CompTIA.Premium.DA0-001.136q

VCEûp

Exam Code: DA0-001
Exam Name: Data+
Certification Provider: CompTIA
Corresponding Certification: CompTIA Data+
Website: https://VCEup.com/
Free Exam: https://vceup.com/exam-da0-001/

Exam A

QUESTION 1
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

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 2
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

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 3
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.

Correct Answer: C D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 4
Which of the following actions should be taken when transmitting data to mitigate the chance of a data leak occurring? (Choose two.)

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

Correct Answer: DE
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 5
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 ✔

Correct Answer: B D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 6
Which of the following BEST describes the issue in which character values are mixed with integer values in a data set column?

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

Correct Answer: A D

Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 7

Extract Transform load

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

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 8

① ② ③

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.

Correct Answer: A C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 9

A data analyst has been asked to derive a new variable labeled "Promotion_flag" based on the total quantity sold by each salesperson. Given the table below:

Store_ID	Item	Salesperson	Quantity_sold	Promotion_flag
104	Pax-2	James	1,000,300	
204	Pax-3	Paul	234,578	
304	Pax-1	Peter	2,000,432	
404	Pax-2	Esther	1,089,678	
204	Pax-3	May	126,578	
304	Pax-1	Park	200,432	
404	Pax-2	Mabel	1,089,000	

Which of the following functions would the analyst consider appropriate to flag "Yes" for every salesperson who has a number above 1,000,000 in the Quantity_sold column?
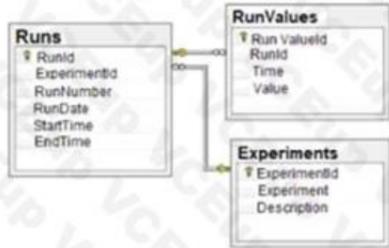
A. Date
B. Mathematical
C. Logical
D. Aggregate

Correct Answer: D C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 10
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 ✓

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 11
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 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~~ Cluster ✓

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 12
A publishing group has requested a dashboard to track submissions before publication. A key requirement is that all changes are tracked, as multiple users will be checking out documents and editing them before submissions are considered final. Which of the following is the BEST way to meet this stakeholder requirement?

A. Display the version number next to each submission on the dashboard. ✓
B. Present a data refresh date at the top of the dashboard.
C. Confirm the dashboard is adhering to the corporate style guide.
D. Use permissions to ensure users only see certain versions of the submissions.

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 13
The number of phone calls that call center receives in a day is an example of:

A. continuous data.
B. categorical data.
C. ordinal data.
D. discrete data. ✓

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 14
A data analyst is asked to create a sales report for the second-quarter 2020 board meeting, which will include a review of the business's performance through the second quarter. The board meeting will be held on July 15, 2020, after the numbers are finalized. Which of the following report types should the data analyst create?

A. Static ✓
B. Real-time
C. Self-service
D. Dynamic

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 15
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

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 16
Which of the following is the correct data type for text?

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

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 17
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.

Correct Answer: D B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 18
Which of the following is a common data analytics tool that is also used as an interpreted, high-level, general-purpose programming language?

A. SAS
B. Microsoft Power BI
C. IBM SPSS
D. Python ✔

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 19
A data analyst needs to present the results of an online marketing campaign to the marketing manager. The manager wants to see the most important KPIs and measure the return on marketing investment. Which of the following should the data analyst use to BEST communicate this information to the manager?

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

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 20
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 ✓

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 21
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.

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 22
An analyst needs to provide a chart to identify the composition between the categories of the survey response data set:

out of 100

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

Which of the following charts would be BEST to use?

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

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 23
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

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 24
Which of the following are reasons to create and maintain a data dictionary? (Choose two.)

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

Correct Answer: AB *BD*
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 25
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.

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 26
Which of the following is an example of a at flat file?

A. CSV file ✔
B. PDF file
C. JSON file
D. JPEG file

Correct Answer: D A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 27
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.

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 28
An analyst is required to run a text analysis of data that is found in articles from a digital news outlet.

Which of the following would be the BEST technique for the analyst to apply to acquire the data?

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

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 29
An analyst runs a report on a daily basis, and the number of datapoints must be validated before the data can be analyzed. The number of datapoints increases each day by approximately 20% of the total number from the day before. On a given day, the number of datapoints was 8,798. Which of the following should be the total number of datapoints on the next day?

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

(handwritten: 8798, 1758, 10556)

Correct Answer: A C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 30
An analyst has been tracking company intranet usage and has been asked to create a chat to show the most-used/most-clicked portions of a homepage that contains more than 30 links. Which of the following visualizations would BEST illustrate this information?

A. Scatter plot
B. Heat map ✓
C. Pie chart
D. Infographic

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 31
An analyst has generated a report that includes the number of months in the first two quarters of 2019 when sales exceeded $50,000:

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

(handwritten: Q1 + Q2 > 50K)

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

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

Correct Answer: D B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 32
While reviewing survey data, an analyst notices respondents entered "Jan," "January," and "01" as responses for the month of January. Which of the following steps should be taken to ensure data consistency?

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

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 33
Which of the following data cleansing issues will be fixed when a DISTINCT function is applied?

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

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 34
A county in Illinois is conducting a survey to determine the mean annual income per household. The county is 427sq mi (2.65q km). Which of the following sampling methods would MOST likely result in a representative sample?

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

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 35
Which of the following statistical methods requires two or more categorical variables?

A. Simple linear regression ✗
B. Chi-squared test ✓
C. Z-test
D. Two-sample t-test

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 36
Which of the following data manipulation techniques is an example of a logical function?

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

Correct Answer: C *D*
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 37
A sales team wants visibility of current sales numbers, pipeline, and team performance. The team would also like to see calculations of individuals' earned commissions and projected commissions based on sales, but they want that information to be kept confidential. Which of the following would be the BEST way to provide this visibility?

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

Correct Answer: B *D*
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 38
Which of the following is a characteristic of a relational database?

A. It utilizes key-value pairs.
B. It has undefined fields.
C. It is structured in nature. ✓
D. It uses minimal memory.

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 39
A data analyst is asked on the morning of April 9, 2020, to create a sales report that identifies sales year to date. The daily sales data is current through the end of the day. Which of the following date ranges should be on the report?

A. January 1, 2020 to April 1, 2020
B. January 1, 2020 to April 7, 2020
C. January 1, 2020 to April 8, 2020
D. January 1, 2020 to April 9, 2020

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 40
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

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 41
Which of the following is used for calculations and pivot tables?

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

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 42

Given the following report:

Quarterly Customer Service Report

Table 1. Frequency of Ticket Statuses

Status	Count
Reported	
In-Progress	32?
Closed	55?

Table 2. Occurrence of Target Phrases

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

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

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

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

Correct Answer: CE EF
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 43
An analyst has been asked to validate data quality. Which of the following are the BEST reasons to validate data for quality control purposes? (Choose two.)

A. Retention
B. Integrity ✔
C. Transmission
D. Consistency ✔
E. Encryption
F. Deletion

Correct Answer: BD
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 44
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

Correct Answer: B C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 45
Which of the following variable name formats would be problematic if used in the majority of data software programs?

A. First_Name_
B. FirstName
C. First_Name
D. First Name ✓

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 46
Which of the following describes the method of sampling in which elements of data are selected randomly from each of the small subgroups within a population?

A. Simple random
B. Cluster
C. Systematic ✓
D. Stratified ✓

Correct Answer: C D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 47
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

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 48
A development company is constructing a new unit in its apartment complex. The complex has the following floor plans:

Unit name	Sq. Ft.	Price	$/Sq. Ft.
Jasmine	1,000	$345,000	$345
Orchid	1,100	$425,000	$386
Azalea	1,300	$460,000	$354
Tulip	1,640	$525,000	$320
Rose	2,000		

Using the average cost per square foot of the original floor plans, which of the following should be the price of the Rose unit?

A. $640,900
B. $690,000
C. $705,200
D. $702,500 ✓

Correct Answer: ✗ D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 49
Which of the following is a control measure for preventing a data breach?

A. Data transmission
B. Data attribution
C. Data retention
D. Data encryption ✓

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 50
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.

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 51
A table in a hospital database has a column for patient height in inches and a column for patient height in centimeters. This is an example of:

A. dependent data.

B. duplicate data.
C. invalid data ✓
D. redundant data ✓

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 52
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.

Correct Answer: ~~C~~ A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 53
Which of the following BEST describes standard deviation?

A. A measure that is used to establish a relationship between two variables
B. A measure of how data is distributed
C. A measure of the amount of dispersion of a set of values ✓
D. A measure that is used to find the significant difference between variables

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 54
A data analyst was asked to create a chart that shows the relationship between study hours and exam scores for each student using the data sets in the table below:

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

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

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

Correct Answer: ~~B~~ A

Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 55
Given the table below:

Transaction ID	Date	Year	Amount
XFW25091	10/1/2019	2019	$100.00
8741STKJG	5/3/2019	2019	$50.00
TIO335AL	8/15/2018	2018	$50.00
53KJNM1C	1/4/2020	2020	$250.00

Which of the following variable types BEST describes the "Year" column?

A. Numeric
B. Date ✓
C. Alphanumeric
D. Text

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 56
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. ✓

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 57
An analysts building a monthly report for production and wants to ensure the audience is aware of its once-a-month cadence. Which of the following is the MOST important to convey that information?

A. The date of the dashboard build
B. The data refresh date ✓
C. A report summary

D. Frequently asked questions

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 58
An analyst is working with the income data of suburban families in the United States. The data set has a lot of outliers, and the analyst needs to provide a measure that represents the typical income.
Which of the following would BEST fulfill the analyst's goal?

A. Median
B. Mean
C. Mode
D. Standard deviation

Correct Answer: B̶ A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 59
Which of the following would be used to store unstructured data from different sources?

A. A data lake ✓
B. A database management system
C. A database
D. A data warehouse

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 60
An analyst is designing a dashboard to determine which site has the highest percentage of new customers. The analyst must choose an appropriate chart to include in the dashboard. The following data is available:

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

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

A. Include a bar chart using the site and the percentage of new customers data. ✓

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

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 61
A cereal manufacturer wants to determine whether the sugar content of its cereal has increased over the years. Which of the following is the appropriate descriptive statistic to use?

A. Frequency
B. Percent change ✓
C. Variance
D. Mean

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 62
The process of performing initial investigations on data to spot outliers, discover patterns, and test assumptions with statistical insight and graphical visualization is called:

A. a t-test.
B. a performance analysis.
C. an exploratory data analysis. ✓
D. a link analysis.

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 63
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

Correct Answer: AE   AE
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 64

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

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

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 65
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

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 66
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

Correct Answer: AC
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 67
Which of the following will MOST likely be streamed live?

A. Machine data ✓ IoT
B. Key-value pairs
C. Delimited rows

D. Flat files

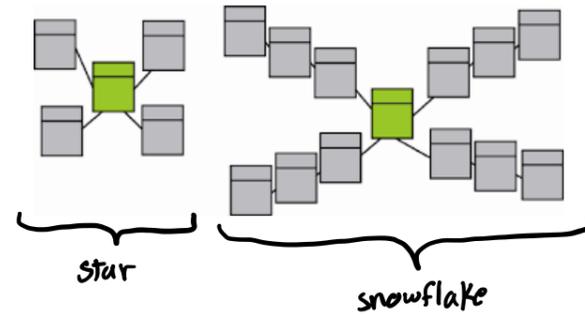**Correct Answer:** ~~B~~ A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 68
A database consists of one fact table that is composed of multiple dimensions. Depending on the dimension, each one can be represented by a denormalized table or multiple normalized tables. This structure is an example of a:

A. transactional schema.
B. star schema.
C. non-relational schema.
D. snowflake schema. ✓

Correct Answer: ~~A~~ D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

star   snowflake

QUESTION 69
A data analyst is designing a dashboard that will provide a story of sales and determine which site is providing the highest sales volume per customer. The analyst must choose an appropriate chart to include in the dashboard. The following data is available:

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

Which of the following types of charts should be considered?

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

Correct Answer: ~~X~~ D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 70
An analyst needs to conduct a quick analysis. Which of the following is the FIRST step the analyst should perform with the data?

A. Conduct an exploratory analysis and use descriptive statistics. ✓
B. Conduct a trend analysis and use a scatter chart.

C. Conduct a link analysis and illustrate the connection points.
D. Conduct an initial analysis and use a Pareto chart.

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 71
A data analyst has been asked to create a sales report that calculates the rolling 12-month average for sales. If the report will be published on November 1, 2020, which of the following months shouts the report cover?

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

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 72
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

Correct Answer: ~~D~~ A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 73
A data analyst needs to create a weekly recurring report on sales performance and distribute it to all sales managers. Which of the following would be the BEST method to automate and ensure successful delivery for this task?

A. Use scheduled report delivery. ✓
B. Implement subscription access delivery.
C. Print out a copy.
D. Upload the report to the server.

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 74
Which of the following is an example of a discrete variable?

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

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 75
Which of the following data types would a telephone number formatted as XXX-XXX-XXXX be considered?

A. Numeric
B. Date
C. Float
D. Text ✓

Correct Answer: ~~A~~ D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 76
The director of operations at a power company needs data to help identify where company resources should be allocated in order to monitor activity for outages and restoration of power in the entire state. Specifically, the director wants to see the following:
* County outages
* Status
* Overall trend of outages
INSTRUCTIONS:
Please, select each visualization to fit the appropriate space on the dashboard and choose an appropriate color scheme. Once you have selected all visualizations, please, select the appropriate titles and labels, if applicable. Titles and labels

may be used more than once.
If at any time you would like to bring back the initial state of the simulation, please click the Reset All button.

Dashboard Editor

Theme Options

Select a title

Select the Appropriate Visualization Depicting **County Outages**

Select a title

Select the Appropriate Visualization Depicting **Status**

Select a title

Select the Appropriate Visualization Depicting the **Number of Outages for the Quarter**

Dashboard Editor

Theme Options

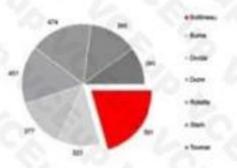**Select a dashboard title**

- Select a dashboard title
- Power Outages Enterprise-wide
- Power Outages Over Time
- EmPOWER Me! Dashboard
- Outages in Sheridan County

Select a title
- Select a title
- Power Outages
- Counties of Outages
- Geographic Area of Outages
- Outages per Month
- Power Outages in the Quarter
- Closed Incidents
- Status of Incidents by County

Select the Appropriate Visualization Depicting **County Outages**

Select a label
- Percentage of Outages
- Percentage of Incidents
- Status of Incidents
- Frequency
- Count of Incidents
- Number of Outages
- Rate of Outages

Select a label
- Year
- Month
- Status
- Number
- County
- Date
- Counts
- Time

Select a label
- Percentage of Outages
- Percentage of Incidents
- Status of Incidents
- Frequency
- Count of Incidents
- Number of Outages
- Rate of Outages

Select a label
- Year
- Month
- Status
- Number
- County
- Date
- Counts
- Time

Select a label
- Percentage of Outages
- Percentage of Incidents
- Status of Incidents
- Frequency
- Count of Incidents
- Number of Outages
- Rate of Outages

Select a label
- Year
- Month
- Status
- Number
- County
- Date
- Counts
- Time

Select a label
- Percentage of Outages
- Percentage of Incidents
- Status of Incidents
- Frequency
- Count of Incidents
- Number of Outages
- Rate of Outages

Select a label
- Year
- Month
- Status
- Number
- County
- Date
- Counts
- Time

Select a title
- Select a title
- Power Outages in the Quarter
- Power Outages
- Closed Incidents
- Geographic Area of Outages
- Status of Incidents by County
- Outages per Month
- Counties of Outages

A. Power outages
B. Power

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 77
Q3 2020 has just ended, and now a data analyst needs to create an ad-hoc sales report that demonstrates how well the Q3 2020 promotion went versus last year's Q3 promotion.
Which of the following date parameters should the analyst use?

A. 2019 vs. YTD 2020
B. Q3 2019 vs. Q3 2020
C. YTD 2019 vs. YTD 2020
D. Q4 2019 vs. Q3 2020

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 78
A data analyst has been asked to create an ad-hoc sales report for the Chief Executive Officer (CEO).
Which of the following should be included in the report?

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

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 79
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.

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:
Data encryption is a way of translating data from plaintext (unencrypted) to ciphertext (encrypted).
Users can access encrypted data with an encryption key and decrypted data with a decryption key.

QUESTION 80
Which of the following is an example of a discrete data type?

A. 8in (20cm)
B. 5 kids ✓
C. 2.5mi (4km)
D. 10.7lbs (4.9kg)

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 81
Which of the following contains alphanumeric values?

A. 10.1?²
B. 13.6
C. 1347
D. A3J7 ✓

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 82
A junior web developer is developing a new application where users can upload short videos. The first task is to create a homepage that shows the headline "Upload Your Short Videos" and a clickable button that says "upload now".
Which of the following HTML commands would help the developer to complete the task successfully?

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

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:
The correct answer is: Upload Your Short Videos upload now The two tags are used to define HTML headings. defines the most important heading. defines the least important heading.
Note: Only use one per page - this should represent the main heading/subject for the whole page.
The tag defines a clickable button.

QUESTION 83
A web developer wants to ensure that malicious users can't type SQL statements when they asked for input, like their username/userid.
Which of the following query optimization techniques would effectively prevent SQL Injection attacks?

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

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:
The correct answer is D: Parametrization. Parameterized SQL queries allow you to place parameters in an SQL query instead of a constant value. A parameter takes a value only when the query is executed, allowing the query to be reused with different values and purposes. Parameterized SQL statements are available in some analysis clients, and are also available through the Historian SDK.
For example, you could create the following conditional SQL query, which contains a parameter for the collector's name: SELECT* FROM ExamsDigest WHERE coursename=? ORDER BY tagname SQL Injection is best prevented through the use of parameterized queries.

QUESTION 84
Consider the following dataset which contains information about houses that are for sale:

```
sonery=# select * from melb limit 5;
 houseid |      address       |      regionname       | type | rooms |    date    | distance |  price
---------+--------------------+-----------------------+------+-------+------------+----------+---------
       1 | 85 Turner St       | Northern Metropolitan | h    |     2 | 2016-03-12 |      2.5 | 1.48e+06
       2 | 25 Bloomburg St    | Northern Metropolitan | h    |     2 | 2016-04-02 |      2.5 | 1.035e+06
       3 | 5 Charles St       | Northern Metropolitan | h    |     3 | 2017-04-03 |      2.5 | 1.465e+06
       4 | 40 Federation La   | Northern Metropolitan | h    |     3 | 2017-04-03 |      2.5 |   850000
       5 | 55a Park St        | Northern Metropolitan | h    |     4 | 2016-04-06 |      2.5 | 1.6e+06
(5 rows)
```

Which of the following string manipulation commands will combine the address and region name columns to create a full address? full_address------------------------ 85 Turner St, Northern Metropolitan 25 Bloomburg St, Northern Metropolitan 5 Charles St, Northern Metropolitan 40 Federation La, Northern Metropolitan 55a Park St, Northern Metropolitan

A. SELECT CONCAT(address, ' , ' , regionname) AS full_address FROM melb LIMIT 5;
B. SELECT CONCAT(address, '-' , regionname) AS full_address FROM melb LIMIT 5;
C. SELECT CONCAT(regionname, ' , ' , address) AS full_address FROM melb LIMIT 5
D. SELECT CONCAT(regionname, '-' , address) AS full_address FROM melb LIMIT 5;

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:
The correct answer is A: SELECT CONCAT(address, ' , ' , regionname) AS full_address FROM melb LIMIT 5; String manipulation (or string handling) is the process of changing, parsing, splicing, pasting, or analyzing strings. SQL is used for managing data in a relational database. The CONCAT() function adds two or more strings together. Syntax CONCAT(stringl, string2,... string_n) Parameter Values Parameter Description stringl, string2, string_n Required. The strings to add together.

QUESTION 85
The ACME Corporation hired an analyst to detect data quality issues in their excel documents. Which of the following are the most common issues? (Select TWO)

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

Correct Answer: DE
Section: (none)
Explanation

Explanation/Reference:
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.

QUESTION 86
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

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
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.

QUESTION 87
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

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
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

QUESTION 88
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

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
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.

QUESTION 89
A data analyst wants to create "Income Categories" that would be calculated based on the existing variable "Income". The "Income Categories" would be as follows:
Income category 1: less than $1.
Income category 2: more than $1 and less than $20,000.
Income category 3: more than $20,001 and less than $40,000.
Income category 4: more than $40,001.
Which of the following data manipulation techniques should the data analyst use to create "Income Categories"?

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

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:
The correct answer is B: Derived variables Derived variables are variables that you create by calculating or categorizing variables that already exist in your data set.
Data merge is incorrect. Data merging is the process of combining two or more data sets into a single data set. Data blending is incorrect.
Data blending involves pulling data from different sources and creating a single, unique, dataset for visualization and analysis.
Data append is incorrect. A data append is a process that involves adding new data elements to an existing database.

QUESTION 90
Angela is aggregating data from CRM system with data from an employee system.
While performing an initial quality check, she realizes that her employee ID is not associated with her identifier in the CRM system.
What kind of issues is Angela facing?
Choose the best answer.

A. ETL process.
B. Record linkage.

C. ELT process.
D. System integration.

Correct Answer: B
Section: (none)
Explanation

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

QUESTION 91
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.

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
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.

QUESTION 92
Amanda needs to create a dashboard that will draw information from many other data sources and present it to business leaders.
Which one of the following tools is least likely to meet her needs?

A. QuickSight.
B. Tableau.
C. Power BI.
D. SPSS Modeler.

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:
SPSS Modeler.
QuickSight, Tableau, and Power BI are all powerful analytics and reporting tools that can pull data from a variety of sources. SPSS Modeler is a powerful predictive analytics platform that is designed to bring predictive intelligence to decisions made by individuals, groups, systems and your enterprise.

QUESTION 93
Daniel is using the structured Query language to work with data stored in relational database.
He would like to add several new rows to a database table.
What command should he use?

A. SELECT.
B. ALTER.
C. INSERT.
D. UPDATE.

Correct Answer: C
Section: (none)

Explanation

Explanation/Reference:
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.

QUESTION 94
Jhon is working on an ELT process that sources data from six different source systems.
Looking at the source data, he finds that data about the sample people exists in two of six systems.
What does he have to make sure he checks for in his ELT process?
Choose the best answer.

A. Duplicate Data.

B. Redundant Data.

C. Invalid Data.

D. Missing Data.

Correct Answer: C A
Section: (none)
Explanation

Explanation/Reference:
Explanation:
Duplicate Data.
While invalid, redundant, or missing data are all valid concerns, data about people exists in two of the six systems. As such, Jhon needs to account for duplicate data issues.

QUESTION 95
Samantha needs to share a list of her organization's top 50 customers with the VP of sales.
She would like to include the name of the customer, the business they represent, their contact information, and their total sales over the past year.
The VP does not have any specialized analytics skills or software but would like to make some personal notes on the dataset.
What would be the best tool for Samantha to use to share this information?

A. Power BI.
B. Microsoft Excel.
C. Minitab.
D. SAS.

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:
Microsoft Excel.
This scenario presents a very simple use case where the business leader needs a dataset in an easyto- access form and will not be performing any detailed analysis.
A simple spreadsheet, such as Microsoft Excel, would be the best tool for this job.
There is no need to use a statistical analysis package, such as SAS or Minitab, as this would likely confuse the VP without adding any value. The same is true of an integrated analytics suite, such as Power BI.

QUESTION 96
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.

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
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.

QUESTION 97
Analytics reports should follow corporate style guidelines.

A. True.
B. False.

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 98
Which one of the following is a measure of dispersion?

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

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 99
Which one of the following in NOT a common data integration tool?

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

Correct Answer: A
Section: (none)
Explanation

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

QUESTION 100

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

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

Correct Answer: A
Section: (none)
Explanation

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

QUESTION 101
What would be an example of an acceptable form of primary identification for the Data+ exam?

A. Passport. ✓
B. School ID card.
C. Employee ID card.
D. Credit card with photo and signature.

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 102
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. ✓

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
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.

QUESTION 103
What SQL command is used to delete an entire table from a database?

A. DROP. ✓
B. MODIFY.
C. DELETE.
D. ALTER.

Correct Answer: A
Section: (none)

Explanation

Explanation/Reference:
Explanation:

QUESTION 104
Which one of the following programming languages is specifically designed for use in analytics applications?

A. Python.
B. R ✓
C. C++
D. Java.

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 105
What role in a data governance is typically responsible for day-to-day oversight of data use?

A. Data processors.
B. Data custodians
C. Data owners.
D. Data stewards. ✓

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 106
What category of data stewardship work is focused on ensuring that the organization respects the wishes of data subjects?

A. Data quality.
B. Data privacy. ✓
C. Data security.
D. Regulatory compliance.

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:
Data privacy defines who has access to data, while data protection provides tools and policies to actually restrict access to the data. Compliance regulations help ensure that user's privacy requests are carried out by companies, and companies are responsible to take measures to protect private user data.
Why is data privacy important?
When data that should be kept private gets in the wrong hands, bad things can happen. A data breach at a government agency can, for example, put top secret information in the hands of an enemy state. A breach at a corporation can put proprietary data in the hands of a competitor.

QUESTION 107
You are working with a dataset and need to swap the values in rows with those in columns.
What action do you need to perform?

A. Recording

B. Filtering.
C. Aggregation.
D. Transposition. ✓

Correct Answer: D
Section: (none)
Explanation

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

QUESTION 108
When analyzing the values of two variables, you decide to convert both variables so they are on a scale of 0 to 1.
What term describes this action?

A. Filtering.
B. Normalization. ✓
C. Transposition.
D. Aggregation.

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:
Normalization is the process of reorganizing data in a database so that it meets two basic requirements: There is no redundancy of data, all data is stored in only one place. Data dependencies are logical, all related data items are stored together.
Put simply, data normalization ensures that your data looks, reads, and can be utilized the same way across all of the records in your customer database. This is done by standardizing the formats of specific fields and records within your customer database.

QUESTION 109
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.

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
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.

QUESTION 110
Emma is working in a data warehouse and finds a finance fact table links to an organization dimension, which in turn links to a currency dimension that not linked to the fact table.
What type of design pattern is the data warehouse using?

A. Star.
B. Sun.
C. Snowflake. ✓

D. Comet.

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:
Correct answer C. Snowflake.
Since the dimension links to a dimension that isn't connected to the fact table, it must be a Snowflake, with a Star, all dimensions link directly to the fact table, Sun and Comet are not data warehouse design patterns.

QUESTION 111
Encryption is a mechanism for protecting data.
When should encryption be applied to data?
Choose the best answer.

A. When data is at rest.
B. When data is at rest or in transit.
C. When data is in transit.
D. When data is at rest, unless you are using local storage.

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:
Correct answer B. When data is at rest or in transit.
To provide maximum protection, encrypt data both in transit and at rest.

QUESTION 112
What subset of Structured Query Language (SQL) is used to add, remove, modify, or retrieve the information stored within a relational database?

A. DDL.
B. DSL.
C. DQL.
D. DML.

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:
Correct answer D. DML.
The Data Manipulation Language (DML) is used to work with the data stored in a database. DML includes the SELECT, INSERT, UPDATE, and DELETE commands.
The Data Definition Language (DDL) contains the commands used to create and structure a relational database. It includes the CREATE, ALTER, and DROP commands.
DDL and DML are the only two sublanguages of SQL.

QUESTION 113
Which of the following roles is responsible for ensuring an organization's data quality, security, privacy, and regulatory compliance?

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

Correct Answer: B
Section: (none)
Explanation

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

QUESTION 114
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.

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
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.

QUESTION 115
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.

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 116
You would like to measure how well an organization is achieving its goals.
What type of analysis should you perform?

A. Performance analysis. ✓
B. Outlier analysis.
C. Predictive analysis.
D. Trend analysis.

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:
Performance analysis is the technique of studying or comparing the performance of a specific situation in contrast to the aim and yet executed. In Human Resource, performance analysis can help to review an employee's contribution towards a project or assignment, which they allotted him or her.

QUESTION 117
Which one of the following would not normally be considered a summary statistic?

Just measure how far a data point is from the mean

A. z-score. ✓
B. Mean.
C. Variance. } *Summary Statistic*
D. Standard deviation.

Correct Answer: A
Section: (none)
Explanation

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

QUESTION 118
What analytics suite is offered by Microsoft and directly integrates with SQL Server Databases?

A. Qlik.
B. Power BI. ✓
C. Domo.
D. Dataroma.

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:
Power BI is a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights. Your data may be an Excel spreadsheet, or a collection of cloud-based and on-premises hybrid data warehouses.

QUESTION 119
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.

Correct Answer: B A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 120
What R package makes it easy to work with dates?

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

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

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

QUESTION 121
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. ✓

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
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.

QUESTION 122
Z →A
Which one of the following values will appear first if they are sorted in descending order?

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

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 123
Which one the following is not considered an aggregate function?

A. SUM
B. MIN
C. SELECT ✓
D. MAX

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 124
You are working with a dataset and want to change the names of categories that you used for different types of books.
What term best describes this action?

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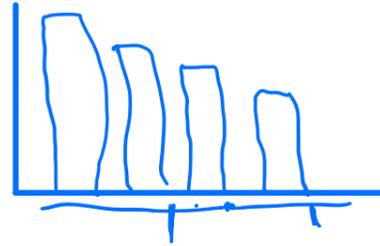C. Aggregating.
D. Filtering.

Correct Answer: A

Explanation/Reference:
Explanation:

QUESTION 125
Joseph is interpreting a left skewed distribution of test scores. Joe scored at the mean, Alfonso scored at the median, and gaby scored and the end of the tail.
Who had the highest score?

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

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:
A left skewed distribution typically has a mean less than the median, with the tail representing the lowest score.

QUESTION 126
Kelly wants to get feedback on the final draft of a strategic report that has taken her six months to develop.
What can she do to get prevent confusion as see seeks feedback before publishing the report?
Choose the best answer.

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

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:
While Kelly needs feedback from the appropriate stakeholders, doing so without a watermark could lead them to believe the report they receive is the final product.

QUESTION 127
Which of the following is the correct data type for text?

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

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 128
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.

Correct Answer: ~~B~~ B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 129
Which of the following is a common data analytics tool that is also used as an interpreted, high-level, general-purpose programming language?

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

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 130
A data analyst needs to present the results of an online marketing campaign to the marketing manager. The manager wants to see the most important KPIs and measure the return on marketing investment. Which of the following should the data analyst use to BEST communicate this information to the manager?

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

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 131
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 ✓

Correct Answer: D
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 132
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.

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 133
An analyst needs to provide a chart to identify the composition between the categories of the survey response data set:

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

Which of the following charts would be BEST to use?

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

Correct Answer: B
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 134
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

Correct Answer: A
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 135
Which of the following are reasons to create and maintain a data dictionary? (Choose two.)

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

Correct Answer: AB BD
Section: (none)
Explanation

Explanation/Reference:
Explanation:

QUESTION 136
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.

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:
Explanation: