Question 1

Which of following is a non-relational database?

Options:

A- Neo4j

B- SQLite

C- MySQL

D- PostgreSQL

Answer:

A

Explanation:

Neo4j is a type of non-relational database that uses a graph model to store dat

a. A graph database is a database that represents data as nodes and edges, where nodes are entities and edges are relationships between them. A graph database can store complex and diverse data that is not easily structured in tables.A graph database can also perform fast and efficient queries on the data by traversing the connections between the nodes

Question 2

Question Type: **MultipleChoice**

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?

Options:

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

Question 3

Question Type: **MultipleChoice**

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?

Options:

A- Order number. salesperson. date shipped, recipient address, and price

B- Item name, salesperson. recipient address, shipping cost. and date shipped

C- Item number, item name, salesperson. date sold. and price

D- Item name. salesperson. price. shipping cost. 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.
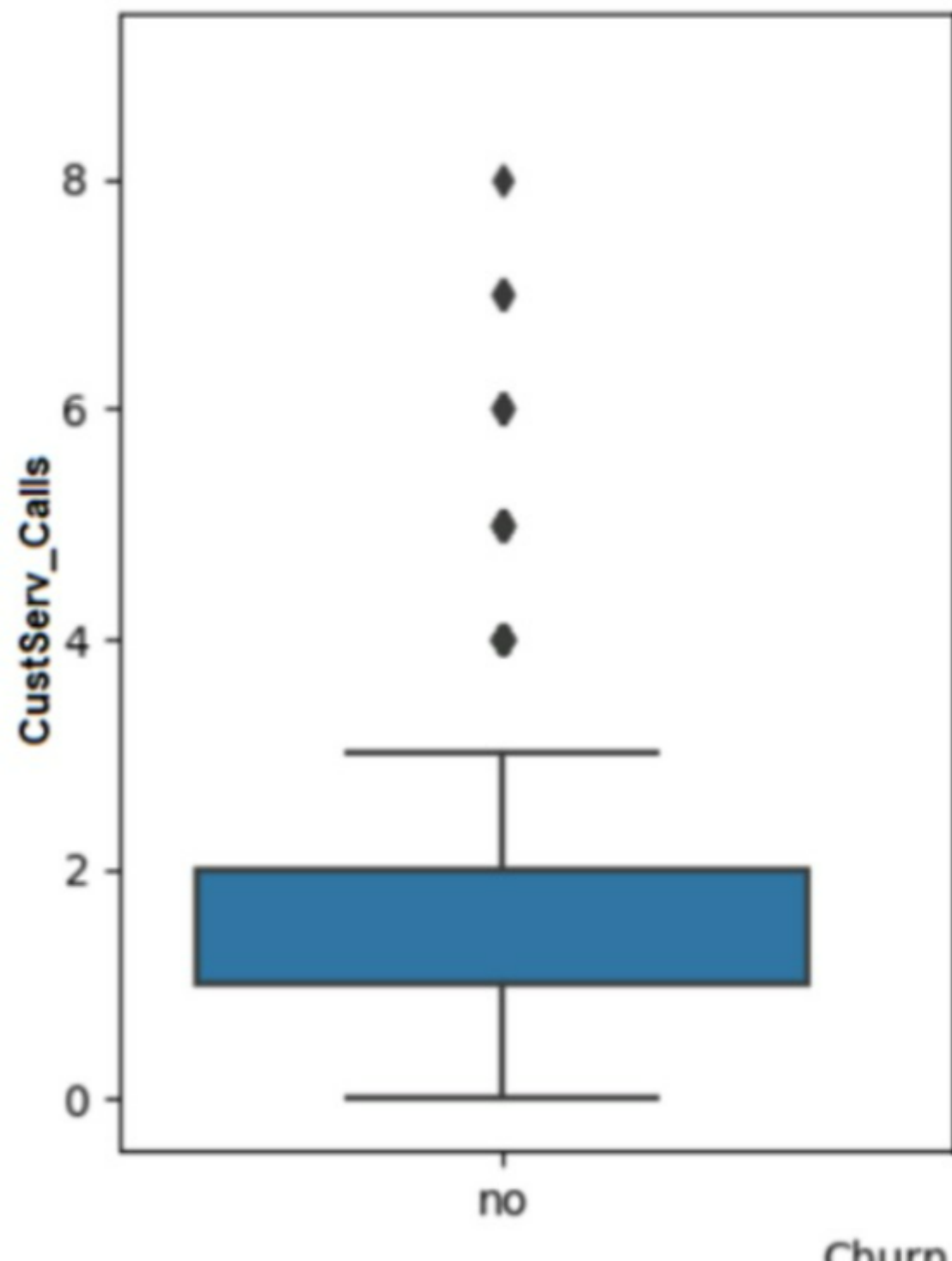
Question 4

Refer to the exhibit.

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.

Question 5

Question Type: MultipleChoice

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?

Options:

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

Question 6

Question Type: MultipleChoice

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?

Options:

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

Question 7

Question Type: **MultipleChoice**

After completing web scraping, which of the following file formats needs to be parsed?

Options:

A- .html

B- .txt

C- .csv

D- .tsv

Answer:

A

Explanation:

The correct answer is .html.

Short explanation: Web scraping is the process of extracting data from websites by parsing the HTML code of the web pages. HTML stands for HyperText Markup Language and it is the standard markup language for creating web pages and web applications. HTML files have the extension .html and they contain tags, elements, attributes, and content that define the structure and appearance of a web page. Web scraping tools need to parse the HTML files to extract the relevant data from the web pages12

Question 8

Question Type: MultipleChoice

A data analyst has been asked to create a daily manufacturing report for the floor manager Which of the following metrics should be included in the report?

Options:

A- Tons of steel produced per hour

B- Annual sales budget

C- End-of-day stock price

D- Daily corporate employee count

Answer:

A

Question 9

Question Type: MultipleChoice

Which of the following best describes an exploratory analysis?

Options:

A- Involves the use of descriptive statistics to understand observations

B- Involves analysis of exploring data sets for performance tracking

C- Involves the testing of specific hypotheses

D- Involves the use of arithmetic algebra to determine the distribution

Answer:

A

Explanation:

A) Involves the use of descriptive statistics to understand observations.

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

Question 10

Question Type: **MultipleChoice**

Which of the following is an example of structured data?

Options:

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.