Phase 1 Weekly Plan: Foundations (Months 1–6)

January 2026 – June 2026 | 4-5 hours/week

Total Time Budget: ~24 weeks × 4.5 hrs = **108 hours**

Goals:

- Master SQL for data analysis and databases
- Learn professional software engineering practices
- Formalize ML knowledge with scikit-learn
- Complete 2 Kaggle competitions
- Build first portfolio projects

Month 1: SQL Fundamentals (Weeks 1–4)

Time Allocation: 4-5 hrs/week = 18-20 total hours

Week 1 | Jan 13–19

Focus: SQL Basics - SELECT, WHERE, ORDER BY

Tasks:

- ☐ Set up environment: Install PostgreSQL or use SQLite Browser (30 min)
- ☐ Mode Analytics Tutorial: Sections 1-3 (2 hrs)
  - Basic SQL
  - Intermediate SQL
  - SQL Aggregations
- ☐ Practice: 10 basic SELECT queries on sample dataset (1.5 hrs)

Resources: - Mode Analytics SQL Tutorial - Dataset: Kaggle Sakila Sample Database

Deliverable: Write and save 10 basic queries with comments

Week 2 | Jan 20–26

Focus: SQL Joins & Subqueries

Tasks:

- ☐ Mode Analytics Tutorial: Sections 4-5 (2 hrs)
  - SQL Joins
  - SQL Subqueries
- ☐ Practice: Write 15 queries using INNER/LEFT/RIGHT JOIN (2 hrs)
- ☐ Learn about foreign keys and database relationships (30 min)
- ☐ Read: "Why physicists struggle with SQL" mindset shift (30 min)

Challenge: Query from 3+ tables with multiple joins

Deliverable: GitHub Gist with annotated JOIN examples

Week 3 | Jan 27–Feb 2

Focus: Aggregations & Window Functions

Tasks:

- ☐ Mode Analytics Tutorial: Advanced Aggregations (1.5 hrs)
- ☐ Learn window functions: ROW_NUMBER(), RANK(), PARTITION BY (2 hrs)
- ☐ Practice: 10 queries with GROUP BY, HAVING, window functions (1.5 hrs)

Resources: - PostgreSQL Window Functions

Deliverable: Solve 5 medium-difficulty SQL problems on LeetCode

---

Week 4 | Feb 3–9

Focus: Real-World SQL Project

Tasks:

- ☐ Download real dataset (e.g., NYC Taxi Data sample) (30 min)
- ☐ Import into SQLite/PostgreSQL (1 hr)
- ☐ Create 10 analytical queries answering business questions (2.5 hrs)
  - Example: "What's the average fare by hour of day?"
  - Example: "Which pickup zones have highest tip percentages?"
- ☐ Document queries with explanations (1 hr)

Deliverable: GitHub repo: `sql-data-analysis-project/` with README

Month 2: Advanced SQL & Git Workflow (Weeks 5–8)

Time Allocation: 4-5 hrs/week = 18-20 total hours

Week 5 | Feb 10–16

Focus: CTEs, Query Optimization

Tasks:

- ☐ Learn Common Table Expressions (CTEs) vs subqueries (1.5 hrs)
- ☐ Query performance: EXPLAIN, indexes, query plans (1.5 hrs)
- ☐ Practice: Refactor previous queries using CTEs (1 hr)
- ☐ Complete 10 LeetCode SQL problems (medium difficulty) (1.5 hrs)

Reading: "Use The Index, Luke" - database performance guide

Deliverable: Comparison document: subquery vs CTE performance

---

Week 6 | Feb 17–23

Focus: Git & GitHub Fundamentals

Tasks:

- ☐ Git basics: clone, commit, push, pull (1 hr)
- ☐ Branching strategy: feature branches, main/develop (1 hr)
- ☐ Practice: Create repo, make 10 commits with good messages (1.5 hrs)
- ☐ Pull request workflow on a practice repo (1 hr)
- ☐ Set up GitHub profile README (30 min)

Resources:

- Git Handbook
- Conventional Commits

Deliverable: Professional GitHub profile with pinned SQL project

---

Week 7 | Feb 24–Mar 2

Focus: Python Project Structure

Tasks:

- ☐ Learn: `pyproject.toml`, virtual environments, dependency management (1.5 hrs)
- ☐ Set up project template with proper structure (1 hr)

```
my-project/
   src/
   tests/
   docs/
   pyproject.toml
   README.md
```

- ☐ Install and use Poetry or uv for dependency management (1 hr)
- ☐ Refactor one old notebook into proper .py modules (1.5 hrs)

Resources: - Real Python - Structuring Projects

Deliverable: Template repo: `python-project-template/`

Week 8 | Mar 3–9

Focus: Testing & Documentation

Tasks:

- ☐ pytest basics: write 5 simple unit tests (1.5 hrs)
- ☐ Learn docstring conventions (NumPy/Google style) (1 hr)
- ☐ Add docstrings to previous project functions (1 hr)
- ☐ Set up GitHub Actions for basic CI (run tests on push) (1.5 hrs)

Resources: - pytest documentation - Google Python Style Guide

Deliverable: Python project with tests + CI badge in README

Month 3: Machine Learning Refresher (Weeks 9–12)

Time Allocation: 4-5 hrs/week = 18-20 total hours

Week 9 | Mar 10–16

Focus: Scikit-learn Fundamentals

Tasks:

- ☐ Scikit-learn tutorial: datasets, train/test split, evaluation (2 hrs)
- ☐ Build first model: linear regression on simple dataset (1.5 hrs)
- ☐ Learn: preprocessing, StandardScaler, one-hot encoding (1 hr)

Hands-on: Predict Boston housing prices or similar regression task

Resources: - Scikit-learn User Guide

Deliverable: Jupyter notebook with annotated regression workflow

---

Week 10 | Mar 17–23

Focus: Classification Algorithms

Tasks:

- ☐ Implement: Logistic Regression, Decision Trees, Random Forest (2 hrs)
- ☐ Evaluate: accuracy, precision, recall, F1, confusion matrix (1.5 hrs)
- ☐ Learn: cross-validation, hyperparameter tuning with GridSearchCV (1 hr)

Hands-on: Binary classification on Titanic dataset

Deliverable: Comparison notebook: 3 algorithms on same dataset

---

Week 11 | Mar 24–30

Focus: Model Selection & Validation

Tasks:

- ☐ Read: Géron's "Hands-On ML" Chapter 2-3 (2 hrs)
- ☐ Implement: train/validation/test split properly (1 hr)
- ☐ Learn: overfitting, bias-variance tradeoff, learning curves (1 hr)
- ☐ Practice: visualize model performance with plots (1 hr)

Deliverable: Blog post draft: "Common ML mistakes physicists make"

Week 12 | Mar 31–Apr 6

Focus: First Kaggle Competition

Tasks:

- ☐ Choose: Kaggle "Getting Started" competition (30 min)
  - Titanic or House Prices
- ☐ Exploratory Data Analysis (2 hrs)
- ☐ Feature engineering (1.5 hrs)
- ☐ Submit first prediction (1 hr)

Goal: Don't aim to win, aim to complete the full workflow

Deliverable: Kaggle submission + GitHub repo with analysis

Month 4: Advanced ML & Portfolio Project (Weeks 13–16)

Time Allocation: 4-5 hrs/week = 18-20 total hours

Week 13 | Apr 7–13

Focus: Ensemble Methods & Feature Engineering

Tasks:

- ☐ Learn: Gradient Boosting (XGBoost, LightGBM) (1.5 hrs)
- ☐ Implement XGBoost on Kaggle competition (2 hrs)
- ☐ Advanced feature engineering techniques (1.5 hrs)
  - Feature interactions
  - Polynomial features
  - Handling missing data strategically

Deliverable: Improve Kaggle leaderboard position

Week 14 | Apr 14–20

Focus: Unsupervised Learning

Tasks:

- ☐ K-means clustering on astronomy or synthetic data (1.5 hrs)
- ☐ PCA for dimensionality reduction (1.5 hrs)
- ☐ DBSCAN for anomaly detection (1 hr)
- ☐ Visualization: t-SNE or UMAP on high-dimensional data (1 hr)

Physics connection: Apply to spectroscopy or simulation data

Deliverable: Clustering analysis notebook

Week 15 | Apr 21–27

Focus: Start Portfolio Project #1

Project: Anomaly Detection (Cosmology → Fraud)

Tasks:

- ☐ Define project scope and dataset (1 hr)
- ☐ Set up GitHub repo with proper structure (30 min)
- ☐ Data exploration and cleaning (2 hrs)
- ☐ First model baseline (1 hr)

Dataset ideas: - Credit Card Fraud - Network Intrusion Detection

Deliverable: Project repo initialized with README and EDA

Week 16 | Apr 28–May 4

Focus: Complete Portfolio Project #1

Tasks:

- ☐ Implement multiple anomaly detection methods (2 hrs)
  - Isolation Forest
  - One-Class SVM
  - Autoencoder (bonus)
- ☐ Compare performance on imbalanced data (1 hr)
- ☐ Create visualizations and write clear README (1.5 hrs)
- ☐ Add docstrings and tests (30 min)

Deliverable: Polished GitHub project ready to showcase

Month 5: Second Kaggle + SQL Portfolio (Weeks 17–20)

Time Allocation: 4-5 hrs/week = 18-20 total hours

Week 17 | May 5–11

Focus: Second Kaggle Competition

Tasks:

- ☐ Choose intermediate competition (time-series or tabular) (30 min)
- ☐ Study winning solutions from past competitions (1 hr)
- ☐ EDA + baseline model (2 hrs)
- ☐ Feature engineering iteration 1 (1.5 hrs)

Deliverable: Initial submission + analysis notebook

Week 18 | May 12–18

Focus: Improve Kaggle Submission

Tasks:

- ☐ Advanced feature engineering (1.5 hrs)
- ☐ Model stacking or ensembling (2 hrs)
- ☐ Hyperparameter tuning (1 hr)
- ☐ Final submission (30 min)

Goal: Top 50% of leaderboard

Deliverable: Complete Kaggle kernel/notebook

Week 19 | May 19–25

Focus: SQL Portfolio Project #2

Project: Retail/Business Analytics with SQL

Tasks:

- ☐ Choose dataset (e.g., Northwind, Adventure Works) (30 min)
- ☐ Set up database and import data (1 hr)
- ☐ Write 15-20 analytical queries (2.5 hrs)
  - Sales trends

  - Customer segmentation
  - Product performance
  - Cohort analysis

☐ Create README with business questions + SQL solutions (1 hr)

Deliverable: GitHub repo: `sql-business-analytics/`

Week 20 | May 26–Jun 1

Focus: Data Visualization

Tasks:

☐ Learn Matplotlib/Seaborn best practices (1.5 hrs)
☐ Create dashboard-style visualizations for SQL project (2 hrs)
☐ Optional: Try Plotly for interactive plots (1 hr)
☐ Add visualizations to project READMEs (30 min)

Resources:

- Matplotlib Gallery
- Seaborn Tutorial

Deliverable: Updated projects with professional visualizations

---

Month 6: Integration & Planning (Weeks 21–24)

Time Allocation: 4-5 hrs/week = 18-20 total hours

Week 21 | Jun 2–8

Focus: Physics-to-Industry Bridge Project Planning

Tasks:

- ☐ Brainstorm: Which simulation code can be adapted? (1 hr)
- ☐ Research: What's the industry equivalent problem? (1.5 hrs)
- ☐ Write project proposal document (1 hr)
- ☐ Sketch out data pipeline architecture (1 hr)

Project ideas:

- Hydro-sim → Predictive maintenance
- N-body methods → Satellite trajectory optimization (aerospace appeal)
- N-body methods → Traffic flow
- Spectral analysis → Sensor data classification
- CFD → Aerodynamic surrogate models (aerospace appeal)

Aerospace Focus: Prioritize projects that demonstrate orbital mechanics, fluid dynamics, or optimization—these directly appeal to Munich aerospace companies (Airbus, OHB, Lilium, IABG).

Deliverable: Project proposal markdown document

Week 22 | Jun 9–15

Focus: Start Physics-Industry Project

Tasks:

- ☐ Set up repo with proper Python project structure (30 min)
- ☐ Identify/generate synthetic dataset (1.5 hrs)
- ☐ Port relevant physics code to clean modules (2 hrs)
- ☐ Initial EDA (1 hr)

Deliverable: Project foundation with simulation code adapted

Week 23 | Jun 16–22

Focus: GitHub Profile Polish

Tasks:

- ☐ Update all project READMEs with:
  - Clear problem statements
  - Results/findings section
  - Professional badges (build status, etc.)
  - Screenshots/plots (1.5 hrs)
- ☐ Pin best 4-6 repositories (15 min)
- ☐ Update GitHub profile README with:
  - Professional summary
  - Tech stack
  - Featured projects (1 hr)
- ☐ Add LinkedIn link, website (if have one) (15 min)
- ☐ Review: Does profile tell a coherent story? (1.5 hrs)

Deliverable: Professional GitHub portfolio ready to share

Week 24 | Jun 23–29

Focus: Phase 1 Review & Phase 2 Prep

Tasks:

- ☐ Complete Phase 1 checklist review (1 hr)
- ☐ Update main career plan document with progress (30 min)
- ☐ Sign up for Deep Learning Specialization on Coursera (15 min)
- ☐ Research Azure DP-100 exam format (1 hr)
- ☐ Set Phase 2 calendar reminders (15 min)
- ☐ Celebrate progress! Write reflection on what you learned (1.5 hrs)

Deliverable: Phase 1 completion report + Phase 2 kickoff plan

---

Weekly Time Breakdown Template

Use this to track your actual time spent each week:

```
### Week X | Date Range

**Planned:** 4.5 hours
**Actual:** _____ hours

| Day | Task | Time | Notes |
|-----|------|------|-------|
| Mon |      |      |       |
| Tue |      |      |       |
| Wed |      |      |       |
| Thu |      |      |       |
| Fri |      |      |       |
| Sat |      |      |       |
| Sun |      |      |       |

**Completed:** [ ] Task 1, [ ] Task 2, etc.
**Blockers:** None / [describe issue]
**Next week prep:** [what to prep in advance]
```

Phase 1 Success Criteria

By end of Week 24, you should have:

- ☐ **SQL:** Completed 50+ problems, built 2 portfolio projects
- ☐ **Git/GitHub:** Professional profile with 6+ repos
- ☐ **Python:** Following best practices (structure, tests, docs)
- ☐ **ML:** Completed 2 Kaggle competitions
- ☐ **Portfolio:** 4-5 projects demonstrating progression
- ☐ **Confidence:** Ready to tackle Deep Learning content

Total Portfolio Projects:

1. SQL Data Analysis (NYC Taxi or similar)
2. Anomaly Detection (Cosmology → Fraud)
3. SQL Business Analytics (Northwind/Adventure Works)

4. Kaggle Competition #1 (complete workflow)
5. Kaggle Competition #2 (better performance)
6. Physics-Industry Bridge Project (foundation started)
    - **Prioritize aerospace-relevant**: Trajectory optimization or CFD-ML surrogate

Aerospace Career Bonus:

☐ **Optional:** Research one Munich aerospace company (Airbus, OHB, Lilium) and tailor project directly to their tech stack

☐ **Optional:** Join Munich Aerospace e.V. mailing list for events

---

Tips for Success

Consistency over intensity: 4-5 hours every week beats 10 hours sporadically.

Track progress publicly: Weekly commits to GitHub show momentum.

Join communities early: Lurking in Data Science Munich meetup Slack/Discord helps you absorb the culture.

Don't get stuck: If spending >2 hours on one problem, move on and ask for help.

Celebrate milestones: Each completed project is real progress.

Buffer time: Some weeks will be harder. It's OK to adjust.

Adjustment Notes

If you have more time (6-7 hrs/week):

- Add optional deep dives into topics
- Start blog posts about learnings
- Contribute to open source projects
- Attend virtual meetups

If you have less time (3 hrs/week):

- Prioritize SQL (Weeks 1-5) and ML refresher (Weeks 9-12)

- Reduce portfolio projects to 3 core ones
- Skip some LeetCode practice
- Extend timeline by 2-4 weeks

If you fall behind:

- Don't restart from scratch
- Skip to the most critical tasks
- Remember: progress > perfection

Last updated: January 13, 2026 Next review: End of Week 4 (Feb 9, 2026)