

GOVERNMENT DEGREE COLLEGE, NANDIKOTKUR

DEPARTMENT OF COMPUTER APPLICATIONS

BIG DATA ANALYTICS USING R

Unit 1: Introduction to Big Data

Multiple Choice Questions (MCQs)

1. Big Data is commonly defined by which three characteristics? a) Volume, Velocity, and Visibility
b) Volume, Veracity, and Velocity
c) Volume, Veracity, and Variety
d) Volume, Velocity, and Variety

Answer: d)

2. Semi-structured data is best described as: a) Data that is fully organized into a predefined structure
b) Data that has no structure
c) Data that has some organizational properties but not fully structured
d) Data that is numeric and stored in databases

Answer: c)

3. The term "Big Data" became popular in: a) 1980s
b) 1990s
c) 2000s
d) 2010s

Answer: c)

4. Which of the following is a major challenge in Big Data management? a) Processing large amounts of data in real-time
b) Storing small datasets
c) Visualizing structured data
d) Organizing metadata

Answer: a)

Fill in the Blanks

1. Big Data analytics aims to uncover _____ from large datasets that were previously hidden or inaccessible.

Answer: insights

2. _____ data is information that is organized and easy to analyze, while _____ data lacks a predefined format.

Answer: Structured, Unstructured

3. Business Intelligence focuses on analyzing _____ data, while Big Data handles a wide range of data formats, including real-time information.

Answer: historical

Short Answer Questions

1. What are the key characteristics of Big Data, and why are they important?
2. Define unstructured data and give two examples of where it can be found.
3. What are the main differences between Business Intelligence (BI) and Big Data?

Essay Questions

1. Discuss the challenges faced in managing Big Data and the solutions that have been developed to address them.
2. Compare structured, semi-structured, and unstructured data, providing real-world examples of each.

Unit 2: Big Data Analytics

Multiple Choice Questions (MCQs)

1. Which of the following is not part of Big Data Analytics? a) Descriptive analytics
b) Predictive analytics
c) Prescriptive analytics
d) Legacy analytics

Answer: d)

2. The primary goal of Big Data Analytics is to: a) Reduce the volume of data
b) Improve business decisions based on data insights
c) Increase the speed of data processing
d) Store large amounts of data

Answer: b)

3. Which technology is most commonly used to process and analyze Big Data? a) Hadoop
b) MySQL
c) Relational Databases
d) JSON

Answer: a)

4. One of the biggest challenges in Big Data Analytics is: a) Collecting enough data
b) Finding skilled personnel to analyze the data
c) Lack of data storage solutions
d) Getting data in real-time

Answer: b)

Fill in the Blanks

1. Predictive analytics uses historical data to predict future _____.

Answer: trends

2. The _____ framework is widely used in Big Data Analytics for processing large datasets.

Answer: Hadoop

3. A major challenge in Big Data Analytics is integrating _____ data from various sources.

Answer: heterogeneous

Short Answer Questions

1. What is the difference between predictive and prescriptive analytics?
2. List and briefly explain the top challenges of Big Data Analytics.
3. Why is Hadoop important in Big Data Analytics?

Essay Questions

1. Discuss the role of Big Data Analytics in driving business innovation and transformation.
2. Explain the technologies required to meet the challenges posed by Big Data Analytics, with a focus on Hadoop, Spark, and cloud computing.

Unit 3: Introduction to R and Getting Started with R

Multiple Choice Questions (MCQs)

1. Which of the following is a feature of R that makes it suitable for statistical computing? a) Open-source and free
b) Lack of community support
c) Limited packages
d) Proprietary software
Answer: a)
2. In R, the data type used to store logical values is: a) Numeric
b) Logical
c) Character
d) Integer
Answer: b)
3. Which command in R is used to list the objects in the environment? a) ls()
b) objects()
c) view()
d) list()
Answer: a)
4. Which of the following is not a data type in R? a) Character
b) Array
c) Integer
d) Function
Answer: d)

Fill in the Blanks

1. R was originally developed for _____ computing and graphics.
Answer: statistical
2. The _____ function in R is used to display all objects stored in the workspace.
Answer: ls()
3. In R, vectors are _____ data structures that can hold elements of the same type.
Answer: homogeneous

Short Answer Questions

1. What are the main data types in R, and how are they used?
2. Explain the advantages of using R over other programming languages for data analysis.

3. How does coercion work in R? Provide an example.

Essay Questions

1. Discuss how R is used for data analysis, highlighting its main features and advantages over other tools like Python or SAS.
2. Explain control structures in R and their role in programming, using examples such as loops and conditional statements.

Unit 4: Exploring Data in R

Multiple Choice Questions (MCQs)

1. The function `summary()` in R is used to: a) Summarize a data frame
b) Create a plot
c) Find the length of a vector
d) Calculate standard deviation

Answer: a)

2. To load a CSV file into R, which function is commonly used? a) `load.csv()`
b) `read.csv()`
c) `write.csv()`
d) `import.csv()`

Answer: b)

3. What does the function `ncol()` return when applied to a data frame? a) Number of rows
b) Number of columns
c) Number of elements
d) Structure of the data frame

Answer: b)

4. Which of the following is used to check the structure of a data frame in R? a) `str()`
b) `dim()`
c) `summary()`
d) `head()`

Answer: a)

Fill in the Blanks

1. The function _____ is used to view the first few rows of a data frame in R.

Answer: head()

2. A data frame is a _____ structure that stores tabular data in R.

Answer: two-dimensional

Short Answer Questions

1. Explain the process of reading data from a CSV file into R.
2. How can you subset a data frame in R? Provide an example.
3. What is the difference between head() and tail() functions in R?

Essay Questions

1. Discuss the various ways to explore and manipulate data in R, using functions such as dim(), str(), summary(), and edit().
2. Provide a detailed explanation of how to load data from different file types (CSV, tab-separated, Excel) into R, with code examples.

Unit 5: Data Visualization using R

Multiple Choice Questions (MCQs)

1. To create a histogram in R, which function is used? a) hist()
b) barplot()
c) plot()
d) pie()
Answer: a)
2. Which file format can be directly imported into R for data analysis? a) CSV
b) JSON
c) Excel
d) All of the above
Answer: d)
3. To create a pie chart in R, the function used is: a) pie()
b) barplot()
c) chart()
d) scatter()
Answer: a)

4. Which of the following data formats is most commonly used for web data in R? a) XML
b) JSON
c) CSV
d) Excel

Answer: b)

Fill in the Blanks

1. In R, a _____ is used to represent categorical data in a circular chart format.

Answer: pie chart

2. To create a scatter plot in R, the function _____ is used.

Answer: plot()

3. R can read external data from various formats such as CSV, JSON, and _____ files.

Answer: XML

Short Answer Questions

1. How do you create a bar chart in R, and when would it be most useful?
2. What are the steps to import a CSV file into R for visualization?
3. Explain the difference between a histogram and a scatter plot in terms of data representation.

Long Answer Questions

1. Describe the process of creating different types of visualizations (e.g., histograms, scatter plots, and pie charts) in R and discuss when to use each type.
2. Explain how to import and visualize external data formats (like JSON and Excel files) in R, including common challenges and solutions.
3. Discuss the importance of customizing visualizations in R (titles, colors, labels) and describe how these customizations can enhance data interpretation.