CompTIA

Exam Questions DA0-001

CompTIA Data+ Certification Exam

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
An analyst wants to check the progress and performance regarding the number of customers an organization served in the last six years. Which of the following represents the type of analysis the analyst should perform?

A. Correlation analysis
B. Trend analysis
C. Regression analysis
D. Descriptive analysis

Answer: B

NEW QUESTION 3
A financial institution is reporting on sales performance to a company at the account level. Due to the sensitive nature of the government the does il with, some account information is not shown. Which of the following fields should be masked?

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

Answer: D

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

NEW QUESTION 4
Five dogs have the following heights in millimeters: 300, 430, 170, 470, 600
Which of the following is the mean height for the five dogs?

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

Answer: A

Explanation:
The mean height for the five dogs is calculated by adding up all the heights and dividing by the number of dogs. The formula is:
mean = (300 + 430 + 170 + 470 + 600) / 5 mean = 1970 / 5 mean = 394
Therefore, option A is correct.
Option B is incorrect because it is the median height, which is the middle value when the heights are arranged in ascending order.
Option C is incorrect because it is the mean height multiplied by 1.25.
Option D is incorrect because it is the mean height multiplied by 1.28.

NEW QUESTION 5
A user receives a large custom report to track company sales across various date ranges. The user then completes a series of manual calculations for each date range. Which of the following should an analyst suggest so the user has a dynamic, seamless experience?

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

Answer: D

Explanation:
Create a dashboard with a date range picker and calculations built in. This is because a dashboard is a type of visualization that displays multiple charts or graphs

on a single page, usually to provide an overview or summary of some data or information. A dashboard can be used to track company sales across various date ranges by showing different metrics and indicators related to sales, such as revenue, volume, or growth. By creating a dashboard with a date range picker and calculations built in, the analyst can suggest a way for the user to have a dynamic, seamless experience, which means that the user can interact with and customize the dashboard according to their needs or preferences, as well as avoid any manual work or errors. For example, a date range picker is a type of feature or function that allows users to select or adjust the time period for which they want to see the data on the dashboard, such as daily, weekly, monthly, or quarterly. A date range picker can make the dashboard dynamic, as it can automatically update or refresh the dashboard with new data based on the selected time period. Calculations are mathematical operations or expressions that can be performed on the data on the dashboard, such as addition, subtraction, multiplication, division, average, sum, etc. Calculations can make the dashboard seamless, as they can eliminate the need for manual calculations for each date range, as well as ensure accuracy and consistency of the results. The other ways are not the best ways to provide a dynamic, seamless experience for the user. Here is why:
? Creating multiple reports, one for each needed date range would not provide a dynamic, seamless experience for the user, but rather create a static, cumbersome experience, which means that the user cannot interact with or customize the reports according to their needs or preferences, as well as have to deal with multiple files or pages. For example, creating multiple reports would make it difficult for the user to compare or contrast the sales across different date ranges, as well as increase the workload and complexity of managing and maintaining the reports.
? Building calculations into the report so they are done automatically would not
provide a dynamic, seamless experience for the user, but rather provide a partial, limited experience, which means that the user can only benefit from one aspect or feature of the report, but not from others. For example, building calculations into the report would help with avoiding manual work or errors, but it would not help with interacting with or customizing the report according to different date ranges.
? Adding macros to the report to speed up the filtering and calculations process
would not provide a dynamic, seamless experience for the user, but rather provide an advanced, complex experience, which means that the user would need to have some technical skills or knowledge to use or apply the macros, as well as face some potential risks or challenges. For example, adding macros to the report would require the user to know how to write or run the macros, which are a type of code or script that automates certain tasks or actions on the report, such as filtering or calculating the data. Adding macros to the report could also expose the user to some security or compatibility issues, such as viruses, malware, or errors.

NEW QUESTION 6
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 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
Jenny wants to study the academic performance of undergraduate sophomores and wants to determine the average grade point average at different points during an academic year.
What best describes the data set she needs?

A. Sample.
B. Observation.
C. Variable.
D. Population.

Answer: A

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

NEW QUESTION 9
A recurring event is being stored in two databases that are housed in different geographical locations. A data analyst notices the event is being logged three hours earlier in one database than in the other database. Which of the following is the MOST likely cause of the issue?

A. The data analyst is not querying the databases correctly.
B. The databases are recording different events.
C. The databases are recording the event in different time zones.
D. The second database is logging incorrectly.

Answer: C

Explanation:
The most likely cause of the issue is that the databases are recording the event in different time zones. A time zone is a region that observes a uniform standard time for legal, commercial, and social purposes. Different time zones have different offsets from Coordinated Universal Time (UTC), which is the primary time standard by which the world regulates clocks and time. For example, UTC-5 is five hours behind UTC, while UTC+3 is three hours ahead of UTC. If an event is being stored in two databases that are housed in different geographical locations with different time zones, it may appear that the event is being logged at different times, depending on how the databases handle the time zone conversion. For example, if one database records the event in UTC-5 and another database records the event in UTC+3, then an event that occurs at 12:00 PM in UTC-5 will appear as 9:00 AM in UTC+3. The other options are not likely causes of the issue, as they are either unrelated or implausible. The data analyst is not querying the databases incorrectly, as this would not affect the time stamps of the events. The databases are not
recording different events, as they are supposed to record the same recurring event. The second database is not logging incorrectly, as there is no evidence or reason to assume that. Reference: [Time zone - Wikipedia]

NEW QUESTION 10
Which of the following is the correct data type for text?

A. Boolean
B. String
C. Integer
D. Float

Answer: B

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

NEW QUESTION 10
Given the following data:

Name	Gender	Age	Annual income
Ralph	M	27	$75,000
Jessie	F	3	$75,000
Monica	F	31	$125,000
Carlos	M	53	$75
Sara	F	43	$0

Which of the following BEST describes the data set?

A. There is data bias.
B. The data is incomplete.
C. The data is inconsistent.
D. The data is outliers.

Answer: C

Explanation:
This is because inconsistency is a type of data quality issue that occurs when the data does not follow a common format, structure, or rule across different sources or systems, which can affect the efficiency and performance of the analysis or process. Inconsistency can be caused by having different spellings, punctuations, capitalizations, or abbreviations for the same or similar values in a data set, such as ??M??, ??m??, ??Male??, or ??male?? for gender in this case. Inconsistency can be eliminated or reduced by using data cleansing techniques, such as standardizing or normalizing the data values. The other options are not correct descriptions of the data set. Here is why:
? Data bias is a type of data quality issue that occurs when the data is not representative or proportional of the population or the parameter, which can affect the validity and reliability of the analysis or process. Data bias can be caused by having a sample that is too small, too large, or too skewed for the population or the parameter, such as having only male customers for a product that targets both genders in this case. Data bias can be eliminated or reduced by using sampling techniques, such as stratified or cluster sampling.
? The data is incomplete is a type of data quality issue that occurs when the data is
absent or missing in a data set, which can affect the accuracy and reliability of the analysis or process. The data is incomplete can be caused by various factors, such as human error, system error, or non-response. The data is incomplete can be addressed by using various methods, such as replacing or imputing the missing values with some reasonable estimates, such as mean, median, mode, or regression.
? The data is outliers is a type of data quality issue that occurs when the data has
values that are unusually high or low compared to the rest of the data set, which can affect the quality and validity of the analysis or process. The data is outliers can be caused by various factors, such as measurement error, natural variation, or extreme events. The data is outliers can be addressed by using various methods, such as removing or filtering out the outliers, or using robust statistics that are less sensitive to outliers, such as median, interquartile range, or box plot.

NEW QUESTION 11
A data analyst for a media company needs to determine the most popular movie genre. Given the table below:

MovieID	Name	Genre	Actors	Rating
01	Ghost Writer	Comedy, Actions	Joshua Wellington, Susana Summons	6.5
02	Life of Suffering	Drama, Foreign, Historical	Shelly May, Rita Moralle, Ethan Warner, Sean Houser	7.2

Which of the following must be done to the Genre column before this task can be completed?

A. Append
B. Merge
C. Concatenate
D. Delimit

Answer: D

Explanation:
The action that must be done to the Genre column before this task can be completed is delimit. Delimit is a process of separating or splitting a string of text into multiple parts based on a delimiter, which is a character or a sequence of characters that marks the boundary between the parts. For example, a comma (,) or a semicolon (;) can be used as a delimiter. In this case, the Genre column contains multiple genres for each movie, separated by commas. To determine the most popular movie genre, the data analyst needs to delimit the Genre column by commas, so that each genre can be counted and compared separately. The other options are not relevant for this task, as they are related to combining or joining strings or tables, not separating them. Append is a process of adding or attaching one string or table to the end of another string or table. Merge is a process of combining or joining two or more tables into one table based on a common column or key. Concatenate is a process of joining or linking two or more strings together into one string. Reference: [How to Split Text in Excel - Exceljet]

NEW QUESTION 13
Which of the following will MOST likely be streamed live?

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

Answer: A

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

NEW QUESTION 14
A data analyst has a set with more than 40.000 rows in the sample schema below:

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

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

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

Answer: D

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

NEW QUESTION 18
A data analyst must separate the column shown below into multiple columns for each component of the name:

Customer_name
Alphonso,Jamie, R.
Benedict,Alice, M.
Smith, Diana, L.

Which of the following data manipulation techniques should the analyst perform?

A. Imputing
B. Transposing
C. Parsing
D. Concatenating

Answer: C

Explanation:
Parsing is the data manipulation technique that should be used to separate the column into multiple columns for each component of the name. Parsing is the process of breaking down a string of text into smaller units, such as words, symbols, or numbers. Parsing can be used to extract specific information from a text column, such as names, addresses, phone numbers, etc. Parsing can also be used to split a text column into multiple columns based on a delimiter, such as a comma, space, or dash1. In this case, the
analyst can use parsing to split the column by the comma delimiter and create three new columns: one for the last name, one for the first name, and one for the middle initial. This will make the data more organized and easier to analyze.

NEW QUESTION 21
Which of the following differentiates a flat text file from other data types?

A. Data is separated by a delimiter.
B. Data is stored in defined rows.
C. Data is defined with key-value pairs.
D. Data is housed in a markup language.

Answer: A

Explanation:
A flat text file is a type of data file that contains only plain text without any formatting or markup. Data in a flat text file is usually separated by a delimiter, which is a character that marks the boundary between different fields or values. For example, a comma-separated values (CSV) file is a flat text file that uses commas as delimiters. Other common delimiters are tabs, spaces, semicolons, and pipes. Therefore, the correct answer is A. References: Plain text - Wikipedia, Comparison of document markup languages - Wikipedia

NEW QUESTION 23
Mario works with a group of R programmers tasked with copying data from an accounting system into a data warehouse.
In what phase are the group's R skills most relevant?

A. Extract.
B. Load.
C. Transform.
D. Purge.

Answer: C

NEW QUESTION 26
A database administrator is required to mask certain table columns containing Pll in order to comply with the company privacy policy. Which of the following are the most likely types of information the administrator should mask? (Select two).

A. Government-issued ID
B. Address
C. Order ID
D. Order date
E. Customer ID
F. Referral number

Answer: AB

NEW QUESTION 31
A data analyst for a media company needs to determine the most popular movie genre. Given the table below:

MovieID	Name	Genre	Actors	Rating
01	Ghost Writer	Comedy, Actions	Joshua Wellington, Susana Summons	6.5
02	Life of Suffering	Drama, Foreign, Historical	Shelly May, Rita Moralle, Ethan Warner, Sean Houser	7.2

Which of the following must be done to the Genre column before this task can be completed?

A. Append
B. Merge
C. Concatenate
D. Delimit

Answer: D

Explanation:
Delimiting is the process of splitting a column of data into multiple columns based on a separator or delimiter character. Delimiting can help separate data that is combined or concatenated in one column into distinct values or categories. For example, if a column contains text values that are separated by commas, such as ??Comedy, Suspense??, delimiting can split this column into two columns, one for ??Comedy?? and one for ??Suspense??. Delimiting is different from other options, such as appending, merging, or concatenating, which are methods of combining or joining data from multiple columns or sources. In this case, the data analyst needs to determine the most popular movie genre based on the Genre column in the table. However, this column contains multiple genres for each movie, separated by commas. Therefore, the data analyst must delimit this column before this task can be completed. Therefore, the correct answer is D. References: Split text into different columns with functions - Office Support, How to Split Text in Excel (Using Formulas & Split Function)

NEW QUESTION 33
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 34
An analyst wants to create a historical data set for the past five years with each year in its own data set. Which of the following methods is the best way to create this historical data set?

A. Data transpose
B. Data concatenation
C. Data append
D. Data normalization

Answer: B

NEW QUESTION 37
Which of the following is a non-parametric test?

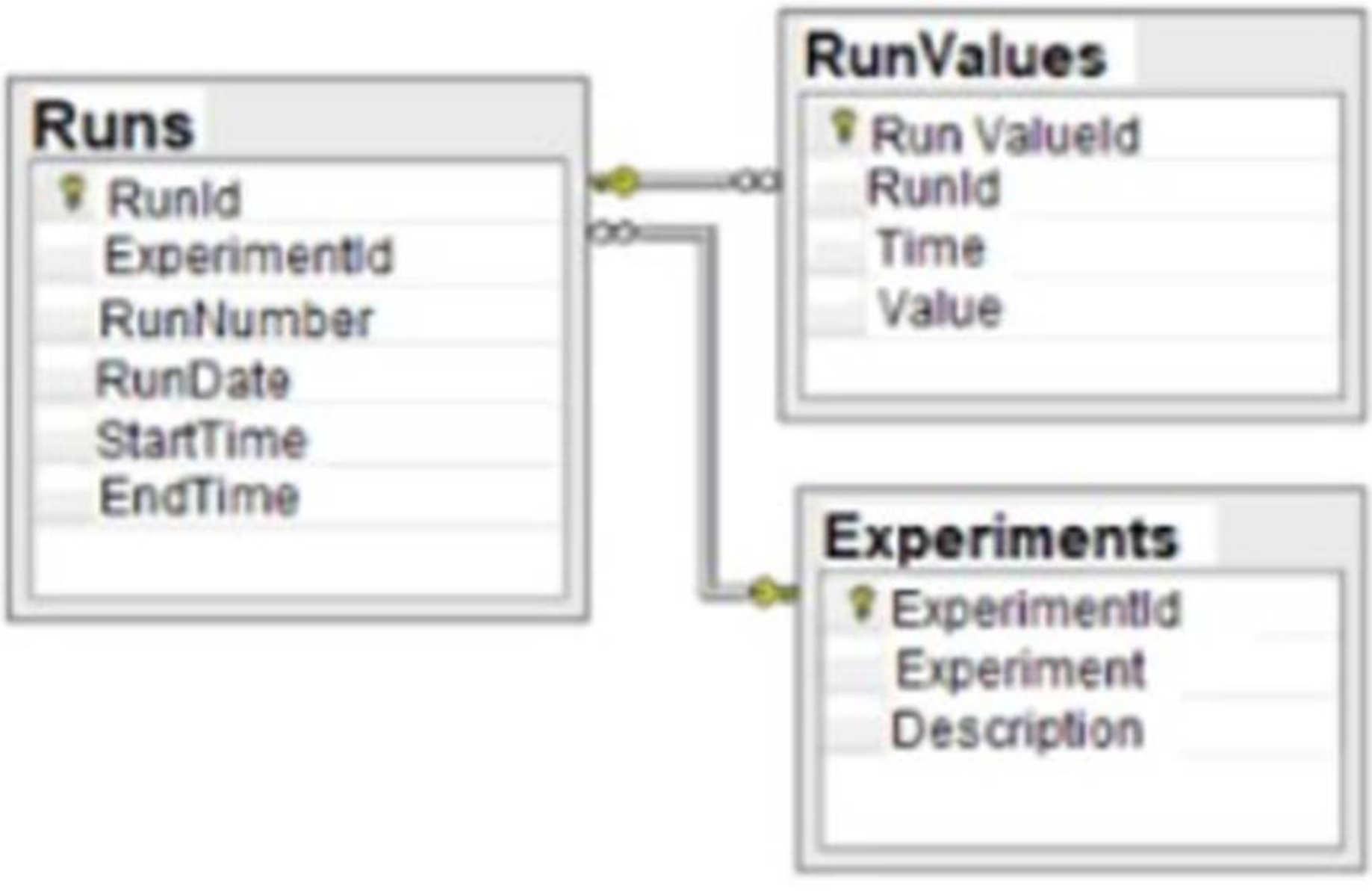
A. One-sample t-test
B. Two-way ANOVA

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 39
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 44
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

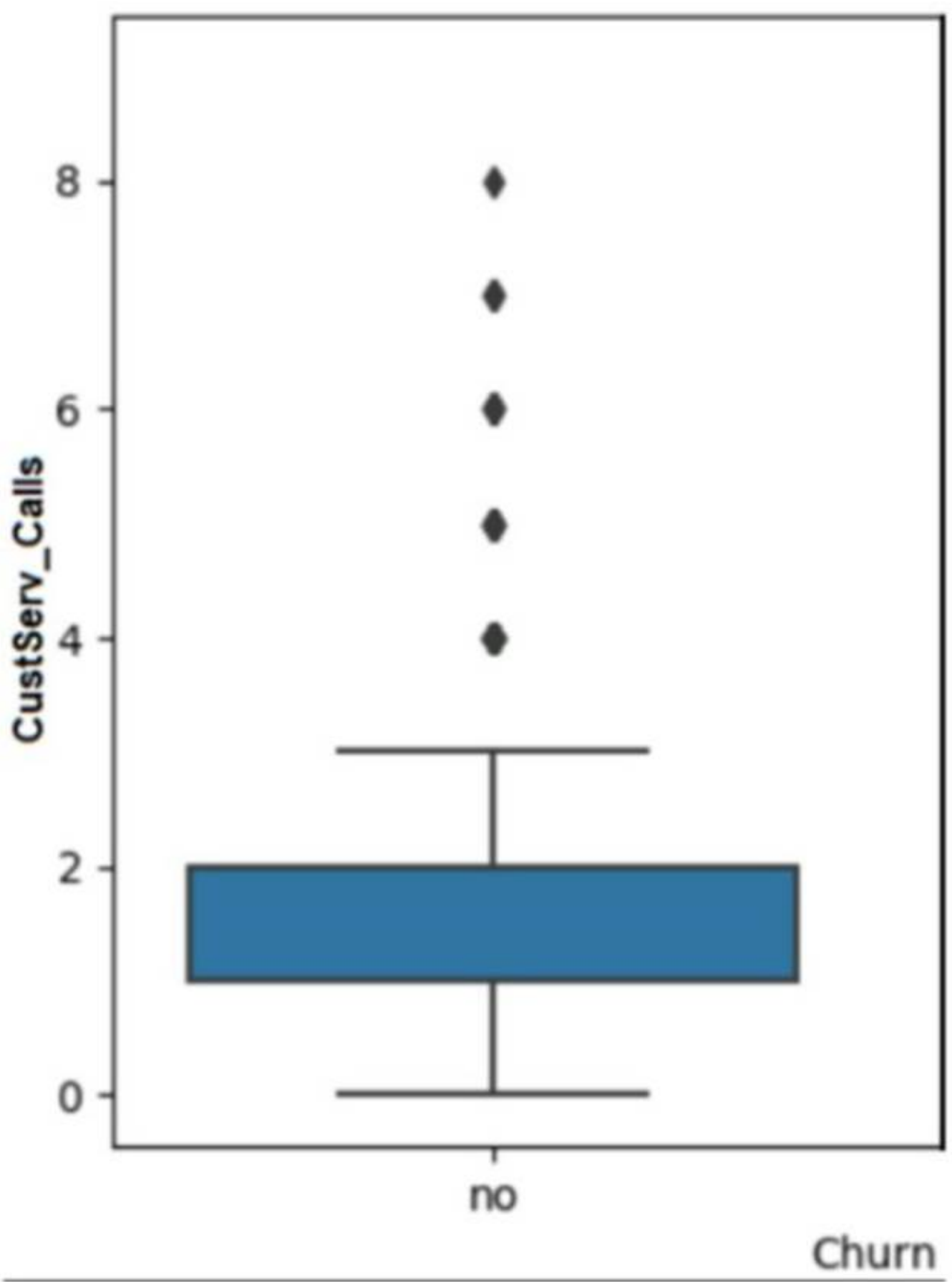
A data analyst wants to create a likely-to-buy score on a scale from 0 to 100, based on an average of the three numerical variables: number of credit cards, age, and income. Which of the following should the analyst do to the variables to ensure they all have the same weight in the score calculation?

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 47
Given the image below:



The data should be cleaned because of the presence of:

A. outlier
B. non-parametric data.
C. multicollinearity.
D. invalid data.

Answer: A

Explanation:
The answer is A. Outlier.
Short Explanation: An outlier is a data point that differs significantly from the rest of the data in a dataset. An outlier can indicate an error, an anomaly, or a rare event in the data. An outlier can affect the statistical analysis and visualization of the data, such as skewing the mean, variance, or distribution of the data. Therefore, data should be cleaned to identify and remove or correct any outliers.
The image below shows a box plot graph with a vertical axis labeled ??Customer Calls?? and a horizontal axis labeled ??Churn??. The box plot is blue in color and the median value is around 2. There are 7 outliers above the box plot, ranging from 4 to 8. image)
A box plot is a type of graph that can show the distribution of data values using five summary statistics: minimum, maximum, median, first quartile, and third quartile. The box represents the interquartile range (IQR), which is the difference between the first and third quartiles. The median is shown as a line inside the box. The whiskers extend from the box to the minimum and maximum values, excluding any outliers. Outliers are shown as dots or circles outside the whiskers.
In this graph, we can see that most of the customer calls are between 0 and 4, with a median of 2. However, there are 7 outliers that have more than 4 customer calls, up to 8. These outliers may indicate some customers who have more issues or complaints than others, or some errors or anomalies in the data collection or recording process. These outliers can affect the analysis and interpretation of the customer calls and churn relationship, such as making it seem that more customer calls lead to less churn, which may not be true for the majority of the customers. Therefore, data should be cleaned to investigate and handle these outliers appropriately.

NEW QUESTION 50
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 55
Which of the following types of analyses should be used to evaluate the connections and anomalies in a data set when either known patterns are being violated or new patterns are emerging?

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

Answer: C

NEW QUESTION 59
A Chief Executive Officer (CEO) is requesting more up-to-date sales data for improved visibility prior to month-end. An analyst must determine the frequency of a sales report that was previously distributed on an as-needed basis. Which of the following would be the most appropriate frequency for this report?

A. Monthly
B. Quarterly
C. Weekly
D. Every other month

Answer: C

Explanation:
The most appropriate frequency for the sales report is weekly, as this will provide the CEO with more up-to-date sales data for improved visibility prior to month-end. A weekly sales report can show the sales performance, trends, and issues of the sales team on a regular basis, and help the CEO to monitor and evaluate the progress and results of the sales activities. A weekly sales report can also help the CEO to identify and address any problems or opportunities that may arise during the month, and to make timely and informed decisions.

NEW QUESTION 64
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. 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:
Year-over-year (YoY) comparison is a method of evaluating two or more measured events to compare the results at one period with those from a comparable period on an annual basis. For a year-over-year comparison of Q2 2020 sales, the analyst should compare the sales figures from Q2 2020 with those from Q2 2019. This comparison will show the growth, stagnation, or decline in sales over the year and is a common practice in financial analysis to assess performance.
References:
? SlideTeam??s article on sales comparison templates1.
? Salesforce help article on calculating YoY or Quarter-over-Quarter (QoQ) in reports2.
? Smartsheet??s content on annual sales report templates3.
? TechRepublic article on creating a YoY comparison chart using a PivotChart in Excel4.

NEW QUESTION 65
A data scientist wants to see which products make the most money and which products attract the most customer purchasing interest in their company.
Which of the following data manipulation techniques would he use to obtain this information?

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

Answer: B

Explanation:

The correct answer is B: Data blending.
Data blending is combining multiple data sources to create a single, new dataset, which can be presented visually in a dashboard or other visualization and can then be processed or analyzed. Enterprises get their data from a variety of sources, and users may want to temporarily bring together different datasets to compare data relationships or answer a specific question. Data append is incorrect. Data append is a process that involves adding new data elements to an

existing database. An example of a common data append would be the enhancement of a company's customer files. A data append takes the information they have, matches it against a larger database of business data, allowing the desired missing data fields to be added. Normalize data is incorrect.
Data normalization is the process of structuring your relational customer database, following a series of normal forms. This improves the accuracy and integrity of your data while ensuring that your database is easier to navigate. Data merge is incorrect. Data merging is the process of combining two or more data sets into a single data set.

NEW QUESTION 67
Alex wants to use data from his corporate sale, CRM, and shipping systems to try and predict future sales.
Which of the following systems is the most appropriate? Choose the best answer.

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

Answer: C

Explanation:

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

NEW QUESTION 72
A research analyst collects ten data points from 1.000 specimens. The analyst will not need any additional data to complete the analysis and will not need to retrieve information by specifier. Which of the following is the best data structure for the analyst to use?

A. NoSQL
B. Flat file
C. JSON
D. Relational database

Answer: B

Explanation:
A flat file is a type of data structure that stores data in a plain text format, such as CSV, TSV, or TXT. A flat file consists of one or more records, each containing one or more fields, separated by a delimiter, such as a comma, tab, or space. A flat file does not have any hierarchical or relational structure, and does not support any complex queries or operations1.
A flat file may be the best data structure for the analyst to use in this scenario, because:
? The analyst collects ten data points from 1,000 specimens, which means the data is relatively small and simple, and can be easily stored and processed in a flat file.
? The analyst will not need any additional data to complete the analysis, which means the data is static and does not require any updates or modifications.
? The analyst will not need to retrieve information by specifier, which means the data
does not require any indexing or searching by key or value.

NEW QUESTION 76
You should always choose the analytics tool that is most appropriate for any given situation, even if that means acquiring a new tool.

A. True.
B. False.

Answer: B

Explanation:
The statement is false. You should not always choose the analytics tool that is most appropriate for any given situation, even if that means acquiring a new tool. Acquiring a new tool can be costly, time-consuming, and risky, as it may not be compatible with your existing data sources, systems, or processes. It may also require additional training, maintenance, and support. Therefore, you should always consider the trade-offs between the benefits and drawbacks of acquiring a new tool versus using an existing one. You should also evaluate the feasibility, availability, and reliability of the new tool before making a decision. Reference: CompTIA Data+ (DA0-001) Practice Certification Exams | Udemy

NEW QUESTION 78
Which of the following is an example of a data-mining ETL tool?

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

Answer: A

Explanation:
A data-mining ETL tool is a software application that performs extract, transform, and load (ETL) operations on data for data mining purposes. Data mining is the process of discovering patterns, trends, and insights from large and complex data sets. ETL tools help to prepare the data for analysis by extracting data from various sources, transforming data into a consistent and suitable format, and loading data into a data warehouse or other destination. SSIS (SQL Server Integration Services) is an example of a data-mining ETL tool that is part of Microsoft SQL Server. SSIS provides graphical tools and wizards for building and debugging ETL packages that can work with various data
sources and destinations. Therefore, the correct answer is A. References: [Data Mining - SQL Server Integration Services (SSIS) | Microsoft Docs], [What Is Data Mining? | Oracle]

NEW QUESTION 81
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 83
Which of the following can be used to translate data into another form so it can only be read by a user who has a key or a password?

A. Data encryption.
B. Data transmission.
C. Data protection.
D. Data masking.

Answer: A

Explanation:

Data encryption can be used to translate data into another form so it can only be read by a user who has a key or a password. Data encryption is a process of transforming data using an algorithm or a cipher to make it unreadable to anyone except those who have the key or the password to decrypt it. Data encryption is a common method of protecting data from unauthorized access, modification, or theft. Reference: Guide to CompTIA Data+ and Practice Questions - Pass Your Cert

NEW QUESTION 88
An analyst has written the following code: SELECT *
FROM Cust_table
WHERE age > 60 AND City = "New York"
Which of the following criteria is the analyst retrieving?

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

Answer: C

Explanation:
The SQL query provided is selecting all records from the Cust_table where the age column has values greater than 60 and the City column matches ??New York??. The > operator selects values that are strictly greater than the comparison value, so it does not include customers aged exactly 60. The term ??New York?? in the context of a city database typically refers to New York City, not the state of New York. Therefore, the correct answer is that the analyst is retrieving data for all customers older than age 60 in New York City.
References:
? The use of the > operator in SQL is to select values greater than the specified value1.
? Understanding the WHERE clause in SQL and its use in filtering records based on specified conditions2.
? Clarification on the distinction between city and state names in database records3.

NEW QUESTION 90
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 92
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 94
A data analyst is helping a retail store categorize its customers into five different groups based on the following information:
• How recently the customers made purchases
• How frequently the customers made purchases
• How much the customers spent Given the following information:

Customer_ID	Channel	Order_Date	Quantity	Territory	Amount ($)
1001	Online	2/11/2020	12	North	1,250
2001	Store	2/10/2020	31	East	5,000
4001	Online	2/09/2020	24	West	2,500
3001	Online	2/11/2020	51	South	6,000
1001	Store	3/10/2020	22	North	2,000
1001	Online	1/09/2020	87	North	8,400
1001	Store	2/09/2020	23	North	2,000

Which of the following would be most important for the analysis?

A. CustomerJ
B. Channel, Order_Date
C. CustomerJD, Territor
D. Amount
E. CustomerJD, Order_Dat
F. Amount
G. CustomerJ
H. Quantity, Amount

Answer: C

NEW QUESTION 98
Consider this dataset showing the retirement age of 11 people, in whole years: 54, 54, 54, 55, 56, 57, 57, 58, 58, 60, 60
This tables show a simple frequency distribution of the retirement age data.

Age	Frequency
54	3
55	1
56	1
57	2
58	2
60	2

A. 56
B. 55
C. 57
D. 54

Answer: D

Explanation:

A measure of central tendency (also referred to as measures of centre or central location) is a summary measure that attempts to describe a whole set of data with a single value that represents the middle or centre of its distribution.
There are three main measures of central tendency: the mode, the median and the mean. Each of these measures describes a different indication of the typical or central value in the distribution.
What is the mode?
The mode is the most commonly occurring value in a distribution.
The most commonly occurring value is 54, therefore the mode of this distribution is 54 years.

NEW QUESTION 103
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 105
An analyst is reporting on the average income for a county and is reviewing the following data:

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

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

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

Answer: B

NEW QUESTION 107
Which of the following describes the use of a representative amount of data from a main repository?

A. Observation
B. Delta load
C. Web scraping
D. Sampling

Answer: D

Explanation:
Sampling refers to the process of selecting a representative subset of data from a larger data set or repository. This technique is used when it is impractical or unnecessary to analyze the entire set of data. A representative sample should accurately reflect the characteristics of the larger population, allowing for analysis and inference about the population as a whole12.
Observation (A) generally refers to the act of monitoring or recording data. Delta load (B) is a term used in data warehousing to describe the process of loading only the changes since the last data extraction, rather than the entire data set. Web scraping © is the process of extracting data from websites.
References:
? Understanding the importance of data sampling1.
? The concept of a representative sample in statistics2.
? Data repository management and usage3.
? Benefits and methods of data sampling4.

NEW QUESTION 111
A data analyst is creating a dashboard and trying to identify the type of information that should be included. Which of the following should the analyst consider first?

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

Answer: D

Explanation:
The answer is D. Data sources and attributes.
Short Explanation: The data analyst should consider the data sources and attributes first when creating a dashboard, because they determine what kind of information can be
included and how it can be displayed. The data sources and attributes define the origin, quality, format, and structure of the data that will be used for the dashboard. They also affect the data refresh rate, the consumer types, and the access permissions of the dashboard12
* A. Data refresh rate is not the first thing to consider, because it depends on the data sources and attributes. The data refresh rate is how often the data in the dashboard is updated or refreshed to reflect the latest changes. The data refresh rate can vary depending on the type, frequency, and availability of the data sources1
* B. Consumer types are not the first thing to consider, because they depend on the data sources and attributes. The consumer types are the intended audiences or users of the dashboard, who may have different needs, preferences, and expectations for the dashboard. The consumer types can influence the design, layout, and functionality of the dashboard. However, the consumer types cannot be determined without knowing what kind of data is available and relevant for them1
* C. Access permissions are not the first thing to consider, because they depend on the data sources and attributes. The access permissions are the rules or policies that govern who can view, edit, or share the dashboard. The access permissions can protect the confidentiality, integrity, and availability of the data in the dashboard. However, the access permissions cannot be set without knowing what kind of data is involved and who needs to access it1

NEW QUESTION 114
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 118
Given the following data tables:

CustomerID	CustomerLastName
01	Manzelli
02	Kraus

SalesRepID	Customer Last Name	Items
01	Poputhopolis	Wagon, Red Paint
02	Smith	Bicycle, Wheels, Handlebars

ItemID	Customer_Last_Name	QuantityPurchased
01	Brown	03
02	Smee	07

Which of the following MDM processes needs to take place FIRST?

A. Creation of a data dictionary
B. Compliance with regulations
C. Standardization of data field names
D. Consolidation of multiple data fields

Answer: A

Explanation:
This is because a data dictionary is a type of document that defines and describes the data elements, attributes, and relationships in a database or a data set. A data dictionary can be used to facilitate the MDM (Master Data Management) process, which is a process that aims to ensure the quality, consistency, and accuracy of the data across different sources and systems. By creating a data dictionary first, the analyst can establish a common understanding and standardization of the data field names, types, formats, and meanings, as well as identify any potential issues or conflicts in the data, such as missing values, duplicate values, or inconsistent values. The other MDM processes can take place after creating a data dictionary. Here is why:
Compliance with regulations is a type of MDM process that ensures that the data meets the legal and ethical requirements and standards of the industry or the

organization.
Compliance with regulations can take place after creating a data dictionary, because the data dictionary can help the analyst to identify and apply the relevant rules and policies to the data, such as data privacy, security, or retention.
Standardization of data field names is a type of MDM process that ensures that the data field names are consistent and uniform across different sources and systems. Standardization of data field names can take place after creating a data dictionary, because the data dictionary can provide a reference and a guideline for naming and labeling the data fields, as well as resolving any discrepancies or ambiguities in the data field names.
Consolidation of multiple data fields is a type of MDM process that combines or merges the data fields from different sources or systems into a single source or system. Consolidation of multiple data fields can take place after creating a data dictionary because the data dictionary can help the analyst to map and match the data fields from different sources or systems based on their definitions and descriptions, as well as eliminating any redundant or duplicate data fields.

NEW QUESTION 123
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 125
A customer's telephone number is in the format 123-456-7890. Which of the following data types is used for the phone number?

A. Boolean
B. Date
C. Text
D. Number

Answer: C

Explanation:
A telephone number, despite being composed of digits, is not used for calculations and often includes formatting characters such as hyphens (-). Therefore, the most appropriate data type for a telephone number is Text (oVr ARCHAR in SQL databases), which can accommodate various formats and lengths, and preserve leading zeros that might be present in some phone numbers. Storing phone numbers as numeric data types would strip away any formatting and could lead to the loss of significant leading digits (like zeros in international numbers).
? Boolean is a binary data type and only represents true or false values.
? Date is a data type used for dates.
? Number could technically store phone numbers, but it is not suitable due to the reasons mentioned above.
References:
? Best Practices for Storing Phone Numbers1
? Data Types in SQL for Phone Numbers2

NEW QUESTION 129
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 133
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 137
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 140
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 143
A data analyst needs to calculate the mean for Q1 sales using the data set below:

Product	Q1 sales
Ground beef	$2,667.60
Crab meet	$1,768.41
Swiss cheese	$3,182.40
Broccoli	$1,509.60
Vegetable spread	$3.202.87

Which of the following is the mean?

A. $2,466.18
B. $2,667.60
C. $3,082.72
D. $12,330.88

Answer: C

Explanation:
The mean is the average of all the values in a data set. To calculate the mean, we add up all the values and divide by the number of values. In this case, the mean for Q1 sales is ($2,000 + $3,000 + $4,000 + $2,500 + $3,500) / 5 = $3,082.72 References: CompTIA Data+ Certification Exam Objectives, page 9

NEW QUESTION 148
Which of the following tools would be best to use to calculate the interquartile range, median, mean, and standard deviation of a column in a table that has 5.000.000 rows?

A. Microsoft Excel
B. R
C. Snowflake
D. SQL

Answer: B

NEW QUESTION 150
Which of the following is a KPI metric for tracking sales performance?

A. Order status percentage
B. Customer acquisition percentage

C. Gross profit percentage
D. Click-through rate percentage

Answer: C

Explanation:
Gross profit percentage is a key performance indicator (KPI) that measures the profitability of a company??s sales by showing the percentage of revenue that exceeds the cost of goods sold (COGS). It is a critical metric for tracking sales performance because it directly reflects the efficiency of a company in managing its production costs and the profitability of its products. This KPI is essential for understanding the financial health of a business and making informed decisions about pricing, cost control, and sales strategies.
References:
? Sales KPIs are essential for measuring the effectiveness of sales activities and the profitability of those efforts1.
? Gross profit percentage is highlighted as a crucial metric for assessing the financial success of sales initiatives2.
? Understanding the difference between sales metrics and KPIs, and the importance of gross profit percentage as a KPI1.
? The significance of gross profit percentage in evaluating sales team performance and guiding business decisions3.

NEW QUESTION 155
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 159
Given the following graph:

Which of the following summary statements upholds integrity in data reporting?

A. Sales are approximately equal for Product A and Product B across all strategies.
B. Strategy 4 provides the best sales in comparison to other strategies.
C. While Strategy 2 does not result in the highest sales of Product D, over all products it appears to be the most effective.
D. Product D should be promoted more than the other products in all strategies.

Answer: B

Explanation:
Strategy 4 provides the best sales in comparison to other strategies. This is because the total sales for Strategy 4 are the highest among all the strategies, as shown by the black line. The other statements are not accurate or do not uphold integrity in data reporting. Here is why:
Statement A is false because sales are not approximately equal for Product A and Product B across all strategies. For example, in Strategy 1, Product A has more sales than Product B, while in Strategy 3, Product B has more sales than Product A.
Statement C is misleading because it does not account for the difference in scale between the products. While Strategy 2 has the highest total sales among all products, it does not necessarily mean that it is the most effective for each product. For instance, Product D has very low sales in Strategy 2 compared to other strategies.
Statement D is biased because it does not provide any evidence or justification for why Product D should be promoted more than the other products in all strategies. It also ignores the fact that Product D has the lowest sales among all products in most of the strategies.

NEW QUESTION 162
A gambler thinks that a coin is fair and is equally likely to turn up heads or tails when the coin is flipped. Which of the following tests should the gambler use to fest this hypothesis?

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

Answer: B

NEW QUESTION 164
Which of the following is an example of structured data?

A. A credit card number
B. An email
C. A photo
D. Social media correspondence

Answer: A

Explanation:
A credit card number is an example of structured data, which is a type of data that conforms to a data model, has a well-defined structure, follows a consistent order, and can be easily accessed and used by a person or a computer program. A credit card number consists of 16 digits that are divided into four groups of four digits each, separated by spaces or hyphens. The first six digits indicate the issuer identification number, the next nine digits indicate the account number, and the last digit is a check digit that validates the
number. A credit card number can be stored and processed in a structured format, such as a database or a spreadsheet1.

NEW QUESTION 166
An analyst has received the requirements for an internal user dashboard. The analyst confirms the data sources and then creates a wireframe. Which of the following is the NEXT step the analyst should take in the dashboard creation process?

A. Optimize the dashboard.
B. Create subscriptions.
C. Get stakeholder approval.
D. Deploy to production.

Answer: C

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

NEW QUESTION 168
An analyst wants to combine two data sets into a single spreadsheet. Column names from
the first spreadsheet are listed in rows in the second spreadsheet. Which of the following is the first step the analyst should take to combine the data sets?

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

Answer: C

NEW QUESTION 170
You have two databases tables that you would like to join together using a foreign key relationship.
What term best describes this action?

A. Blending.
B. Appending.
C. Mixing.
D. Merging.

Answer: D

Explanation:

Data merging is the process of combining two or more data sets into a single data set. Most often, this process is necessary when you have raw data stored in multiple files, worksheets, or data tables, that you want to analyze all in one go.

NEW QUESTION 175
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 179
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 184
An analyst must obtain the average daily sales for the following week:

Date	SalesTotal
2/10/2020	$36,986
2/11/2020	$37,981
2/12/2020	$40,551
2/13/2020	$42,442
2/14/2020	$56,216
2/15/2020	$81,117
2/16/2020	$63,815

Which of the following must the analyst perform to obtain this value?

A. Data normalization
B. Data append
C. Data aggregation
D. Data blending

Answer: C

Explanation:
Data aggregation is the process of compiling data from multiple sources and summarizing it into a single dataset. Data aggregation can be used to calculate statistics, such as averages, sums, counts, or percentages. In this case, the analyst must obtain the average daily sales for the following week, which is a statistic that can be calculated by aggregating the sales data from each day and dividing by the number of days. Data aggregation can be done using various tools and methods, such as spreadsheets, databases, or programming languages.

NEW QUESTION 189
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 191
A company wants to know how its customers interact with an e-commerce website based on clicks over items. Which of the following is the primary requirement for this report?

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

Answer: B

NEW QUESTION 196
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 198
Each month an analyst needs to execute a data pull for the two prior months. Which of the following is the most efficient function for the analyst to use?

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

Answer: B

Explanation:
The most efficient function for an analyst to execute a data pull for the two prior months would be the Date function. This function allows for the manipulation and formatting of date values within a database. Using Date functions, an analyst can dynamically calculate the start and end dates for the previous two months, ensuring that the data pull is accurate and automated without manual intervention.
For example, SQL functions like DATEADD and DATEDIFF can be used to determine the exact range of dates needed for the data pull. These functions can calculate the first and
last day of the previous months relative to the current date, which is essential for monthly reporting and analysis.
References:
? Discussions on Stack Overflow suggest using SQL date functions
like DATEADD and DATEDIFF to dynamically extract data for previous months, which supports the use of Date functions12.
? The use of Date functions is also recommended for ensuring that the data pull is
not only efficient but also accurate, as it avoids potential errors associated with manual date entry3.

NEW QUESTION 199
A data analyst has been asked to organize the table below in the following ways: By sales from high to low -
By state in alphabetic order -

First_name	Last_name	Address	City	State	Sales
Ed	Edens	2851 N. Southport	Chicago	IL	$125,689
Pat	Mudd	710 Bridle Ridge Road	Eagan	MN	$101,259
Katie	Hofstad	2851 S. Windwood Lane	Rosemount	NY	$105,779
Edward	Frank	281 S. Northport	Chicago	IL	$456,231
Rachel	Newman	305 Big Timber Trail	Wheaton	CO	$99,876
Kaylyn	Korth	332 Richfield Drive	Lakeview	MN	$166,874

Which of the following functions will allow the data analyst to organize the table in this manner?

A. Conditional formatting
B. Grouping
C. Filtering
D. Sorting

Answer: D

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

NEW QUESTION 203
The number of phone calls that the call center receives in a day is an example of:

A. continuous data.
B. categorical data.

C. ordinal data.
D. discrete data.

Answer: D

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

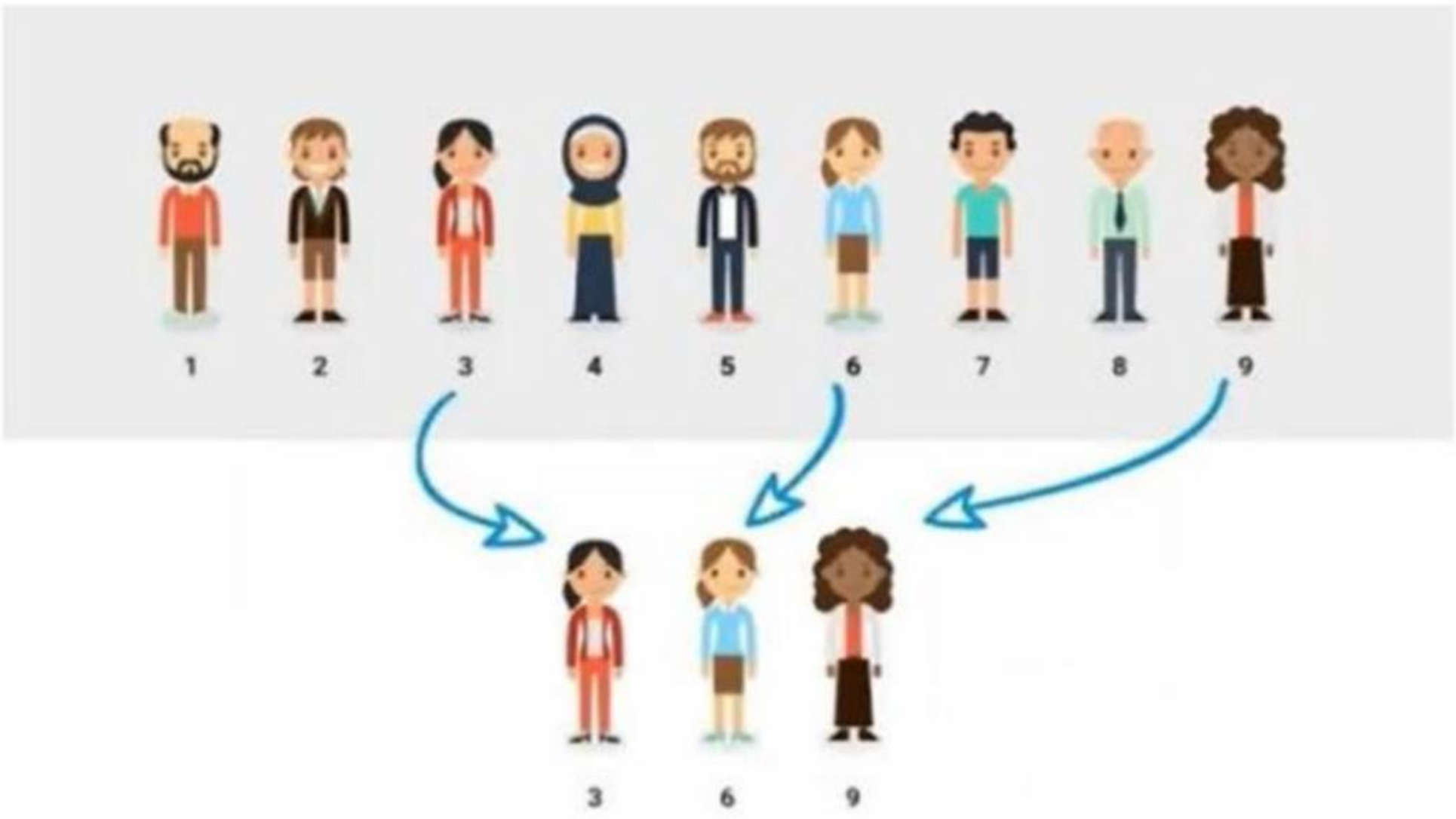
NEW QUESTION 205
A company??s marketing department wants to do a promotional campaign next month. A data analyst on the team has been asked to perform customer segmentation, looking at how recently a customer bought the product, at what frequency, and at what value. Which of the following types of analysis 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 209
Given the diagram below:



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

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

Answer: D

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

NEW QUESTION 211
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 215
Which of the following data governance concepts fits into the security requirements category?

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

Answer: D

NEW QUESTION 220
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 224
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 225
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 227
Andy is a pricing analyst for a retailer. Using a hypothesis test, he wants to assess whether people who receive electronic coupons spend more on average. What should Andy's null hypothesis be?

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

Answer: C

Explanation:

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

NEW QUESTION 228
A data analyst has been asked to merge the tables below, first performing an INNER JOIN and then a LEFT JOIN:

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

Customer Table -
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

Which of the following describes the number of rows of data that can be expected after performing both joins in the order stated, considering the customer table as the main table?

A. INNER: 6 rows; LEFT: 9 rows
B. INNER: 9 rows; LEFT: 6 rows
C. INNER: 9 rows; LEFT: 15 rows
D. INNER: 15 rows; LEFT: 9 rows

Answer: C

Explanation:
An INNER JOIN returns only the rows that match the join condition in both tables. A LEFT JOIN returns all the rows from the left table, and the matched rows from the right table, or NULL if there is no match. In this case, the customer table is the left table and the in-store transactions table is the right table. The join condition is based on the customer_id column, which is common in both tables.
To perform an INNER JOIN, we can use the following SQL query:
SELECT * FROM customer INNER JOIN in_store_transactions ON customer.customer_id
= in_store_transactions.customer_id;
This query will return 9 rows of data, as shown below:
customer_id | name | lastname | gender | marital_status | transaction_id | amount | date 1 | MARC | TESCO | M | Y | 1 | 1000 | 2020-01-01 1 | MARC | TESCO | M | Y | 2 | 5000 | 2020-01-02 2 | ANNA | MARTIN | F | N | 3 | 2000 | 2020-01-03 2 | ANNA | MARTIN | F | N |

4 | 3000 | 2020-01-04 3 | EMMA | JOHNSON | F | Y | 5 | 4000 | 2020-01-05 4 | DARIO | PENTAL | M | N | 6 | 5000 | 2020-01-06 5 | ELENA | SIMSON| F| N|7|6000|2020-01-07 6|TIM|ROBITH|M|N|8|7000|2020-01-08 7|MILA|MORRIS|F|N|9|8000|2020-01-09
To perform a LEFT JOIN, we can use the following SQL query:
SELECT * FROM customer LEFT JOIN in_store_transactions ON customer.customer_id = in_store_transactions.customer_id;
This query will return 15 rows of data, as shown below: customer_id|name|lastname|gender|marital_status|transaction_id|amount|date
1|MARC|TESCO|M|Y|1|1000|2020-01-01 1|MARC|TESCO|M|Y|2|5000|2020-01-02
2|ANNA|MARTIN|F|N|3|2000|2020-01-03 2|ANNA|MARTIN|F|N|4|3000|2020-01-04
3|EMMA|JOHNSON|F|Y|5|4000|2020-01-05 4|DARIO|PENTAL|M|N|6|5000|2020-01-06
5|ELENA|SIMSON||F||N||7||6000||2020-01-07 6||TIM||ROBITH||M||N||8||7000||2020-01-08
7||MILA||MORRIS||F||N||9||8000||2020-01-09
8||JENNY||DWARTH||F||Y||NULL||NULL||NULL
As you can see, the customers who do not have any transactions (customer_id = 8) are still included in the result, but with NULL values for the transaction_id, amount, and date columns.
Therefore, the correct answer is C: INNER: 9 rows; LEFT: 15 rows. Reference: SQL Joins - W3Schools

NEW QUESTION 232
Which of the following is the best description of discrete data types?

A. Non-numeric data used to describe attributes of a population sample
B. The frequency of the number of times each value occurs by using whole numbers
C. Numeric values that can be measured on a continuous scale
D. Non-numeric data used to describe attributes of a population sample ranked in a specific order

Answer: B

NEW QUESTION 237
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 238
A data analyst is performing a data merge within a spreadsheet using the tables below:
https://www.bing.com/images/blob?bcid=S1XCF9p02M4GjpbGxHj0lrIaj9sw.....4c

Table 1

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

Table 2

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

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

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

Answer: D

Explanation:
The error in merging data from Table 2 into Table 1 using last names could be due to discrepancies in spelling or data type between the two tables. It is essential to ensure that the last names are spelled consistently and that the data types are compatible for a successful merge. Option D suggests reviewing these aspects, which can potentially resolve the error, ensuring that each last name in Table 1 accurately corresponds to the same last name in Table 2, allowing for a successful data pull of addresses.
References: This answer is based on general data analytics practices and does not reference a specific document.

NEW QUESTION 241
A research analyst wants to determine whether the data being analyzed is connected to other datapoints. Which of the following is the BEST type of analysis to conduct?

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

Answer: C

Explanation:
This is because link analysis is a type of analysis that determines whether the data being analyzed is connected to other datapoints, such as entities, events, or relationships. Link analysis can be used to identify and visualize the patterns, networks, or associations among the datapoints, as well as measure the strength, direction, or frequency of the connections. For example, link analysis can be used to determine if there is a connection between a customer??s purchase history and their loyalty program status. The other types of analysis are not the best types of analysis to conduct to determine whether the data being analyzed is connected to other datapoints. Here is why:
? Trend analysis is a type of analysis that determines whether the data being analyzed is changing over time, such as increasing, decreasing, or fluctuating. Trend analysis can be used to identify and visualize the patterns, cycles, or movements in the data points, as well as measure the rate, direction, or magnitude of the changes. For example, trend analysis can be used to determine if there is a change in a company??s sales revenue over a period of time.
? Performance analysis is a type of analysis that determines whether the data being
analyzed is meeting certain goals or objectives, such as targets, benchmarks, or standards. Performance analysis can be used to identify and visualize the gaps, deviations, or variations in the data points, as well as measure the efficiency, effectiveness, or quality of the outcomes. For example, performance analysis can be used to determine if there is a gap between a student??s test score and their expected score based on their previous performance.
? Exploratory analysis is a type of analysis that determines whether there are any insights or discoveries in the data being analyzed, such as patterns, relationships, or anomalies. Exploratory analysis can be used to identify and visualize the characteristics, features, or behaviors of the data points, as well as measure their distribution, frequency, or correlation. For example, exploratory analysis can be used to determine if there are any outliers or unusual values in a dataset.

NEW QUESTION 245
An analyst needs to determine the appropriate data type for the following sample data: sample data collected:
Which of the following data types should be used for this data?

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

Answer: B

NEW QUESTION 250
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 253
An analyst in a consumer bank department wants to showcase the concentration of accounts opened in the United States by ZIP Code to describe the effectiveness of the bank's marketing campaigns. Which of the following would be the best way to visualize the data?

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

Answer: D

NEW QUESTION 256
The ACME Corporation hired an analyst to detect data quality issues in their Excel documents. Which of the following are the most common issues? (Select TWO)

A. Apostrophe.
B. Commas.
C. Symbols.
D. Duplicates.
E. Misspellings.

Answer: DE

Explanation:

* 1. Duplicates
* 2. Misspellings
The most common data quality issues are difficult to resolve in Excel because of their rigidity. It forces analysts to do a ton of manual work, which results in a high probability of an error being introduced to the data set. Those common issues include:
- Blanks
- Nulls
- Outliers
- Duplicates
- Extra spaces
- Misspellings
- Abbreviations and domain-specific variations
- Formula error codes
When introduced, these errors can skew or even invalidate the resulting analysis. A smart tool would minimize the possibility of error by automating the manual work. In Excel, you might look for data quality issues in one of two ways. First, you might use auto filters on specific columns to scan for anomalies and blanks or you might use a pivot table to find gaps and discrepancies.
In either case, you're scanning for the anomalies yourself. Suffice it to say that's not a very efficient process. It also means accuracy is only as good as the analyst's eye, so the probability of error varies throughout the day.

NEW QUESTION 261
Which of the following is a process that is used during data integration to collect, blend, and load data?

A. MDM
B. ETL
C. OLTP
D. BI

Answer: B

Explanation:
ETL is a process that is used during data integration to collect, blend, and load data. ETL stands for extract, transform, and load, which are the three main steps involved in moving data from different sources to a common destination, such as a data warehouse or a data lake. ETL helps to consolidate and standardize data for analysis and reporting purposes. References: CompTIA Data+ Certification Exam Objectives, page 12

NEW QUESTION 263
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 264
Which of the following is the first step an analyst should perform upon receiving a business request for analysis?

A. Determine the data needs and sources for analysis.
B. Initiate the analysis for exploratory data analysis.
C. Review the business questions to understand the scope.
D. Finalize the methodology to solve the problem.

Answer: C

Explanation:
Answer C. Review the business questions to understand the scope.

The first step an analyst should perform upon receiving a business request for analysis is to review the business questions to understand the scope of the problem, the objectives, and the expected outcomes. This will help the analyst to define the analytical approach, identify the data needs and sources, and plan the analysis process. Reviewing the business questions will also help the analyst to communicate with the stakeholders and clarify any assumptions or ambiguities1.
Option A is incorrect, as determining the data needs and sources for analysis is not the first step, but rather a subsequent step that depends on the business questions and the analytical approach.
Option B is incorrect, as initiating the analysis for exploratory data analysis is not the first step, but rather a part of the analysis process that involves examining and summarizing the data, identifying patterns and outliers, and testing hypotheses.
Option D is incorrect, as finalizing the methodology to solve the problem is not the first step, but rather a later step that involves selecting and applying the appropriate analytical techniques, tools, and models to answer the business questions.

NEW QUESTION 267
Which of the following statements would be used to append two tables that have the same number of columns?

A. UNION ALL
B. MERGE
C. GROUP BY
D. JOIN

Answer: A

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

NEW QUESTION 269
A data analyst is creating a report that will provide information about various regions, products, and time periods. Which of the following formats would be the MOST efficient way to deliver this report?

A. A workbook with multiple tabs for each region
B. A daily email with snapshots of regional summaries
C. A static report with a different page for every filtered view
D. A dashboard with filters at the top that the user can toggle

Answer: D

Explanation:
A dashboard with filters at the top that the user can toggle would be the most efficient way to deliver this report, because it allows the user to customize the view and explore different combinations of regions, products, and time periods. A workbook with multiple tabs for each region would be cumbersome and repetitive. A daily email with snapshots of regional summaries would not provide enough detail or interactivity. A static report with a different page for every filtered view would be too long and hard to navigate. References: CompTIA Data+ Certification Exam Objectives, page 14

NEW QUESTION 271
Which of the following should be accomplished NEXT after understanding a business requirement for a data analysis report?

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

Answer: B

Explanation:
Exploratory data analysis (EDA) is a process of examining and summarizing a dataset using various techniques, such as descriptive statistics, visualizations, correlations, outliers detection, and hypothesis testing. EDA can help reveal the main characteristics, patterns, trends, and insights from the data, as well as identify any problems or issues with the data quality or structure. EDA is usually performed after understanding a business requirement for a data analysis report and before building a mock dashboard/presentation layout. Therefore, the correct answer is B. References: [What is Exploratory Data Analysis? | Definition and Examples], [Exploratory Data Analysis in Python]

NEW QUESTION 273
Which one of the following is a measure of dispersion?

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

Answer: A

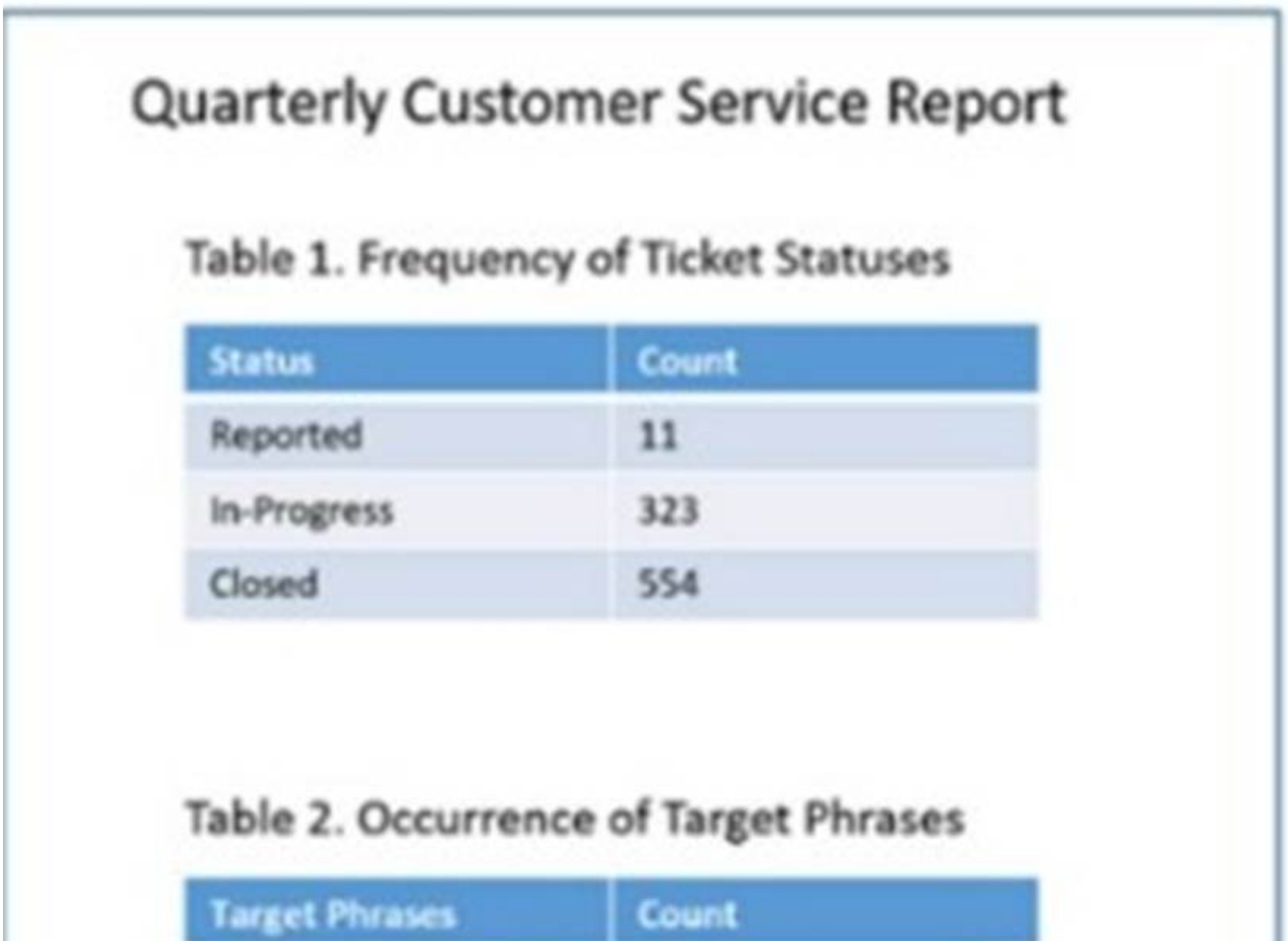
NEW QUESTION 277
Different people manually type a series of handwritten surveys into an online database. Which of the following issues will MOST likely arise with this data? (Choose two.)

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

Answer: AE

Explanation:
? Data accuracy refers to the extent to which the data is correct, reliable, and free of errors. When different people manually type a series of handwritten surveys into an online database, there is a high chance of human error, such as typos, misinterpretations, omissions, or duplications. These errors can affect the quality and validity of the data and lead to incorrect or misleading analysis and decisions.
? Data consistency refers to the extent to which the data is uniform and compatible across different sources, formats, and systems. When different people manually type a series of handwritten surveys into an online database, there is a high chance of inconsistency, such as different spellings, abbreviations, formats, or standards. These inconsistencies can affect the integration and comparison of the data and lead to confusion or conflicts.
Therefore, to ensure data quality, it is important to have clear and consistent rules and procedures for data entry, validation, and verification. It is also advisable to use automated tools or methods to reduce human error and inconsistency.

NEW QUESTION 280
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 284
An analyst is currently working on a ticket for revamping a company-wide dashboard that has been in use for five years. Which of the following should be the first step in the development process?

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

Answer: A

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

NEW QUESTION 289
A sales manager wants quarterly sales reports broken down by unit and week. Which of the following data output lists includes the most necessary information?

A. Order numbe
B. salesperso
C. date shipped, recipient address, and price
D. Item name, salesperso
E. recipient address, shipping cos
F. and date shipped
G. Item number, item name, salesperso
H. date sol
I. and price
J. Item nam
K. salesperso
L. pric
M. shipping cos
N. and date shipped

Answer: C

Explanation:
To create a quarterly sales report broken down by unit and week, the most necessary information is the item number, item name, salesperson, date sold, and price. These data elements can help the sales manager to track the sales volume, revenue, and performance of each unit and each week within a quarter. The item number and item name can identify the products or services sold by each unit. The salesperson can indicate the individual or team responsible for each sale. The date sold can show when each sale occurred and how it relates to the weekly and quarterly goals. The price can show how much revenue each sale generated and how it contributes to the unit and quarterly totals.

NEW QUESTION 292
Which of the following should an analyst do to best summarize the data on a data set?

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

Answer: B

NEW QUESTION 296
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 297
An e-commerce company recently tested a new website layout. The website was tested by a test group of customers, and an old website was presented to a control group. The table below shows the percentage of users in each group who made purchases on the websites:

Conversion	Control group	Test group	p-value
United States	7.8%	8.9%	0.003
Germany	6.3%	7.0%	0.13
United Kingdom	5.3%	9.6%	0.08
France	6.5%	6.7%	0.045
Canada	4.4%	5.1%	0.002

Which of the following conclusions is accurate at a 95% confidence interval?

A. In Germany, the increase in conversion from the new layout was not significant.
B. In France, the increase in conversion from the new layout was not significant.
C. In general, users who visit the new website are more likely to make a purchase.
D. The new layout has the lowest conversion rates in the United Kingdom.

Answer: C

Explanation:
The conclusion that is accurate at a 95% confidence interval is that in general, users who visit the new website are more likely to make a purchase. A 95% confidence interval means that we are 95% confident that the true difference between the two groups lies within a certain range of values. To calculate the 95% confidence interval, we can use the following formula:
CI = (p1 - p2) ?? 1.96 * sqrt(p * (1 - p) * (1/n1 + 1/n2))
where p1 and p2 are the conversion rates for the test and control groups, respectively, p is the pooled conversion rate, n1 and n2 are the sample sizes for the test and control groups, respectively, and 1.96 is the z-score for a 95% confidence level.
Using this formula, we can calculate the 95% confidence interval for each country as follows:
Country | p1 | p2 | n1 | n2 | p | CI United States | 0.12 | 0.11 | 2000 | 2000 | 0.115 | (-0.006, 0.026) Germany | 0.06 | 0.04 | 1000 | 1000 | 0.05 | (-0.002, 0.042) United Kingdom | 0.09 | 0.07 | 1500 | 1500 | 0.08 | (-0.003, 0.053) France | 0.08 | 0.08 | 1200 | 1200 | 0.08 | (-0.024, 0.024) Canada | 0.05 | 0.03 | 800 | 800 | 0.04 | (-0.005, 0.045)
We can see that for all countries except France, the confidence interval does not include zero, which means that the difference between the test and control groups is statistically significant at a 95% confidence level. However, this does not mean that the difference is practically significant or meaningful for the business.
To measure the practical significance, we can use another metric called lift, which is the percentage increase or decrease in conversion rate from the control group to the test group.
Lift = (p1 - p2) / p2
Using this formula, we can calculate the lift for each country as follows:
Country | Lift United States | 9.09% Germany | 50% United Kingdom |28.57% France|0% Canada|66.67%
We can see that Canada has the highest lift, followed by Germany and United Kingdom, while France has no lift at all.
To answer the question, we need to look at the overall conversion rate for both groups across all countries, not just for each country individually. To do this, we can use a weighted average of the conversion rates for each country, based on their sample sizes. Weighted average = (p1 * n1 + p2 * n2) / (n1 + n2)
Using this formula, we can calculate the weighted average conversion rate for both groups as follows:

Group|Weighted average Test|0.084 Control|0.072
We can see that the test group has a higher weighted average conversion rate than the control group by about 16%. We can also calculate the confidence interval and lift for the overall difference as follows:
CI = (p1 - p2) ?? 1.96 * sqrt(p * (1 - p) * (1/n1 + 1/n2)) = (0.084 - 0.072) ?? system The assistant??s response has exceeded the maximum character limit of [500]. Please shorten your response or split it into multiple messages.

NEW QUESTION 301
A military commander would like to see the health scorecards of the troops daily and filter them based on gender and rank. Considering this data is PHI, which of the following would be the best way for the commander to view the information?

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

Answer: B

Explanation:
A password-protected dashboard is a type of web-based application that can display the health scorecards of the troops in a secure and interactive way. A password-protected dashboard can provide the following benefits for the commander:
? It can protect the PHI data from unauthorized access or disclosure by requiring a
valid username and password to log in. This can ensure that only the commander and other authorized personnel can view the information12
? It can allow the commander to filter the data based on gender and rank by using
drop-down menus, sliders, checkboxes, or other controls. This can enable the commander to customize the view and focus on the relevant data13
? It can update the data daily by connecting to a data source that refreshes
automatically or on demand. This can ensure that the commander always sees the latest and most accurate information14
? It can present the data in a visual and intuitive way by using charts, graphs, tables,
or other elements. This can help the commander to understand and analyze the data more easily and effectively1

NEW QUESTION 305
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

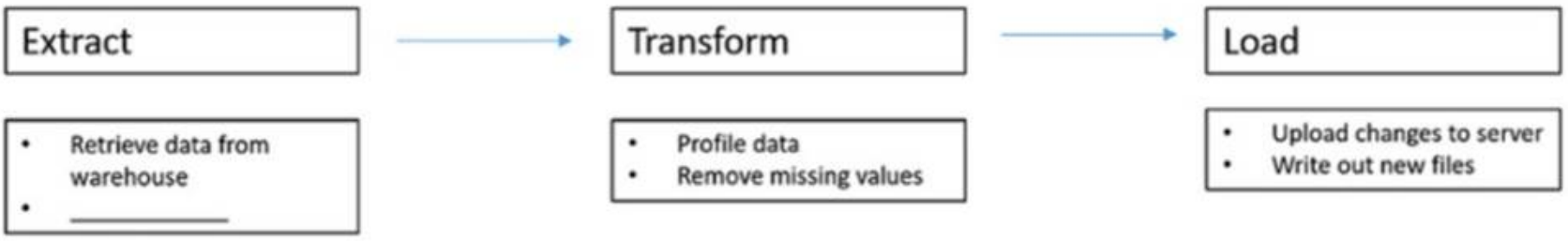
Explanation:
To compare the students?? scores, we need to standardize them by using the z-score formula, which is:
z = (x -) /
where x is the raw score, is the mean, and is the standard deviation. The z-score tells us how many standard deviations a score is above or below the mean. A higher z-score means 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 307
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 311
Given the diagram below:



Which of the following steps is missing?

A. Remove redundant data.
B. Validate the data types.
C. Connect to the data API.
D. Normalize the data.

Answer: A

Explanation:
The missing step in the Extract, Transform, Load (ETL) process is typically the cleaning step, which involves removing redundant data or deduplication. This step is crucial in the ETL process to ensure that the data loaded into the destination is accurate and not inflated by duplicate records. The other options, like validating data types and connecting to the data API, are important but do not fit into the standard ETL process steps as a cleaning operation. Normalizing the data is part of the 'Transform' step, which was already listed.

NEW QUESTION 315
An analyst is creating a resource to improve users' experience when they select specific records based on particular dates. Which of the following should the analyst use to create a resource that best meets user needs?

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

Answer: B

Explanation:
A drop-down menu is a graphical user interface element that allows users to select one option from a list of options that are hidden until the user clicks on the menu. A drop-down menu can be used to create a resource that best meets user needs when they select specific records based on particular dates, because:
? A drop-down menu can provide a predefined list of dates or date ranges that are relevant and valid for the records, such as today, yesterday, last week, last month, custom range, etc. This can help users to avoid typing errors or invalid dates in a text field, and to save time and effort in entering the dates.
? A drop-down menu can also provide a calendar or a date picker that allows users
to select a specific date or a range of dates from a graphical representation of a calendar. This can help users to visualize and compare the dates, and to easily adjust or modify their selection.
? A drop-down menu can improve the user experience by making the interface more
compact and organized, as it only shows one option at a time and hides the rest of the options until the user clicks on the menu. This can help users to focus on their selection and to avoid clutter and distraction.

NEW QUESTION 317
Which of the following is most likely to be used as a data-mining ETL tool?

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

Answer: A

NEW QUESTION 319
Which of the following database schemas features normalized dimension tables?

A. Flat
B. Snowflake
C. Hierarchical
D. Star

Answer: B

Explanation:
The correct answer is B. Snowflake.
A snowflake schema is a type of database schema that features normalized dimension tables. A database schema is a way of organizing and structuring the data in a database. A dimension table is a table that contains descriptive attributes or characteristics of the data, such as product name, category, color, etc. A normalized table is a table that follows the rules of normalization, which is a process of reducing data redundancy and improving data integrity by organizing the data into smaller and simpler tables12
A snowflake schema is a variation of the star schema, which is another type of database
schema that features denormalized dimension tables. A denormalized table is a table that does not follow the rules of normalization, and may contain redundant or duplicated data. A star schema consists of a central fact table that contains quantitative measures or facts, such as sales amount, order quantity, etc., and several dimension tables that are directly connected to the fact table. A snowflake schema differs from a star schema in that the dimension tables are further split into sub-dimension tables, creating a snowflake-like shape13
A snowflake schema has some advantages and disadvantages over a star schema. Some advantages are:
? It reduces the storage space required for the dimension tables, as it eliminates the

redundant data.
? It improves the data quality and consistency, as it avoids the update anomalies that may occur in denormalized tables.
? It allows more detailed analysis and queries, as it provides more levels of dimensions.
Some disadvantages are:
? It increases the complexity and number of joins required to retrieve the data from multiple tables, which may affect the query performance and speed.
? It reduces the readability and simplicity of the schema, as it has more tables and relationships to understand.
? It may require more maintenance and administration, as it has more tables to manage and update13

NEW QUESTION 322
Which of the following is used for calculations and pivot tables?

A. IBM SPSS
B. SAS
C. Microsoft Excel
D. Domo

Answer: C

Explanation:
This is because Microsoft Excel is a type of software application that allows users to create, edit, and analyze data in spreadsheets, which are composed of rows and columns of cells that can store various types of data, such as numbers, text, or formulas. Microsoft Excel can be used for calculations and pivot tables, which are two common features or functions in data analysis. Calculations are mathematical operations or expressions that can be performed on the data in the cells, such as addition, subtraction, multiplication, division, average, sum, etc. Pivot tables are interactive tables that can
summarize and display the data in different ways, such as by grouping, filtering, sorting, or aggregating the data based on various criteria or categories. The other software applications are not used for calculations and pivot tables. Here is why:
IBM SPSS is a type of software application that allows users to perform statistical analysis and modeling on data sets, such as regression, correlation, ANOVA, etc. IBM SPSS does not use spreadsheets or cells to store or manipulate data, but rather uses data views or variable views to display the data in rows and columns. IBM SPSS does not have pivot tables as a feature or function, but rather has output views or charts to display the results of the analysis.
SAS is a type of software application that allows users to perform data management and analysis using a programming language that consists of statements and commands. SAS does not use spreadsheets or cells to store or manipulate data, but rather uses data sets or tables that are stored in libraries or folders. SAS does not have pivot tables as a feature or function, but rather has procedures or macros that can produce summary tables or reports based on the data.
Domo is a type of software application that allows users to create and share dashboards and visualizations that display data from various sources and systems, such as databases, cloud services, or web applications. Domo does not use spreadsheets or cells to store or manipulate data, but rather uses connectors or APIs to access and integrate the data from different sources. Domo does not have pivot tables as a feature or function, but rather has cards or widgets that can show different aspects or metrics of the data.

NEW QUESTION 327
A data analyst needs to apply quality control concepts to a data set for accuracy. Which of the following is the best way to do this?

A. Standardization
B. Parameterization
C. Encryption
D. Cross-validation

Answer: D

NEW QUESTION 330
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 332
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 335
.....

Thank You for Trying Our Product

We offer two products:

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

2nd - Questons and Answers in PDF Format

DA0-001 Practice Exam Features:

* DA0-001 Questions and Answers Updated Frequently

* DA0-001 Practice Questions Verified by Expert Senior Certified Staff

* DA0-001 Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* DA0-001 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year