**🚀 MySQL Roadmap for Data Analysts (5 Weeks)**

**Week 1: Advanced Querying & Data Aggregation**

**Goal:** Cement fundamentals and master summarization.

**Day 1-2: Filtering & Sorting**

* Learn: SELECT, WHERE, ORDER BY, advanced filtering (IN, NOT IN, BETWEEN, LIKE with % and \_), IS NULL, IS NOT NULL.
* Practice: Queries for pattern matching, date range filtering, handling NULLs.
* Apply: Segment users by email provider (LIKE '%gmail.com') or registration date.

**Day 3-4: Aggregations**

* Learn: COUNT, SUM, AVG, MIN, MAX. Difference between WHERE vs HAVING with GROUP BY.
* Practice: Total sales per store, avg order value per month, employees per department.
* Apply: Find customer categories with avg purchase > $100.

**Day 5-6: Conditional Logic**

* Learn: CASE WHEN for segmentation.
* Practice: Tier customers as Gold/Silver/Bronze based on spending.
* Apply: Create a new column categorizing sales into High/Medium/Low.

**Day 7: Review + Mini-Project 1 🏆**

* Project: Analyze a movie dataset. Answer:
  1. How many movies per year?
  2. Top 5 highest-rated movies (≥ 50 ratings).
  3. Year with highest avg rating.

**Week 2: Combining & Shaping Data**

**Goal:** Work with multiple tables effectively.

**Day 8-10: JOINs**

* Learn: INNER, LEFT, RIGHT, FULL OUTER, UNION, UNION ALL.
* Practice: Join multiple tables, check NULL behaviour in joins.
* Apply: List customers + order counts (include those with no orders).

**Day 11-12: Subqueries & CTEs**

* Learn: Subqueries inside SELECT, FROM, WHERE.
* Learn: WITH (CTEs) for readability.
* Practice: Rewrite subqueries as CTEs.
* Apply: Find employees earning above department average using a CTE.

**Day 13-14: Intensive Practice**

* Focus: Solve JOIN & subquery problems on LeetCode/HackerRank.

**Day 15: Review + Mini-Project 2 🚀**

* Project: Using Sakila DB, find top 5 customers by spend (join Customer, Payment, Rental).

**Week 3: Analytical Power Tools**

**Goal:** Go from reporting to analysis.

**Day 16-18: Window Functions ✨**

* Learn: OVER(), PARTITION BY, ROW\_NUMBER(), RANK(), DENSE\_RANK(), LEAD(), LAG(), running totals.
* Practice: Rank employees by salary, calculate MoM sales growth, compute running totals.
* Apply: Find 2nd highest salary per department.

**Day 19-21: Data Cleaning & Transformation**

* Learn:
  + String: CONCAT, LEFT, RIGHT, TRIM, SUBSTRING, REPLACE, POSITION, UPPER, LOWER, CHAR\_LENGTH, REGEXP\_REPLACE, REGEXP\_SUBSTR.
  + Date/Time: DATE\_FORMAT, DATEDIFF, EXTRACT, NOW, CURDATE, TIMESTAMPDIFF, LAST\_DAY, STR\_TO\_DATE.
  + NULL Handling: COALESCE, IFNULL, NULLIF.
  + Conversion: CAST, CONVERT.
* Practice: Clean phone numbers, extract month from dates, merge names.
* Apply: Write a cleaning script to standardize text, format dates, fill NULLs.

**Day 22-23: Advanced Problems + Mini-Project 3 🧠**

* Project: Calculate MoM % growth in user sign-ups with COUNT, date grouping, and LAG.

**Week 4: Performance, Optimization & Projects**

**Goal:** Write efficient, production-ready queries.

**Day 24: Query Optimization**

* Learn: Index basics, clustered vs non-clustered, composite indexes.
* Tools: Use EXPLAIN to view query plan.
* Best Practices: Avoid SELECT \*, filter early, correct data types.
* Apply: Optimize a slow query with indexing.

**Day 25-29: Capstone Project 💼**

* Dataset: Kaggle/Maven Analytics (e.g., e-commerce).
* Define 3–5 questions, e.g.:
  1. Top 10 products by revenue.
  2. Most recent customers + purchases (ROW\_NUMBER).
  3. Avg time between 1st and 2nd order (LAG).
* Process: Clean, analyze, optimize queries.
* Deliverable: Store code + findings on GitHub (with comments).
* Extra: Add ERD diagram + a visualization (Tableau/Power BI).

**Day 30: Final Review 🎉**

* Consolidate notes + mini-projects into portfolio.
* Next Steps:
  + Continue SQL practice (interview-style problems).
  + Connect SQL with BI (Tableau/Power BI).
  + Connect SQL with Python (pandas, SQLAlchemy).

**Week 5 (Optional Extension): Analyst Workflow**

**Goal:** Bridge SQL with analytics ecosystem.

* Stored Procedures & Views → reusable queries.
* User-Defined Functions (UDFs).
* SQL + Python → data extraction + cleaning in SQL, analysis in pandas.
* SQL + BI → dashboards powered by queries.
* Practice real interview case studies (StrataScratch, Mode Analytics).

✅ By the end, you’ll have:

* Strong command of MySQL queries.
* Data cleaning, transformation, and aggregation skills.
* Understanding of optimization and performance.
* 3 mini-projects + 1 capstone (portfolio-ready).
* Skills in SQL → BI/Python integration (extra edge for a data analyst role).