

For Fresher's:

1. Statistics & Probability (Basic)

- What is the difference between mean, median, and mode?
- What is standard deviation, and why is it important?
- Can you explain what a normal distribution is?
- What is the difference between probability and likelihood?
- What are outliers, and how would you handle them?

2. Basic Machine Learning Concepts

- What is the difference between supervised and unsupervised learning?
- What is a training set and a test set? Why do we need them?
- Explain overfitting and underfitting in machine learning.
- What is the difference between classification and regression?
- How does a decision tree work?
- What is K-Nearest Neighbors (KNN), and how does it work?

3. Data Handling & Preprocessing

- How do you deal with missing values in a dataset?
- What is data normalization, and why is it important?
- What are categorical variables, and how would you handle them in a dataset?
- What is the difference between one-hot encoding and label encoding?

4. Basic Algorithms & Models

- What is linear regression, and how is it used in data science?
- What is logistic regression, and how is it different from linear regression?
- What is clustering? Can you explain k-means clustering?
- Explain the concept of cross-validation.
- What is a confusion matrix, and how is it used to evaluate a model?

5. SQL & Database Queries

- What is a SQL JOIN? Can you explain the different types (INNER, LEFT, RIGHT)?
- Write a SQL query to select the top 5 highest-paid employees.
- How would you find duplicate records in a table?
- What is a primary key and foreign key?

6. Python Programming Basics

- What are lists and tuples, and how do they differ?
- How would you reverse a string in Python?
- Explain what a for-loop is and how it works.
- What are functions in Python, and why are they useful?

7. Data Visualization

- What are some common tools used for data visualization (e.g., Matplotlib, Seaborn)?
- How would you choose the right chart type to visualize data?
- Can you explain what a histogram and a bar chart are and the difference between them?

8. Business Understanding & Problem Solving

- If you are given a dataset, what are the first steps you would take to analyze it?
- How would you explain your analysis or model to someone without a technical background?

9. General Questions

- Why did you choose data science as a career?
- Can you describe a project where you used data to solve a problem?
- How do you stay updated with the latest trends in data science?

For one to two years Experience:

1. Statistics

- What is the difference between variance and standard deviation?
- Explain the Central Limit Theorem.
- What is p-value, and why is it important in hypothesis testing?
- Can you explain Type I and Type II errors?
- What is correlation, and how does it differ from causation?
- What are confidence intervals, and how do you interpret them?

2. Probability

- What is Bayes' Theorem? Can you provide an example?
- Explain the difference between conditional probability and joint probability.
- What are the different probability distributions you are familiar with (e.g., normal, binomial, Poisson)?

3. Machine Learning

- Explain the difference between supervised and unsupervised learning.
- What is the bias-variance tradeoff in machine learning?
- Can you describe the difference between L1 and L2 regularization?
- What is overfitting, and how can you prevent it?
- Explain k-fold cross-validation.
- How does a decision tree work, and what are its advantages and disadvantages?
- What is the difference between bagging and boosting?

4. Algorithms and Models

- How does a Random Forest model work?
- Explain the working of Gradient Boosting algorithms like XGBoost or LightGBM.
- What is Support Vector Machine (SVM)? How does it work?
- How does k-means clustering work?
- What are neural networks, and how are they trained?

5. Data Manipulation and Visualization

- What are the key libraries in Python used for data manipulation (e.g., Pandas, NumPy)?
- How would you handle missing data in a dataset?
- What is data normalization, and why is it important?
- Can you explain dimensionality reduction techniques like PCA (Principal Component Analysis)?
- How do you assess the quality of your data?

6. SQL

- What is the difference between JOIN, LEFT JOIN, and RIGHT JOIN in SQL?
- How do you optimize a slow query?
- Write a SQL query to find the second-highest salary in a table.
- What is a primary key and foreign key in SQL?
- Explain the difference between UNION and UNION ALL.

7. Programming

- How do you handle memory management in Python?
- What are Python decorators, and how do they work?
- What is the difference between a list and a tuple in Python?

8. Big Data and Tools

- What is Hadoop, and what are its main components?
- What is Spark, and how is it different from Hadoop?
- What are the pros and cons of using NoSQL vs SQL databases?

9. Business and Problem-Solving

- How would you explain a complex model or finding to a non-technical stakeholder?
- How do you define success for a data science project?
- What steps would you take to clean a large, unstructured dataset?

Project mentioned in Resume:

1) Project: Exploratory Data Analysis on Titanic Dataset

- Performed data cleaning, feature engineering, and data visualization to understand key factors influencing passenger survival rates.
- Used Python libraries (Pandas, Matplotlib) to identify correlations and visualize distributions.

2) Project: Customer Churn Prediction

- Developed a machine learning model to predict customer churn using logistic regression.
- Performed feature selection and model tuning, achieving an accuracy of 85%.

3) Project: Customer Segmentation using K-Means Clustering

- Applied K-Means clustering on retail customer data to segment customers based on purchase history.
- Identified 3 key customer segments, which helped the business target marketing campaigns more effectively.

4) Project: Sentiment Analysis of Amazon Product Reviews

- Processed and analyzed 10,000 customer reviews to classify sentiments using Natural Language Processing.
- Achieved 82% accuracy in sentiment classification using a Naive Bayes classifier.

5) Project: Sales Forecasting for Retail Data

- Developed a time series forecasting model to predict monthly sales using ARIMA.
- Improved forecast accuracy by 10% through seasonal decomposition and hyperparameter tuning.

6) Project: Sales Dashboard using Power BI

- Developed an interactive dashboard to track sales performance across different regions and product categories.
- Utilized various charts and filters, allowing business stakeholders to drill down into specific KPIs.

7) Project: Movie Recommendation System

- Built a movie recommendation system using collaborative filtering techniques.
- Achieved an RMSE of 0.92 on the test data, providing personalized movie recommendations based on user preferences.

8) Project: Automated Data Pipeline for E-commerce Data

- Created a pipeline to collect, clean, and store sales data from multiple sources.
- Automated daily data updates, reducing manual efforts by 80%.

JD for DATA SCIENTIST:

For Fresher:

Job Summary:

We are looking for a motivated and detail-oriented Junior Data Scientist to join our data science team. As a fresher, you will work closely with senior data scientists and analysts to develop data-driven solutions that impact business decisions. You will have the opportunity to apply your data science skills to real-world problems and gain hands-on experience in data analysis, modeling, and visualization.

Key Responsibilities:

- Collaborate with cross-functional teams to gather and analyze large datasets from various sources.
- Perform data cleaning, pre-processing, and exploratory data analysis (EDA) to identify trends and insights.
- Build and evaluate predictive models using machine learning algorithms (e.g., regression, classification, clustering).
- Assist in the development of dashboards and data visualizations to communicate insights to stakeholders using tools like Power BI, Tableau, or Matplotlib.
- Work on feature engineering and optimization of machine learning models.
- Participate in model validation, tuning, and performance evaluation using metrics such as accuracy, precision, recall, and F1 score.
- Assist in the development and maintenance of data pipelines and automated workflows.
- Stay up-to-date with the latest developments in data science, machine learning, and AI technologies.

Key Skills and Qualifications:

- Bachelor's degree in Data Science, Computer Science, Statistics, Mathematics, Engineering, or a related field.
- Strong understanding of basic statistical concepts and probability.
- Knowledge of machine learning algorithms and model evaluation techniques.
- Proficiency in programming languages such as Python or R, and data manipulation libraries like Pandas and NumPy.
- Familiarity with machine learning libraries such as Scikit-learn, TensorFlow, or PyTorch (optional for freshers).
- Experience with SQL for querying databases.
- Ability to work with data visualization tools such as Matplotlib, Seaborn, Power BI, or Tableau.
- Strong problem-solving skills and attention to detail.
- Good communication skills and the ability to work in a team environment.

Preferred Qualifications:

- Experience working on data science projects (either in an academic setting or through internships) and applying machine learning models.
- Knowledge of cloud platforms like AWS, Google Cloud, or Azure is a plus.
- Familiarity with big data technologies like Hadoop or Spark (optional for freshers).
- Strong interest in learning new tools, technologies, and methodologies related to data science.

What We Offer:

- An opportunity to work in a dynamic, innovative, and fast-growing environment.
- Hands-on training and mentorship from senior data scientists.
- A collaborative work environment with a strong focus on personal and professional growth.
- Exposure to diverse and challenging real-world data problems.
- Competitive salary and benefits package.

Application Process:

To apply for this role, please submit your resume along with a brief description of your data science projects and any relevant coursework or certifications. A link to your GitHub or portfolio is a plus.