

DATA GEEK

by **CHERIEDU.COM**

High-Level Learning Book for Modern Data and AI Careers



Index

1. Book Vision and Learning Method
2. Core Foundations (Data + Business + Statistics)
3. SQL for Analytics and Engineering
4. Python and NumPy for Data Work
5. Data Engineering Systems
6. Data Science and Machine Learning
7. Vectors and Similarity Search
8. RAG (Retrieval-Augmented Generation)
9. LLM Engineering
10. Agent Engineering
11. 20+ Model Portfolio Map
12. End-to-End 6-Step Project Architecture
13. Role-wise Roadmap (DA/DE/DS/AIE/MLE/AGE)
14. Career and Portfolio Strategy

Content

1) Book Vision and Learning Method

DATA GEEK is a self-learning roadmap. It focuses on practical understanding and connects each topic to real project use.

2) Core Foundations (Data + Business + Statistics)

All six roles need data thinking, business metric thinking, and statistical reasoning before advanced tools.

3) SQL for Analytics and Engineering

SQL powers extraction, aggregation, KPI creation, data validation, and dataset preparation for analytics and ML systems.

4) Python and NumPy for Data Work

Python builds automation and reproducible workflows. NumPy provides efficient numerical operations for feature and model pipelines.

5) Data Engineering Systems

Ingestion, transformation, storage, orchestration, quality checks, and governance are the core layers of production data platforms.

6) Data Science and Machine Learning

The lifecycle includes framing, feature design, training, evaluation, deployment, and monitoring for drift and performance stability.

7) Vectors and Similarity Search

Vectors convert behavior into numeric representations. Similarity retrieval helps segmentation, peer lookup, and recommendation systems.

8) RAG (Retrieval-Augmented Generation)

RAG improves response quality by retrieving trusted context first, then generating

grounded responses with lower hallucination risk.

9) LLM Engineering

LLM systems require prompt design, context control, tool invocation, evaluation, cost management, and safety guardrails.

10) Agent Engineering

Agents execute multi-step workflows using tools and memory. Reliable agents require planning logic, fallback logic, and strict boundaries.

11) 20+ Model Portfolio Map

This repository includes more than 20 model projects across forecasting, risk, segmentation, churn, copilots, and RAG workloads.

12) End-to-End 6-Step Project Architecture

Using RevenuePrediction planning: ingest, feature engineering, vectorization, RAG retrieval, model training, and prediction with LLM explanation.

13) Role-wise Roadmap

DA: insight and metrics. DE: pipelines and reliability. DS: models and experiments. AIE: LLM apps and RAG. MLE: deployment and monitoring. AGE: autonomous tool workflows.

14) Career and Portfolio Strategy

A strong portfolio should show business impact, technical depth, communication quality, and production readiness.