Data Scientist Internship Preparation Guide

1. Python for Data Science

- Data structures: lists, dicts, sets

- Loops, functions, conditionals

- List comprehensions

- Libraries:

  - NumPy -> arrays, broadcasting

  - Pandas -> reading CSVs, filtering, groupby, merging

  - Matplotlib / Seaborn -> basic plots

2. Statistics & Probability

- Mean, median, mode, variance, standard deviation

- Probability basics (conditional, Bayes' theorem)

- Normal distribution, binomial distribution

- Hypothesis testing (p-values, t-tests, chi-square)

3. Exploratory Data Analysis (EDA)

- Handling missing values, outliers

- Feature scaling and encoding

- Correlation analysis

- Visualizing distributions and trends

4. Machine Learning (Basics)

- Supervised Learning:

  - Linear regression

  - Logistic regression

  - Decision trees

  - KNN

- Unsupervised Learning:

  - K-Means clustering

  - PCA (Principal Component Analysis)

- Model evaluation:

- Confusion matrix, accuracy, precision, recall

- Train/test split and cross-validation

5. SQL & Databases

- SELECT, WHERE, GROUP BY, ORDER BY

- JOINS (Inner, Left, Right)

- Aggregation (COUNT, SUM, AVG)

- Subqueries

6. Mini-Projects or Case Studies

- Use public datasets (e.g., Titanic, Sales, Housing Prices)

- End-to-end projects:

  - Data cleaning -> EDA -> ML modeling -> Evaluation

- Upload to GitHub or make a Jupyter Notebook report

7. Communication & Soft Skills

- Explain your project in simple terms

- Interpret charts and metrics clearly

- Tell a "data story" (problem -> analysis -> solution)

8. Bonus Tools (optional but useful)

- Jupyter Notebook or Google Colab

- Git & GitHub (version control, portfolio)

- Tableau / Power BI (basic visualization tools)

Pro Tips

- Practice Python & SQL on LeetCode, HackerRank, or StrataScratch

- Apply to internships on LinkedIn, Internshala, AngelList

- Create a 1-page resume with:

  - Education

Data Scientist Internship Preparation Guide

- Skills

- 2-3 short projects

- GitHub link