



100+ Common **Data Science** Interview Questions

(Save It Now)

www.cloudymL.com

1. Why do you want to/ why did you choose data science as your career?
2. What is the difference between AI, ML and DL?
3. What is a Python Package, and Have you created your own Python Package?
4. Can you write an inverted star program in python?
5. Write a program to create a data frame and remove elements from it.
6. Write code to find the 8th highest value in the Data Frame.
7. What's the difference between an array and a list?
8. Differentiate between Supervised, Unsupervised and Reinforcement learning with their algorithm example.
9. How would you deal with features of 4 categories and 20% null values?
10. What is central tendency?
11. Which central tendency method is used if there exist any outliers?
12. Explain, What is Central limit theorem?
13. What is the Chi-Square test?
14. What is A/B testing?
15. Tell us the difference between Z and t distribution (Linked to A/B testing)?



16. Tell some outlier treatment methods.
17. What is ANOVA test?
18. What is Cross validation?
19. How will you work in a machine learning project if there is a huge imbalance in the data?
20. Tell the formula of sigmoid function.
21. Can we use sigmoid function in case of multiple classifications?
22. What is Area under the curve (AUC)?
23. Which metric is used to split a node in a Decision Tree?
24. Explain ensemble learning?
25. What is P value?
26. What are histograms?
27. Tell us about the confidence interval?
28. What's the reason for high bias or variance?
29. Which models are generally high biased or high variances?
30. Why do we select validation data other than test data?
31. What are the differences between linear and logistic regression?
32. Why do we take such a complex cost function for logistic regression?
33. Differentiate between random forest and decision tree?
34. How would you decide when to stop splitting the tree?
35. What are the measures of central tendency?

Switch Your Career To

Data Science & Analytics

Get Complete Hands-On Practical Learning Experience



Python



SQL



Statistics



Machine Learning



Deep Learning



ML + DL Projects



DSA



AWS Sagemaker



Tableau



Excel



PowerBI



Interview QnA PDF

Doubt Clearance Support

1 Month Internship

Certification

Lifetime Validity

[Click Here to Visit Our Website](https://www.cloudymL.com)

36. What is the requirement of the k -means algorithm?
37. Which clustering technique uses a combination of clusters?
38. Which is the oldest probability distribution?
39. What all values can a random variable take?
40. What are the different types of random variables?
41. Describe normality of residuals.
42. What is the T-test used for?
43. How do you perform dimensionality reduction?
44. What are the assumptions of linear regression algorithms?
45. Differentiate between Correlation and covariance.
46. How to identify & treat outliers and missing values?

[**www.cloudymL.com**](https://www.cloudymL.com)

47. Explain Box and whisker plot.
48. Explain any unsupervised learning algorithm.
49. Describe Random Forest.
50. What packages in Python can be used for ML? Why do we prefer one over another?
51. What are the Evaluation Metric parameters for testing Logistic Regression?
52. NumPy vs Pandas basic difference.
53. Tuple vs Dictionary. Where do we use them?
54. What is NER(Named Entity Recognition)?
55. Can Linear Regression be used for Classification? If Yes, why if No why?
56. What is Naive Bayes Theorem? Multinomial, Bernoulli, Gaussian Naive Bayes.
57. Differentiate between Over Sampling and Under Sampling.
58. What is the difference between Over Fitting and Under Fitting.
59. Differentiate between Gini Index and Entropy.
60. What are the advantages and disadvantages of PCA?
61. How to deal with imbalance data in classification modelling?
62. What is Gradient Descent? What is the Learning Rate and why do we need to reduce or increase? Tell us why the Global minimum is reached and why it doesn't improve when increasing the LR after that point?
63. What is Log-Loss and ROC-AUC?



64. Two Logistic Regression Models – Which one will you choose – One is trained on 70% and other on 80% data. Accuracy is almost the same?
65. Explain bias – variance trade off. How does this affect the model?
66. What is multi collinearity? How to identify and remove it?
67. Differentiate between Sensitivity and Specificity.
68. What is the difference between K-NN and K-Means clustering?
69. How to handle missing data? What imputation techniques can be used?
70. Explain how you would find and tackle an outlier in the dataset. Follow up: What about inlier?
71. How to determine if a coin is biased? Hint: Hypothesis testing
72. Is interpretability important for machine learning models? If so, ways to achieve interpretability for machine learning models?
73. How would you design a data science pipeline?
74. What does a statistical test do?
75. Explain topic modelling in NLP and various methods in performing topic modelling.
76. Describe back propagation in a few words and its variants?
77. Explain the architecture of CNN.
78. If we put a 3×3 filter over a 6×6 image what will be the size of the output image?
79. What will you do to reduce overfitting in deep learning models?

Switch Your Career To

Data Science & Analytics

Get Complete Hands-On Practical Learning Experience



Python



SQL



Statistics



Machine Learning



Deep Learning



ML + DL Projects



DSA



AWS Sagemaker



Tableau



Excel



PowerBI



Interview QnA PDF

Doubt Clearance Support

1 Month Internship

Certification

Lifetime Validity

[Click Here to Visit Our Website](https://www.cloudymL.com)

80. How would you check if the model is suffering from multiCollinearity?
81. Why is CNN architecture suitable for image classification and not an RNN?
82. What is an RNN (recurrent neural network)?
83. What are the approaches for solving class imbalance problems?
84. Tell us about transfer learning? What are the steps you would take to perform transfer learning?
85. Explain concepts of epoch, batch, and iteration in deep learning.
86. When sampling, what types of biases can be inflicted? How to control the biases?



87. What are some of the types of activation functions and specifically when to use them?
88. Tell us the conditions that should be satisfied for a time series to be stationary?
89. What is the difference between Batch and Stochastic Gradient Descent?
90. What happens when neural nets are too small? Tell us, What happens when they are large enough?
91. Why do we need a pooling layer in CNN? Common pooling methods?
92. Are ensemble models better than individual models? Why/why – not?
93. How is random forest different from Gradient boosting algorithm, given both are tree-based algorithms?
94. Describe steps involved in creating a neural network?
95. In brief, how would you perform the task of sentiment analysis?
96. Is XOR data linearly separable?
97. How do we classify XOR data using logistic regression?
98. LSTM solves the vanishing gradient problem that RNN primarily has. How?
99. GRU is faster compared to LSTM. Why?
100. Use Case – Consider you are working for pen manufacturing company. How would you help sales team with leads using Data analysis?



101. I have 2 guns with 6 holes in each, and I load a single bullet In each gun, what is the probability that if I fire the guns simultaneously,at least 1 gun will fire (at least means one or more than one)?
102. There are 2 groups g1 and g2, g1 will ask g2 members to give them 1 member so that they both will be equal in number, g2 will ask g1 members to give them 1 member so that they will be double in number of g1, how many members are there in each group?
103. Tell the Order of execution of an SQL query.
104. SQL Questions – Group by Top 2 Salaries for Employees – use Row num and Partition.
105. Differentiate between inner join and cross join.
106. What is group-by?
107. Complex sql query– 2 table are there, Table1 (cust_id,Name) and Table2(cust_id,Transaction_amt). Write a query to return the name of customers with 8th highest lifetime purchase.Achieve the same using python.
108. Explain stacking in Data Science.
109. What is the benefit of dimensionality reduction?
110. How is Data modeling different from Database design?
111. What are Loss Function and Cost Functions? Explain the key Difference Between them?

[Click Here to Visit Our Website](#)