**B8136: Intro to Databases for Business Analytics**

**Spring 2019 (B Term)**

**Mattan Griffel**

Email: mattan.griffel@columbia.edu

Office Hours: Available upon request

Section 1: Thursdays 4:00PM – 5:30PM @ Uris 142 (except Thursday, April 23rd – see below.)

Section 2: Thursdays 6:00PM – 7:30PM @ Uris 326 (except Thursday, April 23rd – see below.)

(There will be no class on Thursday, April 23rd. Two makeup sessions will take place on Friday, April 17th from 4:00PM – 5:30PM in Uris 142 or 6:00PM – 7:30PM also in Uris 142. These sessions will also be recorded for those unable to attend.)

***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 7 and Mac OS 10.8).
* 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 Due** |
| **Class 1****(Online)** | **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, Mar 26  | **Querying Bootcamp**: * Running SQL from a file
* Saving to CSV
* SELECT
* Math
* WHERE
* AND
 | Pre-Work Due |
| **Class 2****(Online)** | **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, Apr 2 | **Querying Bootcamp pt. 2**: * Subqueries
* JOIN
* More complex joins
 | Assignment 1 Due |
| **Class 3****(Online)** | **Querying Bootcamp pt. 2**: * Aggregate functions
* GROUP BY
* HAVING
* Extracting from datetime
 |  |
| **Class 3****(In-Person)**Thursday, Apr 9 | **Querying Bootcamp pt. 2**: * Extracting from datetime
* Dates in PostgreSQL
 | Assignment 2 Due |
| **Class 4****(Online)** | **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, Apr 16 | **Creating Databases**: * Cleaning Data
* Updating Tables
* Creating multi-relational tables
	+ One-to-Many
	+ Many-to-Many
 | Assignment 3 Due |
| **Class 5****(Online)** | **Web Apps/Security + Data Analysis**: * Connecting a web app to a database
* SQL Injection
 |  |
| **Class 5****(In-Person)**Friday, Apr 17 | **Web Apps/Security + Data Analysis**: * Data Analysis & Tools
* Yammer Case
 |  |
| **Class 6****(Online)** | **Data Analysis pt. 2*** Yammer Case Continued
 |  |
| **Class 6****(In-Person)**Thursday, Apr 30 | **Data Analysis pt. 2*** Yammer Case Continued
 | Assignment 4 Due &Yammer Case Analysis Due |
| **Due** Sunday, May 10Midnight | **Final Project** | SQL Cheat Sheet Due |

**Grading**

Final grades in the class will be calculated as follows:

*Participation (30%)*

* If you are not present for Day 1, you will not be allowed to add the course. If you are enrolled and you do not attend Day 1 or complete the pre-work, you will be dropped from the course.
* If you 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 should be completed individually.