SQL:

1. CASE WHEN. So many great use cases!
2. Self joins. Common in product user behavior.
3. DISTINCT and GROUP BY
4. Left vs outer joins. Need I say more?
5. UNION. Rarely discussed but frequent.
6. SUM and COUNT. Nail the foundations!
7. Date-time manipulation. This will set you apart.
8. String formatting, substring.
9. Window functions like rank and row. Absolute gold!
10. Subqueries. Because they always show up...
11. HAVING vs WHERE. Do you know why?
12. LAG and LEAD. What do you use these for?
13. Understanding indexing. More intermediate.
14. Running totals. A fun use case to learn.
15. MAX and MIN. More common than anyone says!

1. SELECT and WHERE for filtering and selection
2. COUNT, SUM, MAX, GROUP BY, HAVING for aggregating data"
3. DISTINCT, COUNT DISTINCT for producing useful distinct lists and distinct aggregates
4. OUTER (e.g. LEFT) and INNER JOIN when/where to use them
5. Strings and time conversions
6. UNION and UNION ALL.

Topics that come up regularly (be familiar with them):
1. DML/DDL/DCL concepts
2. Handling NULLs creatively (e.g. with COALESCE)
3. Subqueries and the impact of subqueries on efficiency of the query
4. Temporary tables
5. Self joins
6. Window functions like PARTITION, LEAD, LAG, NTILE
7. UDFs (user defined functions)
8. Use of indexes in querying to make operations faster.

How to select alternate rows in SQL ?
How to modify Sql row value ?
Difference between Where and Having ?
Difference between Union and Union all?
How many Joins in SQL ?

Pre-Interview:

1. CASE WHEN. Shows up all the time.
2. Self joins. Common in product.
3. DISTINCT and GROUP BY

4. Left vs outer joins.
5. UNION. Rarely discussed but frequent.
6. SUM and COUNT
7. Date-time manipulation
8. String formatting, substring
9. Window functions like rank and row
10. Subqueries
11. HAVING vs WHERE
12. LAG and LEAD
13. Understanding indexing
14. Running totals
15. MAX and MIN
16. Using SUM CASE WHEN to count
17. COUNT DISTINCT
18. How to debug a query.
19. How to speed up a query.
20. Rank and dense rank

Resources:

1. Zachary Thomas' SQL Questions https://lnkd.in/g-JJzuD
2. Select * SQL: https://selectstarsql.com/
3. Leetcode: https://lnkd.in/g3c5JGC
4. LinkedIn Learning: https://lnkd.in/gQXFc4n
5. Window Functions: https://lnkd.in/g3RtPCJ
6. HackerRank: https://lnkd.in/grv_9sB
7. W3 Schools: https://lnkd.in/gJPfrrv
8. CodeAcademy: https://lnkd.in/gT5xmpN
9. SQLZoo: https://sqlzoo.net/
10. SQL Bolt: https://sqlbolt.com/

Use a SQL window function to do a cumulative sum.

Try to get the same SQL result with/without a subquery.

R:

Explain the memory allocation done by subset function in dplyr package in R.

Tell me which one out of base and dplyr uses copy mechanism and which one uses a pointer. Which is faster?

Write the code for doing a scatterplot using ggplot in R.

Guesstimates:

1) Case Interviews Cracked
This is a youtube channel run by students of IIT Bombay who are currently working in prestigious firms like Bain&Co, McKinsey, etc.
https://lnkd.in/gn52FUr

2)Consulting and Strategy Club, IIM Lucknow
Excellent channel for practicing a couple of guesstimates.
https://lnkd.in/ghkVVZZ

3) FMS Casebook by The Consulting Club - FMS, Delhi
There are 20 practice guesstimates with well-explained solutions in this book.

Apart from the FMS Casebook, I have also added the following casebooks in the below drive link:
-Consult Club IIMA Casebook
-LBS Casebook
-Case Interviews Cracked book

Drive Link: https://lnkd.in/gD9VCEi

DS Interview:

Why data science as a career?

What is p value?

What is histograms?

What is confidence interval?

You are a Sr data analyst at a new Online Cab booking Startups. How you will do data collection and how you will leverage the data to give useful insights to the Company?

Guestimate: No Of cabs booking per day in Ranchi

You are product head manager(not remember exactly) at a NBFC which gives a Secured loans what factors will you consider giving loan to ?

Inventory Database based on that have to do basic pandas/sql query? Joins / merge to get avg sales, its chart?

You have a list of 3 numbers return the min diff. Can use any python/sql
What is Big Data?

A continuous variable is having missing values, so how will you decide that the missing values should be imputed by mean or median?

What is PCA and what each component means? Also, what is the maximum value for number of components?

What is test of independence? How do you calculate Chi-square value?

When precision is preferred over recall or vice-versa?

Advantages and disadvantages of Random forest over Decision Tree?

What is the c hyperparameter in SVM algorithm and how it affects bias-variance tradeoff?

What are the assumptions of linear regression?

Difference between Stemming and Lemmatization?

Difference between Correlation and Regression?

What is p-value and confidence interval?

What is multicollinearity and how do you deal with multicollinearity? What is VIF?

What is the difference between apply, applymap and map function in python?

Started with Classification particularly Imbalance , oversampling. Which class should i oversample etc.

Telecom Churn Case Study Questions like Evaluation metric for imbalance data what threshold to choose to diving the classes (0.5 in case of balanced else sensitivity / Specifivity etc.

What if i don't use SMOTE() for handling imbalance how should i select the threshold now (messed up by me, roc , auc etc) Ans = Presion - Recall Curve

- NLP Questions
Sentiment analysis, preprocessing like (TFID, BOW), Embeddings, stemming, Lemmatization libraries in know : nltk, spacy

- Regression Preprocessing
answered outlier, missing value immputation, Distribution, dummies, multicolinearity etc

You have two highy co-related columns which one will you drop? : "Based on Business Problem i will see accordingly.", Answered

Naive Bayes Explanation , Drawback of Naive Bayes

Hand Gesture Recognition Techniques (End to End)

- Do you know any Boosting Algorithms : YES
where have you used??

- Gradient Descent (How it works)

- KNN related. How do we choose value of K ??

- Stastical Computing:
Type 1 and Type 2 error
Alternate name of Type 1 error (couldn't answer alternate name of Type 1 error, 'False +ive, him)

What is p-Value (Explained with the example of Linear Regression from statsmodel)

How to rename multiple column names in pandas?

How to add a new column name with all lower case letters in pandas?
How to drop column name in to read json file ?

List comprehension ?

Mutable and immutable ?
List vs tuple ?

How to handle missing values in a dataset?

How to handle imbalance in a dataset ?

What is hyperparameter tuning ?

Overfitting and underfitting ?

Bias variance trade off?

What are the assumptions of Linear regression?

What is multi collinearrity?

What's the threshold you took for vif?
What the formula of vif?

Whats the meaning of precision and recall?

Why is Xgboost better than Gradient boosting?

What are different boosting algorithms? Whats the difference between boosting and bagging?

If R squared is less than adj R squared will you accept the model?

What exactly is the p-value in statsmodel OLS regression package?

You are given an unsorted list of 999,000 unique integers, each from 1 and 1,000,000. Find the missing 1000 numbers. What is the computational and space complexity of your solution?

RNN, NN and CNN difference.

Different feature selection methods

Confusion matrix

How would do knn algorithm for big data?

Suppose you have a fixed number of trucks that you intend to send on different routes for delivery purposes. How would you optimize the number of trucks to be sent to which routes in order to maximize profit? (question on dynamic programming reinforcement learning)

“A word is actually its neighbours in a corpus” - Explain this sentence in machine learning lingo.

Dive deep on linear and logistic regression assumptions.

Revisit important statistical distributions. Learn them again.

Take a short course or read a book on storytelling.

Try to explain your favorite ML model in 10 words or less.

Learn everything you can about sorting algorithms.

Another Interview Experience:

1.Difference between array and list

2.Map function

3. Scenario,
if coupon distributed randomly to customers of swiggy, how to check there buying behaviour.
Use segmenting customers, Compare customers who got coupon and who did not

4. Which is faster dictionary or list for look up

5. How to merge two arrays

6. How much time svm takes to complete if 1 iteration takes 10sec for 1st class.
And there are 4 classes.

7. Kernals in svm, there difference