.
B. Line-item SKU numbers.
C. YTD total sales.
D. The customers' first and last names.

Answer: C

Explanation:
The report for the CEO should include YTD total sales, as this will provide a high-level overview of the sales performance of the company and show how it is meeting its annual goals. The other options are not appropriate for the CEO, as they are either too
detailed or irrelevant for the report. The sales representatives?? home addresses, line-item SKU numbers, and customers?? first and last names are not related to the sales performance and might compromise the privacy and security of the data.
Reference: CompTIA Data+ (DA0-001) Practice Certification Exams | Udemy

NEW QUESTION 227
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 228
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 Charts1
? Choosing the Right Chart for Your Data2
? Sales Dashboards: Examples & Templates3

NEW QUESTION 232
An analyst is updating a customer contacts database with information obtained from a survey of new customers. Which of the following data manipulation techniques should the analyst use?

A. Join
B. Append
C. Transform
D. Blend

Answer: B

NEW QUESTION 235
Joe. an analyst. tests the loading time on a dashboard he is preparing to go live and finds it is slower than he would like. Which of the following must occur to decrease the loading time?

A. Deploy the dashboard to production.
B. Change the field definitions.
C. Update the dashboard subscribers.
D. Optimize the dashboard.

Answer: D

Explanation:
Optimizing the dashboard is the process of improving its performance and reducing its loading time by applying various techniques and best practices. Some of the common ways to optimize a dashboard are:
? Reducing the size and complexity of the data model, such as removing
unnecessary columns, aggregating data at the source, or using data compression techniques12
? Leveraging caching strategies, such as setting appropriate cache refresh intervals
or utilizing Power BI??s built-in caching mechanisms, to minimize data retrieval delays2
? Utilizing query folding, direct query, or live connection to enhance data processing
efficiency and enable real-time data updates23
? Optimizing DAX queries, such as avoiding nested calculations, using variables, or simplifying measures, to improve data calculation speed23
? Reducing visualizations and calculations, such as using fewer or simpler charts, filters, or parameters, to speed up dashboard rendering12
? Evaluating the impact of custom visuals on dashboard load time and avoiding or replacing those that are slow or inefficient2
? Applying aggregation and summarization techniques, such as using extract filters, context filters, or level of detail expressions, to reduce the amount of data displayed on the dashboard1

? Troubleshooting and resolving any issues that may cause slow dashboard load, such as network latency, server overload, or hardware limitations24

NEW QUESTION 237
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 JOI
C. four rows
D. RIGHT JOI
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 table1
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 241
An analyst is designing a dashboard to determine which site has the highest percentage of new customers. The analyst must choose an appropriate chart to include in the dashboard. The following data is available:

Site	Customers	New customers	Percentage of new customers
A1	2236	277	12%
A2	885	300	34%
A3	333	200	60%
B1	483	167	35%
B2	2969	235	8%
B3	2357	153	6%
C1	1524	180	12%
C2	878	150	17%
C3	1925	142	7%

Which of the following types of charts should be considered to best display the data?

A. Include a bar chart using the site and the percentage of new customers data.
B. Include a line chart using the site and the percentage of new customers data.
C. Include a pie chart using the site and percentage of new custorners data.
D. Include a scatter chart using the site and the percent of new customers data.

Answer: A

Explanation:
The best type of chart to display the data is A. Include a bar chart using the site and the percentage of new customers data.
A bar chart is a good choice for comparing categorical data with numerical data, such as the site and the percentage of new customers. A bar chart can show the relative differences between the sites and highlight the site with the highest percentage of new customers. A bar chart can also be easily labeled and formatted to make the data clear and understandable.
A line chart is not suitable for this data, because it is used to show trends or changes over time, which is not relevant for the site and the percentage of new customers data. A line chart would also be confusing and misleading, as it would imply a connection or correlation between the sites that does not exist.
A pie chart is also not a good choice for this data, because it is used to show the proportion of a whole, not the comparison of different categories. A pie chart would also be difficult to read and interpret, as it would require labels or legends to identify the sites and their percentages. A pie chart would also not be able to show the exact values of the percentages, only their relative sizes.
A scatter chart is another inappropriate option for this data, because it is used to show the relationship or correlation between two numerical variables, not between a categorical and a numerical variable. A scatter chart would also be cluttered and unclear, as it would plot each site as a point on a coordinate plane, without any labels or axes. A scatter chart would also not be able to show the differences or rankings between the sites and their percentages.

NEW QUESTION 242
Which of the following actions should be taken when transmitting data to mitigate the chance of a data leak occurring? (Choose two.)

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

Answer: DE

Explanation:
Data encryption 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 243
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+ (DA0-001) Practice Certification Exams | Udemy

NEW QUESTION 246
A data analyst was asked to create a chart that shows the relationship between study hours and exam scores for each student using the data sets in the table below:

Student	Exam score	Study hours
Kim	90	7.5
Leo	80	6
Alpha	60	4
Jude	85	7
Ella	95	8

Which of the following charts would BEST represent the relationship between the variables?

A. A histogram
B. A scatter plot
C. A heat map
D. A bar chart

Answer: B

Explanation:
This is because a scatter plot is a type of chart that shows the relationship between two variables for each observation or unit in a data set, such as study hours and exam scores for each student in this case. A scatter plot can be used to display and analyze the correlation, trend, or pattern among the variables, as well as identify any outliers or clusters in the data. For example, a scatter plot can show if there is a positive, negative, or no correlation between study hours and exam scores, as well as show if there are any students who have unusually high or low exam scores compared to their study hours. The other charts are not the best charts to represent the relationship between the variables. Here is why:
? A histogram is a type of chart that shows the frequency or the count of values in a single variable for different intervals or bins, such as exam scores for different ranges in this case. A histogram can be used to display and analyze the distribution, shape, or spread of the variable, as well as identify any gaps, peaks, or skewness in the data. For example, a histogram can show if most students have high, low, or average exam scores, as well as show if there are any intervals that have no students at all.
? A heat map is a type of chart that shows the intensity or the magnitude of values in two variables for different categories or groups, such as exam scores and study hours for different student names in this case. A heat map can be used to display and analyze the variation, contrast, or comparison among the categories or groups, as well as identify any hot spots, cold spots, or gradients in the data. For example, a heat map can show which students have higher or lower exam scores and study hours than others, as well as show if there is a color pattern that indicates a relationship between exam scores and study hours.
? A bar chart is a type of chart that shows the value or the amount of a single variable for different categories or groups, such as exam scores for different student names in this case. A bar chart can be used to display and analyze the comparison, ranking, or proportion among the categories or groups, as well as identify any differences, similarities, or outliers in the data. For example, a bar chart can show which students have higher or lower exam scores than others, as well as show if there are any students who have exceptionally high or low exam scores.

NEW QUESTION 251
A sales analyst needs to report how the sales team is performing to target. Which of the following files will be important in determining 2019 performance attainment?

A. 2018 goal data
B. 2018 actual revenue
C. 2019 goal data
D. 2019 commission plan

Answer: C

Explanation:
Answer: C. 2019 goal data
To report how the sales team is performing to target, the sales analyst needs to compare the actual sales revenue with the expected or planned sales revenue for the same period. The 2019 goal data is the file that contains the expected or planned sales revenue for the year 2019, which is the target that the sales team is aiming to achieve. By comparing the 2019 goal data with the 2019 actual revenue, the sales analyst can calculate the performance attainment, which is the percentage of the goal that was met by the sales team.
Option A is incorrect, as 2018 goal data is not relevant for determining 2019 performance attainment. The 2018 goal data contains the expected or planned sales revenue for the year 2018, which is not the target that the sales team is aiming to achieve in 2019.
Option B is incorrect, as 2018 actual revenue is not relevant for determining 2019 performance attainment. The 2018 actual revenue contains the actual sales revenue for the year 2018, which is not comparable with the 2019 goal data or the 2019 actual revenue. Option D is incorrect, as 2019 commission plan is not relevant for determining 2019 performance attainment. The 2019 commission plan contains the rules and rates for calculating and paying commissions to the sales team based on their performance attainment, but it does not contain the expected or planned sales revenue for the year 2019.

NEW QUESTION 254
Which of the following data sampling methods involves dividing a population into subgroups by similar characteristics?

A. Systematic
B. Simple random
C. Convenience
D. Stratified

Answer: D

Explanation:
Stratified sampling is a data sampling method that involves dividing a population into subgroups by similar characteristics, such as age, gender, income, etc. Then, a simple random sample is drawn from each subgroup. This method ensures that each subgroup is adequately represented in the sample and reduces the sampling error. References: CompTIA Data+ Certification Exam Objectives, page 11.

NEW QUESTION 257
An analyst notices changes in sales ratios when analyzing a quarterly report. Which of the following is the analyst conducting?

A. A gap analysis
B. A link analysis
C. A trend analysis
D. A statistical analysis

Answer: C

Explanation:
When an analyst observes changes in sales ratios over a period, such as in a quarterly report, they are conducting a trend analysis. Trend analysis is a statistical method used to examine and evaluate the movement of data points over time to identify patterns or trends. This type of analysis is particularly useful for forecasting future events based on historical data. It differs from gap analysis, which assesses the difference between actual performance and potential or desired performance; link analysis, which is used to find associations among data; and statistical analysis, which is a broad term for all types of data analysis methods, including trend analysis.
References:
? Investopedia article on Ratio Analysis1.
? SpringerLink chapter on Financial Ratios Analysis2.
? ExamTopics page mentioning sales ratios in the context of analysis3.
? Investopedia definition of Ratio Analysis4.
? LiveWell article on Financial Ratio Analysis5.

NEW QUESTION 260
A data analyst needs to create a data visualization that aids in un the cumulative impact of sequentially introduced values that are positive or negative. Which of the following data visualization methods should the analyst use?

A. A bubble chart
B. A waterfall chart
C. A scatter plot
D. A line chart

Answer: B

Explanation:
A waterfall chart is a type of data visualization that shows the cumulative impact of sequentially introduced values that are positive or negative. A waterfall chart typically has an initial value and a final value, with intermediate values shown as floating columns that either add to or subtract from the initial value. A waterfall chart can help visualize how different factors contribute to a net change in a value over time. Therefore, the correct answer is B. References: [Waterfall Chart | Definition & Examples - Investopedia], [Waterfall Charts in Excel | How to Create Waterfall Chart in Excel?] 4of30

NEW QUESTION 262
A data analyst is developing a dashboard to track and monitor metrics. Which of the following best practices should be taken into during the FIRST pment process?

A. Create a A Aupirarrame:
B. Deploy to production.
C. Copy a dashboard design from the Internet.
D. Develop a dashboard.

Answer: A

Explanation:
A dashboard is a graphical display that summarizes and presents key performance indicators (KPIs) and metrics for a business or a project. A dashboard should be clear, concise, and easy to understand. To develop a dashboard, one of the best practices is to create a wireframe or a mockup first. A wireframe or a mockup is a low- fidelity sketch or prototype of the dashboard layout and design, which helps to define the scope, requirements, and functionality of the dashboard. Creating a wireframe or a mockup can help to save time and resources, as well as to get feedback from stakeholders and users before deploying the dashboard to production. Therefore, the correct answer is A. References: [Dashboard Design Best Practices: 4 Key Principles | Toptal], [How to Create an Effective Dashboard (with Examples) | Tableau]

NEW QUESTION 263
Which of the following is a common data analytics tool that is also used as an interpreted, high-level, general-purpose programming language?

A. SAS
B. Microsoft Power B1
C. IBM SPSS
D. Python

Answer: D

Explanation:
The option that is a common data analytics tool that is also used as an interpreted, high-level, general-purpose programming language is Python. Python is a

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

NEW QUESTION 268
A financial analyst is creating a daily billing report for a company. One night, the company's data warehouse did not update the data, which caused the data to be reported incorrectly the next day. Which of the following documentation elements should the analyst add to catch this error?

A. Version number
B. Data refresh
C. Frequently asked questions tab
D. Summary

Answer: B

Explanation:
A data refresh is a documentation element that indicates when the data was last updated or refreshed from the source. A data refresh can help the analyst to catch the error of the data warehouse not updating the data, as it will show a discrepancy between the expected and actual date of the data update. A data refresh can also help the users of the report to verify the timeliness and accuracy of the data, and to avoid making decisions based on outdated or incorrect data

NEW QUESTION 270
Which of the following is the best technique for transferring data from one database to another with some data manipulation?

A. Application programming interfaces
B. Delta load
C. Extract, transform, load
D. Export/import

Answer: C

NEW QUESTION 271
An analyst is designing a dashboard to determine which site has the highest percentage of new customers. The analyst must choose an appropriate chart to include in the dashboard. The following data is available:

Site	Customers	New customers	Percentage of new customers
A1	2236	277	12%
A2	885	300	34%
A3	333	200	60%
B1	483	167	35%
B2	2969	235	8%
B3	2357	153	6%
C1	1524	180	12%
C2	878	150	17%
C3	1925	142	7%

Which of the following types of charts should be considered to BEST display the data?

A. Include a bar chart using the site and the percentage of new customers data.
B. Include a line chart using the site and the percentage of new customers data.
C. Include a pie chat using the site and percentage of new customers data.
D. Include a scatter chart using the site and the percent of new customers data.

Answer: A

Explanation:
This is because a bar chart is a type of chart that shows the value or the amount of a single variable for different categories or groups, such as the percentage of new customers for different sites in this case. A bar chart can be used to display and analyze the comparison, ranking, or proportion among the categories or groups, as well as identify any differences, similarities, or outliers in the data. For example, a bar chart can show which site has the highest or lowest percentage of new customers, as well as show how much each site contributes to the total percentage of new customers. The other types of charts are not the best charts to

display the data. Here is why:
? A line chart is a type of chart that shows the change or the trend of a single variable over time, such as the percentage of new customers over months or years in this case. A line chart can be used to display and analyze the movement, cycle, or pattern of the variable, as well as identify any peaks, valleys, or fluctuations in the data. For example, a line chart can show how the percentage of new customers increases or decreases over time, as well as show if there are any seasonal or periodic variations in the data.
? A pie chart is a type of chart that shows the proportion or the percentage of a single variable for different categories or groups, such as the percentage of new customers for different sites in this case. A pie chart can be used to display and analyze the composition, distribution, or share of the variable, as well as identify any segments, slices, or fractions in the data. For example, a pie chart can show how much each site represents of the total percentage of new customers, as well as show if there are any dominant or minor sites in the data.
? A scatter chart is a type of chart that shows the relationship between two variables for each observation or unit in a data set, such as the percentage of new customers and another variable for each site in this case. A scatter chart can be used to display and analyze the correlation, trend, or pattern among the variables, as well as identify any outliers or clusters in the data. For example, a scatter chart can show if there is a positive, negative, or no correlation between the percentage of new customers and another variable, such as sales revenue or customer satisfaction.

NEW QUESTION 272
.....

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual DA0-001 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the DA0-001 Product From:

https://www.2passeasy.com/dumps/DA0-001/

Money Back Guarantee

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