Technical Interviewing Sample Questions

The GBCC asked recent MSBA graduates about interview questions they have received through their interview process. Below are a few examples we heard. Please use this list to help you prepare.

STATISTICS + PREDICTIVE ANALYTICS + EXPLORATORY ANALYTICS

1. Explain R-sq. Explain p-value.
2. What is the difference between Adj. R-sq and R-sq?
3. What are the different types of statistical tests and what packages would you use in R to perform those tests
4. What is a confidence interval? Explain with an example.
5. Explain chi square test and when to use it.
6. What is apriori?
7. What is LDA?
8. In simple layman terms, explain kernel trick and hinge loss in SVM.
9. Elaborate on the pros and cons of at least 2 predictive algorithms.
10. How is hierarchal clustering advantageous over k means clustering?
11. Explain dbscan algorithm. How is it advantageous?
12. Explain how you will use categorical variables in KNN algorithm.
13. Explain neural network clustering logic.
14. What is regularization?
15. What is PCA? What is the first component of the PCA?
16. How do you deal with class imbalance?
17. Would you be looking at Accuracy or Precision Vs Recall in case of Churn?
18. Walkthrough your approach if it's a Credit Fraud and how you will classify it.
19. What is ARIMA and ARMA?
20. Explain bias and variance to business people.
21. You have a lot of correlated features, how do you do feature engineering? The end features should be comprehendible to people.
22. How do you identify outliers numerical feature?
23. If there is a text classification problem without actual labelling/target value. How would you build a classification model without a target class column?
24. What are the hypotheses for linear regression?
25. What are the assumptions of linear regression?
26. Explain Random Forest using layman terms
27. How do Gradient Boosted Tree and Random Forest differ from each other?
28. Why is collinearity a problem in regression?
29. What is undersampling/oversampling?
30. Explain cross validation.

PYTHON

1. Please explain the difference between list, tuple, dictionary and set.
2. How to write a function in python?
3. Write a function to find the maximum in an array.
4. What is the expected number of times you will have to update tracking variable?
5. How would you approach TSP? Can you write the algorithm in Python?
6. Within a string 'x' , print a string that contains the frequency of letters in the string 'x' in any order
7. You are getting a stream of data through a source, how do you divide this stream of data into two streams (one contains 60% of the input stream and 40% in the other stream randomly)
8. What are the packages you used for twitter streaming in python?
9. Write a program for calculating sum of numbers which have 3 factors, except itself. (e.g. 6 = 2,3,1 and 21 = 3,7,1)
10. How do you remove null values in python?
11. How do you merge two datasets?

BIG DATA

1. We have a table with 3 columns - Business Unit, Product & Sales. one Business Unit can contain multiple products. Write a HIVE query/SPARK commands query to get top 5 products by sales within the BU.
2. What would the lambda function on SPARK look like?
3. Write a Linux command line query for opening a csv and replace a column value with some other and save it into a new Excel.
4. Write a Linux command to copy one column of a csv file into a new csv.
5. Write a Linux command to convert one compressed file to another compression file with a different compression algorithm without uncompressing it.
6. Write a Linux command to filter a column based on a regularization expression from a csv.
7. Benefits of Spark over Mapreduce?

SQL

1. We have a table with 3 columns - Business Unit, Product & Sales. Here, 1 business unit can contain multiple products. Please write a SQL query to get the top 5 products by sales within every business unit.
2. We have two tables with product name column (string) in each of them. Join these two tables using these string columns and clean table that has additional characters or words in the product name e.g. 'The' etc.
3. What is the most complicated query you have written? Explain the query and the algorithm that followed.
4. How to find second highest element in a column?
5. How to get the sales of each of the three items separately when you have an Item ID column and Sales column?
6. Calculate the 90th Percentile difference in two columns in POSTGRE SQL.
7. We have four columns (first name, last name, address and gender); there is no unique column in the table. How do you generate a unique column for this table?
8. A table has a single column with 4 values 'a','b','c','d'. Write an SQL query to get combinations of 'ab','ac'... (i.e. unique combination of two alphabets)
9. Find the cumulative sum of a column, without using CUMSUM function.
10. Find the 3rd ranked student for a dataset containing names and marks of students.
11. Calculate percentage of customers in a table where salary > $100,000.
12. Count the number of hattricks (i.e. 3 wins in a row).

R

1. How do you write a function in R?
2. For every question like clustering, classification how would you do it in R? What packages will you use?
3. What are the data manipulation languages you have used?
4. What package and code would you use for cross validation?
5. How would you remove duplicates in dataset at any levels?
6. How to maintain a live connection in R? (e.g. if you want to update the algorithm and get output every week)
7. What are the packages you would require for time series forecasting?
8. How can Matrix multiplication be done in R?
9. Will neural network accept categorical variable?
10. How to scale categorical variables to numeric?

TABLEAU

1. Create a chart in Tableau where you will arrange values by years. Within each year arrange by descending weeks according to their sales.
2. What is context filter?
3. Differentiate between filter and parameters in Tableau.
4. What is the difference between measures and dimensions?
5. What are the various filters in Tableau and how are they different from each other?
6. Which do you prefer, Shiny or Tableau, and why?
7. What is parameter in Tableau?