

The Data Science Roadmap 2026

What to learn, in what order — from zero to job-ready in India.

A free guide by Pankit Kumar

Lead Instructor · Sr. Data Scientist, Parexel (a Goldman Sachs-backed company)

The roadmap, in the order that works

Realistic timings assume ~10–12 focused hours a week. The order matters as much as the list.

0

Start honest

1 week

Assess your real starting point. Never touched data? Do the analyst basics first— skipping the foundation is the #1 reason people stall.

1

Python

4–6 weeks

Syntax → data structures → functions → pandas, NumPy, matplotlib. Goal: load a messy CSV, clean it, answer questions about it.

2

SQL (in parallel)

2–3 weeks

SELECT / WHERE / GROUP BY → joins → subqueries & CTEs → window functions. This is how you get data in every real job.

3

Statistics & EDA

3–4 weeks

Distributions, sampling, correlation vs causation, hypothesis testing— enough to reason honestly, not a maths PhD.

4

Machine Learning

6–8 weeks

Workflow (train/val/test, overfitting) → regression, classification, trees, ensembles → clustering → honest evaluation.

...continued

5

Deep Learning & NLP

4-5 weeks

Neural-net fundamentals, then NLP and transformers — the architecture behind every modern LLM.

6

Modern AI: LLMs & RAG

3-4 weeks

LLM APIs, structured output, retrieval-augmented generation and how to evaluate it. The 2026 differentiator.

7

Deployment

2-3 weeks

FastAPI + Streamlit + Docker → a live URL. The skill most graduates lack, so it's worth the most in interviews.

8

Portfolio & interviews

ongoing

Build 3-5 deployed projects throughout; practise talking through your decisions. Shipping beats watching.

How long does it take?

Realistically 6–9 months of consistent effort from a reasonable starting point — faster with structure and mentorship, slower and with more dead-ends self-teaching from scattered content. The failure mode of self-study isn't lack of material; it's lack of sequencing and feedback.

The mistakes that waste months

- ✗ Tutorial hell — watching endless videos without building anything. Build early, build often.
- ✗ Skipping SQL and fundamentals to jump straight to deep learning. You'll hit a wall.
- ✗ Collecting certificates instead of shipping projects. Employers look at your GitHub.
- ✗ Learning alone with no feedback — a mentor catches in minutes what takes you weeks to notice.

The one rule that beats everything

“The person who ships three deployed projects beats the person who watched forty of tutorials. Build, don't just consume.”

Want this taught with 1:1 review and a portfolio you can show?

Our Data Science programme runs this exact roadmap — deployment and modern AI, not just modelling.

neurapath.in/courses/ds