

**B8136: Intro to Databases for Business Analytics
Spring 2024 (A-Term)**

Mattan Griffel

Email: mattan.griffel@columbia.edu

Office Hours: Available upon request

NOTE: This is a hybrid online/offline course consisting of one and a half (1.5) hours of online video and one and a half (1.5) hours of in-person class time per week. Watching the online videos is mandatory and must be completed before the in-person class each week.

Course Description

We don't think about databases much, right? At least not when they're working right. But they're all around us. They're in every product we use. And when they don't work (think about the iCloud, LinkedIn, or Ashley Madison data breaches in which hundreds of millions of emails and passwords were exposed) the consequences can be extreme.

Every modern company stores their data in a database (it's like a really big version of Excel), and if you want to analyze the data, you may be expected to know how to access it yourself. In fact, at many companies are requiring even their business leaders to have an understanding of databases. At the very least, knowing how to set up and interact with databases will improve your ability to GSD (get stuff done), strengthen your understanding of how technology works, and make you less of a pain for developers to work with.

In this class, we'll explore basic SQL (the most common database language) for business analytics. At the end of the course, students should have a deeper understanding of how databases work, how they fit into the general technology stack, how to connect to databases, and know how to browse and exporting data from databases.

Required Prerequisites

This course assumes no previous knowledge of programming or code.

Required Course Material

- This course does not use a textbook.
- Students must have a laptop that they can bring to class – Mac or PC is fine, as long as your operating system is up to date (at least Windows 11 and Mac OS 14).
- Slides and files will be uploaded to Canvas after class.

Online Video

Each week, students will be expected to watch approximately one and a half hours of additional online video content before attending class. Material in the class will build on the content covered online, and students should be prepared to answer questions related to online material. Video content will be made available via Canvas.

Course Roadmap/Schedule

Session	Topic	Assignment
Class 1 (Recorded Video)	Querying Bootcamp: • Overview of the course • What is SQL? • Installing SQLite, text editor, and command line • Command line crash course • SQLite vs Postgres vs MySQL • SELECT	
Class 1 (In-Person) Thursday, Jan 25	Querying Bootcamp: • Running SQL from a file • Saving to CSV • SELECT • Math • WHERE • AND	Assignment 1 Due night before Class 2
Class 2 (Recorded Video)	Querying Bootcamp pt. 2: • Renaming columns • Concatenating data • String searches using LIKE • Matching multiple values using IN • Searching by dates & times • DISTINCT, ORDER BY, LIMIT • CASE (if...then statements) • Intro to aggregate functions	
Class 2 (In-Person) Thursday, Feb 1	Querying Bootcamp pt. 2: • Subqueries • JOIN • More complex joins	Assignment 2 Due night before Class 3
Class 3 (Recorded Video)	Querying Bootcamp pt. 2: • Aggregate functions • GROUP BY • HAVING • Extracting from datetime	
Class 3 (In-Person)	Querying Bootcamp pt. 2: • Extracting from datetime	Assignment 3 Due night before Class 4

Thursday, Feb 8	Dates in PostgreSQL	
Class 4 (Recorded Video)	Creating Databases: - Setting up a database instance on Amazon - Creating a new database - Creating tables - Loading CSV data - Deleting tables - Inserting data - Column Constraints - Deleting data	
Class 4 (In-Person) Thursday, Feb 15	Creating Databases: - Cleaning Data - Updating Tables - Creating multi-relational tables - One-to-Many - Many-to-Many	Assignment 4 Due night before Class 5
Class 5 (Recorded Video)	Web Apps/Security + Data Analysis: - Connecting a web app to a database - SQL Injection	
Class 5 (In-Person) Thursday, Feb 22	Web Apps/Security + Data Analysis: - Data Analysis & Tools - Yammer Case	
Class 6 (Recorded Video)	Data Analysis pt. 2 Yammer Case Continued	
Class 6 (In-Person) Thursday, Feb 29	Data Analysis pt. 2 Yammer Case Continued	Final Project due Wednesday, Mar 6 at Midnight

Grading

Final grades in the class will be calculated as follows:

Participation (30%)

- If you are on the waitlist or add the course at the last minute, you are expected to complete the pre-work.
- Students are expected to actively participate in class by posting solutions to challenges on a Slack group (an online messaging tool) for the course.

Assignments (40%)

- There will be four homework assignments
- Each assignment should be completed individually.
- Late assignments will be accepted with a 20% penalty any time before the final class. No late assignments will be accepted after the final class.

Final Project (30%)

- There will be a take-home final project.
- The final project must be completed individually.
- **Note:** Completion of the final project will be necessary to receive a passing grade in the course.

AI Policy

You may use AI tools, such as ChatGPT or Github Copilot, as part of your approach to generating solutions for your work in this course (just as any programmer would do), but any code submitted for in-class challenges and assignments must be your own.

This means that it is acceptable to use AI tools to help you understand and come up with a solution, but you must take the time to learn why any bit of code works and rewrite it on your own using style conventions we will learn in this class. Simply copying and pasting code generated by AI or found online is not acceptable (and it is often quite easy to tell).

Please contact me if you have any questions about this policy before submitting your work.