

# Top 20 Data Science Coding Questions and Answers for 2023

[BEGINNER](#)[DATA EXPLORATION](#)[INTERVIEW QUESTIONS](#)[INTERVIEWS](#)

## Introduction

Data structures and algorithms are essential knowledge for every machine learning practitioner. They enable programmers to write efficient code, which is particularly valuable when working with large datasets. Aspiring candidates should have a solid understanding of these fundamentals, as data structure and algorithm questions are frequently posed in data science interviews. To help you prepare, here's a curated list of 15 commonly asked data coding questions.

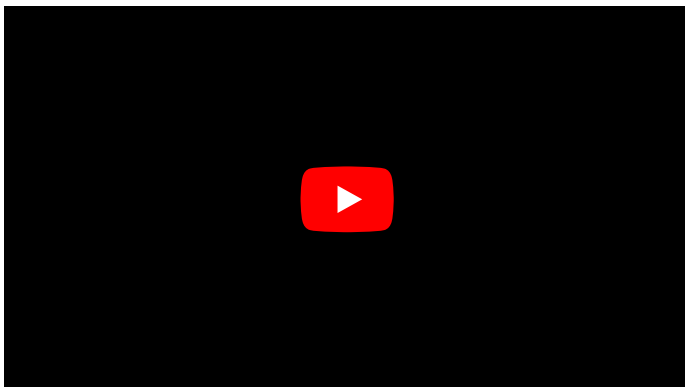
Test your skills by attempting these questions and assessing your proficiency!

*This article was published as a part of the [Data Science Blogathon](#).*

## Table of contents

- [Test your Data Science Knowledge](#)
- [Data Science MCQ Interview Questions](#)
- [Intermediate Data Science Coding Questions](#)
- [Expert Level Data Science Coding Questions](#)
- [Theory Based Data Science Interview Questions](#)
- [Conclusion](#)

## Test your Data Science Knowledge



Source: Analytics Vidhya Youtube Channel

Also Read: [5 Free Data Science Projects With Solutions](#)

## Data Science MCQ Interview Questions

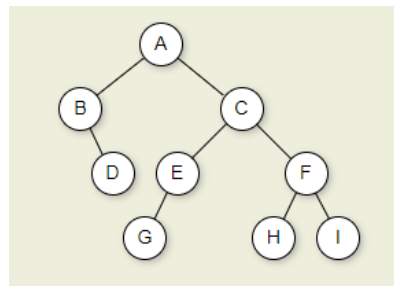
## 1. Which of the following statements are correct about Tree Data Structure?

- (a) It is a non-linear data structure
- (b) In a tree data structure, a node can have any number of child nodes
- (c) There is one and only one possible path between every pair of vertices in a tree
- (d) Any connected graph having  $n$  vertices and  $n$  edges is considered as a tree

**Answer: [ a, b, c ]**

**Explanation:** A graph is a tree if and only if it is minimally connected, which means any connected graph with  $n$  vertices and  $(n-1)$  edges is a tree.

## 2. Which of the following statements are TRUE about Tree Traversals for a given tree?



- (a) The Inorder traversal of the given tree is B D A G E C H F I
- (b) The Preorder traversal of the given tree is A B D C E G F H I
- (c) The Postorder traversal of the given tree is D B G E H I F C A
- (d) The breadth-first traversal of the given tree is A B C D E F G H I

**Answer: [ a, b, c, d ]**

**Explanation:**

- Root → Left → Right
- Inorder: Left → Root → Right
- Postorder: Left → Right → Root

## 3. Which of the following statements are TRUE about Binary Tree?

- (a) In a binary tree, each node must have 2 children
- (b) In a binary tree, nodes are always arranged in a specific order
- (c) It is a special type of tree data structure
- (d) Number of nodes having zero children in any binary tree depends only on the number of nodes with 2 children

**Answer: [ c, d ]**

**Explanation:**

In a binary tree, each node can have at most 2 children. Total Number of nodes having zero children in a Binary Tree = Total Number of nodes having 2 children + 1

#### **4. Which of the following statements are correct about Binary Search Tree(BST)?**

- (a) Binary Search Tree is considered as a special type of binary tree
- (b) Nodes are arranged in a specific order
- (c) Only smaller values in its right subtree
- (d) Only larger values in its left subtree

**Answer: [ a, b ]**

**Explanation:** In a binary search tree (BST), each node contains, only smaller values in its left subtree and only larger values in its right subtree.

#### **5. Which of the following statements are TRUE about AVL Tree?**

- (a) AVL trees are considered as a special kind of binary search tree
- (b) AVL trees are also called self-balancing binary search trees
- (c) In AVL trees, the height of the left subtree and right subtree of every node differs by at least one
- (d) In AVL trees, the balancing factor of each node is either 0 or 1 or -1

**Answer: [ a, b, d ]**

**Explanation:** In AVL trees, the height of the left subtree and the right subtree of every node differs by at most one.

#### **6. Which of the following statements are True about Stack Data Structure?**

- (a) Stack is a type of dynamic set
- (b) It follows the Last-In-First-Out(LIFO) principle
- (c) Stack is a non-linear Data Structure
- (d) The INSERT operation on the stack is often known as PUSH

**Answer: [ a, b, d ]**

**Explanation:** Stack is a linear Data Structure.

Also Read: [Top 10 GitHub Data Science Projects with Source Code](#)

## Intermediate Data Science Coding Questions

**7. The following integers are inserted into an initially empty binary search tree in order: 10, 1, 3, 5, 15, 12, 16. What is the height of the binary search tree formed? (Here, Height is defined as the maximum distance of a leaf node from the root. If the tree has only the root node then the height is 0)**

(a) 2

(b) 3

(c) 4

(d) 5

**Answer: [ b ]**

**Explanation:** The Binary Search Tree formed is shown as below:

**8. Suppose in a Binary tree, the number of the internal nodes having degree-1 is 9 and the number of internal nodes having degree-2 is 16. Then, the number of nodes having 0 children in the binary tree are:**

(a) 10

(b) 17

(c) 25

(d) 7

**Answer: [ b ]**

**Explanation:** Total Number of leaf nodes in a Binary Tree = Total Number of nodes having 2 children + 1

### 9. Which of the following statements are TRUE about Array Data Structure?

- (a) An array is a collection of elements that are stored at contiguous memory locations
- (b) Array can store the elements of different data type
- (c) Array is a Linear Data Structure
- (d) Accessing array elements takes constant time

**Answer:** [ a, c, d ]

**Explanation:** Array contains all the elements of the same data type.

### 10. How many of the following statements are TRUE about Tree Terminology?

- (a) In any tree, there may be more than one root node
- (b) The connecting link between any two nodes in a tree is called an edge
- (c) Nodes that belong to the same parent are called siblings
- (d) Degree of a Tree is the total number of children of any node of a tree

**Answer:** [ b, c ]

**Hint:** Self Explanatory(Basics of Tree Terminology)

### 11. Choose correct output for the following sequence of operations on Stack Data Structure:

push(5) push(8) pop push(2) push(5) pop pop pop push(1) pop

- (a) 8 5 5 2 1
- (b) 8 2 5 5 1
- (c) 8 1 2 5 5
- (d) 8 5 2 5 1

**Answer: [ d ]**

**Explanation:** Stack Data Structure follows the Last-In-First-Out(LIFO) Principle.

## Expert Level Data Science Coding Questions

**12. A binary search tree is formed by inserting the numbers in the given order: 50, 5, 20, 58, 91, 3, 8, 24. Then, Which of the following statements are TRUE about BST formed?**

- (a) The root node in the formed tree is 50
- (b) Number of nodes in the left subtree of the root = 5
- (c) Number of nodes in the right subtree of the root = 2
- (d) Node with label 20 have only 1 child

**Answer: [ a, b, c ]**

**Explanation:** The tree formed after inserting all the elements is shown as below:

**13. Compare the following in the term of increasing time complexity:  $f_1(n)=2^n$ ,  $f_2(n)=n^{3/2}$ ,  $f_3(n)=n\log_2 n$ ,  $f_4(n)=n^{\log_2 n}$**

- (a)  $f_2, f_3, f_4, f_1$
- (b)  $f_2, f_1, f_3, f_4$
- (c)  $f_1, f_2, f_3, f_4$
- (d)  $f_3, f_2, f_4, f_1$

**Answer: [ d ]**

**Explanation:** Comparison of various time complexities:

$O(1) < O(\log(\log n)) < O(\log n) < O(n^{1/2}) < O(n) < O(n \log n) < O(n^2) < O(n^3) < O(n^k) < O(2^n) < O(n^n)$

**14. What is the minimum number of nodes required to construct an AVL Tree of height = 3?**

- (a) 5
- (b) 6
- (c) 7
- (d) 8

**Answer: [ c ]**

**Hint: Using the recursive relation:  $N(h) = N(h-1) + N(h-2) + 1$ , with base condition as  $N(0)=1$  and  $N(1)=2$  and here we have to calculate the value of  $N(3)$ .**

**15. Which of the following properties are correct about Binary Tree?**

- (a) Minimum number of nodes in a binary tree of height  $H = H + 1$
- (b) Maximum number of nodes in a binary tree of height  $H = 2^{H+1} - 1$
- (c) Maximum number of nodes at any level 'L' in a binary tree =  $2^L$
- (d) Maximum number of nodes at any level 'L' in a binary tree =  $2^{L-1}$

**Answer: [ a, b, c ]**

**Hint: Self Explanatory(Take a small tree example and then verifies the options).**

## Theory Based Data Science Interview Questions

### What is Data Science?

Data Science is your guide to the world of numbers and information. It's like the superhero of data, using math, statistics, and computer science to make sense of the vast amounts of information we have today. Think of Data scientists as the detectives of the digital world. They dive into data, clean it up, and uncover hidden treasures that help businesses, researchers, and even your favorite apps make smarter decisions. So, when you hear about Data Science, think of it as the magic that turns data into valuable insights.

## **Differentiate between Data Analytics and Data Science.**

Data analytics and Data science are like two cousins in the data world. While both deal with data, Analytics examines past data to understand what happened. It's like a detective looking at evidence to solve a crime. On the other hand, Data Science takes it a step further. It's not just interested in what happened; it wants to predict the future. It's more like a fortune-teller, using data to anticipate what might happen next.

## **What distinguishes supervised learning from unsupervised learning?**

Think of supervised learning as teaching a child with the answer key. You show the model examples of input and the correct output, and it learns to predict the output for new inputs. On the other hand, unsupervised learning is like giving a child a box of puzzles without a picture on the box. It figures out the patterns and groups the pieces independently, without any guidance.

What are some of the techniques used for sampling? What is the main advantage of sampling?

Sampling is like taking a bite from a large dish to understand its taste. There are various techniques like random, stratified, and cluster sampling. The main advantage is that it saves time and resources. You can get a good sense of the whole dish (or population) without eating the entire thing.

## **What is a Confusion Matrix?**

A Confusion Matrix is like a scorecard for a model's performance. It shows you how well the model can distinguish between different classes. It's called 'confusion' because it might mix things up and this matrix helps you keep track of the mix-ups.

## **How is logistic regression implemented?**

Logistic regression is like fitting a curve to predict yes/no or 1/0 outcomes. It uses a mathematical function to find the relationship between one or more features and a binary outcome. Think of it as drawing a line that best separates the two classes.

## **What is the significance of the p-value?**

The p-value is like a judge in a courtroom. It decides whether the evidence (data) is strong enough to convict a defendant (your hypothesis). If the p-value is low, you have a strong case. If it's high, you might need more evidence.

## **When is the Classification Technique more suitable than the Regression Technique?**

Imagine you have a box of fruits and want to sort them into apples and oranges (Classification). But if you want to predict the weight of the fruits (Regression), you'd be better off using a scale instead of your eyes. Classification is for sorting, and regression is for measuring.

## **Can you recognize the factors that lead to Overfitting and Underfitting?**



Overfitting is like wearing a suit that's too tight; it fits perfectly but leaves no room for movement. It happens when a model learns the training data too well but can't adapt to new data. Conversely, underfitting is like wearing a suit two sizes too big; it's comfortable but looks sloppy. It occurs when a model is too simple and can't capture the complexity of the data. The key is to find the sweet spot in between, just like wearing a suit that fits just right.

## Conclusion

In the ever-evolving landscape of data science, the importance of robust coding skills cannot be overstated. This article has provided a comprehensive array of coding questions and answers designed to empower data scientists in the year 2023. By embracing these challenges, you've fortified your problem-solving abilities and expanded your knowledge in this dynamic field. As data science continues to shape our world, your proficiency in coding is a key asset. So, keep practicing, keep learning, and keep pushing the boundaries of what's possible. With these skills, you're well-equipped to excel in the exciting data science journey that lies ahead.

***The media shown in this article are not owned by Analytics Vidhya and is used at the Author's discretion.***

---

Article Url - <https://www.analyticsvidhya.com/blog/2021/05/popular-coding-questions-asked-in-data-science-interviews/>



### **CHIRAG GOYAL**

I am currently pursuing my Bachelor of Technology (B.Tech) in Computer Science and Engineering from the Indian Institute of Technology Jodhpur(IITJ). I am very enthusiastic about Machine learning, Deep Learning, and Artificial Intelligence. Feel free to connect with me on LinkedIn.