

Exam Preparation

IT Specialist: Data Analytics

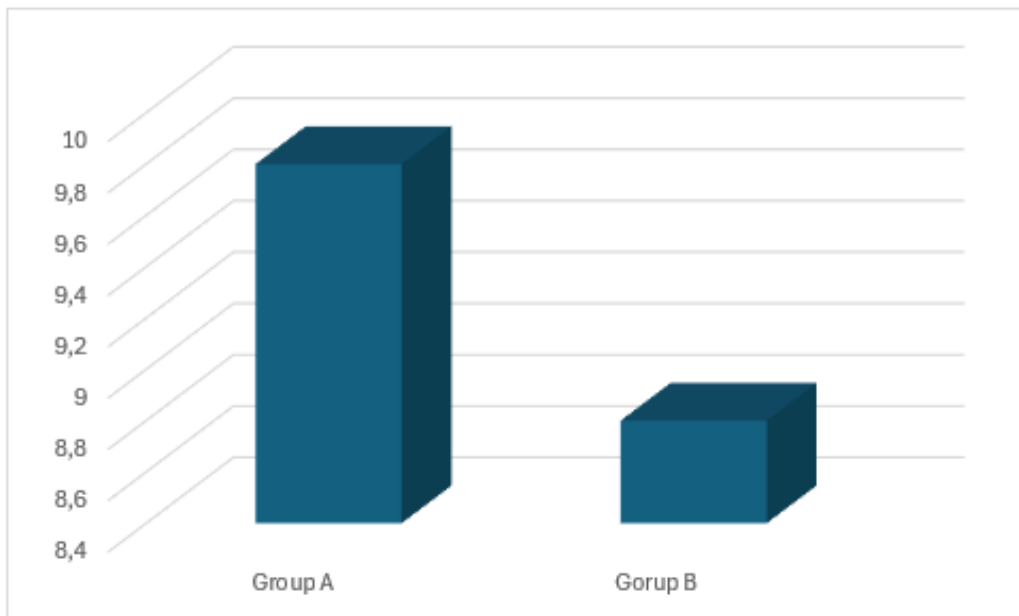


1. The marketing team wants to know which market segment had the highest sales last year.

Which type of data analytics should they use?

- a. Prescriptive analytics
 - b. Diagnostic analytics
 - c. Prescriptive analytics
 - d. Descriptive analytics
2. A colleague shows you the chart below to indicate that Group A has performed significantly better than Group B on a recent assignment. You do not know the sample size or the results of statistical testing.

Which chart element creates the impression of a significant score difference?



- a. The x-axis units of measurement
 - b. The y-axis units of measurement
 - c. The z-axis units of measurement
 - d. The color differentiation
3. Which data type can store a phrase or sentence?
- a. Boolean
 - b. Character
 - c. String
 - d. Integer

4. You have a small dataset that contains personally identifiable information (PII). You need to provide the data to an outside source for additional processing. What could you do to protect the PII but still allow you to eventually relate the additional analysis to your original data
 - a. Employ pseudonymization on PII and use the pseudonym as the key between the new and original datasets.
 - b. Remove every instance of PII in the original dataset and add them back after the new dataset is retrieved.
 - c. Randomly shuffle the original dataset so that each given piece of PII is no longer associated with a particular user
 - d. Retain every text-based PII in the original dataset but convert them to number-based features in the new dataset
5. For each statement about data disaggregation, select **True** or **False**.

Note: You will receive partial credit for each correct selection.

Answer Area	True	False
Data disaggregation provides a summary of the data.		
Data disaggregation combines data from multiple sources.		
Data disaggregation can clarify trends and patterns among subgroups.		

6. A file named `Courses.data` contains the following content:

```
Title|Number|Hours
Algebra|MT101|3
History|HS201|3
Physics|JPS3021|4
Music|MS1011|2
Art|AR201|2
```

You need to use Python to read the data from the file so that you can import it into a database table.

Complete the markup by selecting the correct option from each drop-down list.

Note: You Will receive partial credit for each correct selection.

```
import csv/json/sql/xml

with open('Courses.data', 'r') as file:
```

```
reader = csv.reader(file, column/delimeter/row = '|')
for row in reader:
```

```
# import the data into
```

7. From the data in the table below, you create a PivotTable to show the combined number of certified virtual and In-person teachers for each class at each school.

Move the appropriate labels from the list on the left to the correct locations in the PivotTable on the right You may use each label once or not at all,

Note: You will receive partial credit for each correct answer.

School	Class	Format	Certified Teachers
School A	Networking	In-Person	6
School A	Networking	Virtual	5
School A	Data Analytics	In-Person	2
School A	Data Analytics	Virtual	3
School B	Networking	In-Person	9
School B	Networking	Virtual	7
School B	Data Analytics	In-Person	2
School B	Data Analytics	Virtual	4

Labels
Data Analytics
In-Person
School A
Networking
Virtual
School B

	Label	Label
Label	11	5
Label	16	6

8. Move each function from the list on the left to the correct description on the right.

Functions
COUNT()
MAX()
MIN()
SUM()

Descriptions

Returns the largest value	Function
Returns the smallest value	Function
Returns the number of values	Function
Returns the total of the values	Function

9. Which data structure describes the following data?

["Aabid", "Jesenia", "Mark"]

- Table
 - List
 - Multi-dimensional array
 - Graph
10. You run a t-test with an alpha value of 5% ($\alpha = 0.05$) in order to test an alternative hypothesis (H_1). You finish the analysis and discover that the p-value is 0.017

What can you conclude about the null hypothesis (H_0)?

- You fail to reject the null hypothesis (H_0).
 - You modify the null hypothesis (H_0).
 - You accept the null hypothesis (H_0)
 - You reject the null hypothesis (H_0).
11. What concept allows analysts to drill down into data and examine different levels of information that may be crucial in diagnostic analytics.
- Granularity
 - Transparency
 - Interpretability
 - Completeness

12. Which visualization type is commonly used to display the distribution of a continuous variable, with variable values on the x-axis and corresponding frequencies on the y-axis?

Select the correct visualization type in the answer area.



13. Person A has 5 coins and Person B has 10 coins.

Which type of data does the number of coins represent?

- a. Ordinal data
 - b. Metadata
 - c. Qualitative data
 - d. Quantitative data
14. You are analyzing the sales activity that occurs on national holidays.

What level of data granularity will enable you to perform the most precise analysis?

Data Granularity Options

Years Months Weeks Days Hour

15. What is one goal of data privacy and protection laws such as GDPR, FERPA, and HIPAA?
- a. To hold violators accountable for mishandling data
 - b. To tax companies that use private data
 - c. To ensure that companies openly share industry data

- d. To protect companies from liability related to private data
16. Each month, you need to automatically transform the data from two XML documents into a single flat file with columns and rows that Excel can open and interpret. The document names and structure remain constant. You know the relationship between the two XML documents.

Which two resources can you use? (Choose 2)

Note: You will receive partial credit for each correct selection.

- a. Microsoft Word
 - b. Power Query for Excel (M)
 - c. Python
 - d. JSON
17. You are reviewing the database of restaurant menu items. The table below shows a selection of the data.

You need to display on the Dessert menu with a Type of Cake.

What should you do to nondestructively the data display?

Item	Type	Menu	Gluten-Free	Vegan
Croque Monsieur	Sandwich	Lunch	Optional	No
French Onion Grilled Cheese	Sandwich	Lunch	Optional	No
Lemon Meringue Pie	Pie	Dessert	No	No
Matcha Slice	Cake	Dessert	No	No
Shrimp and Crab Louie	Salad	Lunch; Dinner	Yes	No
Spinach-Artichoke Panini Bites	Sandwich	Lunch	Optional	Yes
Strawberry Pecan Chicken Delight	Salad	Lunch; Dinner	Yes	Optional
Stuffed Portobello	Appetizer	Dinner	Optional	Yes
Vegan Chocolate Cherry Torte	Cake	Dessert	Yes	Yes

- a. Add two slicers, one for Menu and one Type. Set the Menu slicer to Dessert and the Type slicer to Cake.
- b. Group the data by Menu and then group data on the Dessert menu by Type.
- c. Delete all data that has a Menu other than Dessert. Then delete all data that has a Type other than Cake.
- d. Sort the data by Menu and within each Menu, sort by Type

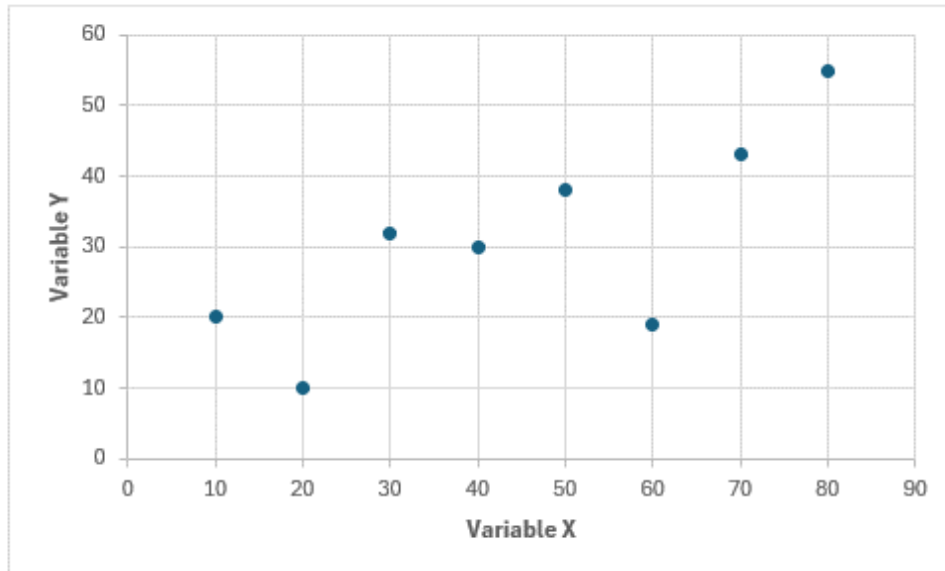
18. You have a dataset that includes product review scores and demographic Information about the reviewers. There are no subcategories associated with the demographic answers. The table shows a selection of the data.

Which scenario is an example of disaggregating the dataset?

Product	Review Score	Reviewer ID	Industry	Ethnicity
AX-150	74	123	Education	Asian
BK-330	82	124	Finance	Latino or Hispanic
BK-315	79	125	Health care	Native Hawaiian or Pacific Islander
CX-290	86	126	Other	African-American
BD-250	61	127	Finance	Other
CD-140	35	128	Food services	Caucasian
AX-310	84	129	Education	Caucasian

- Display the average and mode of the scores for each product grouped by the ethnicity of the reviewers.
 - Display the overall average and mode of all scores on a per-product basis.
 - Display the overall average and mode of all scores and a count of all reviews
 - Display a list of ethnicities that are included in the other option.
19. You are examining a visualization that displays the correlation between Variable X and Variable Y

Evaluate the image and answer the questions by selecting the correct option from each drop-down list.



What is the direction of the correlation between Variable X and Variable Y?

- a. Positive
- b. Negative
- c. Zero

Which correlation range most likely describes the relationship between Variable X and Variable Y?

- a. No Correlation ($r=0.00$)
- b. Some Correlation ($0.00 < r < 0.99$)
- c. Perfect Correlation ($r = 1.00$)

An analyst claims the visualization implies that Variable X causes Variable Y, Is the analyst correct in their assertion?

- a. Yes
- b. No

20. A popular social media site records and counts clicks, likes, and dislikes, and other user interactions.

- a. Big data
- b. Continuous data
- c. Qualitative data
- d. Imputed data

21. Your company has summarized a large data set for the region you live in. You need to compare results from urban and rural communities within your region.

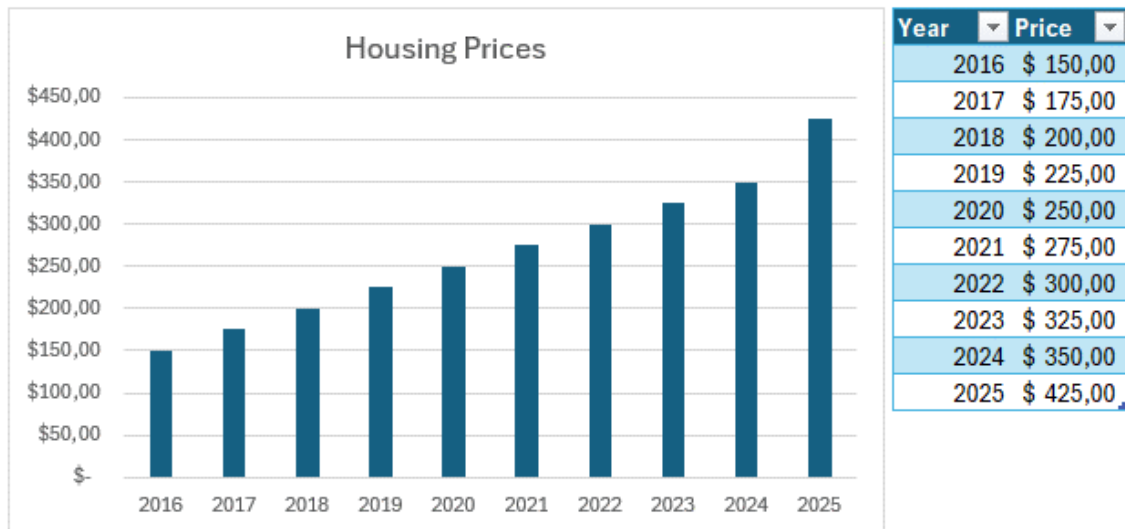
What is the fastest way to obtain this information?

- a. Collect a new data sample.
- b. Review data from neighboring regions.
- c. Aggregate the data.
- d. Disaggregate the data.

22. The visualization and data table depict housing prices in a region.

For each statement about the visualization, select True or False

Note: You will receive partial credit for each correct selection.



	True	False
The visualization uses scaling manipulation.		
An annual increase of \$25,000 occur between 2016 and 2025.		
The visualization accurately depicts the housing prices shown in the table.		

23. What is an example of data cleaning?

- a. Arranging Excel data rows in an order that is easy for a user to read
- b. Removing non-printable characters from a comma-delimited file
- c. Adding quotation marks to the beginning and end of a tab-delimited file
- d. Ensuring that the data a Word table uses a consistent font

24. You believe playing video games increases a person's chance of having a heart attack.

In your research, you notice equal evidence in favor of this hypothesis and opposed to it You spend hours trying to identify the problems with the evidence opposed to your hypothesis, but readily accept the evidence in favor

Which type of bias are you demonstrating?

- a. Confirmation bias
- b. Anchoring bias
- c. Motivated reasoning
- d. Sampling bias

25. You are using data analytics to help answer business questions about a new product your company released.

Move each type of data analytics from the list on the left to the correct question on the right

Note: You will receive partial credit for each correct match.

Why did it happen?

- a. Descriptive analytics
- b. Predictive analytics
- c. Diagnostic analytics
- d. Prescriptive analytics

What action should we take next?

- a. Descriptive analytics
- b. Predictive analytics
- c. Diagnostic analytics
- d. Prescriptive analytics

What might happen in the future?

- a. Descriptive analytics
- b. Predictive analytics
- c. Diagnostic analytics
- d. Prescriptive analytics

What happened in the initial product release?

- a. Descriptive analytics
- b. Predictive analytics
- c. Diagnostic analytics
- d. Prescriptive analytics

26. Which statement correctly assigns a string to the variable that is named score?

- a. `score=string[7];`
- b. `score ="7";`
- c. `score=true;`
- d. `score = 7`

27. For each statement about machine learning, select True or False

Note: You will receive partial credit for each correct selection.

Answer Area	True	False
Machine learning can predict the probability of rain in a region by examining known weather patterns.		
Machine learning can help determine whether a candidate will pass an exam without looking at historical scoring data.		
Machine learning can be used to automatically decline financial purchases based on previous purchase activity.		

28. You conduct a study to identify how many people exercise daily. You recruit all the study participants at gyms.

Which type of bias are you demonstrating

- a. Motivated reasoning
- b. Sampling bias
- c. Anchoring bias
- d. Confirmation bias

29. Which data structure has multiple rows and multiple columns?

- a. List
- b. One-dimensional array
- c. Series
- d. Table

30. You need to create a new data view based on aggregations for further visual analysis. Your data includes sales information for the past five years for food products at your company's stores. Each product belongs to one category. For example, Milk belongs to the Dairy category.

The data view must meet the following requirements:

- Include all products and their associated categories.
- Include sales subtotals for each category and year
- Display a grand total of sales for each category
- Create a summary of each category for every year.

Which two aggregation methods should you use to create the data view? (Choose 2.)

Note: You will receive partial credit for each correct

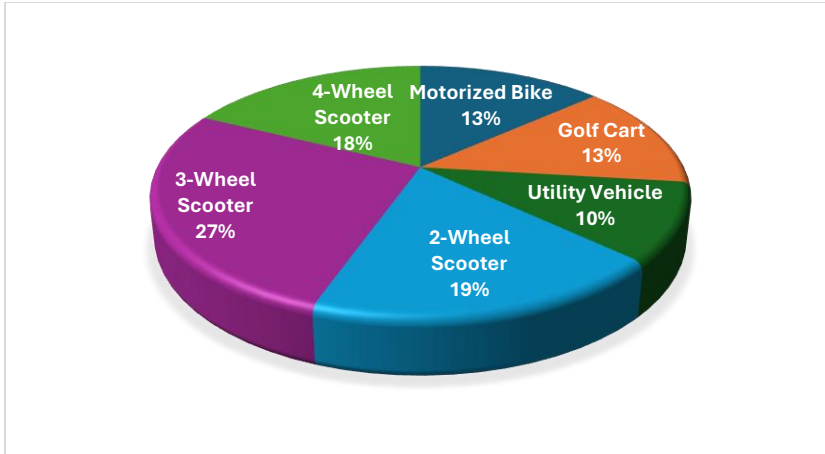
- a. Pivoting
- b. Grouping
- c. Merging
- d. Filtering

31. In which scenario will artificial intelligence (AI) provide the greatest benefit?
- Recording daily sales for three stores owned by one franchise owner
 - Interpreting fundraising sales data for a college soccer team
 - Determining the statistical mean, mode, and standard deviation of the grades for a class
 - Predicting maintenance requirements for a international rental car company's fleet vehicles
32. What is raw data?
- Unprocessed data
 - Visualized data
 - Summarized data
 - Purely numerical data
33. You work for a recreational sports company. The table below shows the company's recreational vehicle sales. You need to demonstrate how each vehicle type contributes to the company's total sales.

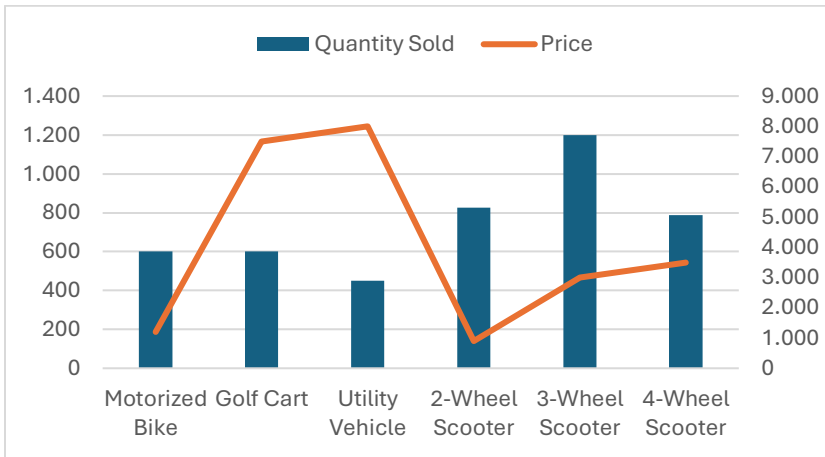
Which visualization should you use?

Select the correct visualization in the answer area

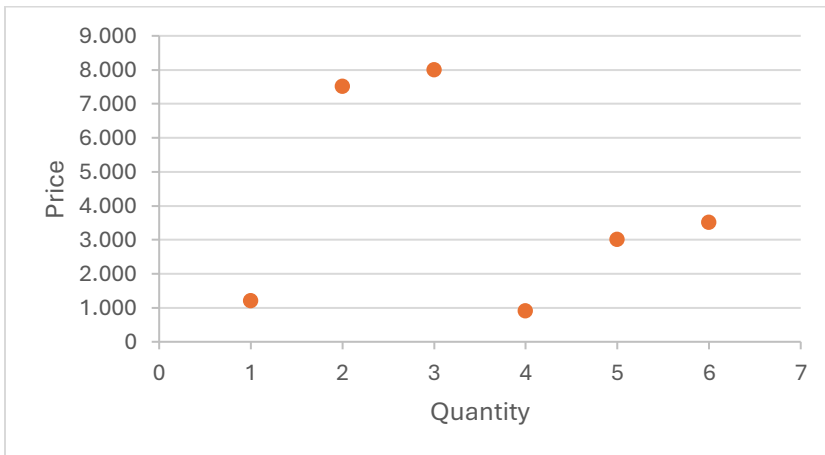
Vehicle Type	Quantity Sold	Price	Total Sales
Motorized Bike	600	1,200	720,000
Golf Cart	600	7,500	4,500,000
Utility Vehicle	450	8,000	3,400,000
2-Wheel Scooter	825	900	765,000
3-Wheel Scooter	1,200	2,999	3,598,800
4-Wheel Scooter	789	3,500	2,761,500
Total	4,464	24,099	15,745,300



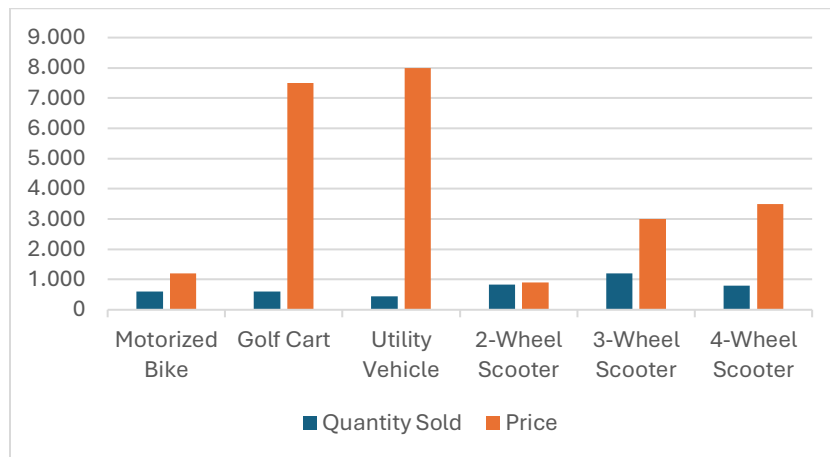
Combo



XY Scater



Column



34. Which two concepts are commonly associated with artificial intelligence (AI) in data analytics? (Choose 2)

Note: You will receive partial credit for each correct selection.

- a. Stakeholders Making
- b. Cost-Benefit Analysis
- c. Automation
- d. Machine Learning

35. What is Metadata?

- a. A context that gives data meaning
- b. The text content of a message
- c. Statistics
- d. Numerical facts

36. You want to show a friend your monthly budget breakdown to prove that most of your expenditure is food costs. You create a table that shows the flow of money as it moves one budget category to the next

Which visualization type should you use to display your analysis based on the table shown?

Source Category	Spend Category	Amount
Total Monthly Budget	Total Monthly Budget	1,000
Total Monthly Budget	Gym Membership	100
Total Monthly Budget	Rent	200

Total Monthly Budget	Food & Entertainment	700
Food & Entertainment	Food Costs	400
Food & Entertainment	Entertainment Costs	300
Entertainment Costs	Movie Theater & Plays	200
Entertainment Costs	Other	100

- a. Time series chart
 - b. Classification tree
 - c. Correlation matrix
 - d. Sankey diagram
37. A coworker is having trouble joining two database tables, TableA and TableB, that were imported from CSV files. They say tables have no common values. You need to troubleshoot the problem. You look at the data in the original CSV files and find that the RowKey values in the TableA file and the RowID values in the TableB file look identical. Both have three numbers followed by a dash (-) and two letters.
- Which two actions should you complete next? (Choose 2.)
- Note: You will receive partial credit for each correct selection.
- a. Trim empty spaces from only the right side of the valid characters
 - b. Visually compare the database values to the CSV values.
 - c. Verify that the data in the database was imported as a numeric data type
 - d. Trim empty spaces from both sides of the valid characters.
38. For which two reasons it is risky to generalize from limited sample data (Choose 2.)
- a. Analyzing data from a smaller sample size is faster
 - b. A limited sample may not represent a larger population
 - c. Limited data samples are easier to collect.
 - d. Findings from a smaller sample size may not be as precise.
39. You have a comma-delimited file with 100,000 rows and 200 columns of phone sales data. One column represents the phone manufacturer.
- You need to analyze all sales data for a specific manufacturer.
- Which technique should you use?
- a. Truncating

- b. Deleting
- c. Filtering
- d. Transposing

40. You are performing descriptive analytics on quarterly sales data.

Move the appropriate statistical metrics from the list on the left to the correct locations on the right. You may use each metric once, more than once, or not at all

Note: You will receive partial credit for each correct response.

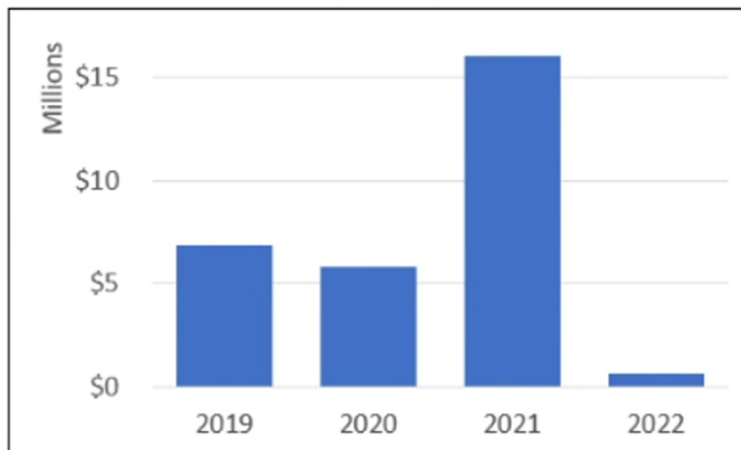
Statistical Metrics
Average
Max
Median
Min
Mode
Sum

Region	Quarter 1	Quarter 2	Quarter 3	Quarter 4
North	25,000	30,000	40,000	50,000
South	35,000	45,000	40,000	55,000
East	35,000	32,500	41,000	52,500
West	34,500	30,000	42,500	55,000
Sum	129,500	137,500	163,500	212,500
Max	35,000	45,000	42,500	55,000
Min	25,000	30,000	40,000	50,000
Average	35,000	30,000	40,000	55,000

41. You create the bar chart below, which shows sales for different years. Management asks for a way to see demographic information associated with the individual sales records for each year.

You have decided to create tables for each year that show the demographic information for the sales in that year. When someone clicks a year in the chart, the associated table with open.

Which reporting technique does this represent?



- a. Unpivoting
 - b. Disaggregating
 - c. Distribution
 - d. Pivoting
42. You need to compare three (3) values of each data point in a series.

Which chart type should you use?

- a. Waterfall chart
 - b. Area chart
 - c. Bubble chart
 - d. Scatter chart
43. Which data type results from processing a conditional statement?
- a. String
 - b. Character
 - c. Integer
 - d. Boolean

44. You are responsible for e-commerce sales at your company. You need to present the quarterly data shown in the table to upper management using the most accurate, unbiased visualization.

Which visualization should you choose?

Select the correct visualization in the answer area.

Year	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
2025	50,000	55,000	60,000	65,000	230,000
2024	45,000	50,000	55,000	60,000	210,000
2023	40,000	45,000	50,000	55,000	190,000
2022	35,000	40,000	45,000	50,000	170,000
2021	30,000	35,000	40,000	45,000	150,000
Break-Even Point	37,500	42,500	47,500	52,500	180,000



45. You ran a t-test with an alpha value of 1% ($\alpha = 0.01$).

Which p-value would cause you to reject the null hypothesis?

- a. 0.001
- b. 0.011
- c. 0.09
- d. 0.10

46. Which sentence most accurately describes the relationship between data and statistics?
- a. All statistics are data, but not all data are statistics.
 - b. All data are statistics, but not all statistics are data
 - c. Data and statistics are both purely quantitative in nature
 - d. Data and statistics are both purely qualitative in nature
47. Which scenario is an example of predictive analytics?
- a. A real estate agent mails flyers to homeowners in a neighborhood after reviewing a sales report
 - b. A car salesperson approaches every fifth person who walks into a car dealership to attempt to make a sale.
 - c. A student tosses a coin to determine how to answer a True or False question
 - d. A stock trader examines financial reports of a company to determine whether the stock price is likely to keep rising
48. Your marketing department attends a variety of events each year and distributes promotional items to event participants. The table below shows the quantity distributed of each promotional item

You are performing analysis on the data. Complete the sentences about data organization by selecting the correct option from each drop-down list.

Note. You will receive partial credit for each correct selection.

Promotional Item	Quantity Distributed
T-shirts	600
Stuffed Animals	425
Drinkaware	55
Backpacks	100
Blankets	55
Magnets	250
Gift Cards	50
Candy	500
Notebooks	450

... can arrange distributed items from highest to lowest.

- a. Appending
- b. Filtering
- c. Sorting
- d. Truncating

... can limit the display of distributed Items to greater than 500.

- a. Appending
- b. Filtering
- c. Sorting
- d. Truncating

... can limit the display of promotional Items to stuffed animals and t-shirts

- a. Appending
- b. Filtering
- c. Sorting
- d. Truncating

49. In the United States and Europe, which two data points are considered non-sensitive PII (personal identifiable information)? (Choose 2)

- a. Gender
- b. Medical records
- c. Mailing address
- d. Date of birth

50. You are tasked with completing a data analysis project for a large organization_ During the project you must handle personally identifiable information (PII).

Throughout the project, which three principles should you follow? (Choose 3.)

- a. Retain only the PII that you might need for future analysts.
- b. Limit your handling of the PII to only what is necessary for the current analysts
- c. Remove all PII from your computer after the analysis IS complete
- d. Keep track of the PII that you have during the analysis.
- e. Request all data from the database that contains PII.

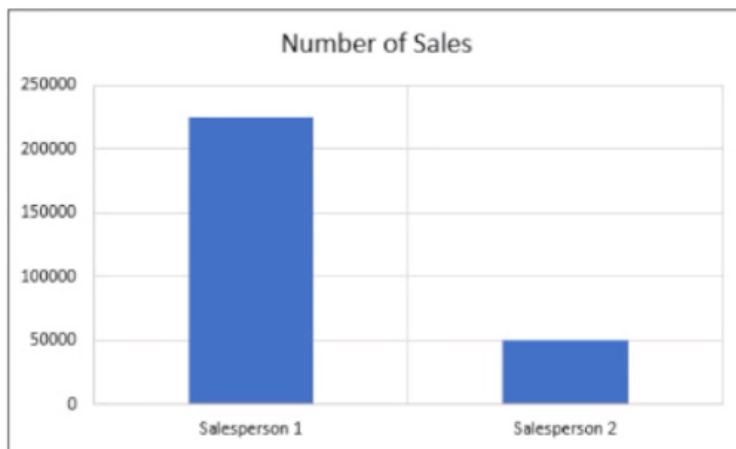
51. What are two goals of data privacy and protection laws such as GDPR, FERPA, and HIPAA? (Choose 2)

- a. To regulate how data is sorted
- b. To regulate how data is shared
- c. To regulate data storage
- d. To regulate data validation

52. The visualization below displays sales data for two salespeople. A conclusion indicates that Salesperson 1 has a higher lead-to-sales rate than Salesperson 2. (A lead-to-sales rate is the number of actual sales divided by the number of attempted sales)

You need to determine the accuracy of this conclusion

What should you conclude?



- a. The conclusion is accurate.
- b. The conclusion is inaccurate because the visualization is missing sales lead data
- c. The conclusion is inaccurate because the visualization uses scale manipulation.

- d. The conclusion is inaccurate because the visualization uses size manipulation.

53. For each statement about data organization, select True or False

Note. You will receive partial credit for each correct selection.

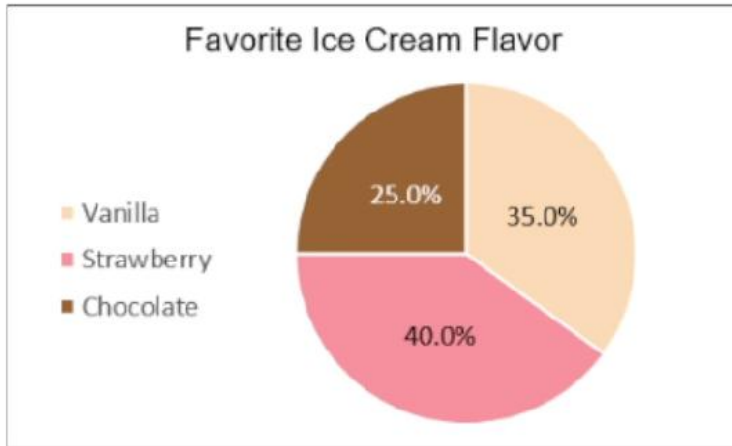
	True	False
Slicers can be used to filter data		
Sorts can be used to display a subset of data.		
Filters can be used to display a subset of data		

54. Which two chart types should you use to rank values in ascending or descending order? (Choose 2)

Note. You Will receive partial credit for each correct selection.

- a. Line chart
 - b. Column chart
 - c. Bubble chart
 - d. Bar chart
55. Which concept most comprehensively describes the general meaning of data in the context of data analytics?
- a. Meaningful statistics
 - b. Unprocessed facts
 - c. Interpreted evidence
 - d. Analyzed details
56. A group of students is asked about their favorite flavor of ice cream. The pie chart below illustrates the proportions of each response.

What can you conclude from the visualization below about ice cream preferences for this group of students?



- a. The most students chose chocolate.
 - b. The fewest students chose strawberry
 - c. The most students chose vanilla.
 - d. The fewest students chose chocolate
57. You have been given a large data set that includes location, income, and age.
- Why should you disaggregate the data?
- a. To hide differences among subgroups
 - b. To form generalizations about the entire data set
 - c. To analyze income within different age groups or locations
 - d. To combine data sets and present a summary of your findings
58. Which data type is appropriate for a phone number using the format (###) ### - ####?
- a. Numeric
 - b. Boolean
 - c. String
 - d. Binary
59. You are preparing to export data from a database to a flat file to be used by another system. Each field in the file should be separated by a comma.
- You notice that the data in several columns includes commas. You decide to enclose the values in each of these columns in double quotes (“).
- What feature of delimited files defines enclosing column data in double quotes?
- a. Row delimiter
 - b. Column delimiter
 - c. Data formatter
 - d. Text qualifier

60. You need to create a data view based on aggregations for further visual analysis. Your data includes sales information for the past five years for food products at your company's stores. Each product belongs to one category for example, Milk belongs to the Dairy category

The data View must meet the following requirements.

- Include all products and their associated categories.
- Include sales subtotals for each category and year.
- Display a grand total of sales for each category.
- Create a summary of each category for every year.

Which two aggregation methods should you use to create the data view (Choose 2)

Note You Will receive partial credit for each correct selection

- a. Grouping
- b. Pivoting
- c. Filtering
- d. Merging

61. A data set indicates that global temperature increases 0.03 degrees Celsius every 10 years.

You use this data and a linear regression model to determine the Earth's temperature in the year 2050.

What type of data analytes are you using?

- a. Prescriptive analytics
- b. Diagnostic analytics
- c. Predictive analytics
- d. Descriptive analytics

62. For each statement about data disaggregation, select **True** or **False**.

Note: You will receive partial credit for each correct selection.

Answer Area	True	False
Data mining is used to find anomalies.		
Data mining is used to summarize raw data from large data sets.		
Data mining is used to review underlying details in a given table		

63. You have a data set of 100,000 rows. The data values fall within a standard range. The data has been cleaned to remove outliers.

Approximately 100 rows of the data set contain NULL values in a numeric data column. You need to determine a best practice for handling the NULL values.

For each statement about handling NULL, select Yes if it is a best practice or No if it is not.

Note: You will receive partial credit for each correct selection.

	Yes	No
Remove the rows that contain NULL values.		
Replace each NULL value with a random value.		
Use a statistic such as average to account for the NULL values.		

64. What is an example of machine learning in predictive analysis?

- Your streaming service suggests a category of movies based on the last 10 movies you watched.
- Your vehicle turns on a warning sensor because one of its components requires maintenance.
- Your thermostat adjusts to a higher temperature because you programmed it based on the time of day.
- Your computer automatically goes into sleep mode because the battery has less than 10 percent power.

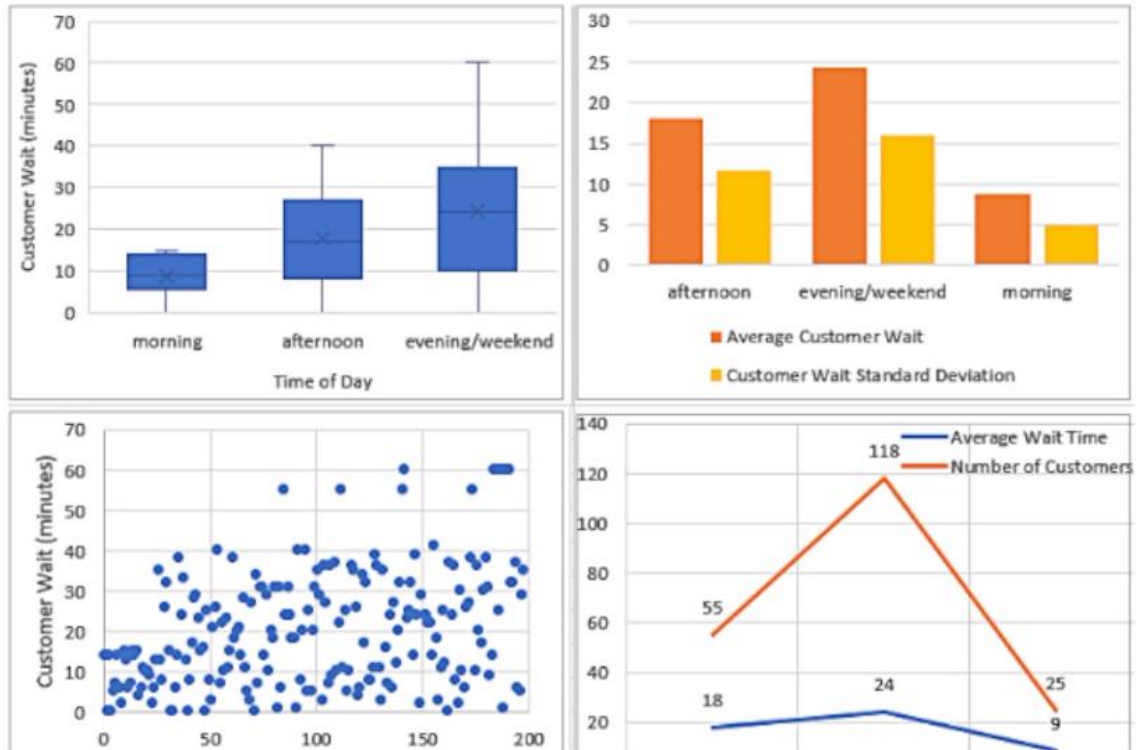
65. For which scenario should you use a line chart to represent the data?

- The binned distribution for the heights of different students.
- The proportion of Yes and No answers to a survey question.
- The maximum, minimum, and average values for a set of data.
- The weekly average stock price during a one-year period.

66. You are given a data set displaying the time of day and number of minutes customers waited in line for service. You need to remove bias from the results by eliminating outliers.

Which visualization illustrates outliers in your dataset?

Select the correct visualizing in the answer area.



67. How is an unstructured data set different from a structured data set?
- An unstructured data set has a predefined data model.
 - An unstructured data set can be quickly searched without manipulation
 - The data organization of an unstructured data set IS explicitly defined
 - An unstructured data set can store large amounts of unrelated data
68. What type of data is too complex to be stored in a traditional database management system (DBMS)?
- Qualitative data
 - Big data
 - Metadata
 - Imputed data