

## Data Science MCQ Questions and Answer PDF

### 1. Which one is NOT from Phase 1 of Data Science Life Cycle

1. Learning the target domain
2. Developing initial hypothesis
3. Visualize initial hypothesis
4. Identifying key variables

**Answer: Visualize initial hypothesis**

### 2. Which of the following is the most important language for Data Science?

1. Ruby
2. R
3. Java
4. None

**Answer: R**

### 3. A collection of information about a related topic is referred to as a \_\_\_\_\_

1. Visualisation
2. Analysis
3. Conclusion
4. Data

**Answer: Visualisation**

### 4. To find the \_\_\_\_\_ you add up all the numbers and then divide by how many numbers you have.

1. Median
2. Mean
3. Mode
4. Range

**Answer: Mean**

### 5. Which of the following is performed by Data Scientist ?

1. Create reproducible code
2. Challenge results
3. Define the question
4. All of the above

**Answer: Challenge results**

### 6. Which is not a tool for Statistical Data Analysis?

1. Logistic Regression
2. Linear & Non-linear Regression
3. Histogram
4. ANOVA

**Answer: Histogram**

**7. What is the mean of test scores?{70, 70, 80, 85, 85, 90, 95, 95, 100, 100}**

1. 85, 95, and 100
2. 30
3. 87
4. None

**Answer: 87**

---

**8. Choose the correct keyword for this definition: A graphical representation of a data set**

1. Data Set
2. Investigative Cycle
3. Visualisation
4. None

**Answer: Visualisation**

**9. To find the \_\_\_\_\_ you put all numbers in order from least to greatest and find the number that is in the middle.**

1. Median
2. Mode
3. Mean
4. Range

**Answer: Median**

**10. R is an interpreted language so it can access through \_\_\_\_\_?**

1. Command line interpreter
2. Disk operating system
3. Operating system
4. User interface operating system

**Answer: Command line interpreter**

**11. Data has been collected on visitors' viewing habits at a bank's website. Which technique is used to identify pages commonly viewed during the same visit to the website?**

1. Clustering

2. Classification
3. Association Rules
4. Regression

**Answer: Association Rules**

**12. A relationship between two or more variables is referred to as a \_\_\_\_\_**

1. Trend
2. Spike
3. All of above
4. None of above

**Answer: Trend**

---

**13. A graphical representation of a data set is referred to as a \_\_\_\_\_**

1. Visualization
2. Data Set
3. Investigative Cycle
4. None

**Answer: Visualization**

**14. Which of the following step is performed by data scientist AFTER acquiring the data?**

1. Data Integration
2. Data Replication
3. Data Cleansing
4. All of the above

**Answer: Data Cleansing**

---

**15. Data that sits outside the trend is referred to as a \_\_\_\_\_**

1. Outlier
2. Trend
3. Spike
4. Both 1 & 2

**Answer: Both 1 & 2**

**16. Which of the following approach should be used to ask Data Analysis question?**

1. Find out the question which is to be answered
2. Find only one solution for particular problem
3. Find out answer from dataset without asking question
4. None

**Answer: Find out the question which is to be answered**

**17. Which of the following is NOT a machine learning algorithm?**

1. SVG
2. Random Forest
3. SVM
4. None

**Answer: SVG**

**18. What is Big Data?**

1. Data with the word 'big' in it
2. Data about people who are big
3. Data with a large size
4. Data made with a big purpose

**Answer: Data with a large size**

**19. What is R an implementation of?**

1. Logical Scoping
2. S Programming Language
3. Lexical Scoping
4. Q Programming Language

**Answer: S Programming Language**

**20. The 5 steps required to identify a problem and come up with a solution are referred to as the \_\_\_\_\_ Cycle**

1. Visualization
2. Investigative
3. Conclusion
4. None

**Answer: Investigative**

**21. Which of the following is characteristic of Processed Data?**

1. Hard to use for data analysis
2. Data is not ready for analysis
3. All steps should be noted
4. None of the above

**Answer: None of the above**

**22. Which was not mentioned as a latest trend tool\_\_\_\_\_**

1. Excel
2. Pentaho
3. SPSS
4. Notepad

**Answer: Notepad**

**23. Which of the following is one of the key data science skill ?**

1. Machine Learning
2. Statistics
3. Data Visualization
4. All of the above

**Answer: All of the above**

**24. Which of the following is not a stage in the Investigative Cycle?**

1. Investigate
2. Analysis
3. Conclusion
4. None

**Answer: Investigate**

**25. Vectors come in two parts \_\_\_\_\_ and \_\_\_\_\_**

1. Atomic vectors and list
2. Atomic vectors and array
3. Atomic vectors and matrix
4. None

**Answer: Atomic vectors and list**

**26. Choose the correct keyword for this definition: A collection of information about a related topic**

1. Trend
2. Spike
3. Data Set
4. None

**Answer: Data Set**

**27. The process of evaluating data through analytical and statistical tools.**

1. Data Mining
2. Data Exploration
3. Data Analysis
4. Data Visualization

**Answer: Data Analysis**

**28. Which of the following is key characteristic of hacker ?**

1. Willing to find answers on their own
2. Afraid to say they don't know the answer
3. Not Willing to find answers on their own
4. All of the mentioned

**Answer: Willing to find answers on their own**

**29. Which of the following characteristic of big data is relatively more concerned to data science ?**

1. Variety
2. Volume
3. Velocity
4. None

**Answer: Variety**

**30. R is an\_\_\_\_\_ programming language?**

1. GPL
2. Open source
3. Closed source
4. Definite source

**Answer: Open source**