viva answers

Descriptive Analytics: It involves summarizing historical data to gain insights, often through statistical analysis, visualization, or data aggregation, to understand what happened in the past.

Predictive Analytics: It uses historical data and statistical algorithms to forecast future outcomes or trends. It involves predictive modeling and analysis to anticipate what might happen in the future.

Prescriptive Analytics: It suggests actions to take advantage of predictions and insights. It goes beyond predicting future outcomes and recommends decisions or actions based on those predictions.

Strengths and Weaknesses as a Data Analyst: Strengths might include proficiency in data manipulation, statistical analysis, and problem-solving. Weaknesses could be lack of domain knowledge, limited experience with certain tools, or challenges in communicating complex findings effectively.

Common Data Analytics Tools: Some commonly used data analytics tools include R, Python, SQL, Tableau, Power BI, Excel, SAS, SPSS, and Apache Spark, among others.

Differences between Data Mining and Data Profiling: Data mining involves discovering patterns or relationships within large datasets, aiming to extract useful information. Data profiling, on the other hand, is about examining data quality, consistency, and structure to understand its characteristics, such as missing values or outliers.

Advantages and Drawbacks of R:
Advantages: Rich package ecosystem, statistical modeling capabilities, excellent visualization tools, active community support.
Drawbacks: Steeper learning curve for beginners, slower execution for certain tasks compared to compiled languages, memory management challenges.

Memory Limit of R: The memory limit in R is determined by the physical memory (RAM) available on the system. However, R might face limitations based on the system's architecture and settings.

Data Frame: A data frame is a two-dimensional data structure in R, similar to a table or spreadsheet. It consists of rows and columns where each column can be of a different data type.

Different Data Types in R: R includes various data types such as numeric, integer, character, factor, logical, date, and complex.

Finding Missing Values in R: You can find missing values in R using functions like is.na() or complete.cases() to identify and handle missing values in datasets.

Difference between Vector and List: Vectors contain elements of the same data type while lists can hold elements of different data types. Lists are also

recursive structures in R, allowing lists within lists.

Differences between Matrix and Data Frame: Matrices in R contain elements of the same data type arranged in rows and columns, whereas data frames can store different data types in columns similar to a table in a database.

Difference between summary() and str() Function: summary() provides summary statistics of numerical variables in a dataset, while str() displays the structure of an R object including data types and dimensions.

Using ggplot2: ggplot2 is an R package for data visualization based on the grammar of graphics. It involves creating plots by specifying data, aesthetics, and layers using the ggplot() function.

Concatenating Strings in R: Strings can be concatenated in R using the paste() function or using the paste0() function to join strings without any separator.

Disadvantages of R Programming: Slower execution speed for certain tasks, memory management issues with larger datasets, and a steep learning curve for beginners.

Comparison between R and Python: Both are popular programming languages in data analytics. R is preferred for statistical analysis and data visualization, while Python is versatile, used for data manipulation, machine learning, and web development.

Data Structures in R: R includes various data structures like vectors, matrices, lists, data frames, factors, arrays, and tables (via packages like data.table or tibble).

Use of the Melt Function: The melt() function in R (often in the reshape2 or tidyr package) transforms data from a wide format to a long format, making it easier for analysis or plotting.

Data Analysis Process: It involves defining the problem, data collection, data cleaning, exploratory data analysis (EDA), data modeling, interpreting results, and making data-driven decisions.

Data Cleansing: Data cleansing is the process of detecting and correcting errors or inconsistencies in a dataset, involving activities like handling missing values, removing duplicates, correcting errors, etc.

****

same answers in one word

****

Here are simplified answers to the viva questions in a straightforward method.

Descriptive Analytics: It's about summarizing what happened in the past using data.

Predictive Analytics: Predicts future outcomes based on historical data.

Prescriptive Analytics: Recommends actions based on predictions to make better decisions.

Strengths and Weaknesses as a Data Analyst: Good at problem-solving, might need more experience in certain areas or tools.

Common Data Analytics Tools: R, Python, Excel, Tableau, and SQL are commonly used tools.

Data Mining vs. Data Profiling: Data mining finds patterns in data, while data profiling checks data quality.

Advantages and Drawbacks of R:

Advantages: Lots of tools, graphs, and a strong community.
Drawbacks: A bit hard for beginners, slower for some tasks, and can face memory issues.

Memory Limit of R: Depends on computer memory (RAM).

Data Frame: It's like a table in R holding data in rows and columns.

Different Data Types in R: R has numbers, text, dates, and true/false types.

Finding Missing Values in R: Use functions like is.na() or complete.cases() to find missing data.

Difference between Vector and List: Vectors are for the same type of data, lists can have different types.

Differences between Matrix and Data Frame: Matrix has one data type, while a data frame can hold different types.

Difference between summary() and str() Function: summary() shows stats for numbers, str() shows data structure.

Using ggplot2: ggplot2 is for making cool charts in R.

Concatenating Strings in R: Use paste() or paste0() to join words.

Disadvantages of R Programming: Slower for big data and harder to learn initially.

Comparison between R and Python: R is great for stats, Python is versatile for

many things like web, data, and AI.

Data Structures in R: R has many structures like lists, tables, and arrays for different types of data.

Use of the Melt Function: melt() changes data from wide to long for better analysis.

Data Analysis Process: Starts by defining the problem, then collecting, cleaning, exploring, and modeling data.

Data Cleansing: Fixes errors or mistakes in data like missing values or duplicates.