

# **Data Science Interview Questions**

Published on September 25, 2019



I prepared some common Data Science interview questions. Being prepared for these can help you make a good impression.

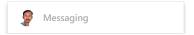
## Some of my tips are.

- 1. Have your resume with good projects. Be clear and thorough about in and around the projects you have mentioned.
- 2. Be confident, even if you don't know any answer have guts to say." I don't know or I did not get the oppurtunity to know about it. But given time. I can learn."

I am listing down some of the interview questions. If I were an Interviewer.

# **Model Deployment**

- 1. Describe the Project you have worked and mention which ML technique you have used?
- 2. How will you deploy the model in real-time?





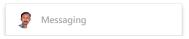


#### **Statistics**

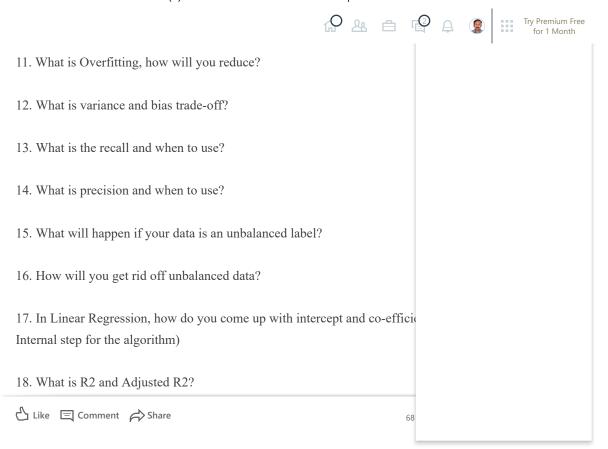
- 1. What kind of Statistics you have used in your project?
- 2. What all descriptive statistics you have used, with the scenario?
- 3. What is Z-Score?
- 4. What is Standard Deviation?
- 5. Difference between Z-Test and T-Test?
- 6. Different types of the t-test?
- 7. What is Chi-Square?
- 8. When will you go for Chi-Square?
- 9. What is Anova?
- 10. What is F-Test?
- 11. Difference between Parametric and Non-Parametric Test?

## **Machine Learning**

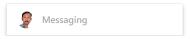
- 1. What all preprocessing steps you do, given a dataset?
- 2. How will you transform the data?
- 3. How will you find the which feature is significant for you?
- 4. Which algorithm do you like most? (So, that I will question from that )
- 5. What is Gradient Descent?
- 6. Explain the steps in Gradient Descent?
- 7. How will you evaluate a model?
- 8. What error metrics you use for Regression?
- 9. What error metrics you use for Classification?







- 20. What is Null Deviance and AIC in Logistic Regression?
- 21. How do you construct the Decision Tree?
- 22. The formula for Entropy and Gini Index?
- 23. Which method of the Decision tree is more efficient in terms of time?
- 24. What is the problem in the decision tree?
- 25. How will you avoid Overfitting in a Decision tree?
- 26. What is Pruning in Decision Tree?
- 27. What will happen internally when you have numeric column while constructing Decision Tree?
- 28. Why do we need to go Decision Tree when we have Logistic Regression?
- 29. What is Bagging?
- 30. What is the problem in Decision Tree?
- 31. Explain the internal working of Random Forest?



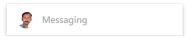




- 33. Random Forest uses entropy or Gini by default and why?
- 34. Advantage of Random Forest?
- 35. What is the Difference between Bagging and Boosting?
- 36. Explain anyone Boosting Algorithm?
- 37. What algorithm you have worked on Unsupervised Learning?
- 38. How K-means works internally?
- 39. In K-means how do you find optimal K value?
- 40. What is the disadvantage of K-means?
- 41. What is K-means++?
- 42. How will you find optimal value in Hierarchical cluster?
- 43. How will you construct Dendrogram?

### Deep learning

- 1. How good are you in Deep learning?
- 2. In which project did you use the neural network?
- 3. Why do we use Activation function in Neural networks?
- 4. List some Activation function?
- 5. What is a Back Propagation?
- 6. What is the problem of using Deep Learning?
- 7. What is Vanishing and Exploding Gradient Descent?
- 8. How will you avoid Vanishing and Exploding Gradient Descent?
- 9. What is Batch Normalization?
- 10. What is Gradient Clipping?

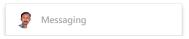




- 12. How will you avoid Overfitting in Deep learning?
- 13. What Is Dropout?
- 14. What is the epoch, iteration, batch?
- 15. What is CNN
- 16. What is RNN
- 17. Why LSTM on RNN?
- 18. What are pre-trained models?
- 19. What is transfer learning?

#### **Text Analytics**

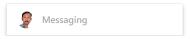
- 1. Have ever done a project in text analytics?
- 2. What is the difference between Bag of words and TF-IDF?
- 3. What is Vectorization of words?
- 4. How do you do unsupervised learning on the text?
- 5. What all library u have used in NLP?
- 6. What is Word-embeddings?
- 7. What are word2vec, glove and Fast-text?
- 8. How LSTM is helping in text analytics?
- 9. How genism word2vec works?
- 10. what are pretrained models you worked on text analytics?
- 11. Transfer Learning with text analytics?
- 12. Topic Model in NLP using LDA, NMF, LSI?
- 13. Why Cosine Similarity?



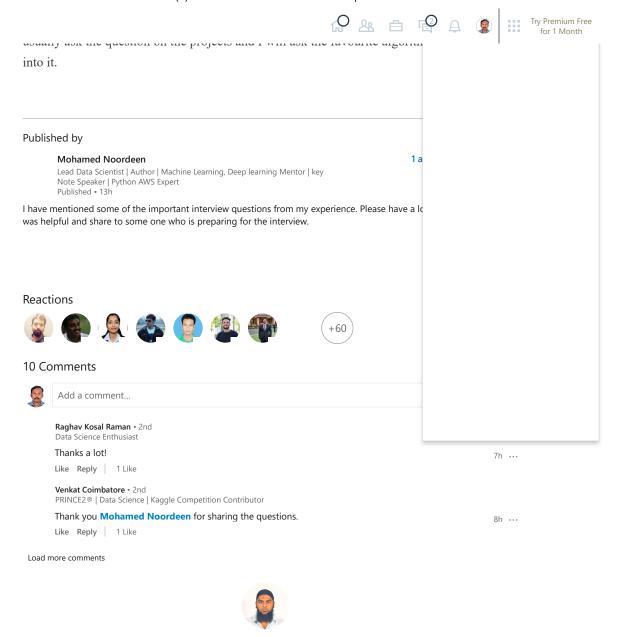




- 1. How much do you rate yourself in python?
- 2. Which IDE do you use?
- 3. What is the difference between list and tuple?
- 4. What is the attribute of set and String?
- 5. If the list is there why tuple. Being immutable what is the need of having
- 6. What is dynamic typing?
- 7. Have you ever used the pointer in python?
- 8. How do you create a class and object in python?
- 9. If we have an array in python, Why do we need Numpy array?
- 10. What is the underlying program architecture of numpy?
- 11. Why numpy is time and space-efficient?
- 12. Can I use more than one data type in numpy array?
- 13. If I want to use different datatype in each column which features to use in python?
- 14. In pandas, how to check null values and delete only all rows is na?
- 15. How will you find the duplicate value in pandas?
- 16. How will you impute the missing value in pandas?
- 17. What do value\_counts() function does?
- 18. How will you remove outlier?
- 19. How will you convert the non-normal data to normal distribution data?
- 20. How will you plot a histogram in python?
- 21. How do you group by a column in pandas?
- 22. How will you serialize the model?







## Mohamed Noordeen

Lead Data Scientist | Author | Machine Learning, Deep learning Mentor | key Note Speaker | Python AWS Expert

