

# **INTERVIEW QUESTIONS**

## **DATA ANALYTICS**

### **Technical Knowledge**

#### **1. What is the difference between supervised and unsupervised learning?**

- Company: Facebook (2022)
- Answer: Supervised learning uses labelled data to predict outcomes, while unsupervised learning identifies hidden patterns without labels.

#### **2. What is overfitting, and how do you prevent it?**

- Company: Amazon (2023)
- Answer: Overfitting occurs when a model performs well on training data but poorly on new data. It can be mitigated using techniques like cross-validation and regularization.

#### **3. Can you explain p-value in statistical analysis?**

- Company: Google (2021)
- Answer: The p-value measures the probability of observing the data if the null hypothesis is true. A p-value  $< 0.05$  typically indicates statistical significance.

#### **4. How do you calculate correlation between variables?**

- Company: Netflix (2022)
- Answer: Use Pearson's or Spearman's correlation coefficient, depending on the nature of the data.

**5. What is A/B testing, and where have you applied it?**

- Company: Uber (2023)

- Answer: A/B testing compares two versions of a variable to determine which performs better. It is often used in website optimization.

**Data Analytics Tools and Software**

**6. What SQL queries do you commonly use?**

- Company: IBM (2022)

- Answer: Queries include `JOIN`, `GROUP BY`, and `HAVING` for manipulating datasets.

**7. Explain how you use Python for data manipulation.**

- Company: PayPal (2023)

- Answer: Python libraries like Pandas help clean and transform data efficiently.

**8. What is your experience with Power BI or Tableau?**

- Company: Microsoft (2023)

- Answer: I've built dashboards, used DAX in Power BI, and designed reports in Tableau.

**9. What is ETL, and how does it function in data analytics?**

- Company: Deloitte (2021)

- Answer: ETL stands for Extract, Transform, Load, a process used to gather data from multiple sources and prepare it for analysis.

**10. Describe a time you used Apache Spark or Hadoop.**

- Company: Airbnb (2023)
- Answer: Apache Spark enabled me to process large datasets quickly for customer segmentation.

**Problem Solving and Critical Thinking**

**11. How do you handle missing or incomplete data?**

- Company: Twitter (2022)
- Answer: Techniques include imputing missing values with means or using predictive models.

**12. Describe a project where you identified an outlier.**

- Company: Accenture (2023)
- Answer: I used Z-scores to detect outliers in customer purchase data, helping optimize inventory.

**13. How do you determine if a model is good?**

- Company: EY (2022)
- Answer: Evaluate metrics like R-squared for regression or accuracy for classification models.

**14. Describe a situation where you faced conflicting data.**

- Company: PwC (2022)
- Answer: I resolved discrepancies by cross-referencing data from other sources and consulting with stakeholders.

**15. How do you optimize a machine learning model?**

- Company: Spotify (2023)
- Answer: I fine-tune hyperparameters and use techniques like cross-validation.

**Visualization and Reporting**

**16. What makes a good data visualization?**

- Company: IBM (2023)
- Answer: Clarity, relevance, and simplicity, ensuring the visualization conveys the intended message.

**17. Can you explain the difference between a report and a dashboard?**

- Company: Microsoft (2022)
- Answer: Reports provide static summaries, while dashboards offer real-time, interactive data insights.

**18. How do you design dashboards for non-technical stakeholders?**

- Company: Airbnb (2021)
- Answer: Use simple visualizations and ensure that key metrics are highlighted.

**19. What's your approach to storytelling with data?**

- Company: Google (2023)
- Answer: I build narratives around data trends and ensure the audience understands insights easily.

**20. How do you manage large datasets in visualization tools?**

- Company: Facebook (2023)
- Answer: I optimize queries and use data sampling to improve visualization performance.

**Soft Skills and Behavioural Questions**

**21. How do you handle tight deadlines in data projects?**

- Company: Uber (2022)
- Answer: I prioritize tasks, automate workflows, and communicate effectively with team members.

**22. Describe a time when your insights influenced business decisions.**

- Company: PayPal (2021)
- Answer: My analysis of customer churn helped the marketing team improve retention strategies.

**23. How do you communicate technical insights to non-technical stakeholders?**

- Company: Spotify (2022)
- Answer: I focus on simplifying jargon and using visual aids like charts to convey key messages.

**24. What are some challenges in data analytics projects?**

- Company: Accenture (2023)
- Answer: Common challenges include data quality issues and resistance to change from stakeholders.

**25. How do you stay updated with data analytics trends?**

- Company: Google (2023)
- Answer: I follow blogs, attend conferences, and take online courses regularly.

**Advanced Topics and Concepts**

**26. What is the curse of dimensionality?**

- Company: Netflix (2021)
- Answer: It refers to the difficulty of analysing high-dimensional data due to sparsity.

**27. Explain the importance of feature selection in machine learning.**

- Company: IBM (2023)
- Answer: Feature selection reduces model complexity and improves performance by removing irrelevant variables.

**28. How do you implement time series forecasting?**

- Company: Twitter (2022)
- Answer: I use ARIMA models and exponential smoothing for predicting trends.

**29. What are the differences between classification and regression?**

- Company: Amazon (2022)
- Answer: Classification predicts categorical outcomes, while regression predicts continuous values.

**30. How do you measure the success of a data analytics project?**

- Company: Deloitte (2021)

- Answer: Success is measured by how well insights align with business goals and drive actionable decisions.