PassQueen
a.k.a dumpsqueen.com

CompTIA Data+ Exam (2026)

CompTIA DA0-002

Version Demo

Total Demo Questions: 14

Total Premium Questions: 142

Buy Premium PDF

https://passqueen.com

support@passqueen.com

Topic Break Down

Topic	No. of Questions
Topic 1, Data Concepts and Environments	23
Topic 2, Data Acquisition and Preparation	29
Topic 3, Data Analysis	24
Topic 4, Data Visualization and Reporting	35
Topic 5, Data Governance, Quality, and Controls	31
Total	142

QUESTION NO: 1

A monthly revenue dashboard shows a 35% increase for one product category after a new sales extract is loaded. The business has not identified any corresponding change in demand or pricing. Which of the following actions should the analyst take first to validate the result? (Select two).

A. Check the new extract for duplicate transaction records.

B. Confirm that the extract has the expected transaction-level grain.

C. Replace the dashboard bar chart with a line chart.

D. Calculate a forecast using the new revenue total.

E. Remove product-category filters from the dashboard.

ANSWER: A B

Explanation:

The analyst should check for duplicate transaction records and confirm the grain of the new extract. Duplicate rows can directly overstate an aggregated revenue measure. A grain mismatch can also inflate results, such as when transaction-level data is joined to a table containing multiple rows per product or customer.

Changing the visualization or forecasting future sales does not validate the current figure. Comparing the result with a budget may identify an anomaly, but it does not determine whether the loaded data has caused the increase.

QUESTION NO: 2

A data analyst needs to get an accurate idea of how data components are automated. Which of the following types of documentation should the analyst review first?

A. Data flow diagram

B. Data explainability report

C. Data dictionary

D. Data lineage

ANSWER: A

Explanation:

Data flow diagram is the best documentation to review first because it presents a visual model of how data moves between sources, processes, data stores, and destinations. For an analyst trying to understand automation, this view makes the operational path visible: where data enters a workflow, which processing activities handle it, what systems exchange it, and where the resulting data is delivered or retained. Automated pipelines commonly include scheduled extracts, transformations, validations, loads, and integrations; a data flow diagram provides a concise, system-level picture of those activities and their relationships. This makes it especially useful as an initial artifact before examining detailed implementation material such as job configurations, orchestration schedules, scripts, or interface specifications. A well-maintained diagram also supports effective communication with data engineers and business stakeholders by establishing a shared understanding of the workflow boundaries and handoffs. IBM describes a data flow diagram as a visual representation of the flow of information through a system, including processes and data stores, which directly supports this need: IBM: What is a data flow diagram?. CompTIA's published exam-objective resources also emphasize interpreting data environments and data processes as part of certification preparation: CompTIA Exam Objectives.

QUESTION NO: 3

A data breach occurs at a company. Which of the following actions should be taken?

A. Make an announcement on social media so customers are aware as soon as possible.

B. Tell the company management team and then tell regulatory agencies.

C. Keep the incident a secret until the issue is resolved.

D. Inform the entire IT sector, but ask for discretion.

ANSWER: B

Explanation:

Tell the company management team and then tell regulatory agencies is the appropriate action because an effective breach response begins with escalation through the organization's established incident-response and data-governance processes. Management must be informed promptly so authorized stakeholders—such as legal counsel, privacy personnel, security leadership, communications staff, and executive leadership—can assess the incident, contain it, preserve evidence, determine the types and scope of affected data, and coordinate required notifications. Regulatory obligations differ by jurisdiction and the nature of the data involved, but many privacy laws require notification to a supervisory authority within a defined time frame. For example, the GDPR generally requires notification to the relevant supervisory authority without undue delay and, where feasible, within 72 hours after becoming aware of a qualifying personal-data breach. A coordinated management-led response also helps ensure that communications are accurate, timely, documented, and approved by the appropriate parties. CompTIA Data+ data-governance objectives emphasize applying privacy and governance practices to data handling; escalation and legally appropriate reporting are essential parts of those practices. See the GDPR breach-notification requirements at GDPR Article 33 and incident-response guidance from CISA.

QUESTION NO: 4 - (SIMULATION)

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

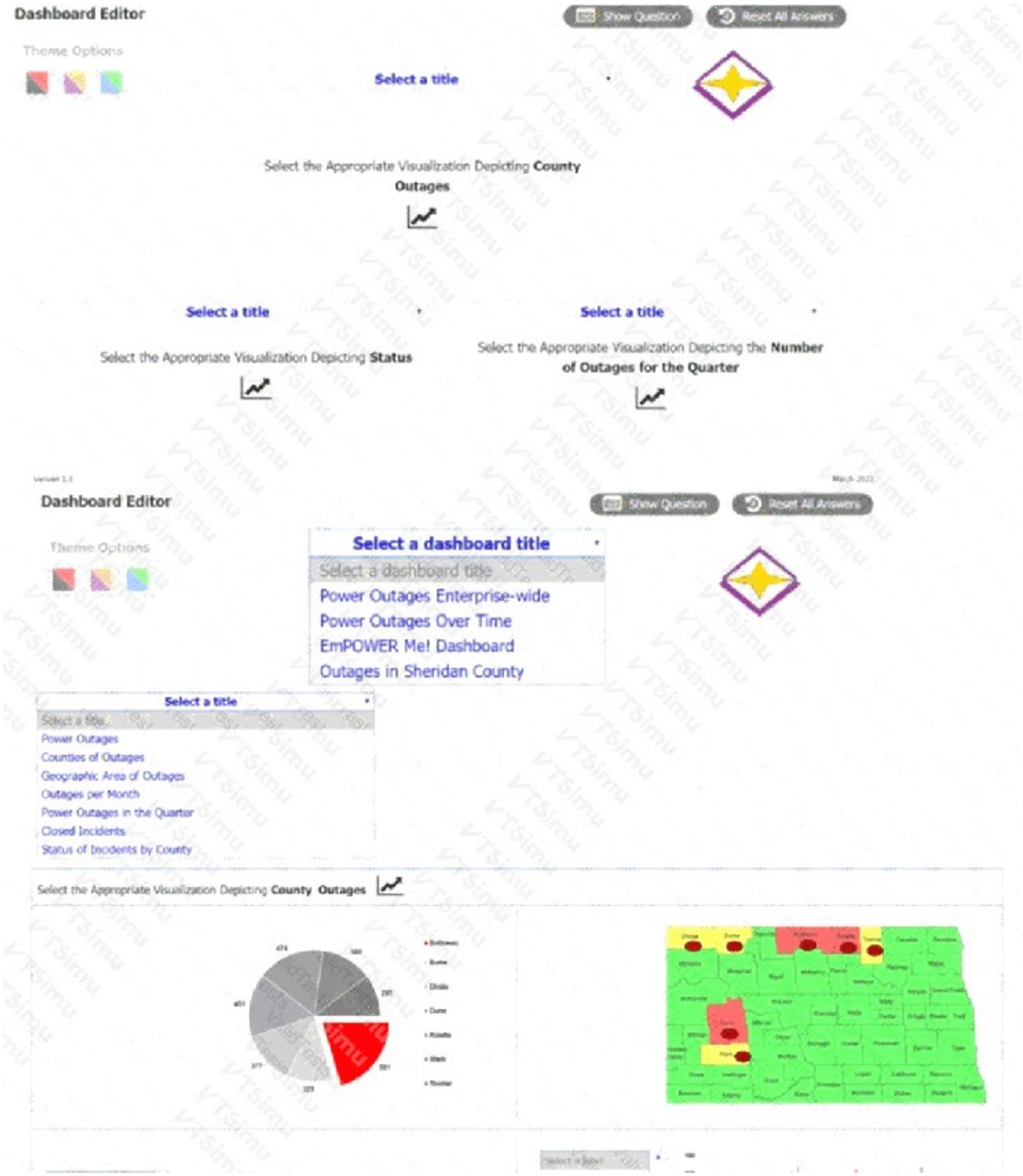
* Status

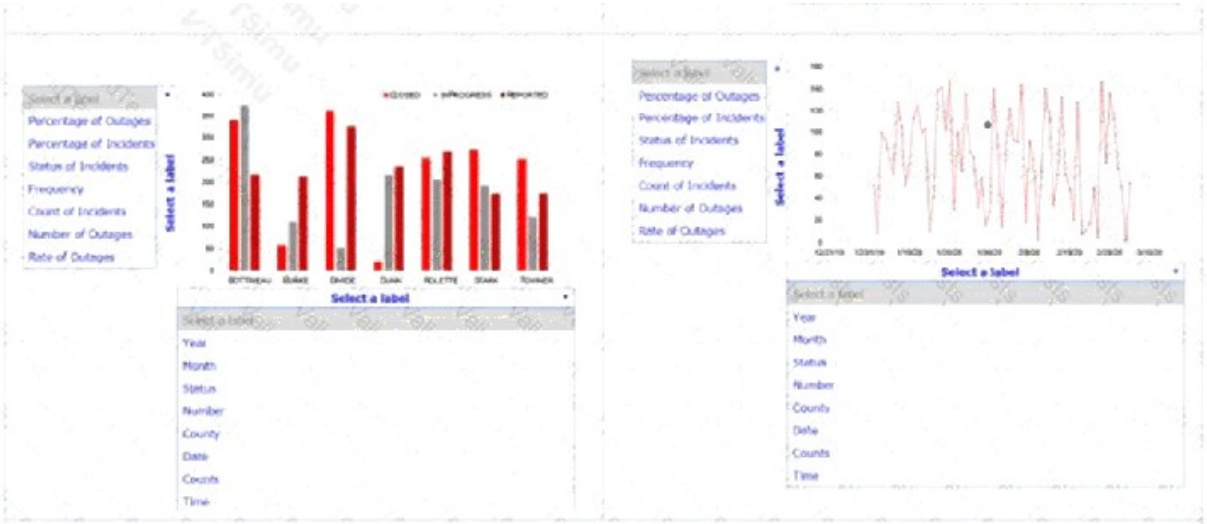
* Overall trend of outages

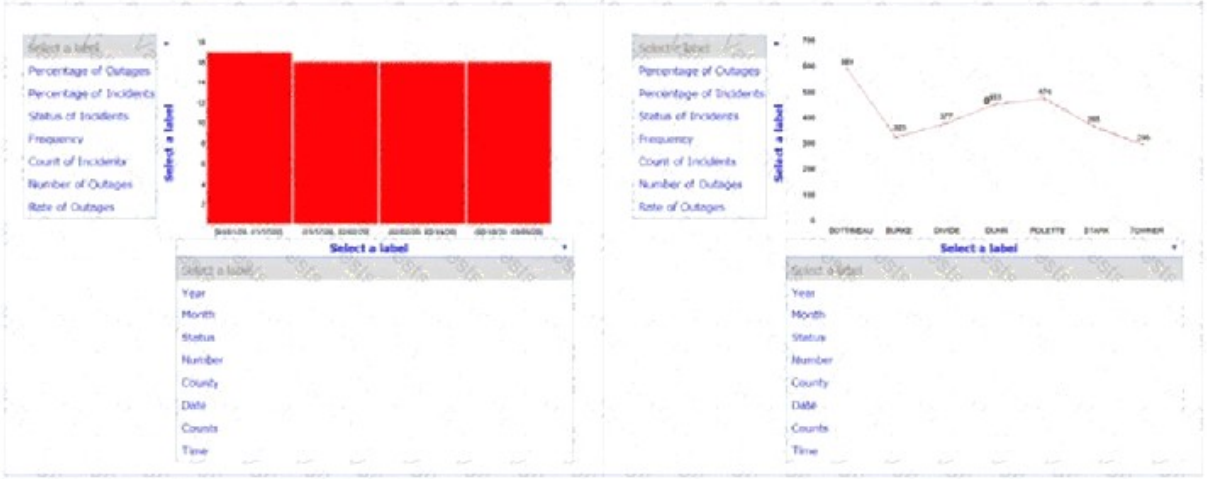
INSTRUCTIONS:

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
Show Question
Reset All Answers
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
Show Question
Reset All Answers
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

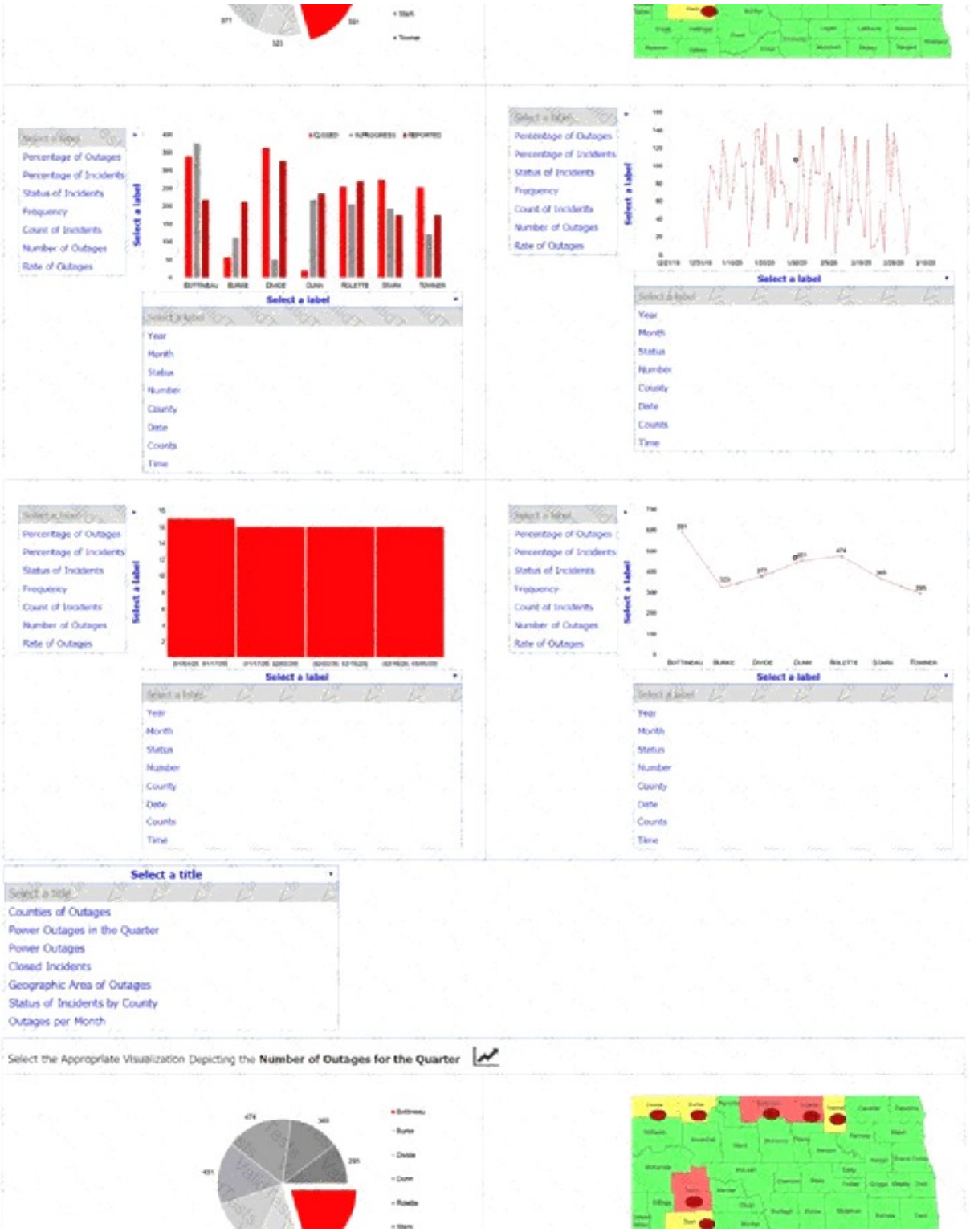

Select a label
Percentage of Outages
Percentage of Incidents
Status of Incidents
Frequency
Count of Incidents
Number of Outages
Rate of Outages
Select a label
Year
Month
Status
Number
County
Date
Counts
Time

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


Select the Appropriate Visualization Depicting Status


Select a label
Percentage of Outages
Percentage of Incidents
Status of Incidents
Frequency
Count of Incidents
Number of Outages
Rate of Outages
Year
Month
Status
Number
County
Date
Counts
Time
Select a title
Counties of Outages
Power Outages in the Quarter
Power Outages
Closed Incidents
Geographic Area of Outages
Status of Incidents by County
Outages per Month
Select the Appropriate Visualization Depicting the Number of Outages for the Quarter



ANSWER: See the explanation for the answer

Explanation:

Configure the dashboard as follows:

This combination gives operations a geographic county-level outage view, a concise status breakdown, and a time-based outage trend for the quarter.

QUESTION NO: 5

A company has a document that includes the names of key metrics and the standard for how those metrics are calculated company-wide. Which of the following describes this documentation?

A. Data dictionary

B. Data explainability report

C. Data lineage

D. Data flow diagram

ANSWER: A

Explanation:

Data dictionary is the correct documentation because it provides a shared, governed definition of data items used throughout an organization. For business metrics, this includes the metric name, its business meaning, the calculation formula or logic, relevant source fields, units of measure, and applicable rules. Establishing these definitions in a centrally maintained dictionary ensures that teams calculate and interpret measures consistently. For example, a metric such as customer retention can have a single approved definition specifying the population, time period, exclusions, and formula, rather than allowing each department to derive a different result from the same underlying data. This supports data governance, improves reporting quality, and creates a common vocabulary for analysts and business stakeholders. A data dictionary can also document technical metadata, such as field names and data types, alongside business definitions and calculation standards. CompTIA Data+ expects candidates to understand documentation that defines data elements and promotes consistent use of organizational data. See CompTIA's Data+ certification overview and IBM's overview of data dictionaries.

QUESTION NO: 6

An organization plans to publish a public dashboard showing employee turnover trends by department. The underlying dataset includes employee identifiers and some departments have very few employees. Which of the following actions should the analyst take before publishing the dashboard? (Select two).

A. Encrypt the source dataset while retaining all employee-level fields in the public dashboard

B. Remove direct employee identifiers from the published data

C. Publish all row-level employee records to maximize transparency

D. Add employee birth dates to distinguish similar records

E. Aggregate or suppress results for very small departments

ANSWER: B E

Explanation:

The analyst should remove direct employee identifiers because individual identities are not needed to communicate department-level turnover trends. The analyst should also aggregate or suppress results for very small groups to reduce the risk that viewers could infer information about an individual employee from a small population.

Encrypting the source dataset does not prevent sensitive information from being displayed in a public dashboard. Publishing all row-level records creates unnecessary privacy exposure. Adding more detailed employee attributes would increase, rather than reduce, the risk of re-identification.

QUESTION NO: 7

A table contains several rows of cellular numbers with call timestamps, call durations, called numbers, and carriers of the called number. Which of the following allows a data analyst to sort the cellular numbers based on the carriers of the called numbers and include the total call durations?

A. SELECT cellular_number, called_number_carrier, SUM(call_duration) FROM calls GROUP BY cellular_number ORDER BY called_number_carrier;

B. SELECT cellular_number, SUM(call_duration) FROM calls GROUP BY call_duration ORDERBY called_number_carrier;

C. SELECT cellular_number, called_number_carrier, SUM(call_duration) FROM calls GROUP BY cellular_number, called_number_carrier ORDER BY called_number_carrier;

D. SELECT cellular_number, called_number_carrier, SUM(call_duration) FROM calls GROUP BY call_duration ORDER BY called_number_carrier;

ANSWER: C

Explanation:

`SELECT cellular_number, called_number_carrier, SUM(call_duration) FROM calls GROUP BY cellular_number, called_number_carrier ORDER BY called_number_carrier;` is correct because it performs both the required aggregation and sorting at the appropriate levels of detail. Each result row represents a unique combination of `cellular_number` and `called_number_carrier`. Grouping by those two fields makes it possible to calculate `SUM(call_duration)` as the total duration for calls placed by each cellular number to each carrier. This preserves the carrier information needed for the requested report while producing a meaningful aggregate rather than an individual call-level result.

The `ORDER BY called_number_carrier` clause then sorts the grouped output according to the carrier of the called number. SQL processing conceptually groups and aggregates the source rows before applying the final ordering, so the displayed results show the total call duration for each cellular-number/carrier combination in carrier order. This is a standard analytical SQL pattern: select categorical dimensions, group by every nonaggregated selected dimension, calculate an aggregate such as `SUM`, and order the final result set by the reporting dimension. PostgreSQL's documentation describes `GROUP BY` as condensing rows sharing grouped values and permits aggregate expressions such as `SUM`; its `ORDER BY` documentation describes sorting the returned rows by the specified expression. See PostgreSQL GROUP BY documentation and PostgreSQL ORDER BY documentation.

QUESTION NO: 8

An analyst must provide a third-party consultant with a customer dataset for a short-term churn analysis. The consultant needs purchase history and support-case counts but does not need names, email addresses, or phone numbers. Which of the following actions best support appropriate data handling? (Select two).

A. Remove direct identifiers before creating the extract.

B. Limit access to the approved consultant for the defined engagement period.

C. Include all customer fields in case additional analysis is requested later.

D. Publish the extract to a shared location accessible to all analysts.

E. Create a permanent account so the consultant can access future extracts.

ANSWER: A B

Explanation:

Removing direct identifiers applies data minimization by providing only the information required for the stated analysis. Restricting access to the approved consultant and defined engagement period applies least-privilege access and limits unnecessary exposure of customer information.

Including all available fields increases privacy risk without supporting the requirement. Publishing the extract to a shared location and using a permanent access account do not provide appropriate control over sensitive data.

QUESTION NO: 9

A regional director needs to compare actual quarterly revenue with quarterly targets for 12 sales regions. The director must be able to identify regions that exceeded or missed their targets without focusing on a time-series trend. Which of the following visualizations is the most appropriate?

A. Grouped bar chart

B. Line chart

C. Pie chart

D. Scatter plot

ANSWER: A

Explanation:

A grouped bar chart is the most appropriate visualization because it enables direct comparison of two measures, actual revenue and target revenue, across a discrete set of regions. Each region can display two adjacent bars, making positive and negative variances easy to identify.

A line chart is better suited to showing change over an ordered sequence, such as time. A pie chart emphasizes part-to-whole composition, and a scatter plot is intended to show the relationship between two quantitative variables rather than performance against a target by category.

QUESTION NO: 10

A data analyst is designing an executive dashboard that will be used by employees with varying visual abilities. Which of the following practices should the analyst use to improve accessibility? (Select two).

A. Use red and green as the only indicators of performance status

B. Use text labels or symbols in addition to color

C. Use the smallest possible font to display more metrics

D. Select colors with sufficient contrast between visual elements

E. Add decorative images to each dashboard page

ANSWER: B D

Explanation:

Using clear text labels in addition to color ensures that users do not need to distinguish colors alone to interpret a result. Selecting colors with sufficient contrast improves readability for users with low vision and helps make labels, values, and visual elements easier to distinguish.

Relying only on red and green to communicate status can exclude users with color-vision deficiencies. Small text reduces readability, and automatically refreshing a dashboard without allowing users to pause or control movement can create usability issues. Decorative images do not improve the accessibility of the analytical content.

QUESTION NO: 11

An organization stores employee compensation data for workforce planning. The data governance team must enforce least-privilege access and ensure the data is not retained beyond the approved business period. Which of the following controls should the organization implement? (Select two).

A. Apply role-based access controls to the dataset

B. Publish the dataset in a shared location for all employees

C. Define and enforce a retention and disposal schedule

D. Allow all managers to download the complete dataset

E. Retain the data indefinitely to support future analysis

ANSWER: A C

Explanation:

Role-based access control supports least privilege by granting access according to job responsibilities rather than making compensation data broadly available. A documented retention and disposal schedule ensures that records are retained only for the approved period and are disposed of according to organizational policy.

Publishing the data in a shared location conflicts with the sensitive nature of compensation information. Allowing all managers to download the complete dataset exceeds least-privilege requirements. Keeping data indefinitely does not meet a defined retention requirement and can increase privacy and governance risk.

QUESTION NO: 12

A data analyst needs to join together a table data source and web API data source using Python. Which of the following is the best way to accomplish this task?

A. Convert the data from the API and database to a varchar format and convert them to pandas DataFrames that are then merged together.

B. Convert the data from the API and database to a JSON format and convert them to pandas DataFrames that are then merged together.

C. Convert the data from the API and database to a TXT format and convert them to pandas DataFrames that are then merged together.

D. Convert the data from the API and database to a string format and convert them to pandas DataFrames that are then merged together.

ANSWER: B

Explanation:

Convert the data from the API and database to a JSON format and convert them to pandas DataFrames that are then merged together. This is the best choice because JSON is a common structured interchange format for web services and can represent records, fields, nested objects, and arrays in a form Python can parse reliably. In a typical workflow, the analyst retrieves JSON from the web API, loads or transforms the tabular source into a comparable structured representation, and creates pandas DataFrames with consistent column names and compatible data types for the join key. pandas provides functions for constructing DataFrames from JSON-like data and for combining datasets with database-style joins. The `merge()` operation can join the two DataFrames on one or more shared columns, while supporting common join types such as inner, left, right, and outer joins. This approach preserves tabular structure and makes it practical to inspect, clean, standardize, validate, and reconcile the data before merging. See the pandas documentation for DataFrame.merge() and json_normalize().

QUESTION NO: 13

A database administrator needs to implement security triggers for an organization ' s user information database. Which of the following data classifications is the administrator most likely using? (Select two).

A. Public

B. Open

C. Sensitive

D. Non-Sensitive

E. Private

F. Encrypted

ANSWER: C E

Explanation:

Sensitive and **Private** are the appropriate classifications for an organization's user-information database. User records commonly include personally identifiable information, such as names, email addresses, phone numbers, account identifiers, authentication-related attributes, and possibly role or access information. Exposure or improper modification of these records can create privacy, security, and operational risks, so the data warrants classification that drives stronger handling requirements.

Classifying the information as sensitive establishes that it requires safeguards based on the potential impact of unauthorized disclosure, alteration, or access. A private classification identifies information intended for restricted organizational use and, particularly where it contains PII, supports controls that limit access to authorized personnel and systems. These classifications provide the business context for security triggers, such as alerting on unusual account changes, logging access to user records, or detecting unauthorized privilege modifications.

CompTIA's Data+ objectives include data-governance concepts and data classifications as mechanisms for applying appropriate management and protection practices. NIST also explains that PII requires protection because it can distinguish or trace an individual's identity; see NIST's definition of PII. Additional guidance on safeguarding sensitive data is available from CISA.

QUESTION NO: 14

While interacting with a data set, a data analyst learns that copies of the data are distributed across several data repositories around the globe. Which of the following concepts best describes this practice?

A. Redundancy

B. High availability

C. Replication

D. Duplication

ANSWER: C

Explanation:

Replication is the deliberate process of maintaining synchronized copies of data in multiple repositories, servers, regions, or geographic locations. In this scenario, the defining detail is that copies of the same data set are distributed globally. A replication process transfers data from a source to one or more target locations and may operate synchronously, where changes are committed across locations together, or asynchronously, where updates reach replicas after a short delay. Data analysts should recognize replication because it can affect the freshness, consistency, and provenance of the data they query. For example, a report generated from a geographically local replica might not yet contain the most recent updates if asynchronous replication is used. Geographic replication is commonly implemented to improve access performance for users in different regions and to support resilient data operations if a location becomes unavailable. The exact behavior, including latency and consistency guarantees, depends on the platform and replication configuration. Microsoft describes geo-replication as creating readable secondary copies of data in another region, while AWS documentation describes cross-Region replication as automatically copying data to a destination in a different AWS Region. See Microsoft Learn: Active geo-replication and AWS: Replicating objects.