100

Data Scientist

Interview

Questions

Interview Questions asked in FAANGs,
startups and consulting firms

Applied Statistics

1. Can you explain the Central Limit Theorem and its importance in statistics?
2. What is the difference between Type I and Type II errors in hypothesis testing?
3. Can you explain what is meant by p-value?
4. How do you assess the normality of a dataset?
5. Can you describe the difference between correlation and causation?
6. How do you handle missing or corrupted data in a dataset?
7. What are the assumptions required for linear regression? What if some of these assumptions are violated?
8. Explain the concept of power in statistical tests.
9. How would you explain to a non-technical team member what a confidence interval is?
10. What is multiple regression and when do we use it?
11. Can you describe the difference between cross-sectional and longitudinal data?
12. What is the role of data cleaning in data analysis?
13. Can you explain the difference between ANOVA and t-test?
14. How do you interpret the coefficients of a logistic regression model?
15. What is multicollinearity and how do you detect and deal with it?
16. Can you describe the difference between a parametric and a non-parametric test?
17. What are some of the methods you would use to detect outliers?
18. What is survival analysis and when can it be useful?
19. Can you explain what is meant by the term 'residual' in regression analysis?
20. Describe a situation where you used statistical analysis to make a decision or solve a problem

Machine Learning

1. What is the difference between supervised and unsupervised learning?
2. Can you explain the concept of overfitting and underfitting in machine learning models?
3. What is cross-validation? Why is it important?
4. Describe how a decision tree works. When would you use it over other algorithms?
5. How do you handle missing or corrupted data in a dataset?
6. What is the bias-variance tradeoff?
7. What is the difference between bagging and boosting?
8. How would you validate a model you created to generate a predictive analysis?
9. Can you explain the principle of a support vector machine (SVM)?
10. What are some of the advantages and disadvantages of a neural network?
11. How does the k-means algorithm work?
12. Can you explain the difference between L1 and L2 regularization methods?
13. What is principal component analysis (PCA) and when is it used?
14. Can you describe what an activation function is and why it is used in an artificial neural network?
15. How would you handle an imbalanced dataset?
16. Can you explain the concept of "feature selection" in machine learning?
17. What is the difference between stochastic gradient descent (SGD) and batch gradient descent?
18. Can you describe how a convolutional neural network (CNN) works?
19. How do you handle categorical variables in your dataset?
20. What is reinforcement learning? Can you give an example of where it could be used?

AB Testing

1. How would you design an experiment to test a new feature?
2. Can you explain the concept of A/B testing and how it can be useful for product development?
3. Describe a product that you think is particularly well-designed. What makes it effective?
4. How would you measure the success of a new feature launch?
5. How would you determine the optimal sample size for an A/B test?
6. What factors would you consider before rolling out a new feature?
7. How would you interpret the results of an A/B test that showed no significant difference between the control and treatment groups?
8. Can you explain the concept of statistical power in the context of A/B testing?
9. Can you describe a time when you used data to make a decision about a product?
10. How would you approach balancing the need for innovation with the potential risk to user experience?
11. What is a p-value? How would you explain it to a non-technical stakeholder?
12. Describe a situation where A/B testing would not be appropriate.
13. What metrics would you look at to understand user engagement for a mobile app?
14. What are some pitfalls or common mistakes in A/B testing?
15. How would you prioritize features for a new product?
16. How do you deal with seasonality in A/B testing?
17. Describe how you would validate the results of an A/B test.
18. How would you design a recommendation system for a new product?
19. How would you use data to improve an existing product's user experience?
20. What could be some potential issues with running multiple A/B tests at the same time?

Coding

1. Write a Python function to find the factorial of a number.
2. How would you write a function to check if a string is a palindrome in Python?
3. How can you handle missing values in a dataset using Pandas?
4. Given an array of integers, find the two numbers that sum up to a specific target number.
5. Write a Python function that sorts the elements in a list.
6. How do you merge two data frames in Python using Pandas?
7. Given a list of numbers, write a function to find the mean, median, and mode.
8. How do you create a binary variable in Python from a numerical variable?
9. Write a Python function to implement a linear regression model.
10. Write a function to calculate the Euclidean distance between two points.
11. How can you transpose a DataFrame in Pandas?
12. How do you calculate percentiles with numpy?
13. Given a string, write a function to check if it's a permutation of a palindrome.
14. Write a function to implement the K-nearest neighbors (KNN) algorithm.
15. How would you write a function that performs a SQL-like groupby operation on a Pandas DataFrame?
16. Given a list of numbers, write a function to find the outliers.
17. How can you visualize a decision tree in Python?
18. Write a Python function to calculate the root mean square error (RMSE) of a machine learning model.
19. Write a Python function that performs a binary search in a sorted array.
20. How do you filter rows in a Pandas DataFrame?

SQL & Databases

1. How can you combine data from multiple tables in SQL? What is the difference between SQL and NoSQL databases?

2. What is a relational database? What about non-relational database?

3. Can you explain the concept of database normalization?

4. What is a primary key in SQL? How does it differ from a foreign key?

5. What is a transaction in SQL? Can you describe a scenario where you would need to use transactions?

6. Can you describe how indexing works in SQL? Why is it important for database performance?

7. What are some strategies you would use to optimize a slow-performing SQL query?

8. Can you explain what ACID properties are in the context of databases?

9. How can denormalization improve the performance of a SQL database?

10. Can you explain the role of a database management system (DBMS)?

11. What are some common causes of database performance issues and how would you address them?

12. Can you explain the concept of sharding in databases? What problems does it solve?

13. What is an ORM (Object-Relational Mapping)? What are the pros and cons of using an ORM?

14. Can you explain the CAP theorem in the context of databases?

15. How does data replication work in SQL databases? What are its benefits?

16. How would you handle concurrency issues in a SQL database?

17. Can you explain the concept of a database lock?

18. What is a database view and what are its uses?

19. How do you use the LIKE operator in SQL? What is the difference between INNER JOIN and OUTER JOIN in SQL?

20. What is the difference between a left join, a right join, and an inner join?