**Data Scientist Profile**

**Statistics & Probability**

**Interview Questions**

* + Explain mean, median, mode.
  + Difference between variance and standard deviation.
  + Hypothesis testing (p-value, confidence intervals).
  + Probability puzzles (cards, dice, conditional probability).

**Related Questions**

* + Central Limit Theorem.
  + Bayes’ Theorem (applications in spam filtering, recommendations).
  + A/B testing – design & evaluation.

**Algorithms & Math**

**Interview Questions**

* + Gradient Descent – how it works.
  + Cost functions in ML.
  + PCA (dimensionality reduction)

**Machine Learning**

**Interview Questions**

* + Explain linear regression vs logistic regression.
  + Decision Trees vs Random Forest.
  + What is overfitting and how to prevent it?
  + Hyperparameter tuning methods (Grid search, Random search).
  + How do you handle class imbalance?

**Related Questions**

* + Regularization (L1, L2).
  + SVM – intuition & kernel trick.
  + Ensemble learning methods (bagging, boosting).
  + Feature selection methods.

**ML Ops / Deployment:**

* + Model versioning & monitoring.
  + Handling data drift & concept drift.
  + CI/CD pipelines for ML.

**Deep Learning**

**Interview Questions**

* + Difference between CNN and RNN.
  + What is backpropagation?
  + Use cases of Transformers

**Related Questions**

* + Activation functions (ReLU, Sigmoid, Softmax).
  + Transfer learning.
  + Attention mechanism in NLP.

**Programming (Python/SQL)**

**Interview Questions**

* + Write SQL query for nth highest salary.
  + Detect and treat missing values in Pandas.
  + Outlier detection methods in Python.
  + Implement logistic regression from scratch in Python.

**Related Questions**

* + Python: list comprehensions, decorators, multiprocessing.
  + SQL: window functions, joins, group by with aggregations.

**Projects / Case Studies**

* + Explain an ML model you built end-to-end.
  + How would you scale your model for millions of users?
  + How do you ensure fairness and avoid bias in ML models?

**HR**

* + Why Data Science?
  + Example of solving a real-world problem using data.
  + Collaboration with cross-functional teams.

**Preparation Strategy**

1. **Mathematics & Statistics**
   * Revise probability, statistics, distributions, and hypothesis testing.
   * Practice probability puzzles and A/B testing design.
2. **Machine Learning Fundamentals**
   * Be able to explain common algorithms intuitively (Linear/Logistic Regression, Trees, SVM, Clustering).
   * Revise overfitting, bias-variance tradeoff, hyperparameter tuning.
   * Practice implementing ML algorithms in Python (scikit-learn).
3. **Deep Learning**
   * Revise CNN, RNN, LSTMs, Transformers basics.
   * Understand backpropagation and activation functions.
   * Work with frameworks like TensorFlow or PyTorch.
4. **Programming (Python + SQL)**
   * Be fluent in Python libraries: Pandas, NumPy, Scikit-learn, Matplotlib/Seaborn.
   * Practice SQL queries: joins, window functions, aggregation.
   * Implement small ML models end-to-end.
5. **Projects / Case Studies**
   * Prepare to explain at least 2 projects in detail (data → preprocessing → model → results).
   * Be ready with scaling and deployment strategies.
   * Show awareness of real-world constraints (imbalanced data, drift, fairness).
6. **HR / Behavioral**
   * Prepare STAR stories about applying data science to business impact.
   * Be ready to explain work with non-technical stakeholders.
   * Stay updated on trends (LLMs, Generative AI, FinTech/Healthcare applications, etc.).