PassQueen
a.k.a dumpsqueen.com

CompTIA Data+ Certification Exam

CompTIA DA0-001

Version Demo

Total Demo Questions: 46

Total Premium Questions: 464

Buy Premium PDF

https://passqueen.com

support@passqueen.com

Topic Break Down

Topic	No. of Questions
Topic 1, Data Concepts and Environments	113
Topic 2, Data Mining	7
Topic 3, Data Analysis	163
Topic 4, Visualization	73
Topic 5, Data Governance, Quality, and Controls	105
Topic 6, Mix Questions	3
Total	464

QUESTION NO: 1

An employer needs to maintain adequate office staffing during the winter and wants to track storm data. Which of the following data collection methods should the employer use?

A. Web scraping

B. Public databases

C. Observations

D. Weather surveys

ANSWER: B

Explanation:

Public databases are the appropriate data collection method because weather and storm information is already collected, maintained, and made available by authoritative public agencies. For a staffing decision, the employer needs timely, geographically relevant information such as winter-storm warnings, forecasts, observed snowfall, ice conditions, and historical climate patterns. Established public weather databases provide these data in standardized formats and are generally updated by organizations with meteorological expertise and formal quality-control processes. This makes the information practical for defining weather-related staffing policies, monitoring conditions before a workday, and recognizing recurring seasonal disruption risks.

For example, the U.S. National Weather Service publishes active alerts, forecasts, and local weather information that can support operational decisions about office closures, remote-work activation, or staggered staffing. The National Centers for Environmental Information also provides access to historical weather and climate records that can help an employer evaluate past winter conditions and plan for likely staffing impacts. Using these sources allows the employer to obtain relevant external data without having to create a new primary-data collection effort. See the National Weather Service and NOAA's National Centers for Environmental Information.

QUESTION NO: 2

A data analyst is using a two-tailed, independent t-test to determine whether the type of stretching, dynamic or static, has any influence on a dancer's flexibility. Which of the following is the alternative hypothesis?

A. A dancer's flexibility is improved through static stretching.

B. The change in a dancer's flexibility is not equal to zero.

C. There is a difference in a dancer's flexibility between static and dynamic stretching.

D. The means of the static and dynamic stretching groups do not differ from each other.

ANSWER: C

Explanation:

The correct alternative hypothesis is: "There is a difference in a dancer's flexibility between static and dynamic stretching." An independent-samples t-test evaluates whether the population means of two separate, independently sampled groups differ. Here, the groups are dancers who use static stretching and dancers who use dynamic stretching, while flexibility is the measured outcome. Because the test is two-tailed, the hypothesis is nondirectional: it tests for any difference in mean flexibility, whether static stretching produces a higher mean, a lower mean, or simply a mean that is not equal to that of dynamic stretching. In statistical notation, this is expressed as H_1: μstatic ≠ μdynamic. This hypothesis directly matches the analyst's question about whether stretching type has an influence on flexibility. The test compares the observed difference between the group means against the amount of variability expected from sampling. If the resulting evidence is sufficiently strong under the selected significance level, the analyst rejects the no-difference hypothesis and concludes that flexibility differs by stretching type. See the NIST Engineering Statistics Handbook's discussion of two-sample t-tests and Scribbr's overview of independent-samples t-tests.

QUESTION NO: 3

A data analyst has been asked to organize the table below in the following ways:

By sales from high to low -

By state in alphabetic order -

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

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

A. Conditional formatting

B. Grouping

C. Filtering

D. Sorting

ANSWER: D

Explanation:

Sorting is the correct function because it changes the order of complete records in a table according to the values in one or more columns. The analyst can apply a descending sort to the Sales column so that the largest sales values appear first, then apply an alphabetical ascending sort to the State column as an additional sort level. This is commonly called a multi-level, multi-column, or custom sort. It preserves the relationship between each sales value, its state, and every other field in the associated row, rather than treating a column as isolated data. A spreadsheet or analytics tool typically lets the analyst choose the priority of the sort keys: Sales can be the primary key in descending order, with State as a secondary ascending key for records with the same sales value; alternatively, State can be primary if the requested report layout requires alphabetic state groupings first. Sorting is a foundational data-preparation activity because orderly data makes comparisons, ranking, review, and subsequent analysis more reliable. Microsoft's documentation describes sorting data by one or multiple columns and selecting ascending or descending order: Sort data in a range or table. CompTIA lists data manipulation and analysis skills within the Data+ exam objectives: CompTIA exam objectives.

QUESTION NO: 4

A data scientist wants to see which products make the most money and which products attract the most customer purchasing interest in their company.

Which of the following data manipulation techniques would he use to obtain this information?

A. Data append

B. Data blending

C. Normalize data

D. Data merge

ANSWER: B

Explanation:

Data blending is the appropriate technique because the analyst needs a unified view of product performance drawn from related business measures that may reside in separate datasets or systems. For example, revenue or profit information may be held in product, invoice, or financial records, while customer purchasing interest may be represented by transaction counts, orders, browsing activity, wish lists, or other customer-engagement data. Blending brings these complementary sources together around a shared business attribute, such as product ID, SKU, or product name, so the analyst can compare products using both financial contribution and demand-related measures.

Once the data is blended, the analyst can aggregate revenue by product and calculate interest indicators such as number of purchases or customers purchasing each product. This supports identifying products that generate the greatest monetary value, as well as products with high customer interest that may warrant inventory, marketing, or pricing attention. CompTIA's Data+ exam objectives include data manipulation and transformation concepts used to prepare datasets for analysis, including combining data appropriately to produce useful analytical views. See the CompTIA exam objectives resource and CompTIA Data+ certification page.

QUESTION NO: 5

Which of the following data types best describe 4Ac1? (Select two).

A. Alphanumeric

B. Symbolic

C. Numeric

D. Float

E. Boolean

F. String

ANSWER: A F

Explanation:

Alphanumeric and **String** best describe `4Ac1`. Alphanumeric identifies the composition of the value: it contains both alphabetic characters (`A` and `c`) and numeric characters (`4` and `1`). This description is especially useful in data analysis when identifying fields such as product codes, account identifiers, postal codes, or other labels that combine letters and digits.

String identifies how the value should be represented and handled in a dataset or programming environment. A string is a sequence of characters, and character sequences may include letters, digits, spaces, or symbols. Although `4Ac1` contains digits, it must be stored as text because the embedded letters mean it cannot be evaluated as a numeric quantity. Treating this value as a string preserves all four characters and supports operations appropriate for identifiers, such as matching, filtering, joining, validation, and text parsing. In data work, recognizing that a value's visible characters and its storage type can be described differently is important: "alphanumeric" describes the character pattern, while "string" describes the data representation. See the Python documentation for text strings and the IBM documentation on character strings.

QUESTION NO: 6

A quality assurance manager is examining tolerances in Internet of Things sensors. Which of the following is the best measure for the manager to calculate?

A. Standard deviation

B. Quartile range

C. Median

D. Mean

ANSWER: A

Explanation:

Standard deviation is the best measure because sensor tolerance concerns the amount by which repeated readings vary around an expected or average value. A quality assurance manager can use standard deviation to quantify that typical spread in the sensor measurements in the same units as the readings themselves. A small standard deviation indicates that readings are tightly clustered and the sensor is producing consistent results; a larger standard deviation signals greater variability and a higher risk that measurements will fall outside the permitted tolerance. This makes it useful for comparing sensor consistency across devices, production batches, environmental conditions, or calibration periods. When the expected measurement and acceptable tolerance limits are known, the manager can evaluate whether the observed variation is compatible with the required level of quality and investigate unusual process variation. Standard deviation is also a foundational descriptive statistic for quality-control and statistical-process-monitoring work. The NIST Engineering Statistics Handbook describes standard deviation as a measure of data dispersion and its role in assessing process variation: NIST: Standard Deviation. CompTIA's Data+ objectives likewise include applying descriptive statistical methods in data analysis: CompTIA Exam Objectives.

QUESTION NO: 7 - (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

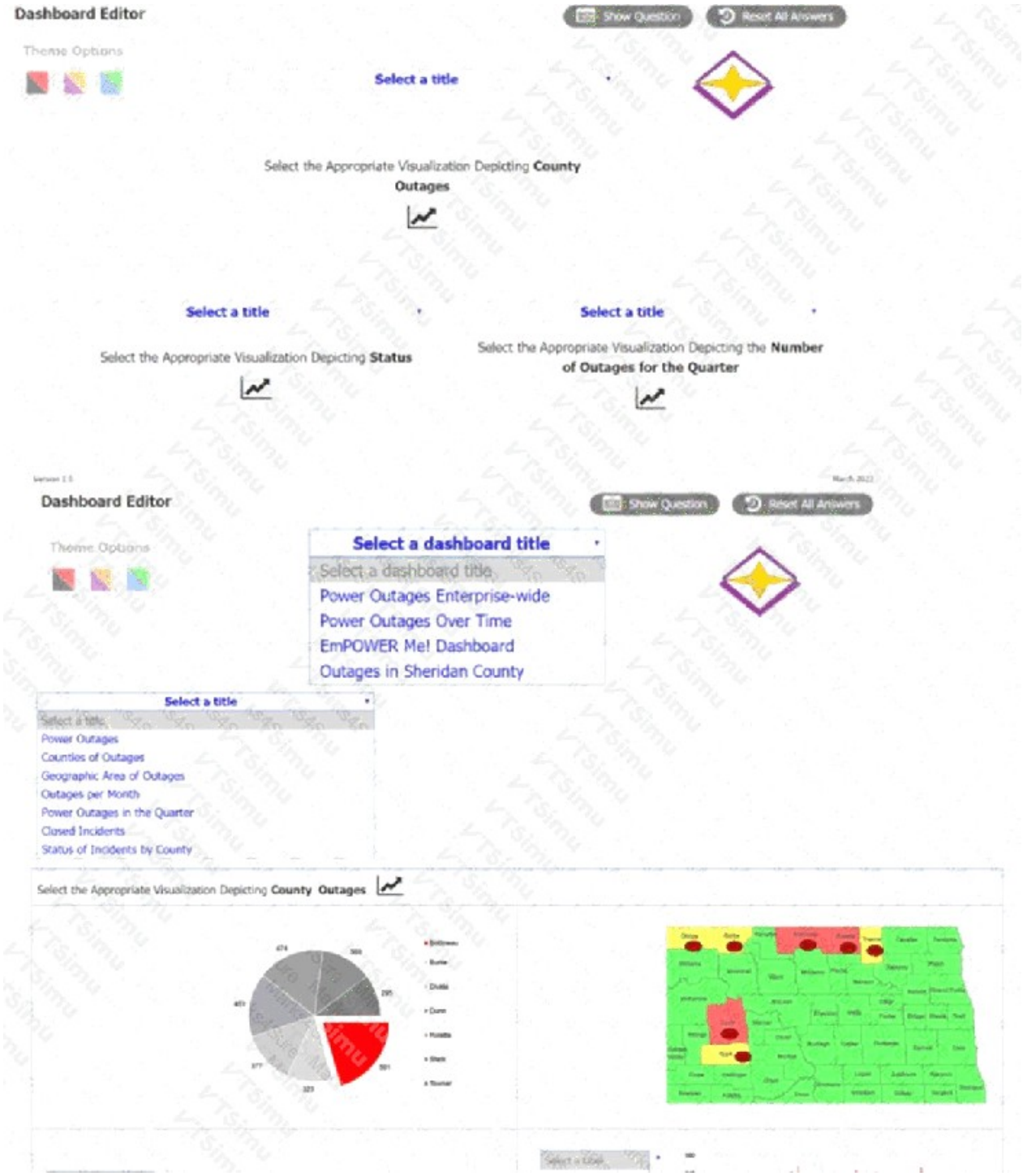
* Status

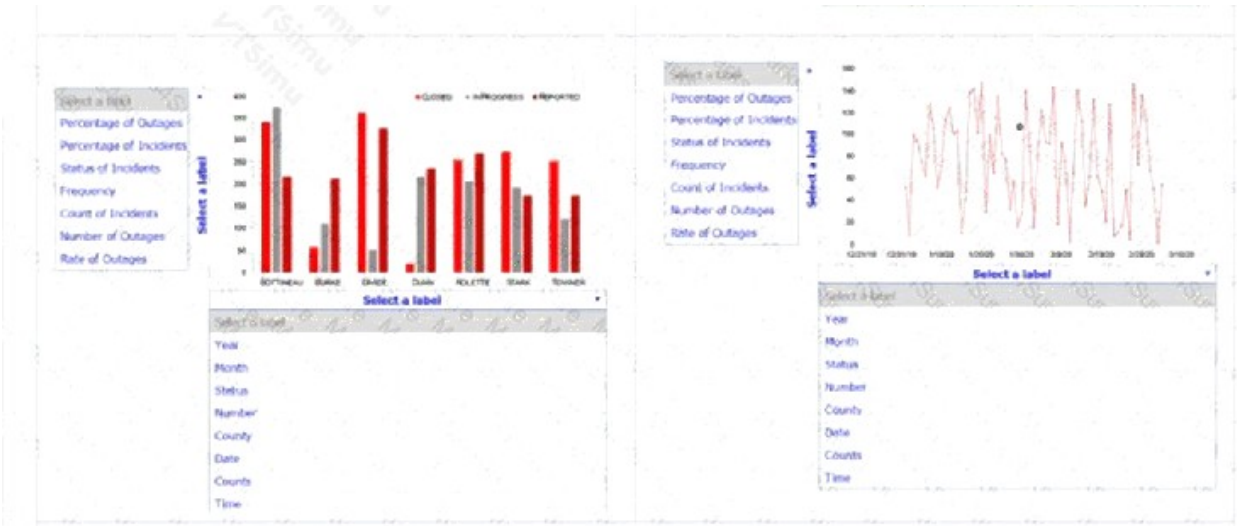
* Overall trend of outages

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.

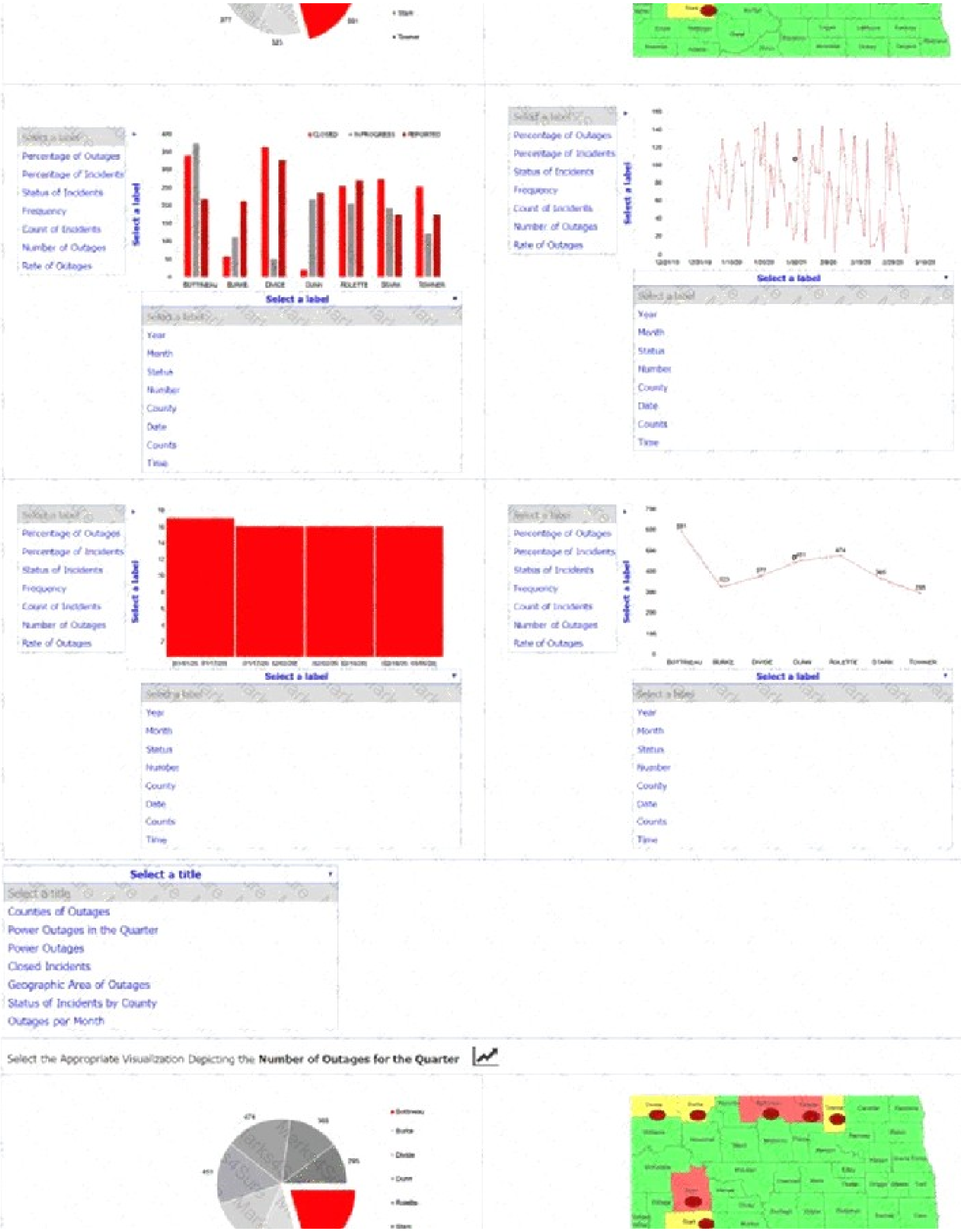

Dashboard Editor
Show Question
Reset All Answers
Theme Options
Select a title
Select the Appropriate Visualization Depicting County Outages
Select a title
Select the Appropriate Visualization Depicting Status
Select a title
Select the Appropriate Visualization Depicting the Number of Outages for the 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
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

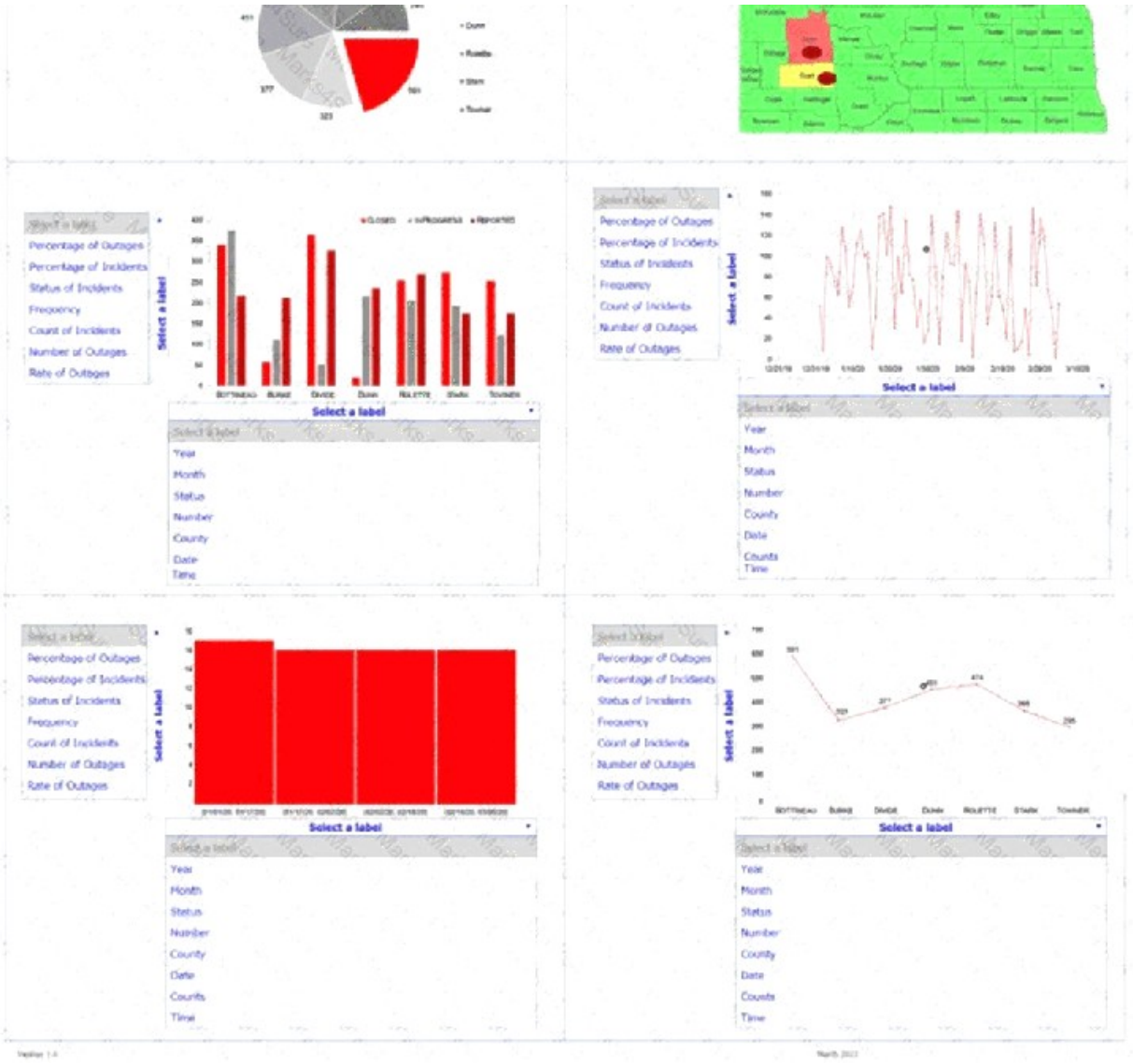

Select a label
Percentage of Outages
Percentage of Incidents
Status of Incidents
Frequency
Count of Incidents
Number of Outages
Rate of Outages
Select a label
Year
Month
Status
Number
County
Date
Counts
Time


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 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:

Dashboard title: Select `Power Outages Entire State`.

Theme: Use the red outage-oriented theme (the leftmost red/pink theme option).

This provides geographic county-level outage visibility, a categorical status breakdown, and a time-based outage trend for the quarter, enabling statewide resource-allocation decisions.

QUESTION NO: 8

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

A. To improve data acquisition

B. To remember specifics about data fields

C. To specify user groups for databases

D. To provide continuity through personnel turnover

E. To confine breaches of PHI data

F. To reduce processing power requirements

ANSWER: B D

Explanation:

A data dictionary is maintained documentation describing the structure and meaning of data elements, such as field names, definitions, permitted values, formats, data types, and business rules. “To remember specifics about data fields” is therefore a core purpose: it provides a reliable, shared record of what each field represents and how it should be interpreted. This reduces reliance on individual memory and helps analysts, database administrators, developers, and business users apply fields consistently when querying, reporting, integrating, or modifying data.

“To provide continuity through personnel turnover” is also a key reason to maintain a data dictionary. Data knowledge often resides with the people who designed a system or have worked with it over time. Documenting field-level metadata preserves that organizational knowledge when those people change roles or leave. New personnel can use the dictionary to understand the available data, its intended use, and relevant constraints without having to rediscover definitions through trial and error. This supports more dependable long-term data governance and more consistent use of organizational data assets. See IBM’s overview of a data dictionary and Microsoft’s guidance on managing and discovering data metadata through the Microsoft Purview Data Catalog.

QUESTION NO: 9

An analyst is preparing a dataset that contains customer email addresses and payment card numbers for a third-party vendor that only needs to evaluate purchase patterns. Which of the following actions would BEST support data minimization? (Choose two.)

A. Remove payment card numbers.

B. Replace email addresses with a nonidentifying customer ID.

C. Include full mailing addresses for all customers.

D. Add customers' dates of birth to the dataset.

E. Provide the vendor with production-system credentials.

ANSWER: A B

Explanation:

Removing payment card numbers supports data minimization because the vendor does not need those highly sensitive values to evaluate purchase patterns. Excluding unnecessary sensitive fields reduces the potential impact of an unauthorized disclosure.

Replacing email addresses with a nonidentifying customer ID also supports the vendor's analytical purpose while reducing direct exposure of personally identifiable information. The vendor can still analyze repeat purchases and customer-level behavior using the replacement identifier. Data minimization involves limiting data collection and sharing to information that is relevant and necessary for a defined purpose. See the FTC privacy and security guidance.

QUESTION NO: 10

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

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

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

A. Append

B. Merge

C. Concatenate

D. Delimit

ANSWER: D

Explanation:

Delimit is correct because the Genre column contains values that must be separated into distinct genre entries before their frequencies can be counted accurately. A single movie can be assigned more than one genre, commonly stored in one field using a separator such as a comma, slash, or pipe. The analyst must identify that separator and split, or delimit, the combined value so each genre becomes independently available for analysis. For example, a value containing both Action and Adventure needs to contribute to the count for Action and also to the count for Adventure, rather than being treated as one combined, unique category. Once the values are delimited into individual genres, the analyst can group the normalized genre values, calculate counts, and identify the genre with the highest frequency. This is a data-cleaning and transformation step that supports valid aggregation of categorical data, a core Data+ skill area. Microsoft's guidance on splitting a column by a delimiter illustrates this transformation process: Split a column by delimiter. CompTIA's Data+ certification overview also identifies data mining, manipulation, and visualization as key analysis competencies: CompTIA Data+.

QUESTION NO: 11

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

A. Mean

B. Minimum

C. Mode

D. Variance

E. Correlation

F. Maximum

ANSWER: A C

Explanation:

Mean and mode are measures of central tendency because each describes a typical or central value within a data set. The mean is calculated by adding all observed values and dividing the sum by the number of observations. It provides a single average that represents the data set's overall numerical center and is especially useful for quantitative data when extreme values do not unduly influence the result. The mode identifies the value or category that occurs most frequently. It is useful for finding the most common observation and can be applied to categorical data as well as numerical data. A data set may have one mode, multiple modes when several values share the highest frequency, or no mode when no value repeats. In descriptive analysis, these statistics help an analyst summarize a distribution concisely and communicate common patterns before conducting further analysis. CompTIA Data+ objectives include using descriptive statistical methods to summarize data and identify patterns. For further background, see the Encyclopaedia Britannica overview of measures of central tendency and the NIST Engineering Statistics Handbook discussion of measures of location.

QUESTION NO: 12

An analyst is creating a visualization to compare monthly revenue for four product lines over one year. Which of the following visualization techniques would be appropriate? (Choose two.)

A. Line chart

B. Grouped bar chart

C. Pie chart

D. Scatter plot

E. Word cloud

ANSWER: A B

Explanation:

A **line chart** is appropriate because it displays changes in revenue across ordered monthly periods and allows the viewer to compare trends for multiple product lines. A **grouped bar chart** is also appropriate because it supports direct comparisons of product-line revenue within each month.

A pie chart is not well suited to showing changes across many time periods, and a scatter plot is generally used to evaluate the relationship between two numerical variables rather than compare monthly category totals. See the Data to Viz chart-selection guidance.

QUESTION NO: 13

An analyst needs to summarize the number of people in Chicago in 2022 using the following set of data:

Name	City	Year	Grade
Chloe	Chicago	2022	A
Blake	Chicago	2023	B
Carter	Chicago	2022	A
Kim	Detroit	2021	C

Which of the following steps should the analyst use to provide results? (Select two).

A. Aggregation

B. Sorting

C. Filtering

D. Indexing

E. Cleaning

F. Replacing

ANSWER: A C

Explanation:

Filtering and **Aggregation** are the required steps to produce the requested result. Filtering first restricts the source data to only records that meet both stated conditions: Chicago as the location and 2022 as the time period. Applying these conditions ensures that records for other cities or years do not affect the result. In a spreadsheet, database query, or business-intelligence tool, this can be implemented with criteria such as a city field equal to Chicago and a year field equal to 2022.

Once the relevant records have been isolated, aggregation summarizes them into the requested number of people. The specific aggregation may be a count of person records or a sum of a population-related numeric field, depending on how the displayed data represents people. For example, SQL commonly uses a `WHERE` clause to filter rows and an aggregate function such as `COUNT()` or `SUM()` to generate the summary value. This sequence—limit the dataset to the requested scope, then calculate the summary—is a core data-analysis practice because it produces a result aligned precisely with the business question. See the Microsoft documentation for SELECT and WHERE and the Microsoft documentation for aggregate functions.

QUESTION NO: 14

Which of the following best describe qualitative data? (Select two).

A. Discrete

B. Ordinal

C. Batch

D. Continuous

E. Nominal

F. Real-time

ANSWER: B E

Explanation:

Qualitative data describes attributes, labels, or categories rather than measured numeric quantities. **Ordinal** and **Nominal** are the two standard categorical scales used for qualitative variables. Ordinal data places categories in a meaningful relative order, such as service ratings of poor, fair, good, and excellent. The ordering conveys information about rank or position, although it does not establish that the difference between adjacent categories is equal. This makes ordinal values appropriate when an analyst needs to preserve category order without treating the values as precise measurements.

Nominal data groups observations into distinct named categories with no inherent sequence, such as department, product type, country, or account status. A nominal value identifies membership in a class; it does not indicate magnitude, rank, or numerical distance. Data analysts commonly summarize nominal data through counts, proportions, and frequency distributions, while ordinal data can additionally be summarized according to ordered category distributions. Both forms are fundamental to selecting appropriate visualizations, aggregations, and statistical methods for categorical fields. See the NIST Engineering Statistics Handbook for measurement-scale definitions and the Encyclopaedia Britannica overview of qualitative data for categorical-data context.

QUESTION NO: 15

An analyst is importing sales data from multiple regional offices. Which of the following actions should be taken to improve the consistency of the data before it is combined? (Choose two.)

A. Standardize date formats.

B. Use approved values for region names.

C. Increase the number of records.

D. Convert all numeric values to text.

E. Remove all columns with null values.

ANSWER: A B

Explanation:

Standardizing date formats ensures that equivalent dates are represented consistently across all source files. For example, converting values such as 03/04/2024 and 2024-03-04 to one documented format prevents incorrect sorting, filtering, and date-based aggregation.

Using approved values for region names prevents the same region from being recorded under inconsistent labels, such as Northeast, N.E., and NE. Mapping values to a controlled list allows the analyst to group and compare regional results correctly. Data-quality preparation commonly includes standardizing formats and validating values against defined rules. See Microsoft Purview data quality documentation.

QUESTION NO: 16

Given a product sales table, which of the following aggregate functions are suitable for retrieving the total quantity sold by each employee? (Select two).

A. MAX

B. SELECT

C. HAVING

D. GROUP BY

E. ORDER BY

F. SUM

ANSWER: D F

Explanation:

`SUM` is used to calculate the total quantity sold because it adds the values in the quantity column across the applicable sales records. To return a separate total for every employee rather than one total for the entire table, `GROUP BY` organizes the rows into employee-specific groups before the aggregate is calculated. In practice, these work together in a query such as `SELECT employee_id, SUM(quantity) AS total_quantity FROM sales GROUP BY employee_id;`. The resulting dataset contains one row per employee, with the associated summed quantity. Although `GROUP BY` is technically a SQL grouping clause rather than an aggregate function itself, it is required alongside `SUM` to fulfill the stated requirement of retrieving totals *by each employee*. This is a core data-manipulation pattern: aggregate a numeric measure and group the results by a categorical identifier. SQL documentation describes `SUM` as returning the sum of values and explains that `GROUP BY` places rows sharing a grouping value into groups for aggregate processing. See the Microsoft SUM documentation and the PostgreSQL GROUP BY documentation.

QUESTION NO: 17

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

A. SSIS

B. Stata

C. SPSS

D. Cognos

ANSWER: A

Explanation:

SSIS is correct because SQL Server Integration Services is Microsoft's platform for building and running extract, transform, and load workflows. An SSIS package can connect to heterogeneous sources, extract data, apply transformations such as cleansing, type conversion, merging, aggregation, and derived-column calculations, and load the prepared results into destinations such as SQL Server databases and data warehouses. These capabilities make SSIS an ETL tool used to prepare, integrate, and deliver data for downstream analytics, reporting, and data-mining work. SSIS provides a visual development environment in SQL Server Data Tools, along with reusable connection managers, control-flow tasks, data-flow components, logging, error handling, and scheduling support. In a data-mining context, reliable preparation of source data is essential: analytical models depend on data that has been consistently formatted, validated, and consolidated before patterns can be identified. SSIS is therefore an appropriate example of a tool supporting the ETL stage of a data-mining pipeline. Microsoft describes Integration Services as a platform for building enterprise-level data integration and data transformation solutions, including loading data warehouses and cleaning and standardizing data. See Microsoft Learn: SQL Server Integration Services and Microsoft Learn: SSIS Data Flow.

QUESTION NO: 18

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

A. To perform web scraping

B. To track KPIs

C. To improve accuracy

D. To review data sets

E. To increase the sample size

F. To calculate trends

ANSWER: C F

Explanation:

To improve accuracy is a core reason for data cleansing. Cleansing identifies and corrects inaccurate, inconsistent, duplicated, incomplete, or improperly formatted records so that the resulting dataset is more reliable for analysis. Accurate data supports trustworthy calculations, reporting, and business decisions, which aligns with the Data+ focus on improving data quality before it is used to generate insights.

To calculate trends is also correct because trend analysis depends on comparable, consistent observations over time. Data cleansing addresses issues such as invalid values, duplicate records, inconsistent date or category formats, and missing data that could otherwise distort measures, create artificial spikes or declines, or prevent meaningful comparisons across periods. Once the relevant data has been standardized and validated, an analyst can aggregate it by time period and calculate trends with greater confidence. CompTIA's Data+ certification covers data-quality activities as part of preparing data for analysis; see CompTIA Data+. For additional best-practice context on correcting errors and standardizing data before analysis, see IBM's data-cleaning overview.

QUESTION NO: 19

Which of the following programming languages are best suited for analysis and machine-learning applications? (Select two).

A. Ruby

B. Rust

C. PHP

D. Python

E. Kotlin

F. R

ANSWER: D F

Explanation:

Python and R are the best-suited choices because both have mature ecosystems designed around data analysis, statistics, visualization, and machine learning. Python is widely used throughout the data lifecycle: analysts can prepare and manipulate data with pandas and NumPy, build visualizations, and develop predictive models with libraries such as scikit-learn. Its readable syntax and broad integration capabilities also make it practical for automating analytical workflows and deploying models into business applications. The Python Software Foundation's data-science resources describe the language's extensive use across scientific and analytical computing: https://www.python.org/about/apps/.

R was created for statistical computing and graphics, making it especially well suited to exploratory data analysis, statistical testing, modeling, and clear analytical reporting. Its large package ecosystem, including packages distributed through CRAN, provides established implementations for statistical and machine-learning methods. R also supports reproducible analysis through tools such as R Markdown and is commonly used when rigorous statistical analysis is central to the task. The R Project describes R as a free software environment specifically for statistical computing and graphics: https://www.r-project.org/about.html. Together, Python and R directly align with the analytical programming skills covered in data-focused roles and the CompTIA Data+ exam domain.

QUESTION NO: 20

Given the following report:

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

A. A control group for the phrases

B. A summary of the KPIs

C. Filter buttons for the status

D. The date when the report was last accessed

E. The time period lhe report covers

F. The date on which the report was run

ANSWER: E F

Explanation:

A point-in-time, static report must clearly identify both the period represented by its data and when the report output was generated. "The time period lhe report covers" establishes the reporting window or as-of period, such as a particular month, quarter, or a defined start and end date. This lets readers interpret the metrics in their proper historical context and distinguish the report from a view of current, continuously changing operational data.

"The date on which the report was run" records when the report was produced. Together with the covered period, this provides essential provenance: readers can determine the intended data timeframe and identify the specific reporting instance used for analysis, review, or audit evidence. For a report to remain static in practice, its underlying output should also be retained as an exported or otherwise immutable snapshot rather than recalculated dynamically when opened. Microsoft's guidance on Power BI report pages describes report filtering and report content, while its refresh documentation distinguishes refreshed data from a fixed report output: Power BI filters and report views and Power BI data refresh.

QUESTION NO: 21

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

A. Government-issued ID

B. Address

C. Order ID

D. Order date

E. Customer ID

F. Referral number

ANSWER: A B

Explanation:

Government-issued ID and **Address** are the most likely fields to require masking because both are directly associated with an identifiable individual and can expose sensitive personal information if displayed to unauthorized users. Government-issued identifiers, such as a Social Security number, passport number, or driver's license number, are particularly sensitive because they may be used for identity verification and can contribute to identity theft when exposed. An address is also personally identifiable information because it identifies or helps locate a person or household, especially when combined with other customer data.

Data masking protects privacy by replacing all or part of a sensitive value with a non-sensitive representation while preserving an appropriate format for authorized business, testing, reporting, or analytical use. For example, a government-issued identifier might be displayed only as its final few characters, and a street address might be partially obscured. The U.S. National Institute of Standards and Technology describes personally identifiable information as information that can distinguish or trace an individual's identity, including information linked or linkable to a specific person. See NIST's PII definition and OWASP's data-masking guidance.

QUESTION NO: 22

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

A. Government-issued ID

B. Address

C. Order ID

D. Order date

E. Customer ID

F. Referral number

ANSWER: A B

Explanation:

Government-issued ID and Address are the most likely columns to require masking because both are directly associated with an identifiable individual and commonly receive heightened protection under organizational privacy policies. Government-issued identifiers, including items such as Social Security, passport, national identity, or driver's license numbers, are highly sensitive personal data. Exposure can enable identity theft, fraud, or unauthorized account access, so masking them in database views, reports, and nonproduction datasets is an important privacy control.

Address is also personally identifiable information because it identifies or helps locate a specific person or household. Even where an address is not used alone, combining it with other customer or transaction data can make an individual readily identifiable. A database administrator can mask all or part of the value—for example, displaying only a city or postal-code prefix—while preserving enough utility for approved analysis. The U.S. National Institute of Standards and Technology describes PII as information that can distinguish or trace an individual's identity, including information linked or linkable to that person. Guidance on protecting PII is available from NIST's PII glossary and NIST SP 800-122.

QUESTION NO: 23

A JSON file is an example of:

A. structured data.

B. web data.

C. machine data.

D. processed data.

E. semi-structured data.

ANSWER: E

Explanation:

JSON is an example of **semi-structured data**. It represents information using a defined, self-describing arrangement of objects, arrays, and name/value pairs, but it does not require the rigid, fixed schema associated with traditional structured data in relational database tables. A JSON document can nest objects and arrays, and individual records can contain different fields when needed. This flexibility makes JSON useful for exchanging data between applications, APIs, and web services while retaining enough organization for software to parse and analyze it.

The JSON specification defines its core structures as objects, which contain unordered collections of name/value pairs, and arrays, which contain ordered collections of values. Those structures provide metadata-like labels and hierarchy within the document rather than enforcing columns and data types through a predefined table schema. For data analytics purposes, JSON commonly requires parsing or flattening before it can be loaded into tabular systems for reporting or SQL analysis. See the JSON Data Interchange Format (RFC 8259) and IBM's overview of semi-structured data.

QUESTION NO: 24

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.

ANSWER: D

Explanation:

snowflake schema. is correct because it is a dimensional data-warehouse design built around a central fact table and its related dimensions, while allowing dimension data to be normalized into additional related tables when appropriate. The fact table stores measurable business events, such as sales amounts or quantities, and contains keys that link those events to descriptive dimensions such as date, product, customer, or location. In a snowflake design, a dimension can remain in a single denormalized table where that is useful, or its hierarchical attributes can be separated into multiple normalized tables. For example, a product dimension may link to separate subcategory and category tables. This arrangement gives the schema its branching, snowflake-like structure. It can reduce repeated descriptive data and can help maintain shared

hierarchies consistently, while retaining the fact-and-dimension model used for analytics. Microsoft's dimensional-modeling guidance describes a snowflake dimension as a dimension whose attributes are stored across multiple related tables and notes that the model remains centered on fact and dimension tables. See Microsoft Learn: Understand star schema and the importance for Power BI.

QUESTION NO: 25

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

A. A control group for the phrases

B. A summary of the KPIs

C. Filter buttons for the status

D. The date when the report was last accessed

E. The time period the report covers

F. The date on which the report was run

ANSWER: E F

Explanation:

The time period the report covers and the date on which the report was run establish the temporal context needed for a point-in-time, static report. A coverage period tells the reader exactly which interval is represented in the reported measures, such as a calendar month, quarter, or defined start and end dates. This makes the scope of aggregations, trends, and KPI values unambiguous.

The run date records when the report's underlying data was extracted, calculated, or published. Because source systems can be updated after a reporting period closes—for example, through late-arriving transactions, corrections, or record reclassification—the generation date identifies the specific snapshot represented by the report. Together, these elements let a reader distinguish the reporting period from the date the snapshot was produced and provide essential context when the report is stored, shared, or reviewed later.

This aligns with Data+ reporting concepts involving report types, report distribution, and communicating data context. CompTIA describes Data+ as covering data reporting and visualization skills at CompTIA Data+. General reporting guidance also recognizes that a report's displayed data can depend on execution timing and parameterized date ranges; see Microsoft's report-parameters documentation.

QUESTION NO: 26

Which of the following are appropriate reasons to retain a record of data transformations performed during an analysis? (Choose two.)

A. Supporting reproducibility

B. Providing an audit trail

C. Eliminating the need to validate results

D. Increasing the number of source systems

E. Preventing all future data-quality issues

ANSWER: A B

Explanation:

Supporting reproducibility is an appropriate reason because another analyst can repeat the documented transformation steps and obtain the same prepared dataset from the original source data. This makes the analysis more dependable and easier to maintain when data is refreshed.

Providing an audit trail is also appropriate because transformation documentation identifies what changed, why it changed, and which rules were applied. This helps stakeholders review the quality and lineage of the data used to produce a report or recommendation. Documentation is an important component of sound data governance and analytical work. See the Microsoft Purview Data Catalog documentation.

QUESTION NO: 27

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

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

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

A. A histogram

B. A scatter plot

C. A heat map

D. A bar chart

ANSWER: B

Explanation:

A scatter plot is the best visualization for showing the relationship between study hours and exam scores because both fields are quantitative variables measured for each individual student. Each plotted point represents one student, with study hours positioned on one axis and the corresponding exam score on the other. This makes the overall relationship directly visible: an analyst can assess whether scores generally rise as study time increases, whether the association is weak or strong, and whether the pattern appears linear or has a different form. Scatter plots also make unusual observations easy to identify, such as a student whose score is much higher or lower than would be expected given their study hours. This is an important exploratory-data-analysis use case, since the chart preserves the paired nature of the observations rather than summarizing either variable separately. CompTIA Data+ objectives include selecting appropriate visualizations to communicate relationships and trends in data. Guidance from the U.S. National Institute of Standards and Technology describes scatter plots as graphs used to examine the relationship between two variables. See the NIST scatter plot reference and CompTIA's Data+ certification overview.

QUESTION NO: 28

A data analyst is compiling a report that a Chief Executive Officer needs for an impromptu meeting. The report should include information on the previous day's performance. Which of the following reports should the analyst provide?

A. Tactical

B. Ad hoc

C. Dynamic

D. Recurring

ANSWER: B

Explanation:

Ad hoc is the appropriate report because it is created in response to a specific, immediate business request rather than produced according to a predefined schedule. The Chief Executive Officer needs information for an impromptu meeting, which indicates that the analyst must quickly prepare a targeted view of the previous day's performance for the decision at hand. An ad hoc report can be scoped to the executive's current needs, including the relevant performance measures, time period, level of detail, and presentation format. This makes it particularly useful when leaders need timely information to discuss an unexpected issue, validate a recent result, or make a near-term decision. In data analytics practice, report selection should account for the audience, business purpose, delivery timing, and requested reporting period. Here, the audience is an executive and the timing is immediate, while the requested data covers a completed daily period. CompTIA's Data+ certification objectives include communicating business insights and supporting business requirements through appropriate reporting and visualization practices. See the CompTIA Data+ certification overview and the CompTIA exam objectives resources.

QUESTION NO: 29

Given the data below:

In which of the following file formats is the data presented?

A. Xs

B. CSV

C. RIF

D. XML

ANSWER: B

Explanation:

CSV is the appropriate answer for tabular data represented as plain text with values separated by commas. A CSV file organizes information into records, usually with one record per line, while commas delimit the individual fields within each record. When a first row is present, it commonly supplies column headings—for example, labels such as name, age, gender, or occupation—followed by rows containing the corresponding values. This simple row-and-column structure makes CSV a common interchange format for spreadsheets, databases, analytics tools, and data-preparation workflows. Data analysts should also recognize that CSV is a text-based format rather than a spreadsheet file format: it does not inherently retain worksheet formatting, formulas, or data types. Applications infer or assign those properties when importing the file. The IETF specification describes the CSV format as records separated by line breaks and fields separated by commas, with quoting rules for values that themselves contain commas, quotation marks, or line breaks. See RFC 4180 for the traditional CSV format description and IETF CSV specification draft for updated guidance.

QUESTION NO: 30

An analyst is reviewing a data set before calculating quarterly sales totals. Which of the following issues could cause the totals to be overstated? (Choose two.)

A. Duplicate transaction records

B. Incorrect decimal placement

C. Missing transaction amounts

D. Inconsistent capitalization of product names

E. Multiple date formats

ANSWER: A B

Explanation:

Duplicate transaction records can cause totals to be overstated because the same sale may be included more than once in an aggregation. **Incorrect decimal placement** can also overstate totals when, for example, a value intended to be 125.00 is recorded as 1,250.00. Both issues should be identified and corrected before reporting.

Missing values may understate a total, while inconsistent capitalization and date formatting can affect grouping and filtering but do not inherently increase sales amounts. Data profiling and validation help identify these quality issues before analysis.

QUESTION NO: 31

Given the following data:

Which of the following BEST describes the data set?

A. There is data bias.

B. The data is incomplete.

C. The data is inconsistent.

D. The data is outliers.

ANSWER: C

Explanation:

The data is inconsistent. The intended data values represent the same gender category with differing capitalization and formats, such as “M,” “m,” “Male,” and “male.” Although these entries convey equivalent meaning to a person, they are distinct values to many data tools unless they are standardized. This is a data-quality consistency issue: a field that should follow one agreed convention instead contains multiple representations of the same concept. Consistency is important because grouping, filtering, joining, and calculating category counts depend on values being represented uniformly. For example, a report that groups records by the raw gender field could display separate categories for “M,” “m,” and “Male,” producing fragmented and misleading results. The appropriate preparation activity is to cleanse and standardize the field according to a defined data rule, such as converting all values to a single approved label or code before analysis. This supports reliable analysis and repeatable reporting. Microsoft’s data-quality guidance discusses standardization and validation as key practices for improving the reliability of data used in analytics: Microsoft data quality guidance. CompTIA’s Data+ exam objectives also include identifying data-quality issues and applying data-cleaning techniques: CompTIA exam objectives.

QUESTION NO: 32

An analyst is working on a project for a director. During this process. the analyst pulled the data. created summarized tables and graphs with descriptions, created a report summary, and inserted all items into a report. After writing the report, which of the following would be the most appropriate next step?

A. Complete an audit on the data pulled for the report.

B. Complete a check for quality in the report.

C. Complete a review of the data and a check for consistency

D. Complete a trend analysis to be included in the report.

ANSWER: B

Explanation:

Complete a check for quality in the report. At this stage, the analysis, visualizations, descriptions, summary, and report content have already been assembled, so the appropriate next activity is quality assurance before the report is delivered to the director. A report-quality check confirms that tables and graphs accurately reflect the underlying results, calculations and labels are correct, narrative statements agree with the visuals, and the executive summary communicates the findings clearly. It should also verify completeness, consistent formatting and terminology, readable visual design, and that the report directly addresses the stated business purpose. This final review helps ensure that decision-makers receive a dependable, understandable deliverable and reduces the likelihood that an avoidable presentation or interpretation issue affects a business decision. CompTIA Data+ covers communicating insights and applying data-governance and quality practices throughout the analytics process; a final quality review is the practical control that validates the finished reporting artifact before it is shared. See the CompTIA Data+ (DA0-001) exam objectives and the CompTIA Data+ certification overview.

QUESTION NO: 33

An analyst needs to summarize the number of people in Chicago in 2022 using the following set of data:

Name	City	Year	Grade
Chloe	Chicago	2022	A
Blake	Chicago	2023	B
Carter	Chicago	2022	A
Kim	Detroit	2021	C

Which of the following steps should the analyst use to provide results? (Select two).

A. Aggregation

B. Sorting

C. Filtering

D. Indexing

E. Cleaning

F. Replacing

ANSWER: A C

Explanation:

Filtering is needed to isolate only the records relevant to the requested summary: records whose location is Chicago and whose date or year is 2022. Applying those criteria first ensures that the calculation uses the intended subset of the dataset rather than including people from other cities or years. Filtering is a core data-analysis activity because it narrows a dataset to records that meet defined conditions.

Aggregation is then needed to summarize the filtered people values into one result. Depending on how the source data is structured, this typically means calculating a total with a sum operation across the qualifying records. Aggregation converts detailed, row-level observations into a meaningful summary metric, which directly answers the request for the number of people in Chicago during 2022. This reflects the Data+ focus on using data-manipulation techniques to prepare and summarize data for analysis. Microsoft's documentation also describes filtering as displaying only rows meeting specified criteria and provides standard methods for summing values: Filter data in a range or table and SUM function.

QUESTION NO: 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 countyâ€™s office

D. Surveys sent to 100 randomly selected homes that are reflective of the population

ANSWER: D

Explanation:

Surveys sent to 100 randomly selected homes that are reflective of the population is the best method because it uses probability-based selection across the target population of county households. Random selection gives eligible households throughout the county an opportunity to be included rather than restricting selection to a particular time, housing type, or small geographic area. The added requirement that the selected homes be reflective of the population means the sample should mirror important population characteristics relevant to household income, such as the distribution of households across the county and its housing arrangements. This reduces selection bias and makes the resulting sample mean a more credible estimate of the county's mean annual household income. Although the county's physical size is provided, representativeness depends primarily on how the sampling frame and selection process cover the population, not simply on the area sampled. CompTIA Data+ objectives include evaluating data quality and applying appropriate statistical concepts when analyzing data. For additional background on probability sampling and representative survey design, see the U.S. Census Bureau's sampling-design guidance and the CompTIA exam objectives resource.

QUESTION NO: 35

Which of the following are benefits of applying data-validation rules during data entry? (Choose two.)

A. Preventing invalid values

B. Improving data consistency

C. Eliminating the need for data backups

D. Guaranteeing that all data is complete

E. Replacing the need for access controls

ANSWER: A B

Explanation:

Preventing invalid values is a primary benefit of data validation. A rule can restrict a field to an allowed range, format, or set of values, such as requiring a quantity to be nonnegative or requiring a status to be selected from an approved list.

Improving data consistency is also a benefit because validation rules encourage users to enter data in a standardized form. For example, a drop-down list for department names prevents users from entering variations of the same department. These controls improve the reliability of later reporting and analysis. See Microsoft Support: Apply data validation to cells.

QUESTION NO: 36

A data analyst needs to identify the percentage of orders delivered after their promised delivery date. Which of the following calculations should be used?

A. Late orders ÷ total orders × 100

B. Total orders ÷ late orders × 100

C. Late orders − total orders

D. Late orders + on-time orders

ANSWER: A

Explanation:

The appropriate calculation is the number of late orders divided by the total number of orders, multiplied by 100. This produces a percentage that describes the share of all orders that failed to meet the promised delivery date. For example, 25 late orders out of 500 total orders results in a late-delivery rate of 5%.

Using total orders as the denominator ensures that the metric reflects the full delivery population for the reporting period.

QUESTION NO: 37

A data analyst reviews the following data set:

Which of the following is the range value?

A. 9

B. 10

C. 12

D. 13

E. Cannot be determined because the data set is not provided.

ANSWER: E

Explanation:

Cannot be determined because the data set is not provided. The range is a measure of dispersion calculated by subtracting the smallest observed value in a data set from the largest observed value: range = maximum − minimum. Although the question states that a data analyst is reviewing a data set, the space where those values should appear is blank. Without at least the minimum and maximum values, there is no valid calculation that can establish a numeric range. A range value such as 9, 10, 12, or 13 could only be confirmed after the underlying observations are supplied. This is an important data-analysis practice: measures of spread must be traceable to the actual data and reproducible by another analyst. The U.S. National Institute of Standards and Technology describes range as the difference between the maximum and minimum observations in a sample. See NIST: Range. For an overview of range as a descriptive statistic, see Statistics Canada: Measures of dispersion.

QUESTION NO: 38

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

A. The temperature of a hot tub

B. The height of a horse

C. The time to complete a task

D. The number of people in an office

ANSWER: D

Explanation:

The number of people in an office is a discrete variable because it is a count of individual entities and can take only distinct, countable whole-number values, such as 0, 1, 2, or 25 people. A fraction or decimal value is not meaningful when counting people in this context: an office cannot contain 3.6 people. Discrete variables commonly represent counts, including the number of transactions, defects, customers, or items sold. This distinction is useful in data analysis because the type of variable affects how data is summarized, visualized, and modeled. For example, counts may be displayed in a frequency

table or bar chart and may be modeled using count-focused statistical distributions when appropriate. The U.S. National Institute of Standards and Technology describes discrete data as data that can take only particular values, often integers, and distinguishes it from continuously measured quantities. For further background, see the NIST Engineering Statistics Handbook on discrete data and Encyclopaedia Britannica's definition of a discrete variable.

QUESTION NO: 39

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

A. Data refresh rate

B. Consumer types

C. Access permissions

D. Data sources and attributes

ANSWER: B

Explanation:

Consumer types should be considered first because a dashboard must be designed around the people who will use it and the decisions they need to make. Identifying the intended consumers establishes the dashboard's purpose, appropriate level of detail, relevant metrics, terminology, and visual presentation. For example, operational users may need timely, actionable measures that support daily work, while executives generally need concise summaries of strategic performance indicators and trends. Once the analyst understands the consumers' roles, goals, data literacy, and decision-making responsibilities, the analyst can select information that is meaningful rather than merely available. This audience-first approach also helps prevent clutter and ensures that the dashboard communicates a focused story aligned to business needs. CompTIA's Data+ certification covers transforming business requirements into data-driven insights and communicating those insights through appropriate visualizations and reports. Guidance for dashboard design likewise emphasizes defining the audience and its goals before choosing the information and visuals to present. See the CompTIA Data+ certification overview and Microsoft guidance on Power BI dashboards.

QUESTION NO: 40

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.

ANSWER: D

Explanation:

Parametrization. is the correct answer because parameterized queries separate the SQL command structure from values supplied by a user. Rather than concatenating a username, user ID, or other input directly into a query string, the application sends the SQL statement with placeholders and provides user input as bound parameter data. The database interprets that supplied value only as data for the designated field, not as executable SQL syntax. This prevents malicious input from altering clauses, adding commands, or otherwise changing the query's intended logic.

For example, an application can define a query such as `SELECT * FROM users WHERE username = ?` and bind the submitted username to the placeholder through the database driver or framework. This practice should be applied consistently to all database operations that accept untrusted input, including searches, login workflows, inserts, updates, and deletes. Parameterization is also commonly called prepared statements or parameterized statements. It is a primary SQL-injection defense recommended by OWASP and is relevant even when input validation is also implemented, since validation complements rather than replaces safe query construction. See the OWASP SQL Injection Prevention Cheat Sheet and the OWASP SQL Injection overview.

QUESTION NO: 41

Which of the following concepts should be applied if a data set with 40 fields needs to be pared down to 20 fields and contains similar data across multiple fields?

A. Duplication

B. Consolidation

C. Compliance

D. Standardization

ANSWER: B

Explanation:

Consolidation is the appropriate concept because the objective is to reduce the number of fields by combining information that is similar, overlapping, or redundant across the data set. In this situation, the analyst would review the 40 fields to identify fields that represent the same business concept, closely related attributes, or duplicate measures, then combine them into a smaller set of meaningful fields. For example, several columns that capture related categories or repeated versions of an attribute may be mapped into one standardized consolidated field. This reduces structural complexity while preserving the information needed for analysis.

Consolidation is an important data-preparation activity because data sets assembled from different systems, teams, or collection processes often contain overlapping attributes. Reducing those overlaps can make reporting and downstream analysis easier to manage, improve consistency in how a concept is represented, and lessen unnecessary processing and maintenance. The resulting 20 fields should be documented clearly, including the business rule used to merge source values, so users can understand the new field definitions and retain data lineage. CompTIA Data+ includes data cleaning, transformation, and preparation practices as core analytics skills; see the CompTIA Data+ certification overview and the DA0-001 exam objectives.

QUESTION NO: 42

An analyst has been asked to publish a dashboard containing employee compensation data. Which of the following actions should be taken to protect sensitive information? (Choose two.)

A. Implement role-based access controls

B. Apply row-level security

C. Publish the dashboard through a public link

D. Enable unrestricted data export

E. Display the data refresh date

ANSWER: A B

Explanation:

Role-based access controls restrict dashboard access according to a user's job responsibilities, helping ensure that compensation details are available only to authorized users. **Row-level security** can further limit the records each authorized user can view, such as allowing managers to see compensation information only for employees in their own departments.

Publishing a public link or sharing an unrestricted export would increase exposure. A data refresh date is useful report metadata but does not protect confidential information. See NIST guidance on role-based access control and least privilege.

QUESTION NO: 43

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

A. Structural equation modeling

B. Transcription

C. Sequential analysis

D. Sampling

ANSWER: B

Explanation:

Transcription is the necessary step because multimedia-recorded data, particularly audio or video, commonly must be converted into a usable textual representation before it can be systematically coded, searched, categorized, compared, and interpreted. A transcript creates a durable, reviewable record of spoken content and can include relevant contextual details such as pauses, speaker turns, or notable nonverbal events when the analytical objective requires them. This transformation makes unstructured multimedia material more accessible to analysts and supports consistent examination of the evidence. It also enables use of common data-analysis workflows, including text search, annotation, coding, quality review, and documentation of findings. CompTIA Data+ emphasizes preparing and cleaning data so it is in an appropriate format for analysis; converting source material into a format that supports analysis is consistent with this preparation activity. The U.S. National Archives similarly describes transcription as producing a text version of recorded material, which improves access to the information in those records. For further context on the Data+ exam's data-preparation and analysis domains, see CompTIA Data+. Guidance on the value of transcripts for access to audiovisual records is available from the U.S. National Archives.

QUESTION NO: 44

An analyst is explaining the company's financial systems and reporting tools to a new coworker. Which of the following data quality dimensions are the most important? (Select three).

A. Data formatting

B. Data accuracy

C. Data maturity

D. Data field

E. Data completeness

F. Data consistency

G. Data diversity

ANSWER: B E F

Explanation:

Data accuracy, data completeness, and data consistency are the key data-quality dimensions for financial systems and reporting. Financial reports support operational planning, management decisions, external reporting, and compliance activities, so the values reported must faithfully represent the underlying transactions and balances. Data accuracy ensures amounts, dates, account classifications, and calculations correctly reflect the real-world financial events they describe. This is essential for dependable statements, reconciliations, forecasts, and audit evidence.

Data completeness ensures all required transactions, fields, records, and reporting periods are included. A report may contain individually correct records but still produce an unreliable total or conclusion if invoices, ledger entries, adjustment data, or required attributes are absent. Data consistency ensures the same financial entities, definitions, formats, and values are represented uniformly across source systems, dashboards, reports, and time periods. Consistent treatment of items such as account codes and currency values enables reconciliation and gives stakeholders confidence that reports can be compared and combined. These dimensions align with the Data+ objective area covering data quality and data governance; see the CompTIA exam objectives resources. General data-quality guidance also identifies accuracy, completeness, and consistency as core characteristics for trustworthy data: Microsoft Purview data quality overview.

QUESTION NO: 45

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

A. Email

B. Scheduled

C. Subscription

D. Static

ANSWER: C

Explanation:

Subscription is the best delivery method because it provides the client with continuing access to the dashboard through the reporting or business-intelligence service, rather than treating the dashboard as a one-time exported artifact. This model supports a client who needs to return to the dashboard whenever business questions arise, view the current published content, and use available dashboard interactions such as filtering or drilling into visualizations. It also supports ongoing availability as the analyst refreshes the underlying dataset and republishes updated dashboard content.

In practical dashboard deployments, subscriptions commonly complement browser-based access by notifying stakeholders that refreshed content is available or by sending a scheduled snapshot. The important delivery-design principle is that the client has an established, persistent relationship to the published dashboard and can access the maintained reporting environment as needed. This aligns with the distinction between an interactive, service-hosted dashboard and a static file delivery. Microsoft describes dashboards as content viewed in the Power BI service and documents subscriptions as a way to keep users informed about report or dashboard content: Microsoft Power BI dashboards documentation and Microsoft Power BI subscription documentation.

QUESTION NO: 46

Which of the following query statements would be used when filtering data in a relational database management system? (Select two).

A. ORDER BY

B. HAVING

C. WHERE

D. SELECT

E. INSERT

F. GROUP BY

ANSWER: B C

Explanation:

`WHERE` and `HAVING` are both SQL clauses used to filter query results, but they operate at different stages of query processing. `WHERE` filters individual rows before grouping and aggregation occur. For example, a query can use `WHERE` to limit results to transactions from a specified date range, customers in a particular region, or records whose status meets a condition. This makes it the standard clause for applying conditions to source rows in a relational database query.

`HAVING` filters groups after rows have been grouped, typically with `GROUP BY`, and it is especially useful for conditions involving aggregate values. For example, it can return only product categories whose total sales exceed a specified threshold or departments with more than a given number of employees. In practical data analysis, these clauses are often used together: `WHERE` reduces the input records first, and `HAVING` retains only grouped results that meet an aggregate-based condition. Microsoft's SQL documentation describes the use of `WHERE` for search conditions and `HAVING` for conditions applied to groups: WHERE (Transact-SQL) and HAVING (Transact-SQL).