□ Ultimate Data Science Roadmap

This roadmap is structured to not only tell you **what to learn**, but also **why it matters** in Data Science, **what exactly to focus on**, and most importantly, **how to practice it**. It's designed for absolute beginners and can guide you to a level where you're confident working on real-world ML projects.

---

1. Python Programming □

Why It Matters:

Python is the **foundation** of Data Science. It's easy to learn, widely used, and comes with a massive ecosystem of libraries for analysis, modeling, and deployment.

What to Learn:

- Variables, Data Types, Operators
- Loops and Control Statements (if-else, for, while, break)
- Functions & Lambda Functions
- f-string & doc strings
- Lists, Tuples, Dictionaries, Sets
- File Handling
- Raising Custom Errors & Exception Handling (with Finally Keyword)
- Enumerate and its Use Cases
- List Comprehensions
- Understanding `if __name__ == "__main__"`
- Object-Oriented Programming (OOP)
- Working with external libraries

How to Practice:

- Solve 50+ Python practice problems on your choice of platform.
- Build mini projects: To-Do List, Calculator, Simple Quiz App
- Use platforms like HackerRank, LeetCode (easy category)

---

2. Python for Data Analysis

Why It Matters:

You'll use libraries like **NumPy**, **Pandas**, and **Matplotlib** for manipulating, cleaning, and visualizing data—skills required before any modeling.

What to Learn:

- NumPy: Arrays, indexing, broadcasting
- Pandas: DataFrames, filtering, merging, groupby
- Matplotlib & Seaborn: Line, bar, scatter, histograms
- Exploratory Data Analysis (EDA)

How to Practice:

- Kaggle's Titanic or IRIS dataset
- Download CSVs and perform complete EDA on them

---

3. SQL for Data Science

Why It Matters:

SQL is essential when working with structured data in databases. Most interviews ask SQL questions.

What to Learn:

- SELECT, WHERE, GROUP BY, ORDER BY
- Joins (INNER, LEFT, RIGHT)
- Window functions
- Subqueries and CTEs

How to Practice:

- Use platforms like Mode Analytics, StrataScratch, LeetCode (SQL section)
- Work on sample HR or sales datasets

---

4. Math for Data Science □

Why It Matters:

Math is the backbone of ML models. It helps you **understand what's happening under the hood.**

What to Learn:

a. Linear Algebra

Required to learn **Linear Regression**, **Dimensionality Reduction**

- Vectors, Matrices, Operations
- Dot Product, Matrix Multiplication
- Eigenvalues and Eigenvectors

b. Statistics

Required for **Model Evaluation**, **Probability Models**

- Mean, Median, Mode, Variance
- Probability, Bayes Theorem
- Hypothesis Testing
- Distributions (Normal, Binomial, etc.)

c. Calculus

Required for **Gradient Descent**, **Neural Networks**

- Derivatives
- Chain Rule
- Partial Derivatives

How to Practice:

- YouTube - 3Blue1Brown
- By studying Machine Learning

5. Machine Learning □

Why It Matters:

This is the heart of Data Science—teaches machines to learn patterns from data.

What to Learn:

- Supervised vs Unsupervised Learning
- Regression Algorithms (Linear, Ridge, Lasso)
- Classification Algorithms (Logistic, Decision Trees, KNN, SVM)
- Clustering (KMeans, Hierarchical)
- Evaluation Metrics (Accuracy, Precision, Recall, F1, AUC)

How to Practice:

- Work on datasets from Kaggle
- Build projects: House Price Prediction, Spam Detection
- Time to parallelly get started with MLOps

---

6. Deep Learning □

Why It Matters:

Deep Learning enables powerful applications in **NLP**, **Computer Vision**, and **LLMs**.

What to Learn:

- Basics of Neural Networks
- Forward & Backward Propagation
- CNN (For Image Data) → Foundation for Computer Vision
- RNN, LSTM (For Sequential Data) → Foundation for NLP
- Transformers → Foundation for LLMs

How to Practice:

- Use TensorFlow or PyTorch on small projects
- MNIST digit classification, Sentiment analysis

7. Natural Language Processing (NLP)

Why It Matters:

NLP allows machines to understand and work with human language. Critical for LLM-based applications.

What to Learn:

- Text Cleaning & Tokenization
- Bag of Words, TF-IDF
- Word Embeddings (Word2Vec, GloVe)
- Transformers (BERT, GPT family)
- Named Entity Recognition, Text Summarization

How to Practice:

- Sentiment Analysis using Twitter data
- Build a simple chatbot using rules + BERT

---

8. GenAI & LLMs

Why It Matters:

LLMs are at the **frontier** of AI. LangChain, OpenAI APIs, and prompt engineering are reshaping industries.

What to Learn:

- Prompt Engineering Basics
- OpenAI API (GPT Models)
- LangChain for chaining LLM calls
- LangGraph for structured LLM workflows
- RAG (Retrieval Augmented Generation)

How to Practice:

- Use OpenAI API to build a resume review bot
- Create a LangChain-based PDF QnA chatbot

Final Advice □

- Don’t try to learn everything at once. Go step by step.
- Focus on understanding the **why**, not just the **how**.
- Build projects after every major concept.
- Keep updating your LinkedIn, GitHub, and resume as you grow.

□ Resources to Kickstart:

- **Free MLOps Playlist** → here
- **Docker Masterclass** →here
- **Follow me on LinkedIn** → here
- **1-1 Mentorship Sessions** → Book a Call