SQL for Data Science

Course Overview

A comprehensive, industry-standard SQL for Data Science course outline is designed to take a learner from absolute database basics to advanced analytical data wrangling.This curriculum structures the core competencies required by modern data scientists into six progressive modules

Course Description

This course provides a comprehensive introduction to SQL (Structured Query Language) tailored specifically for data analysis and data science. Students will transition from writing basic database queries to mastering advanced data manipulation techniques used by industry professionals.Participants will learn how to extract, filter, clean, and aggregate complex datasets from relational databases. The curriculum emphasizes real-world application, teaching students how to join disparate data sources, write efficient analytical queries using window functions, and engineer features for predictive modeling. By the conclusion of this course, learners will seamlessly bridge the gap between relational databases and data science environments by integrating SQL workflows directly into Python and Pandas.

Learning Curriculum

Module 1

Database Fundamentals & Data Retrieval

Introduction to Relational Databases: Understanding Relational Database Management Systems (RDBMS) like PostgreSQL, MySQL, and SQLite.The Basic Query: Master SELECT and FROM syntax to pull columns from data tables.Limiting and Unique Records: Use LIMIT to truncate row counts and DISTINCT to extract unique records.Comment Hygiene: Writing clean code using inline (--) and block (/* */) commentary

Module 2

Filtering, Sorting, & Formatting

Conditional Filtering: Deploying the WHERE clause with comparison operators (=, >, <).Logical Operators: Combining filters using AND, OR, and NOT.Pattern Matching & Ranges: Finding explicit ranges using BETWEEN, testing sets via IN, and string matching with LIKE wildcards.Sorting Output: Utilizing ORDER BY for ascending (ASC) and descending (DESC) organization

Module 3

Aggregations & Data Summarization

Aggregate Functions: Calculating essential metrics via COUNT(), SUM(), AVG(), MIN(), and MAX().Data Segmentation: Grouping categorical subsets using the GROUP BY clause.Filtering Aggregates: Applying the HAVING clause to filter post-aggregated data frames (unlike WHERE).Handling Nulls: Basic data profiling and managing missing data spaces

Module 4

Relational Joins & Set Operations

Primary and Foreign Keys: Understanding the relational bridges that link database schemas together.Inner Joins: Merging rows from multiple tables matching on specified shared entities.Outer Joins: Retaining unmatched records across distinct data sets via LEFT JOIN, RIGHT JOIN, and FULL OUTER JOIN.Set Operations: Stacking query outputs on top of one another using UNION and UNION ALL

Module 5

Advanced Analytics & Feature Engineering

Subqueries: Embedding queries inside queries (Nested and Correlated Subqueries).Common Table Expressions (CTEs): Writing cleaner code using WITH statements for temporal isolation.Window Functions: Performing localized metrics (e.g., ROW_NUMBER(), RANK(), LEAD(), LAG()) without collapsing rows.Conditional Logic: Creating dummy features or flags using CASE WHEN ... THEN logic.Data Transformations: Basic manipulation techniques for strings, dates, and geographic timestamps

Module 6

Data Science Ecosystem Integration

Connecting SQL to Python/R: Executing active SQL chunks inside interactive environment engines like Jupyter Notebooks.Bridging Databases to Pandas: Querying databases to ingest tables directly into Pandas DataFrames.Performance Tuning: Understanding EXPLAIN syntax to discover query bottlenecks on massive tables.

© 2026 Proasia Upskilling Sdn Bhd | www.proasia.my