1. **Which of the following options is true for a vector?**
2. A vector has only a magnitude.
3. A vector has a shape and weight.
4. A vector has a magnitude and direction.
5. A vector has only direction.

A vector has a magnitude and direction.

1. You have a sample of size 20 from a population with unknown mean and standard deviation. You measured that the **sample mean** X¯=50 and the **sample standard deviation** is s=10. A confidence interval of 95% of confidence level is given by:

Hint: t0.025=2.093

* (48.95,51.05)
* (45.32,54.68)
* (45.2,54.8)
* (48.9,51.1)

(45.32,54.68)

**3.** To calculate one confidence interval for the mean of a population with unknown distribution, what assumptions we must assure (check all that apply)?

* The sample is a not random sample.
* We can only have a confidence interval if the population is known as having a Normal distribution.
* If the distribution is not Normal, the sample size must be big enough (usually over 30).
* The sample must have mean 0 and standard deviation 1.

If the distribution is not Normal, the sample size must be big enough (usually over 30)

4.**. In statistical hypothesis testing, which of the following statements correctly defines Type Ⅰ and Type Ⅱ errors?**

* Type Ⅰ errror occurs when we reject a null hypothesis that is true, while Type Ⅱ error occurs when we do not reject a null hypothesis that is false.
* Type Ⅰ errror occurs when we do not reject a null hypothesis that is true, while Type Ⅱ error occurs when we reject a null hypothesis that is false.
* Type Ⅰ errror occurs when we reject a null hypothesis that is false, while Type Ⅱ error occurs when we do not reject a null hypothesis that is true.
* Type Ⅰ errror occurs when we do not reject a null hypothesis that is false, while Type Ⅱ error occurs when we reject a null hypothesis that is true.

 Type Ⅰ errror occurs when we reject a null hypothesis that is true, while Type Ⅱ error occurs when we do not reject a null hypothesis that is false.

**Which visualization technique is best for displaying the distribution of a single variable?**

* A) Line Chart
* B) Scatter Plot
* C) Histogram
* D) Pie Chart  
  **Answer:** C) Histogram

**Which of the following would describe a dataset with a right (positive) skew?**

* A) Mean < Median
* B) Mean > Median
* C) Mean = Median
* D) Mean > Mode  
  **Answer:** B) Mean > Median

**What does the Central Limit Theorem state?**

* A) The distribution of the sample mean approaches a normal distribution as the sample size increases
* B) The population mean will always be normally distributed
* C) The sample data will always follow a binomial distribution
* D) The variance of the sample is always smaller than the population variance  
  **Answer:** A) The distribution of the sample mean approaches a normal distribution as the sample size increases

1. **What is data in the context of analytics?**
   * A) Information collected but not organized
   * B) Raw facts and figures
   * C) A set of unchanging facts
   * D) Information presented in reports

*Answer:\** B) Raw facts and figures

1. **Which chart is best for showing trends over time?**
   * A) Bar Chart
   * B) Pie Chart
   * C) Line Chart
   * D) Scatter Plot
   * *Answer:\** C) Line Chart
2. **What is the primary purpose of a scatter chart?**
   * A) To compare parts of a whole
   * B) To show frequency distributions
   * C) To observe relationships between two variables
   * D) To show data trends over time
   * *Answer:\** C) To observe relationships between two variables
3. **What is the main difference between a bar graph and a histogram?**
   * A) Bar graphs show continuous data, histograms show categorical data
   * B) Histograms display continuous data, bar graphs display categorical data
   * C) Bar graphs show relationships, histograms show time-based data
   * D) There is no difference
   * *Answer:\** B) Histograms display continuous data, bar graphs display categorical data
4. **Which chart is most appropriate for showing the proportion of categories in a dataset?**
   * A) Line Chart
   * B) Pie Chart
   * C) Scatter Plot
   * D) Histogram
   * *Answer:\** B) Pie Chart

**Characterizing Data (Descriptive Statistics)**

1. **Which measure of central tendency is the middle value in an ordered dataset?**
   * A) Mean
   * B) Mode
   * C) Median
   * D) Standard Deviation
   * *Answer:\** C) Median
2. **What does the standard deviation measure?**
   * A) The average of the dataset
   * B) The central value of the dataset
   * C) The amount of variation or dispersion in a dataset
   * D) The most frequently occurring value
   * *Answer:\** C) The amount of variation or dispersion in a dataset
3. **Which of the following is a measure of association between two variables?**
   * A) Correlation
   * B) Skewness
   * C) Variance
   * D) Mean
   * *Answer:\** A) Correlation
4. **What does a negative skew in a dataset indicate?**
   * A) The tail is on the left side of the distribution
   * B) The tail is on the right side of the distribution
   * C) The data is symmetrically distributed
   * D) The data has no central tendency
   * *Answer:\** A) The tail is on the left side of the distribution

**Probability Basics**

1. **What is the probability of an event in a uniform distribution?**
   * A) Depends on the event's frequency
   * B) The same for all events
   * C) Higher for frequent events
   * D) Lower for infrequent events
   * *Answer:\** B) The same for all events
2. **In a binomial distribution, what does "n" represent?**
   * A) The number of successes
   * B) The probability of failure
   * C) The number of trials
   * D) The expected value
   * *Answer:\** C) The number of trials
3. **What is the mean of a standard normal distribution?**
   * A) 0
   * B) 1
   * C) 2
   * D) It varies depending on the dataset
   * *Answer:\** A) 0

**Predicting from Data (Inferential Statistics)**

1. **Which sampling technique ensures every individual in the population has an equal chance of being selected?**
   * A) Stratified Sampling
   * B) Random Sampling
   * C) Cluster Sampling
   * D) Systematic Sampling
   * *Answer:\** B) Random Sampling
2. **What is the null hypothesis in hypothesis testing?**
   * A) A statement that no relationship or effect exists
   * B) A statement that an effect exists
   * C) A claim to be proven
   * D) A statement predicting a positive outcome
   * *Answer:\** A) A statement that no relationship or effect exists
3. **According to the Central Limit Theorem, what happens as the sample size increases?**
   * A) The mean of the sample becomes closer to the population mean
   * B) The sample variance increases
   * C) The data becomes more skewed
   * D) The distribution remains unchanged
   * *Answer:\** A) The mean of the sample becomes closer to the population mean
4. **Which error occurs when the null hypothesis is wrongly rejected?**
   * A) Type I Error
   * B) Type II Error
   * C) Sampling Error
   * D) Confidence Interval Error
   * *Answer:\** A) Type I Error

**Crafting Your Data**

1. **What is the first step in data collection?**
   * A) Cleaning the data
   * B) Defining the objectives
   * C) Preparing a report
   * D) Visualizing the data
   * *Answer:\** B) Defining the objectives
2. **What does data cleaning involve?**
   * A) Collecting raw data
   * B) Transforming raw data into a structured form
   * C) Correcting or removing errors in the data
   * D) Analyzing the processed data
   * *Answer:\** C) Correcting or removing errors in the data
3. **Which of the following is an example of a data preparation step?**
   * A) Data storage
   * B) Data wrangling
   * C) Model deployment
   * D) Data archiving
   * *Answer:\** B) Data wrangling
4. **In the data collection phase, which method involves gathering data from different segments of a population?**
   * A) Random Sampling
   * B) Stratified Sampling
   * C) Cluster Sampling
   * D) Convenience Sampling
   * *Answer:\** B) Stratified Sampling

**Visualizing Data**

1. **Explain the differences between a bar chart, a pie chart, and a histogram. In which scenarios would you use each type of chart?**
2. **Discuss the advantages and disadvantages of using scatter plots to visualize relationships between two variables. Provide an example scenario where a scatter plot would be effective.**
3. **What are the key elements of an effective data visualization? Explain how each element contributes to clear communication of data insights.**
4. **Describe how a frequency distribution can be represented graphically. How does a frequency graph help in understanding data distributions?**
5. **Explain the process of creating a line chart and discuss when it is appropriate to use a line chart for data analysis.**

**Characterizing Data (Descriptive Statistics)**

1. **Define and differentiate between mean, median, and mode. How do outliers affect each of these measures of central tendency?**
2. **What is standard deviation, and why is it important in understanding the variability of a dataset? Provide an example to illustrate your explanation.**
3. **Explain what skewness represents in a dataset. Provide an example of a positively skewed and negatively skewed dataset.**
4. **What is a box plot, and how does it help in visualizing the distribution of a dataset? Explain the significance of each component of a box plot.**
5. **Describe the concept of correlation. How does it help in understanding the relationship between two variables? Include an explanation of positive, negative, and zero correlation.**

**Probability Basics**

1. **Define the concept of probability. How is probability applied in data science and analytics? Provide a real-world example.**
2. **Explain the difference between a uniform distribution and a binomial distribution. Provide an example of each.**
3. **What are the properties of a normal distribution? Why is the normal distribution so important in statistics and data analysis?**
4. **Describe the concept of conditional probability. How does it differ from simple probability? Provide a real-world scenario where conditional probability would be used.**
5. **Discuss the role of probability distributions in predictive analytics. How do different types of distributions (e.g., normal, binomial, Poisson) impact the analysis of data?**

**Predicting from Data (Inferential Statistics)**

1. **What is hypothesis testing, and what are the key steps involved in conducting a hypothesis test? Provide an example of a null hypothesis and an alternative hypothesis.**
2. **Explain the Central Limit Theorem and its importance in inferential statistics. How does this theorem help in making inferences about population parameters from sample data?**
3. **Describe the difference between Type I and Type II errors in hypothesis testing. What are the consequences of each type of error?**
4. **What is a p-value in the context of hypothesis testing? How is it used to determine whether to accept or reject the null hypothesis? Provide an example.**
5. **Explain the difference between random sampling and stratified sampling. When would you choose one sampling method over the other in data analysis?**

**Visualizing Data**

1. **Differences between a bar chart, pie chart, and histogram:**
   * **Bar Chart:** Compares categorical data using rectangular bars with heights proportional to values. Used when categories are distinct, e.g., comparing sales across different products.
   * **Pie Chart:** Represents parts of a whole using slices of a circle. Useful for showing proportions, e.g., market share percentages.
   * **Histogram:** Displays the frequency distribution of continuous data using adjacent bars. Useful for visualizing the spread of data, e.g., income ranges in a population.
2. **Scatter plots advantages/disadvantages and use:**
   * **Advantages:** Shows relationships between two variables, helps identify correlations, trends, or clusters.
   * **Disadvantages:** Can be unclear if data points overlap, and does not handle categorical variables well.
   * **Example:** A scatter plot could be used to explore the relationship between marketing spend and sales revenue.
3. **Key elements of effective data visualization:**
   * **Title:** Clearly describes the content and purpose of the visualization.
   * **Axis Labels:** Identify what each axis represents.
   * **Legend:** Explains what different colors or symbols mean.
   * **Scale:** Proportional and accurate representation of data.
   * **Annotations:** Highlight important data points or trends. These elements contribute by ensuring clarity, readability, and focus on the important data insights.
4. **Frequency distribution representation:**
   * Frequency distributions can be represented using a **histogram** or **frequency polygon**. A frequency graph allows for visualization of how frequently each value or range of values appears, helping identify the shape of the data distribution (e.g., normal, skewed).
5. **Process of creating a line chart:**
   * A line chart is created by plotting data points for two variables (typically time on the x-axis and a quantitative variable on the y-axis) and connecting them with a line. It is appropriate for displaying trends or changes over time, such as tracking monthly sales figures.

**Characterizing Data (Descriptive Statistics)**

1. **Mean, median, and mode and effect of outliers:**
   * **Mean:** The average of all data points. Outliers can significantly affect the mean.
   * **Median:** The middle value when data is ordered. Outliers have little to no effect.
   * **Mode:** The most frequent value. Outliers have no effect on the mode.
2. **Standard deviation:**
   * Standard deviation measures the dispersion or spread of a dataset. A low standard deviation indicates data points are close to the mean, while a high standard deviation indicates greater variability. For example, the standard deviation of exam scores shows how much students’ performances vary from the average score.
3. **Skewness in a dataset:**
   * Skewness represents the asymmetry of a distribution:
     + **Positive skew:** Tail on the right, e.g., income data where most people earn less, but a few earn very high amounts.
     + **Negative skew:** Tail on the left, e.g., age at retirement where most people retire older, but a few retire very early.
4. **Box plot explanation:**
   * A **box plot** displays the minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum. The "box" shows the IQR (Q1 to Q3), while "whiskers" extend to the min and max. Outliers may be plotted as individual points. It is useful for comparing distributions and spotting outliers.
5. **Correlation:**
   * Correlation measures the strength and direction of a linear relationship between two variables:
     + **Positive correlation:** As one variable increases, the other also increases.
     + **Negative correlation:** As one variable increases, the other decreases.
     + **Zero correlation:** No relationship between variables.

**Probability Basics**

1. **Definition of probability:**
   * Probability is the likelihood or chance of an event occurring, ranging from 0 to 1. In data science, probability is used to predict outcomes, such as in classification models or risk assessment. **Example:** Predicting the probability of rain based on weather data.
2. **Uniform vs. binomial distribution:**
   * **Uniform distribution:** All outcomes have equal probability, e.g., rolling a fair die (1 to 6).
   * **Binomial distribution:** Describes the number of successes in a fixed number of trials with only two possible outcomes (success/failure), e.g., flipping a coin 10 times.
3. **Properties of a normal distribution:**
   * A normal distribution is symmetrical and bell-shaped, with most data points clustered around the mean. The mean, median, and mode are all equal. It is important because many statistical tests assume normality due to the Central Limit Theorem.
4. **Conditional probability:**
   * Conditional probability is the probability of an event occurring given that another event has already occurred. It differs from simple probability as it depends on prior information. **Example:** Given that a customer bought a product, what’s the probability they will buy again?
5. **Role of probability distributions in predictive analytics:**
   * Probability distributions model uncertainty in data. For example, a **normal distribution** models continuous data, while a **binomial distribution** models binary outcomes. They guide decision-making by estimating the likelihood of future outcomes.

**Predicting from Data (Inferential Statistics)**

1. **Hypothesis testing:**
   * Hypothesis testing is used to test assumptions (hypotheses) about population parameters. Steps:
     + Formulate **null (H₀)** and **alternative (H₁)** hypotheses.
     + Choose significance level (α).
     + Collect sample data and calculate test statistic.
     + Compare test statistic to critical value to accept or reject H₀.
     + **Example:** H₀: The new drug has no effect, H₁: The new drug has a positive effect.
2. **Central Limit Theorem (CLT):**
   * The CLT states that as the sample size increases, the distribution of the sample mean approaches a normal distribution, regardless of the population's distribution. This helps in making inferences about population parameters using sample data.
3. **Type I vs. Type II errors:**
   * **Type I error:** Rejecting a true null hypothesis (false positive). Consequence: Believing a treatment works when it doesn't.
   * **Type II error:** Failing to reject a false null hypothesis (false negative). Consequence: Missing the effect of a treatment that works.
4. **P-value in hypothesis testing:**
   * The p-value is the probability of observing a test statistic as extreme as the one observed, assuming the null hypothesis is true. A small p-value (below α, typically 0.05) leads to rejection of the null hypothesis.
   * **Example:** If a p-value = 0.03, and α = 0.05, we reject the null hypothesis.
5. **Random vs. stratified sampling:**
   * **Random sampling:** Every member of the population has an equal chance of being selected. Used when the population is homogenous.
   * **Stratified sampling:** The population is divided into strata (groups), and random samples are taken from each group. Used when the population has distinct subgroups (e.g., age or gender).

**Introduction to Power BI**

1. **What does "SSBI" stand for in the context of Power BI?**
   * a) Self-Service Business Insights
   * b) Self-Service Business Intelligence
   * c) Software Solution Business Integration
   * d) Strategic Service Business Intelligence
   * *Answer:\** b) Self-Service Business Intelligence
2. **Which of the following is NOT a key benefit of Power BI?**
   * a) Real-time data access
   * b) High cost of deployment
   * c) Interactive visualizations
   * d) Collaboration capabilities
   * *Answer:\** b) High cost of deployment
3. **What is Power BI primarily used for?**
   * a) Website development
   * b) Business Intelligence and data analysis
   * c) Software programming
   * d) Cloud storage solutions
   * *Answer:\** b) Business Intelligence and data analysis
4. **What are the core building blocks of Power BI?**
   * a) Reports, Tables, Databases
   * b) Visualizations, Datasets, Reports, Dashboards
   * c) Queries, Formulas, Workbooks
   * d) Models, Relationships, Filters
   * *Answer:\** b) Visualizations, Datasets, Reports, Dashboards
5. **Which of the following is part of Power BI architecture?**
   * a) Power BI Desktop
   * b) Power Query
   * c) Power BI Service
   * d) All of the above
   * *Answer:\** d) All of the above

**Power BI Desktop & Power Query**

1. **What is Power BI Desktop used for?**
   * a) Designing reports and dashboards
   * b) Data entry
   * c) Building mobile applications
   * d) Writing SQL queries
   * *Answer:\** a) Designing reports and dashboards
2. **Which feature allows you to merge multiple datasets in Power BI Desktop?**
   * a) Data Modelling
   * b) Data Merging
   * c) Query Editor
   * d) Data Blending
   * *Answer:\** c) Query Editor
3. **What does the 'Append' function in Power BI do?**
   * a) Combines data from multiple queries into one
   * b) Deletes duplicate records
   * c) Joins tables by matching keys
   * d) Changes data types of columns
   * *Answer:\** a) Combines data from multiple queries into one
4. **What is the Query Editor used for in Power BI?**
   * a) Writing SQL queries
   * b) Cleaning and transforming data
   * c) Importing images
   * d) Creating dashboards
   * *Answer:\** b) Cleaning and transforming data
5. **What is the primary purpose of "Data Sources" in Power BI?**
   * a) To connect to and retrieve data from external systems
   * b) To manage reports and dashboards
   * c) To edit visualizations
   * d) To schedule data refreshes
   * *Answer:\** a) To connect to and retrieve data from external systems

**Clean and Transform Data**

1. **Which operation in Power BI helps you convert columns into rows?**
   * a) Pivoting
   * b) Unpivoting
   * c) Merging
   * d) Appending
   * *Answer:\** b) Unpivoting
2. **What is the purpose of "Splitting Columns" in Power BI?**
   * a) To combine multiple columns into one
   * b) To divide data in one column into multiple columns
   * c) To delete empty columns
   * d) To create a calculated column
   * *Answer:\** b) To divide data in one column into multiple columns
3. **How can you replace specific values in a column in Power BI?**
   * a) Using DAX
   * b) Using Replace Values function in Query Editor
   * c) Using Conditional Formatting
   * d) Using the Data Model
   * *Answer:\** b) Using Replace Values function in Query Editor
4. **What is a conditional column in Power BI?**
   * a) A column added based on specific conditions or logic
   * b) A column generated from appending queries
   * c) A column that automatically deletes duplicates
   * d) A column that splits into two or more columns
   * *Answer:\** a) A column added based on specific conditions or logic

**Data Analysis Expressions (DAX)**

1. **What does DAX stand for in Power BI?**
   * a) Data Analytics Exchange
   * b) Data Analysis Expressions
   * c) Data Access Execution
   * d) Data Aggregation Expressions
   * *Answer:\** b) Data Analysis Expressions
2. **Which DAX function is used to calculate the sum of a column?**
   * a) SUMX
   * b) COUNT
   * c) SUM
   * d) AVERAGE
   * *Answer:\** c) SUM
3. **What is the difference between ROW context and FILTER context in DAX?**
   * a) ROW context refers to rows; FILTER context applies only to columns
   * b) ROW context refers to a single row; FILTER context applies filters across rows and columns
   * c) ROW context refers to calculated measures; FILTER context refers to data tables
   * d) There is no difference between ROW and FILTER context
   * *Answer:\** b) ROW context refers to a single row; FILTER context applies filters across rows and columns
4. **Which function is considered the most versatile in DAX?**
   * a) CALCULATE
   * b) SUMMARIZE
   * c) FILTER
   * d) IF
   * *Answer:\** a) CALCULATE

**Data Visualization**

1. **Which of the following visualizations in Power BI is best suited to track Key Performance Indicators (KPIs)?**
   * a) Scatter Plot
   * b) Pie Chart
   * c) KPI Visual
   * d) Bar Chart
   * *Answer:\** c) KPI Visual
2. **What is a Decomposition Tree used for in Power BI?**
   * a) To drill down into hierarchical data for detailed analysis
   * b) To show geographical data on a map
   * c) To combine multiple visuals into one
   * d) To create a new column from existing data
   * *Answer:\** a) To drill down into hierarchical data for detailed analysis

**Modelling with Power BI**

1. **What is a "Star Schema" in Power BI?**
   * a) A single table data model
   * b) A data model with a central fact table connected to dimension tables
   * c) A circular data model
   * d) A method to import data from multiple sources
   * *Answer:\** b) A data model with a central fact table connected to dimension tables
2. **In Power BI, what does a "Data Relationship" represent?**
   * a) A connection between different reports
   * b) A link between two or more tables based on a common field
   * c) A way to combine measures and calculated columns
   * d) A method to format visuals
   * *Answer:\** b) A link between two or more tables based on a common field
3. **Which of the following types of relationships is NOT possible in Power BI?**
   * a) One-to-Many
   * b) Many-to-One
   * c) One-to-One
   * d) Many-to-Many
   * *Answer:\** b) Many-to-One
4. **What is the purpose of the "Manage Data Relationships" feature in Power BI?**
   * a) To delete unused tables
   * b) To join tables from different databases
   * c) To create and manage relationships between tables in the data model
   * d) To merge and append tables
   * *Answer:\** c) To create and manage relationships between tables in the data model
5. **Which schema is more normalized, leading to complex queries but efficient storage?**
   * a) Star Schema
   * b) Snowflake Schema
   * c) Flat Schema
   * d) Network Schema
   * *Answer:\** b) Snowflake Schema

**Data Analysis Expressions (DAX)**

1. **What does the DAX function CALCULATE() allow you to do in Power BI?**
   * a) Aggregate values based on row context
   * b) Apply filters to modify the context of a calculation
   * c) Calculate column totals
   * d) Create new calculated tables
   * *Answer:\** b) Apply filters to modify the context of a calculation
2. **Which DAX function can be used to create a running total in Power BI?**
   * a) SUMMARIZE
   * b) CALCULATE
   * c) ALL
   * d) SUMX
   * *Answer:\** d) SUMX
3. **What does the DAX function FILTER() do?**
   * a) Filters rows in a table
   * b) Modifies visual elements
   * c) Filters out empty columns
   * d) Sorts data based on a condition
   * *Answer:\** a) Filters rows in a table
4. **Which DAX function can return the rank of a number in a dataset?**
   * a) RANK.EQ
   * b) TOPN
   * c) RANKX
   * d) CALCULATE
   * *Answer:\** c) RANKX
5. **What is the purpose of the ALL() function in DAX?**
   * a) To remove all filters from a table or column
   * b) To aggregate all data in a table
   * c) To append data from multiple tables
   * d) To create multiple measures in one column
   * *Answer:\** a) To remove all filters from a table or column

**Data Visualization**

1. **What is the purpose of a slicer in Power BI?**
   * a) To filter data visually by selecting different values
   * b) To create a hierarchy in the dataset
   * c) To merge multiple tables into one
   * d) To add new calculated columns
   * *Answer:\** a) To filter data visually by selecting different values
2. **Which Power BI visualization is best for showing geographical data?**
   * a) Bar Chart
   * b) Line Chart
   * c) Map Visualization
   * d) KPI Visual
   * *Answer:\** c) Map Visualization
3. **What are "Custom Visuals" in Power BI?**
   * a) Visualizations created using Power Query
   * b) Visualizations built by third-party developers that can be added to reports
   * c) Visualizations that can only be used in Power BI Service
   * d) Default charts provided by Power BI
   * *Answer:\** b) Visualizations built by third-party developers that can be added to reports
4. **Which visualization is commonly used to display hierarchical data in Power BI?**
   * a) Scatter Plot
   * b) Decomposition Tree
   * c) Pie Chart
   * d) Line Chart
   * *Answer:\** b) Decomposition Tree
5. **What is a KPI visual used for in Power BI?**
   * a) To compare multiple data points
   * b) To display a Key Performance Indicator, such as progress towards a target
   * c) To display historical trends over time
   * d) To perform calculations on data
   * *Answer:\** b) To display a Key Performance Indicator, such as progress towards a target

**Introduction to Power BI Dashboard, Q&A, and Data Insights**

1. **What is the purpose of the "Q&A" feature in Power BI dashboards?**
   * a) To ask questions and receive feedback from Power BI support
   * b) To generate natural language insights from the data
   * c) To answer questions about data relationships
   * d) To create custom queries in Power BI Desktop
   * *Answer:\** b) To generate natural language insights from the data
2. **What is the main purpose of a dashboard in Power BI?**
   * a) To create complex data models
   * b) To display reports and visualizations in an interactive format
   * c) To manage data relationships
   * d) To schedule data refreshes
   * *Answer:\** b) To display reports and visualizations in an interactive format
3. **Which feature of Power BI allows users to explore trends in a decomposition tree?**
   * a) Smart Narrative
   * b) KPI Visual
   * c) Q&A
   * d) Decomposition Tree
   * *Answer:\** d) Decomposition Tree
4. **How does the "Key Influencers" visual in Power BI help users?**
   * a) It highlights the highest and lowest data points in a chart
   * b) It helps identify factors that influence a particular metric
   * c) It is used to sort data in hierarchical order
   * d) It visualizes the data on a map
   * *Answer:\** b) It helps identify factors that influence a particular metric
5. **What is a "Smart Narrative" in Power BI?**
   * a) A feature that provides automated textual summaries of data insights
   * b) A visual that combines multiple charts into one
   * c) A data modeling tool
   * d) A method to create pivot tables
   * *Answer:\** a) A feature that provides automated textual summaries of data insights

**Introduction to Power BI**

1. **Explain the concept of Self-Service Business Intelligence (SSBI) and how Power BI facilitates it.**
2. **What are the key components of Power BI architecture, and how do they interact with each other?**
3. **Discuss the benefits of using Power BI in comparison to other business intelligence tools. Provide examples where Power BI would be particularly useful.**
4. **Describe the primary building blocks of Power BI and explain how each one contributes to effective data analysis.**
5. **What are the different roles available in a Power BI environment, and how does each role contribute to the overall business intelligence process?**

**Power BI Desktop & Power Query**

1. **How does Power BI Desktop differ from Power BI Service? In what scenarios would you use one over the other?**
2. **Explain how to clean and transform data using Power BI's Query Editor. Provide examples of common data preparation tasks such as removing duplicates, splitting columns, or changing data types.**
3. **Describe the process of combining multiple datasets in Power BI Desktop. When would you use 'merge' versus 'append' queries?**
4. **Power BI allows you to connect to a wide range of data sources. Explain how to connect to different data sources and discuss the advantages of connecting to live data sources versus importing data.**
5. **What are the key features of Power Query in Power BI? Explain the role of M Query language in Power Query.**

**Data Modelling in Power BI**

1. **Explain the concept of data modeling in Power BI. Why is creating relationships between tables essential for building effective data models?**
2. **Compare and contrast the Star Schema and Snowflake Schema in Power BI data modeling. What are the advantages and disadvantages of each?**
3. **What are calculated columns and measures in Power BI? How do they differ, and when would you use each? Provide examples.**
4. **Describe the steps for managing relationships between tables in Power BI. How can you handle one-to-many and many-to-many relationships in a data model?**
5. **What is a surrogate key, and why is it important in Power BI data modeling? Explain with an example.**

**Data Analysis Expressions (DAX)**

1. **What is DAX (Data Analysis Expressions), and why is it important in Power BI? Discuss how DAX differs from traditional Excel formulas.**
2. **Explain the difference between row context and filter context in DAX. Provide examples to demonstrate how each context works.**
3. **Describe how the CALCULATE() function works in DAX and explain why it is considered one of the most powerful functions in Power BI. Provide a scenario where CALCULATE() would be essential.**
4. **What are time intelligence functions in DAX? How can they be used to analyze trends over time? Provide examples of commonly used time intelligence functions.**
5. **Discuss the use of DAX aggregation functions like SUMX, AVERAGEX, and COUNTX. When would you use these functions over their standard counterparts (SUM, AVERAGE, COUNT)? Provide an example of each.**

**Introduction to Power BI**

1. **Explain the concept of Self-Service Business Intelligence (SSBI) and how Power BI facilitates it.**
   * **Answer:** SSBI refers to the ability of business users to access and analyze data without needing deep technical skills or assistance from IT departments. Power BI facilitates SSBI by providing intuitive drag-and-drop features, easy data connection, and powerful visualizations. Users can create reports, dashboards, and insights without relying on specialized developers, enabling faster decision-making.
2. **What are the key components of Power BI architecture, and how do they interact with each other?**
   * **Answer:** The key components of Power BI architecture are Power BI Desktop, Power BI Service, Power BI Mobile, and Power BI Gateway. Power BI Desktop is used for data preparation and report creation, while Power BI Service hosts reports online for sharing and collaboration. Power BI Mobile allows for viewing reports on mobile devices, and the Power BI Gateway connects on-premise data sources to Power BI Service.
3. **Discuss the benefits of using Power BI in comparison to other business intelligence tools. Provide examples where Power BI would be particularly useful.**
   * **Answer:** Power BI offers several advantages such as ease of use, integration with Microsoft products, cost-effectiveness, and a large variety of visualizations. It is particularly useful for organizations using the Microsoft ecosystem (e.g., Excel, Azure) and those needing rapid deployment of business intelligence solutions. Power BI's real-time dashboard capabilities and natural language query feature (Q&A) also make it a strong choice for interactive analytics.
4. **Describe the primary building blocks of Power BI and explain how each one contributes to effective data analysis.**
   * **Answer:** The primary building blocks of Power BI include datasets, reports, dashboards, and tiles. Datasets are the underlying data sources, reports are multi-page visual representations of the data, dashboards are single-page collections of visualizations, and tiles are individual visuals pinned to dashboards. Each building block contributes by allowing users to visualize, interact with, and drill into data for deeper insights.
5. **What are the different roles available in a Power BI environment, and how does each role contribute to the overall business intelligence process?**
   * **Answer:** Key roles in a Power BI environment include data analysts (who create reports and dashboards), data engineers (who prepare and manage data), IT administrators (who manage security and access control), and end-users (who consume and interact with reports). Each role ensures data is accessible, secure, and actionable for informed decision-making.

**Power BI Desktop & Power Query**

1. **How does Power BI Desktop differ from Power BI Service? In what scenarios would you use one over the other?**
   * **Answer:** Power BI Desktop is an on-premises application used for data modeling, report creation, and transforming data. Power BI Service is a cloud-based platform where reports are published for sharing and collaboration. Power BI Desktop is used for developing reports, while Power BI Service is used for sharing and viewing reports with stakeholders.
2. **Explain how to clean and transform data using Power BI's Query Editor. Provide examples of common data preparation tasks such as removing duplicates, splitting columns, or changing data types.**
   * **Answer:** Power BI's Query Editor allows users to clean and transform data by applying various steps. For example, to remove duplicates, users can select columns and choose "Remove Duplicates". Splitting columns can be done by selecting "Split Column" based on delimiters or characters. Data types can be changed by selecting the column and using the "Data Type" dropdown to convert data to numeric, text, or date formats.
3. **Describe the process of combining multiple datasets in Power BI Desktop. When would you use 'merge' versus 'append' queries?**
   * **Answer:** In Power BI Desktop, 'merge' is used to join two datasets based on a common column (like SQL joins), which is useful when combining related information from different sources. 'Append' is used when combining datasets with the same structure (same columns) into one table. Merging is used when datasets are related, while appending is for stacking datasets on top of each other.
4. **Power BI allows you to connect to a wide range of data sources. Explain how to connect to different data sources and discuss the advantages of connecting to live data sources versus importing data.**
   * **Answer:** Power BI Desktop allows users to connect to data sources like Excel, SQL databases, Azure, Web data, and more via the "Get Data" option. Live connections provide real-time data updates without importing data into Power BI, ideal for dashboards requiring up-to-the-minute information. Importing data creates a static snapshot that can be refreshed at intervals, which is more efficient for smaller datasets or reports not needing real-time updates.
5. **What are the key features of Power Query in Power BI? Explain the role of M Query language in Power Query.**
   * **Answer:** Power Query's key features include data connection, cleaning, transformation, and reshaping. It offers functions like merging, appending, filtering, and grouping data. M Query is the underlying language used by Power Query to record and apply these transformations. Users can modify M Query directly for more complex data manipulations.

**Data Modelling in Power BI**

1. **Explain the concept of data modeling in Power BI. Why is creating relationships between tables essential for building effective data models?**
   * **Answer:** Data modeling in Power BI involves creating relationships between tables so that users can analyze data from multiple sources efficiently. Relationships allow tables to interact with one another based on key fields, enabling users to slice and filter data across different tables and obtain meaningful insights without duplicating data.
2. **Compare and contrast the Star Schema and Snowflake Schema in Power BI data modeling. What are the advantages and disadvantages of each?**
   * **Answer:** A Star Schema consists of a central fact table linked to multiple dimension tables, while a Snowflake Schema normalizes the dimension tables into smaller sub-tables. Star Schemas are simpler and faster for querying but may contain redundant data. Snowflake Schemas reduce redundancy but are more complex and can result in slower query performance due to multiple joins.
3. **What are calculated columns and measures in Power BI? How do they differ, and when would you use each? Provide examples.**
   * **Answer:** Calculated columns are new columns added to a table using DAX formulas, calculated row by row. Measures are calculations performed on aggregations (e.g., sum, average), often used in visualizations. Calculated columns are used for row-specific calculations (e.g., "Total Price" = "Quantity" \* "Unit Price"), while measures are used for summary data (e.g., "Total Sales" = SUM(Sales)).
4. **Describe the steps for managing relationships between tables in Power BI. How can you handle one-to-many and many-to-many relationships in a data model?**
   * **Answer:** To manage relationships in Power BI, navigate to the "Model" view, select tables, and drag fields to create a relationship. For one-to-many relationships, a single unique value in one table is linked to multiple values in another. For many-to-many relationships, use a bridging table or relationship settings that allow bi-directional filtering to resolve ambiguity.
5. **What is a surrogate key, and why is it important in Power BI data modeling? Explain with an example.**
   * **Answer:** A surrogate key is an artificial key added to a table to uniquely identify records when natural keys are absent or complex. In Power BI, surrogate keys help simplify relationships between tables. For example, a sales table might use a surrogate key like "OrderID" to link sales details to a customer table.

**Data Analysis Expressions (DAX)**

1. **What is DAX (Data Analysis Expressions), and why is it important in Power BI? Discuss how DAX differs from traditional Excel formulas.**
   * **Answer:** DAX is a formula language used in Power BI to create custom calculations, aggregations, and measures. Unlike Excel formulas, which operate cell-by-cell, DAX works on columns and tables, making it more suited for large datasets and complex calculations, particularly in business intelligence contexts.
2. **Explain the difference between row context and filter context in DAX. Provide examples to demonstrate how each context works.**
   * **Answer:** Row context refers to calculations performed on individual rows in a table (e.g., calculated columns). Filter context applies filters across tables or rows during aggregations (e.g., in measures). For example, a calculated column like "Total Sales" = "Quantity" \* "Unit Price" operates in row context, while "Total Sales for 2023" = CALCULATE(SUM(Sales), Year = 2023) uses filter context.
3. **Describe how the CALCULATE() function works in DAX and explain why it is considered one of the most powerful functions in Power BI. Provide a scenario where CALCULATE() would be essential.**
   * **Answer:** CALCULATE() modifies the filter context in which a DAX expression is evaluated, allowing for custom aggregations. It is powerful because it can change the context dynamically. For example, to calculate total sales for a specific year, you can use CALCULATE(SUM(Sales), Year = 2022), overriding the global filters in the report.
4. **What are time intelligence functions in DAX? How can they be used to analyze trends over time? Provide examples of commonly used time intelligence functions.**
   * **Answer:** Time intelligence functions in DAX allow users to perform calculations based on dates and time periods. They are used to analyze trends like year-over-year growth or moving averages. Common functions include SAMEPERIODLASTYEAR(), TOTALYTD(),

and DATEADD(), which allow users to compare current sales to previous periods or calculate cumulative totals.

1. **Discuss the use of DAX aggregation functions like SUMX, AVERAGEX, and COUNTX. When would you use these functions over their standard counterparts (SUM, AVERAGE, COUNT)? Provide an example of each.**
   * **Answer:** SUMX, AVERAGEX, and COUNTX iterate over a table or expression, performing calculations on each row before aggregating the results. They are used when calculations need to happen row-by-row before summarizing. For example, SUMX can calculate "Total Sales" by multiplying quantity by price for each row, then summing the results.

**Power BI Desktop & Power Query**

1. **What are the key steps involved in installing Power BI Desktop?**
   * **Answer:** The steps to install Power BI Desktop include:
     + Downloading the installation file from the Microsoft Power BI website or the Microsoft Store.
     + Running the installer and following the on-screen instructions.
     + After installation, launching the application to begin data connections and report creation.
2. **How can you connect Power BI Desktop to a SQL Server database, and what are the common connection methods?**
   * **Answer:** In Power BI Desktop, use the “Get Data” feature, select "SQL Server" as the data source, and provide the server name and database name. Connection methods include importing data, using DirectQuery for real-time queries, or using a live connection.
3. **Explain the difference between query duplicate and query reference in Power BI. When would you use each?**
   * **Answer:** Query duplicate creates a full copy of the query, while query reference creates a reference or pointer to the original query. Use duplicate when you want to create an independent copy for transformations, and reference when you want to make a version that reflects changes to the original query without duplicating the data.
4. **How does Power BI handle irregularly formatted data, and what are the techniques to clean it?**
   * **Answer:** Power BI uses Power Query for data cleaning. Techniques include removing duplicates, replacing values, transforming data types, splitting and merging columns, and unpivoting columns to normalize the data.

**Data Cleaning and Preparation**

1. **Describe the process of pivoting and unpivoting columns in Power BI and provide use cases for each.**
   * **Answer:** Pivoting in Power BI turns row values into column headers (e.g., creating a pivot table), often used for summarizing data. Unpivoting turns columns into rows, which is useful for denormalizing data when column headers are categorical data points.
2. **How do you handle missing or null values in Power BI?**
   * **Answer:** Power BI allows users to replace missing or null values using Power Query. The "Replace Values" function can fill in missing values with a default value, or users can remove rows or filter them based on null conditions.
3. **What is the importance of changing data types in Power BI? Explain the potential issues that can arise if data types are not managed correctly.**
   * **Answer:** Changing data types ensures that Power BI handles data correctly, especially during calculations and visualizations. Incorrect data types can lead to errors in aggregation (e.g., treating numbers as text) and visual formatting, causing inaccurate reporting.
4. **Explain the process of adding conditional columns in Power BI and provide an example where it would be useful.**
   * **Answer:** Conditional columns are created using if-then-else logic in Power Query. For example, if you want to categorize sales data based on thresholds (e.g., "High," "Medium," "Low" sales categories based on sales amount), you can use a conditional column.

**Data Modeling and Relationships**

1. **What are the benefits of using star schema in data modeling for Power BI?**
   * **Answer:** The star schema simplifies queries, optimizes performance, and makes it easier to understand relationships between fact and dimension tables. This schema is ideal for analysis since it reduces redundancy and speeds up queries.
2. **How do you create and manage data relationships in Power BI Desktop?**
   * **Answer:** Data relationships in Power BI are managed in the "Model" view. Relationships can be created by dragging a column from one table to a related column in another table. The relationships can be one-to-one, one-to-many, or many-to-many, and users can define cross-filtering behavior and cardinality.
3. **What are bidirectional filters in Power BI, and when would you enable them?**
   * **Answer:** Bidirectional filters allow data to be filtered in both directions between related tables, enhancing interaction between visualizations. It is used when you want changes in a dimension table to reflect in the fact table and vice versa, but should be used cautiously to avoid performance issues.

**Data Analysis Expressions (DAX)**

1. **What is the difference between calculated measures and calculated columns in DAX? Provide an example of when you would use each.**
   * **Answer:** Calculated columns operate at the row level and are created for each row in a table. Measures are aggregations calculated dynamically based on the context of the report or visualization. Use a calculated column for static row-level operations (e.g., calculating profit for each row), and a measure for aggregating data (e.g., total sales across regions).
2. **Explain the use of the FILTER() function in DAX. How does it differ from using a regular filter in Power BI?**
   * **Answer:** The FILTER() function in DAX allows users to apply a filter condition to a table or a dataset in a calculated measure or column. It is different from the visual filters in Power BI since it applies directly within a DAX expression, allowing more control and flexibility. For example, you can create a measure that only sums sales greater than a specific amount using FILTER().
3. **What are row context and filter context in DAX? How do they impact the calculation of measures?**
   * **Answer:** Row context refers to the current row of the table during calculations, affecting calculated columns. Filter context refers to the set of filters applied to a calculation in a measure. Row context operates on individual rows, while filter context limits the data considered in aggregations.
4. **Explain the role of CALCULATE() in DAX. Provide an example of how it can be used to change the filter context.**
   * **Answer:** CALCULATE() in DAX is used to change the filter context of a calculation. For example, you can use CALCULATE() to calculate total sales for a specific region by overriding the global filter: CALCULATE(SUM(Sales[Amount]), Sales[Region] = "East").

**Data Visualization**

1. **What are the advantages of using custom visuals in Power BI? Provide an example of a scenario where custom visuals would be useful.**
   * **Answer:** Custom visuals provide flexibility and offer visualizations not available in standard Power BI visuals. For example, if an organization needs a specialized KPI chart or radar chart, custom visuals allow them to create or import these from the marketplace. Custom visuals are especially useful for specific industries or unique reporting requirements.
2. **Explain the importance of color formatting and conditional formatting in Power BI. How can it enhance data interpretation?**
   * **Answer:** Color formatting and conditional formatting allow users to highlight important data trends or outliers visually. For instance, using red for negative values and green for positive values in a financial report immediately draws attention to areas of concern or success, enhancing the user's ability to interpret the data quickly.
3. **How can you create an interactive dashboard in Power BI using slicers and filters?**
   * **Answer:** Interactive dashboards in Power BI are created by adding slicers and filters that allow users to dynamically interact with the report data. Slicers can be added to filter visualizations based on categories like date, product, or region, while report-level filters apply to all pages within a report.

**Power BI Service and Power BI Direct Connectivity**

1. **What is the difference between importing data and using DirectQuery in Power BI?**
   * **Answer:** Importing data brings the entire dataset into Power BI, allowing users to perform transformations and calculations on it. DirectQuery establishes a live connection to the data source without importing, ensuring real-time data but with some limitations on functionality and performance, particularly in large datasets.
2. **Explain the purpose of Power BI Data Gateways and how they facilitate on-premises data connectivity.**
   * **Answer:** Power BI Data Gateways enable secure data transfer between on-premises data sources (such as SQL Server) and Power BI Service or Power BI in the cloud. They facilitate real-time data refresh and ensure that Power BI can access enterprise data without exposing it to the cloud, maintaining data security and compliance.

**Introduction to Python**

1. **Which of the following is a feature of Python?**
   * a) Strongly typed language
   * b) Interpreted language
   * c) Requires compilation before execution
   * d) Low-level programming
   * **Answer:** b) Interpreted language
2. **Python can be used in which of the following application areas?**
   * a) Web development
   * b) Data analysis
   * c) Machine learning
   * d) All of the above
   * **Answer:** d) All of the above
3. **What command is used to execute a Python program from the command line?**
   * a) run python filename.py
   * b) python filename.py
   * c) execute filename.py
   * d) start filename.py
   * **Answer:** b) python filename.py

**Data Types & Variables**

1. **Which of the following is NOT a valid Python data type?**
   * a) int
   * b) float
   * c) real
   * d) bool
   * **Answer:** c) real
2. **What is the correct syntax to create a variable in Python?**
   * a) var x = 10
   * b) int x = 10
   * c) x = 10
   * d) define x = 10
   * **Answer:** c) x = 10
3. **Which of the following is a valid identifier in Python?**
   * a) 2name
   * b) name2
   * c) @name
   * d) name$
   * **Answer:** b) name2
4. **What is the output of the following code: print(type(10))?**
   * a) <class 'int'>
   * b) <class 'float'>
   * c) <class 'str'>
   * d) <class 'bool'>
   * **Answer:** a) <class 'int'>
5. **Which function is used to convert a value to an integer in Python?**
   * a) to\_int()
   * b) int()
   * c) convert()
   * d) integer()
   * **Answer:** b) int()

**Conditional Statements**

1. **Which of the following is a valid conditional statement in Python?**
   * a) if (a == b) then
   * b) if a == b:
   * c) if (a == b):
   * d) if a == b then:
   * **Answer:** b) if a == b:
2. **What is the purpose of the else statement in Python?**
   * a) To execute a block of code when the if condition is true
   * b) To execute a block of code when the if condition is false
   * c) To repeat a block of code
   * d) To break out of a loop
   * **Answer:** b) To execute a block of code when the if condition is false
3. **Which of the following operators is used to combine multiple conditions in an if statement?**
   * a) and
   * b) +
   * c) or
   * d) Both a and c
   * **Answer:** d) Both a and c
4. **Which of the following is a logical operator in Python?**
   * a) +
   * b) ==
   * c) or
   * d) %
   * **Answer:** c) or

**Loops in Python**

1. **Which loop is guaranteed to execute at least once in Python?**
   * a) for loop
   * b) while loop
   * c) do-while loop
   * d) None of the above
   * **Answer:** d) None of the above (Python does not have a do-while loop)
2. **What does the range() function do in a for loop?**
   * a) Iterates over a sequence of numbers
   * b) Iterates over a list
   * c) Returns the last element of a list
   * d) Sorts a sequence of numbers
   * **Answer:** a) Iterates over a sequence of numbers
3. **Which keyword is used to skip the current iteration of a loop in Python?**
   * a) break
   * b) continue
   * c) pass
   * d) return
   * **Answer:** b) continue
4. **What is the output of the following code?**

for i in range(5):  
    if i == 3:  
        break  
    print(i)

* + a) 0 1 2 3 4
  + b) 0 1 2
  + c) 0 1 2 3
  + d) None of the above
  + **Answer:** b) 0 1 2

**More on Data Types**

1. **Which of the following methods can be used to convert a string to lowercase in Python?**
   * a) lower()
   * b) down()
   * c) to\_lowercase()
   * d) str\_lower()
   * **Answer:** a) lower()
2. **Which data type allows you to store a collection of key-value pairs in Python?**
   * a) List
   * b) Tuple
   * c) Dictionary
   * d) Set
   * **Answer:** c) Dictionary
3. **What is the correct way to create a tuple in Python?**
   * a) my\_tuple = {1, 2, 3}
   * b) my\_tuple = [1, 2, 3]
   * c) my\_tuple = (1, 2, 3)
   * d) my\_tuple = <1, 2, 3>
   * **Answer:** c) my\_tuple = (1, 2, 3)

**Functions and Modules**

1. **Which of the following is a built-in Python module?**
   * a) math
   * b) custom
   * c) string\_ops
   * d) user\_module
   * **Answer:** a) math

**Data Types & Variables**

1. **What will be the output of the following code?**

 x = 5  
 y = 3  
 print(x \*\* y)

* + a) 8
  + b) 125
  + c) 15
  + d) 243
  + **Answer:** b) 125

1. **Which of the following statements is true about Python lists?**
   * a) Lists are mutable
   * b) Lists are immutable
   * c) Lists can store only one data type
   * d) Lists cannot be indexed
   * **Answer:** a) Lists are mutable
2. **Which method is used to add an element at a specific position in a list?**
   * a) insert()
   * b) append()
   * c) add()
   * d) extend()
   * **Answer:** a) insert()
3. **Which of the following is the correct syntax for creating a set in Python?**
   * a) my\_set = {1, 2, 3}
   * b) my\_set = [1, 2, 3]
   * c) my\_set = (1, 2, 3)
   * d) my\_set = 1, 2, 3
   * **Answer:** a) my\_set = {1, 2, 3}
4. **What will be the output of the following code?**

 a = "Hello"  
 b = "World"  
 print(a + b)

* + a) Hello World
  + b) HelloWorld
  + c) "Hello" "World"
  + d) Error
  + **Answer:** b) HelloWorld

**Conditional Statements**

1. **Which of the following operators is used to check if two values are equal in Python?**
   * a) =
   * b) ==
   * c) !=
   * d) <>
   * **Answer:** b) ==
2. **What will be the output of the following code?**

 x = 10  
 if x > 5:  
     print("Greater")  
 else:  
     print("Lesser")

* + a) Greater
  + b) Lesser
  + c) Error
  + d) Nothing
  + **Answer:** a) Greater

1. **Which of the following is a valid elif statement in Python?**
   * a) elif (x == 10):
   * b) else if (x == 10):
   * c) elif x == 10 then:
   * d) else if x == 10:
   * **Answer:** a) elif (x == 10):
2. **Which of the following logical operators is used to check if both conditions are true?**
   * a) and
   * b) or
   * c) not
   * d) xor
   * **Answer:** a) and
3. **What is the output of the following code?**

x = 5  
y = 10  
if x > 3 and y < 15:  
    print("Condition True")  
else:  
    print("Condition False")

* + a) Condition True
  + b) Condition False
  + c) Error
  + d) No output
  + **Answer:** a) Condition True

**Loops in Python**

1. **Which of the following loops will execute the body at least once even if the condition is false?**
   * a) while loop
   * b) do-while loop
   * c) for loop
   * d) Python does not have a loop that guarantees at least one execution
   * **Answer:** d) Python does not have a loop that guarantees at least one execution
2. **What will be the output of the following code?**

for i in range(2, 5):  
    print(i)

* + a) 1 2 3 4
  + b) 2 3 4
  + c) 2 3 4 5
  + d) 0 1 2 3 4
  + **Answer:** b) 2 3 4

1. **Which of the following is the correct syntax to create an infinite loop in Python?**
   * a) for i in range():
   * b) while True:
   * c) while (i > 0):
   * d) for i in range(infinite):
   * **Answer:** b) while True:
2. **What will be the output of the following code?**

for i in range(5):  
    if i == 3:  
        continue  
    print(i)

* + a) 0 1 2 4
  + b) 0 1 2 3 4
  + c) 0 1 2
  + d) 3 4
  + **Answer:** a) 0 1 2 4

**More on Data Types**

1. **Which of the following methods is used to remove an item from a list by its index?**
   * a) remove()
   * b) delete()
   * c) pop()
   * d) clear()
   * **Answer:** c) pop()
2. **Which of the following is an immutable data type in Python?**
   * a) List
   * b) Set
   * c) Tuple
   * d) Dictionary
   * **Answer:** c) Tuple
3. **What will be the output of the following code?**

s = {1, 2, 3, 4}  
s.add(5)  
print(s)

* + a) {1, 2, 3, 4, 5}
  + b) [1, 2, 3, 4, 5]
  + c) (1, 2, 3, 4, 5)
  + d) Error
  + **Answer:** a) {1, 2, 3, 4, 5}

**Functions and Modules**

1. **Which keyword is used to define a function in Python?**
   * a) func
   * b) function
   * c) def
   * d) define
   * **Answer:** c) def
2. **What is the purpose of the return statement in a function?**
   * a) To define the function body
   * b) To terminate the function execution
   * c) To exit the program
   * d) To return a value from the function to the caller
   * **Answer:** d) To return a value from the function to the caller
3. **Which of the following functions returns the length of a list in Python?**
   * a) length()
   * b) len()
   * c) size()
   * d) count()
   * **Answer:** b) len()

**NumPy**

1. **Which of the following methods is used to reshape a NumPy array?**
   * A) np.resize()
   * B) np.shape()
   * C) np.reshape()
   * D) np.reform()
   * *Answer:\** C) np.reshape()
2. **How do you concatenate two NumPy arrays along a specific axis?**
   * A) np.concat()
   * B) np.concatenate()
   * C) np.merge()
   * D) np.append()
   * *Answer:\** B) np.concatenate()
3. **Which of the following functions can be used to generate a sequence of numbers in NumPy with a fixed step size?**
   * A) np.linspace()
   * B) np.arange()
   * C) np.range()
   * D) np.step()
   * *Answer:\** B) np.arange()
4. **What will np.array([1, 2, 3]) + 5 return?**
   * A) [1, 2, 3, 5]
   * B) [6, 7, 8]
   * C) 5
   * D) None
   * *Answer:\** B) [6, 7, 8]
5. **Which function in NumPy can be used to compute the dot product of two arrays?**
   * A) np.multiply()
   * B) np.product()
   * C) np.dot()
   * D) np.cross()
   * *Answer:\** C) np.dot()

**Pandas**

1. **Which Pandas function is used to convert a DataFrame to a NumPy array?**
   * A) to\_numpy()
   * B) convert()
   * C) np.array()
   * D) to\_ndarray()
   * *Answer:\** A) to\_numpy()
2. **How can you rename columns in a Pandas DataFrame?**
   * A) df.columns.rename()
   * B) df.rename(columns=...)
   * C) df.rename\_col()
   * D) df.modify\_columns()
   * *Answer:\** B) df.rename(columns=...)
3. **Which of the following methods is used to concatenate multiple DataFrames?**
   * A) concat()
   * B) join()
   * C) merge()
   * D) append()
   * *Answer:\** A) concat()
4. **What does the value\_counts() function in Pandas do?**
   * A) It returns the unique values in a DataFrame
   * B) It counts the frequency of unique values in a column
   * C) It counts the number of rows in a DataFrame
   * D) It counts the non-null values in a column
   * *Answer:\** B) It counts the frequency of unique values in a column
5. **How can you sort a Pandas DataFrame by values in a specific column?**
   * A) df.sort()
   * B) df.sort\_values()
   * C) df.order()
   * D) df.arrange()
   * *Answer:\** B) df.sort\_values()

**Matplotlib**

1. **Which argument in the plot() function is used to specify the color of the line in Matplotlib?**
   * A) color
   * B) linecolor
   * C) c
   * D) linestyle
   * *Answer:\** A) color
2. **How do you set the labels for the x and y axes in Matplotlib?**
   * A) plt.axis()
   * B) plt.xlabel() and plt.ylabel()
   * C) plt.label()
   * D) plt.axes()
   * *Answer:\** B) plt.xlabel() and plt.ylabel()
3. **What is the purpose of the legend() function in Matplotlib?**
   * A) To set the title of the plot
   * B) To display a legend for labeled data
   * C) To format the axes
   * D) To add a grid to the plot
   * *Answer:\** B) To display a legend for labeled data
4. **Which parameter in the scatter() function sets the size of the markers?**
   * A) s
   * B) size
   * C) scale
   * D) marker\_size
   * *Answer:\** A) s

**Seaborn**

1. **Which Seaborn function is used to create box plots?**
   * A) boxplot()
   * B) barplot()
   * C) violinplot()
   * D) scatterplot()
   * *Answer:\** A) boxplot()
2. **Which Seaborn function is used to create a bar plot?**
   * A) countplot()
   * B) barplot()
   * C) lineplot()
   * D) catplot()
   * *Answer:\** B) barplot()
3. **How do you change the color palette in Seaborn plots?**
   * A) sns.color\_palette()
   * B) sns.set\_palette()
   * C) sns.change\_palette()
   * D) sns.use\_palette()
   * *Answer:\** B) sns.set\_palette()
4. **What does the pairplot() function in Seaborn do?**
   * A) Creates a pair of bar plots
   * B) Creates a matrix of scatter plots for pairwise variable comparisons
   * C) Creates a heatmap for correlations
   * D) Creates a single line plot for two variables
   * *Answer:\** B) Creates a matrix of scatter plots for pairwise variable comparisons

**EDA (Exploratory Data Analysis)**

1. **What does the .corr() method in Pandas return?**
   * A) Covariance of variables
   * B) Correlation matrix of numerical columns
   * C) Counts of unique values
   * D) The range of numerical values
   * *Answer:\** B) Correlation matrix of numerical columns
2. **In which step of EDA is the identification of outliers most important?**
   * A) Data collection
   * B) Data visualization
   * C) Data cleaning
   * D) Data modeling
   * *Answer:\** B) Data visualization

**Introduction to Python**

1. **Explain the key features of Python that make it popular for programming.**
2. **Describe the process of installing Python on your system and running a basic Python program using IDLE.**

**Data Types & Variables**

1. **Discuss the differences between mutable and immutable data types in Python. Provide examples of each.**
2. **Explain the concept of type casting in Python. How can you convert between different data types?**
3. **What are literals in Python? Explain with examples.**

**Conditional Statements**

1. **Write a Python program that checks if a number is positive, negative, or zero using if...elif...else statements.**
2. **Explain the difference between the if, elif, and else statements in Python. Provide a code example.**
3. **What are comparison and logical operators in Python? Explain their role in conditional statements.**

**Loops in Python**

1. **Describe the difference between a for loop and a while loop in Python. When would you use each?**
2. **Explain the purpose of the break and continue statements in Python loops. Provide examples of each.**
3. **Write a Python program that prints the Fibonacci series up to 50 using a while loop.**
4. **What is a nested loop in Python? Write a program that uses nested loops to print a pattern (e.g., a triangle of stars).**

**More on Data Types**

1. **Describe the differences between lists, tuples, sets, and dictionaries in Python. Provide examples of when to use each.**
2. **What are list comprehensions in Python? Write a program that uses list comprehensions to create a list of squares of numbers from 1 to 10.**
3. **Explain the concept of string slicing in Python. Provide examples of how to extract substrings from a given string.**

**Functions and Modules**

1. **Write a Python function that takes a list of numbers as input and returns the sum of all even numbers in the list.**
2. **Explain the concept of variable scope in Python. What is the difference between local and global variables?**
3. **What are lambda functions in Python? Provide examples of how they can be used in conjunction with functions like map() and filter().**
4. **Explain the concept of recursion in Python. Write a recursive function to calculate the factorial of a given number.**

**Object-Oriented Programming (OOP)**

1. **What is Object-Oriented Programming (OOP) in Python? Explain the key principles of OOP, such as inheritance, polymorphism, and encapsulation, with examples.**

**Introduction to Python**

1. **Compare Python with other programming languages like Java and C++. What are the major differences and similarities?**
2. **Explain the role of Python in data science and machine learning. How is Python used in these fields?**

**Data Types & Variables**

1. **Discuss the significance of variables in Python. How are variables declared and assigned in Python compared to other programming languages?**
2. **Write a Python program to demonstrate the difference between shallow and deep copy for lists.**
3. **Explain how Python handles memory management with respect to variable assignment and data types.**

**Conditional Statements**

1. **Explain the concept of short-circuit evaluation in Python with respect to logical operators. How does it impact conditional statements?**
2. **Write a Python program that determines if a year is a leap year using conditional statements.**

**Loops in Python**

1. **What are infinite loops? Explain how you can avoid creating infinite loops in Python.**
2. **Write a Python program that calculates the factorial of a number using a for loop.**
3. **Describe how the else statement works with loops in Python. Write a program to demonstrate its usage.**

**More on Data Types**

1. **What is a dictionary in Python? Explain how you can add, update, and remove items in a dictionary with examples.**
2. **Compare tuples and lists in Python. What are the key differences, and when should you prefer one over the other?**
3. **Write a Python program that counts the frequency of elements in a list using a dictionary.**
4. **Explain the difference between a set and a list. How would you use a set to remove duplicates from a list?**
5. **What are the different string methods in Python? Write a program that demonstrates how to manipulate strings using these methods.**

**Functions and Modules**

1. **Explain the difference between positional and keyword arguments in Python functions. Provide examples for each.**
2. **What is the purpose of \*args and `**kwargs` in Python functions? Write a function to demonstrate their use.\*\*
3. **How can you create and import user-defined modules in Python? Write a Python script to demonstrate module creation and usage.**
4. **Discuss the importance of documentation in Python functions. How can you document a function and view its docstring?**

**Object-Oriented Programming (OOP)**

1. **What is inheritance in Python? Write a Python program to demonstrate single and multiple inheritance.**

**NumPy**

1. **Explain the role of NumPy in Python and its advantages over Python lists.**
2. **How would you create a NumPy array of zeros, ones, and random numbers? Provide examples.**
3. **What is broadcasting in NumPy, and how does it work with arrays of different shapes?**
4. **Write a Python program to find the mean, median, and standard deviation of a NumPy array.**
5. **Explain how slicing and indexing work in NumPy arrays with an example. How is it different from Python lists?**

**Pandas**

1. **What is the primary purpose of the Pandas library, and how does it facilitate data analysis?**
2. **How can you handle missing data in Pandas? Write a Python script to fill or drop missing values in a DataFrame.**
3. **Explain the difference between loc[] and iloc[] in Pandas. Provide examples for each.**
4. **Write a Python program using Pandas to load a CSV file and perform basic data exploration (e.g., shape, summary statistics, and column information).**
5. **How can you perform groupby operations in Pandas, and why are they useful? Provide an example.**

**Matplotlib**

1. **What is the purpose of the Matplotlib library in Python, and how does it differ from Seaborn?**
2. **Write a Python program to create a simple line plot using Matplotlib. Customize the title, labels, and grid.**
3. **How can you create subplots in Matplotlib? Demonstrate with an example where you display a line plot and a bar plot side by side.**
4. **Explain the different types of charts you can create using Matplotlib. Provide code for creating a histogram and a scatter plot.**

**Seaborn**

1. **What is Seaborn, and how does it extend Matplotlib for data visualization?**
2. **Write a Python program to create a heatmap using Seaborn, and explain its use cases in EDA.**
3. **How can you visualize relationships between multiple variables in a dataset using Seaborn’s pairplot()? Provide an example.**
4. **Explain the difference between a barplot() and a countplot() in Seaborn. When should you use each?**

**Exploratory Data Analysis (EDA)**

1. **What is Exploratory Data Analysis (EDA), and why is it an important step in the data science process?**
2. **Write a Python program that demonstrates the process of EDA on a dataset using Pandas, Matplotlib, and Seaborn. Include steps such as data cleaning, visualizing distributions, and identifying correlations.**

Here are the \*\*answers to the 20 subjective questions\*\* on Python:

---

### \*\*Introduction to Python\*\*

1. \*\*Compare Python with other programming languages like Java and C++. What are the major differences and similarities?\*\*

- \*\*Answer:\*\* Python is an interpreted, high-level, and dynamically typed language, whereas Java and C++ are compiled, statically typed languages. Python is known for its simplicity and readability, while Java has stricter syntax and C++ provides more control over hardware with low-level features. Python does not require explicit declaration of variables, unlike Java and C++. In terms of similarities, all three support object-oriented programming, have large standard libraries, and are widely used in software development.

2. \*\*Explain the role of Python in data science and machine learning. How is Python used in these fields?\*\*

- \*\*Answer:\*\* Python plays a significant role in data science and machine learning due to its ease of use, readability, and extensive libraries like NumPy, pandas, scikit-learn, TensorFlow, and PyTorch. These libraries simplify tasks such as data manipulation, statistical analysis, and building machine learning models, making Python the preferred language in these domains.

---

### \*\*Data Types & Variables\*\*

3. \*\*Discuss the significance of variables in Python. How are variables declared and assigned in Python compared to other programming languages?\*\*

- \*\*Answer:\*\* In Python, variables are created by assignment (e.g., `x = 5`). Python is dynamically typed, so there is no need to declare the type of a variable explicitly. This differs from statically typed languages like C++ and Java, where variable types must be declared (e.g., `int x = 5;` in C++). Variables in Python can hold any type of data and can be reassigned to different types during program execution.

4. \*\*Write a Python program to demonstrate the difference between shallow and deep copy for lists.\*\*

- \*\*Answer:\*\*

```python

import copy

# Original list

original\_list = [[1, 2, 3], [4, 5, 6]]

# Shallow copy

shallow\_copy = copy.copy(original\_list)

# Deep copy

deep\_copy = copy.deepcopy(original\_list)

# Modifying original list

original\_list[0][0] = 'X'

print("Original List:", original\_list)

print("Shallow Copy:", shallow\_copy)

print("Deep Copy:", deep\_copy)

```

In a shallow copy, changes to the original list affect the copied list if the elements are mutable objects like lists within lists. In a deep copy, changes to the original list do not affect the copied list.

5. \*\*Explain how Python handles memory management with respect to variable assignment and data types.\*\*

- \*\*Answer:\*\* Python handles memory management through an internal system that includes automatic garbage collection. Variables are references to objects in memory, and when a variable is reassigned, Python's memory manager ensures that unused objects are automatically cleaned up (garbage collected). Python uses reference counting to track how many references an object has, and when an object's reference count reaches zero, it is deallocated.

---

### \*\*Conditional Statements\*\*

6. \*\*Explain the concept of short-circuit evaluation in Python with respect to logical operators. How does it impact conditional statements?\*\*

- \*\*Answer:\*\* Short-circuit evaluation means that Python stops evaluating a logical expression as soon as the result is determined. For the `and` operator, if the first condition is `False`, the whole expression is `False`, and the second condition is not evaluated. For the `or` operator, if the first condition is `True`, the whole expression is `True`, and the second condition is not evaluated. This can improve efficiency in conditional statements.

7. \*\*Write a Python program that determines if a year is a leap year using conditional statements.\*\*

- \*\*Answer:\*\*

```python

year = int(input("Enter a year: "))

if (year % 4 == 0 and year % 100 != 0) or (year % 400 == 0):

print(f"{year} is a leap year")

else:

print(f"{year} is not a leap year")

```

---

### \*\*Loops in Python\*\*

8. \*\*What are infinite loops? Explain how you can avoid creating infinite loops in Python.\*\*

- \*\*Answer:\*\* An infinite loop occurs when the condition in a `while` or `for` loop never becomes `False`, causing the loop to run endlessly. To avoid infinite loops, ensure that the loop's condition is modified within the loop so that it will eventually evaluate to `False`. For example, make sure counters are incremented or conditions are updated correctly.

9. \*\*Write a Python program that calculates the factorial of a number using a `for` loop.\*\*

- \*\*Answer:\*\*

```python

num = int(input("Enter a number: "))

factorial = 1

for i in range(1, num + 1):

factorial \*= i

print(f"Factorial of {num} is {factorial}")

```

10. \*\*Describe how the `else` statement works with loops in Python. Write a program to demonstrate its usage.\*\*

- \*\*Answer:\*\* The `else` clause in a loop is executed after the loop finishes running, but only if the loop was not terminated by a `break` statement. For example:

```python

for i in range(5):

print(i)

else:

print("Loop completed without break")

```

---

### \*\*More on Data Types\*\*

11. \*\*What is a dictionary in Python? Explain how you can add, update, and remove items in a dictionary with examples.\*\*

- \*\*Answer:\*\* A dictionary in Python is a collection of key-value pairs. You can add an item by assigning a value to a new key, update an item by reassigning a value to an existing key, and remove an item using the `del` keyword or `pop()` method.

```python

my\_dict = {"a": 1, "b": 2}

my\_dict["c"] = 3 # Add

my\_dict["a"] = 10 # Update

del my\_dict["b"] # Remove

```

12. \*\*Compare tuples and lists in Python. What are the key differences, and when should you prefer one over the other?\*\*

- \*\*Answer:\*\* The key difference is that lists are mutable, meaning their contents can be changed, while tuples are immutable. Use tuples when you want a collection of items that should not change (e.g., coordinates), and lists when you need a dynamic data structure.

13. \*\*Write a Python program that counts the frequency of elements in a list using a dictionary.\*\*

- \*\*Answer:\*\*

```python

lst = [1, 2, 2, 3, 3, 3]

freq\_dict = {}

for item in lst:

freq\_dict[item] = freq\_dict.get(item, 0) + 1

print(freq\_dict)

```

14. \*\*Explain the difference between a set and a list. How would you use a set to remove duplicates from a list?\*\*

- \*\*Answer:\*\* A set is an unordered collection of unique elements, while a list can contain duplicates. To remove duplicates from a list:

```python

lst = [1, 2, 2, 3, 3]

unique\_lst = list(set(lst))

```

15. \*\*What are the different string methods in Python? Write a program that demonstrates how to manipulate strings using these methods.\*\*

- \*\*Answer:\*\* Python has several string methods like `upper()`, `lower()`, `strip()`, `replace()`, `find()`, and `split()`. Example:

```python

text = " Hello World "

print(text.strip().lower().replace("hello", "hi"))

```

---

### \*\*Functions and Modules\*\*

16. \*\*Explain the difference between positional and keyword arguments in Python functions. Provide examples for each.\*\*

- \*\*Answer:\*\* Positional arguments are passed in the same order as parameters are defined. Keyword arguments are passed by explicitly specifying the parameter name, which allows passing arguments in any order.

```python

def greet(name, age):

print(f"Hello {name}, you are {age} years old")

greet("Alice", 25) # Positional

greet(age=30, name="Bob") # Keyword

```

17. \*\*What is the purpose of `\*args` and `\*\*kwargs` in Python functions? Write a function to demonstrate their use.\*\*

- \*\*Answer:\*\* `\*args` allows a function to accept a variable number of positional arguments, and `\*\*kwargs` allows a function to accept a variable number of keyword arguments.

```python

def my\_function(\*args, \*\*kwargs):

print("Positional arguments:", args)

print("Keyword arguments:", kwargs)

my\_function(1, 2, 3, name="Alice", age=25)

```

18. \*\*How can you create and import user-defined modules in Python? Write a Python script to demonstrate module creation and usage.\*\*

- \*\*Answer:\*\* A module is a Python file containing functions and variables. To create a module, save a Python script with functions, then import it into another script.

```python

# my\_module.py

def greet(name):

print(f"Hello, {name}!")

# main.py

import my\_module

my\_module.greet("Alice")

```

19. \*\*Discuss the importance of documentation in Python functions. How can you document a function and view its docstring?\*\*

- \*\*Answer:\*\* Documentation is crucial for code readability and maintainability. You can document a Python function using docstrings, which are written within triple quotes just after the function definition. The docstring can be viewed using the `help()` function or `\_\_doc\_\_` attribute.

```python

def greet(name):

"""This function greets the person with the given name."""

print(f"Hello, {name}!")

print(greet.\_\_doc\_\_)

```

---

### \*\*Object-Oriented Programming (OOP)\*\*

20. \*\*What is inheritance in Python? Write a Python program to demonstrate single and multiple inheritance.\*\*

- \*\*Answer:\*\*

Inheritance is an OOP concept where a new class (derived/child class) inherits attributes and methods from an existing class (base/parent class).

```python

# Single Inheritance

class Animal:

def speak(self):

print("Animal speaks")

class Dog(Animal):

def bark(self):

print("Dog barks")

dog = Dog()

dog.speak() # Inherited method

dog.bark()

# Multiple Inheritance

class Bird:

def fly(self):

print("Bird flies")

class FlyingDog(Dog, Bird):

pass

flying\_dog = FlyingDog()

flying\_dog.bark()

flying\_dog.fly()

```

---

These answers provide in-depth explanations and examples for each question, covering a range of Python topics such as functions, loops, data types, and object-oriented programming.

**SQL Setup and Basics**

1. **Which of the following is a relational database management system?**
   * A) Python
   * B) MySQL
   * C) HTML
   * D) Git
   * *Answer:\** B) MySQL
2. **What does SQL stand for?**
   * A) Structured Query Language
   * B) Simplified Query Language
   * C) Sequential Query Language
   * D) Systematic Query Language
   * *Answer:\** A) Structured Query Language
3. **Which of the following is NOT a SQL data type?**
   * A) INT
   * B) VARCHAR
   * C) DECIMAL
   * D) STRING
   * *Answer:\** D) STRING
4. **In SQL, what is a collection of related data called?**
   * A) Database
   * B) Table
   * C) Row
   * D) Column
   * *Answer:\** A) Database
5. **Which SQL statement is used to extract data from a database?**
   * A) GET
   * B) EXTRACT
   * C) SELECT
   * D) FETCH
   * *Answer:\** C) SELECT

**Crafting SQL Databases (DDL) & Mastering Data Control (DML)**

1. **Which SQL command is used to create a new table?**
   * A) CREATE TABLE
   * B) ADD TABLE
   * C) MAKE TABLE
   * D) INSERT TABLE
   * *Answer:\** A) CREATE TABLE
2. **What does the INSERT INTO statement do in SQL?**
   * A) It updates existing records
   * B) It adds new records to a table
   * C) It deletes records
   * D) It creates a new table
   * *Answer:\** B) It adds new records to a table
3. **Which SQL statement is used to modify the structure of an existing table?**
   * A) MODIFY TABLE
   * B) CHANGE TABLE
   * C) ALTER TABLE
   * D) UPDATE TABLE
   * *Answer:\** C) ALTER TABLE
4. **Which of the following is NOT a DDL command?**
   * A) CREATE
   * B) INSERT
   * C) ALTER
   * D) DROP
   * *Answer:\** B) INSERT
5. **In SQL, which constraint ensures that all values in a column are different?**
   * A) NOT NULL
   * B) UNIQUE
   * C) PRIMARY KEY
   * D) DEFAULT
   * *Answer:\** B) UNIQUE

**Querying into Data (DQL)**

1. **Which SQL clause is used to filter the result set based on a condition?**
   * A) GROUP BY
   * B) HAVING
   * C) WHERE
   * D) ORDER BY
   * *Answer:\** C) WHERE
2. **Which operator in SQL is used to check for a range of values?**
   * A) IN
   * B) LIKE
   * C) BETWEEN
   * D) NOT
   * *Answer:\** C) BETWEEN
3. **What does the DISTINCT keyword do in a SQL query?**
   * A) Selects unique values from a column
   * B) Deletes duplicate values
   * C) Counts the total rows in a table
   * D) Groups values together
   * *Answer:\** A) Selects unique values from a column
4. **What will SELECT \* FROM Employees WHERE salary IS NULL; do?**
   * A) Retrieve rows where salary is equal to 0
   * B) Retrieve rows where salary is empty or not entered
   * C) Retrieve all rows in the table
   * D) Retrieve rows where salary is not empty
   * *Answer:\** B) Retrieve rows where salary is empty or not entered
5. **Which SQL clause is used to sort the result-set in ascending or descending order?**
   * A) GROUP BY
   * B) ORDER BY
   * C) WHERE
   * D) HAVING
   * *Answer:\** B) ORDER BY

**Unlocking Built-in Functions**

1. **Which SQL function is used to count the number of rows in a table?**
   * A) COUNT()
   * B) SUM()
   * C) AVG()
   * D) MAX()
   * *Answer:\** A) COUNT()
2. **What will the query SELECT UPPER(name) FROM Employees; return?**
   * A) Lowercase values of the "name" column
   * B) Capitalized first letters in the "name" column
   * C) Uppercase values of the "name" column
   * D) Encrypted values of the "name" column
   * *Answer:\** C) Uppercase values of the "name" column
3. **Which function is used to return the current date and time in MySQL?**
   * A) CURRDATE()
   * B) TIME()
   * C) NOW()
   * D) DATETIME()
   * *Answer:\** C) NOW()
4. **What does the GROUP BY clause do in SQL?**
   * A) It sorts rows in ascending order
   * B) It groups rows that have the same values into summary rows
   * C) It filters rows based on a condition
   * D) It limits the number of rows returned
   * *Answer:\** B) It groups rows that have the same values into summary rows
5. **What is the purpose of the HAVING clause in SQL?**
   * A) To apply a filter on aggregate functions
   * B) To sort data in descending order
   * C) To return rows with duplicate values
   * D) To limit the number of rows returned
   * *Answer:\** A) To apply a filter on aggregate functions

**Crafting SQL Databases (DDL) & Mastering Data Control (DML)**

1. **What happens when you issue the DROP TABLE statement in SQL?**
   * A) The table data is deleted, but the table structure remains.
   * B) The table structure and data are permanently deleted.
   * C) The table is archived but can be restored.
   * D) The table data is hidden but not deleted.
   * *Answer:\** B) The table structure and data are permanently deleted.
2. **Which of the following SQL commands is used to modify an existing row in a table?**
   * A) MODIFY
   * B) UPDATE
   * C) ALTER
   * D) REPLACE
   * *Answer:\** B) UPDATE
3. **Which SQL command is used to remove records from a table without removing the table itself?**
   * A) DROP
   * B) TRUNCATE
   * C) DELETE
   * D) CLEAR
   * *Answer:\** C) DELETE
4. **What is a foreign key in SQL?**
   * A) A key used to uniquely identify each row in a table.
   * B) A key used to identify rows in another table.
   * C) A key used to reference the primary key of another table.
   * D) A key used to limit duplicates in a table.
   * *Answer:\** C) A key used to reference the primary key of another table.
5. **Which SQL command is used to change an existing table's name?**
   * A) MODIFY TABLE
   * B) ALTER TABLE
   * C) RENAME TABLE
   * D) CHANGE TABLE
   * *Answer:\** C) RENAME TABLE

**Querying into Data (DQL)**

1. **What will the SQL query SELECT name, age FROM Students WHERE age > 20; return?**
   * A) All rows where age is less than 20
   * B) All students with an age greater than 20
   * C) All students with an age of 20
   * D) All students in the Students table
   * *Answer:\** B) All students with an age greater than 20
2. **Which operator would you use in a SQL query to find rows containing a specific pattern in a column?**
   * A) BETWEEN
   * B) LIKE
   * C) IN
   * D) EXISTS
   * *Answer:\** B) LIKE
3. **What does the SELECT COUNT(\*) query return?**
   * A) The sum of values in all rows
   * B) The number of rows in a table
   * C) The total number of distinct rows
   * D) The maximum value in a column
   * *Answer:\** B) The number of rows in a table
4. **Which SQL clause is used to group rows that have the same values in specified columns?**
   * A) GROUP BY
   * B) HAVING
   * C) WHERE
   * D) ORDER BY
   * *Answer:\** A) GROUP BY
5. **Which SQL operator is used to combine multiple conditions in a WHERE clause?**
   * A) AND
   * B) JOIN
   * C) IF
   * D) MATCH
   * *Answer:\** A) AND

**SQL Functions**

1. **Which function in SQL is used to calculate the average value of a numeric column?**
   * A) AVG()
   * B) MEAN()
   * C) SUM()
   * D) MEDIAN()
   * *Answer:\** A) AVG()
2. **How can you extract the year from a date in SQL?**
   * A) EXTRACT(YEAR FROM date\_column)
   * B) DATE(YEAR date\_column)
   * C) YEAR(date\_column)
   * D) GETYEAR(date\_column)
   * *Answer:\** A) EXTRACT(YEAR FROM date\_column)
3. **What does the SQL function LEN() return?**
   * A) The total number of rows in a table
   * B) The length of a string in a column
   * C) The total size of a database
   * D) The distinct values in a column
   * *Answer:\** B) The length of a string in a column
4. **Which SQL aggregate function is used to find the largest value in a column?**
   * A) MAX()
   * B) MIN()
   * C) GREATEST()
   * D) HIGHEST()
   * *Answer:\** A) MAX()
5. **What is the difference between HAVING and WHERE clauses in SQL?**
   * A) HAVING is used before WHERE.
   * B) HAVING is used with aggregate functions, WHERE is not.
   * C) WHERE is used to filter after aggregation, HAVING is not.
   * D) HAVING only works with numeric data types.
   * *Answer:\** B) HAVING is used with aggregate functions, WHERE is not.

**Joins and Subqueries**

1. **Which of the following SQL keywords is used to combine rows from two or more tables based on a related column?**
   * A) MERGE
   * B) JOIN
   * C) COMBINE
   * D) ATTACH
   * *Answer:\** B) JOIN
2. **Which type of SQL JOIN returns all records from the left table and matching records from the right table, with NULL where there are no matches?**
   * A) INNER JOIN
   * B) FULL JOIN
   * C) LEFT JOIN
   * D) CROSS JOIN
   * *Answer:\** C) LEFT JOIN
3. **What does the following query return? SELECT \* FROM Employees E WHERE EXISTS (SELECT \* FROM Departments D WHERE E.dept\_id = D.dept\_id);**
   * A) All employees with a matching department
   * B) All employees without a department
   * C) All departments with employees
   * D) All employees who do not belong to a department
   * *Answer:\** A) All employees with a matching department
4. **Which type of subquery can be used inside a SELECT statement to return a single value?**
   * A) Correlated subquery
   * B) Scalar subquery
   * C) Multi-row subquery
   * D) Independent subquery
   * *Answer:\** B) Scalar subquery
5. **In SQL, which type of JOIN returns only the records that have matching values in both tables?**
   * A) LEFT JOIN
   * B) RIGHT JOIN
   * C) FULL JOIN
   * D) INNER JOIN
   * *Answer:\** D) INNER JOIN

**1. Explain the different data types available in SQL.**

* Discuss the primary data types (e.g., INT, VARCHAR, DATE, etc.) and how they are used in creating tables and managing data in SQL.

**2. Describe the purpose of the CREATE and ALTER statements in SQL.**

* Explain how to create a new table and how to modify an existing table using SQL commands.

**3. What are constraints in SQL, and why are they important? Give examples of different constraints.**

* Explain constraints like PRIMARY KEY, FOREIGN KEY, UNIQUE, NOT NULL, and CHECK, and describe how they ensure data integrity.

**4. Differentiate between the DELETE, TRUNCATE, and DROP statements in SQL.**

* Explain the differences between these commands and the scenarios in which each one would be used.

**5. How does the GROUP BY clause work in SQL? Provide an example where GROUP BY is used along with an aggregate function.**

* Describe how data is grouped using GROUP BY and demonstrate its use in conjunction with functions like COUNT() or SUM().

**6. Explain the concept of joins in SQL. What are the different types of joins, and how do they differ from each other?**

* Discuss INNER JOIN, LEFT JOIN, RIGHT JOIN, and FULL JOIN with examples to illustrate how they retrieve data from multiple tables.

**7. What is a subquery in SQL? Provide an example of a query that uses a subquery to filter data.**

* Describe what subqueries are and provide an example where a subquery is used within a SELECT or WHERE clause.

**8. Describe how the HAVING clause differs from the WHERE clause. Provide an example.**

* Explain the difference between these clauses, focusing on how HAVING is used with aggregate functions.

**9. What is a view in SQL, and why would you use it? Provide an example of creating a view.**

* Explain the purpose of views, how they can be used to simplify queries or protect data, and show how to create a view.

**10. Explain how transactions work in SQL. What are the benefits of using transactions, and how do the COMMIT and ROLLBACK statements function?**

* Discuss the concept of transactions, atomicity, and how they ensure data integrity during multiple operations.

**11. What is a stored procedure in SQL? Provide an example of how to create and execute a stored procedure.**

* Define stored procedures, their advantages, and give a practical example of how to write and call a procedure.

**12. How does indexing work in SQL? What are the benefits and drawbacks of using indexes?**

* Explain how indexes improve query performance and discuss potential downsides, such as the overhead of maintaining indexes.

**13. Describe the difference between UNION and UNION ALL. Provide an example where each would be used.**

* Compare these two SQL operators and explain how UNION ALL includes duplicates, while UNION removes them.

**14. What are SQL triggers? Explain with an example how to create a trigger that automatically logs data changes.**

* Define triggers, their types (BEFORE, AFTER), and give an example of using a trigger for logging data changes.

**15. Discuss the role of the CASE statement in SQL. Provide an example of using CASE to return custom output based on column values.**

* Explain how conditional logic is applied in SQL queries using CASE statements, with a sample query.

**16. What are aggregate functions in SQL? Describe the most commonly used aggregate functions and provide examples.**

* Explain the purpose of aggregate functions such as SUM(), AVG(), COUNT(), MAX(), and MIN() with appropriate examples.

**17. How does the EXISTS operator work in SQL? Provide an example of using EXISTS in a query.**

* Explain the functionality of the EXISTS operator, typically used for checking the existence of rows in a subquery, with an example.

**18. What are the differences between scalar, inline, and multi-statement table-valued functions in SQL?**

* Compare these types of functions, how they are used, and provide examples of each type.

**19. Explain the concept of normalization in database design. What are the advantages of normalizing a database?**

* Discuss the process of organizing a database into tables to reduce redundancy and improve data integrity, focusing on different normal forms (1NF, 2NF, 3NF).

**20. How would you optimize an SQL query for better performance? Provide several techniques used in query optimization.**

* Discuss methods like indexing, using proper join conditions, limiting result sets, avoiding unnecessary subqueries, and analyzing query execution plans.

**1. What is the difference between a primary key and a unique key in SQL? Can a table have multiple unique keys? Explain with an example.**

* Discuss the differences in uniqueness and nullability, and provide an example of a table having both primary and unique keys.

**2. Explain the different types of subqueries in SQL. Provide examples of correlated and non-correlated subqueries.**

* Define and differentiate between correlated and non-correlated subqueries, including practical examples of each.

**3. How can SQL be used to perform data security? Discuss the roles of GRANT and REVOKE statements.**

* Explain how SQL handles permissions, and how GRANT and REVOKE control user access to databases and tables.

**4. What is the difference between a clustered index and a non-clustered index in SQL? When would you use one over the other?**

* Discuss how each index type affects table storage and retrieval, and when each type should be applied for performance optimization.

**5. Explain the concept of a database schema in SQL. How does a schema differ from a database, and what are its advantages?**

* Clarify the distinction between a database and a schema, and describe the benefits of using schemas to organize data.

**6. How do foreign keys maintain referential integrity in SQL databases? Provide an example of how a foreign key works across two related tables.**

* Explain referential integrity and demonstrate how a foreign key enforces it between parent and child tables.

**7. Describe how you would implement pagination in SQL queries to handle large datasets.**

* Discuss techniques for paginating query results using LIMIT and OFFSET, or equivalent clauses.

**8. What are the ACID properties of a transaction in SQL? Provide an example of a transaction that demonstrates these properties.**

* Define Atomicity, Consistency, Isolation, and Durability, and give an example of how a transaction fulfills these properties.

**9. What is the purpose of normalization in database design? Explain the first three normal forms (1NF, 2NF, 3NF) with examples.**

* Describe how normalization reduces redundancy and dependency, and provide practical examples of each normal form.

**10. How does SQL handle concurrency in transactions? Explain the concept of isolation levels and their impact on data consistency.**

* Discuss different isolation levels (READ UNCOMMITTED, READ COMMITTED, REPEATABLE READ, SERIALIZABLE) and how they manage concurrent transactions.

**11. What are materialized views, and how do they differ from regular views? When would you use a materialized view?**

* Define materialized views and explain the performance benefits they offer compared to standard views.

**12. Explain how the MERGE statement works in SQL. Provide an example of how to use it to update and insert data in a single operation.**

* Describe the MERGE statement, often used for performing UPSERT operations, and give a real-world example.

**13. What are user-defined functions (UDFs) in SQL? How do they differ from stored procedures, and when would you use each?**

* Compare UDFs and stored procedures, explaining when it’s more appropriate to use one over the other.

**14. Discuss the benefits and limitations of using stored procedures in SQL. Provide an example of a simple stored procedure.**

* Explain the advantages (e.g., reusability, security) and drawbacks (e.g., portability) of stored procedures, with a sample procedure.

**15. What is partitioning in SQL, and how can it improve query performance in large databases? Provide an example of range partitioning.**

* Explain the concept of partitioning (horizontal or vertical) and how it divides large datasets to improve performance.

**16. How would you write a recursive query in SQL using Common Table Expressions (CTE)? Provide an example.**

* Describe how recursive CTEs work and provide a query that demonstrates a recursive operation, such as calculating a hierarchical relationship.

**17. Explain the concept of a cursor in SQL. Provide an example where a cursor is useful in handling row-by-row operations.**

* Define cursors, their advantages in row-by-row data processing, and provide an example of using cursors in a loop.

**18. What is an execution plan in SQL, and how does it help in optimizing queries? Provide an example of how to interpret an execution plan.**

* Discuss the purpose of execution plans and explain how they can be used to identify query bottlenecks and optimize performance.

**19. How would you handle null values in SQL queries? Discuss the use of the COALESCE and NULLIF functions.**

* Explain how SQL handles null values, and describe how COALESCE and NULLIF can be used to manage or replace them.

**20. What are cross joins in SQL? How do they differ from other types of joins, and when would you use them? Provide an example.**

* Define cross joins, their purpose in creating Cartesian products, and compare them with inner and outer joins, including an example query.