



Data Camp Live Training:

Getting Started in SQL





KELSEY MCNEILLIE

INSTRUCTOR RECRUITER, DATA CAMP



Live Training Session Agenda

- Introduction
- Context
- Coding Activity
- Recap/Closing Notes
- Next Steps/Take home assignment

SQL stands for 'Structured Query Language'

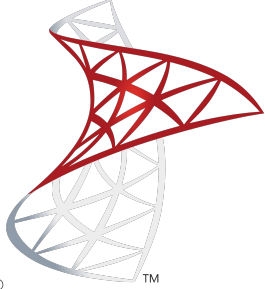


- Originally developed in the 1970s
 - Designed to isolate and return data from an IBM relational database management system (RDMS)
 - Has undergone many evolutions and updates
- Strengths
 - SQL can handle much, much larger databases than Excel, which can only handle 1,048,576 rows
 - SQL can also aggregate and clean data much faster
- Weaknesses
 - SQL does not produce data visualizations - to do this, you will need to export your cleaned data to another program

Why Learn SQL?



- SQL remains a vitally important data coding language
- Used by data analysts, data scientists, and database administrators
- Can be used to query, clean, and organize data quickly and efficiently



Microsoft®
SQL Server®



PostgreSQL

ORACLE®
PL/SQL



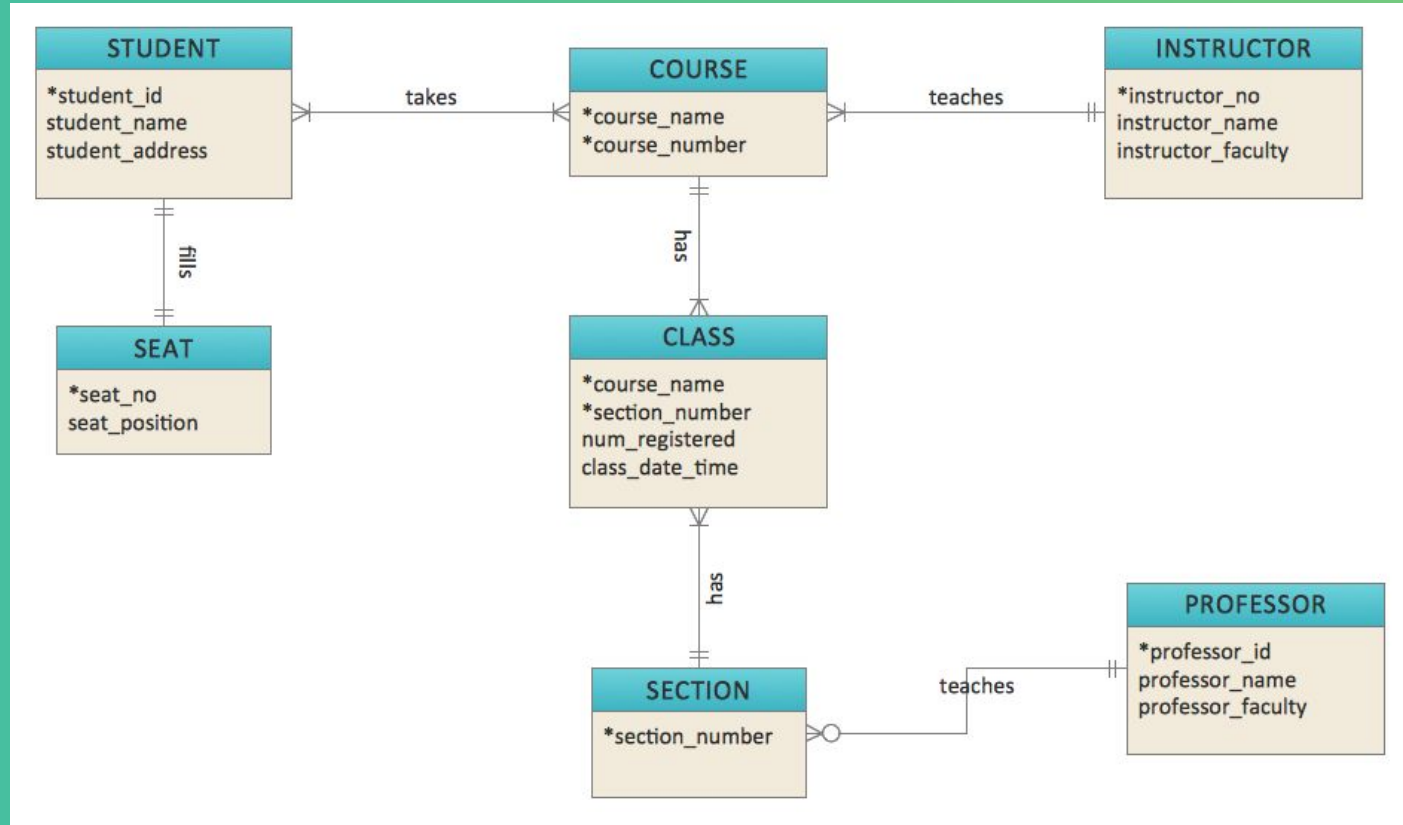
MySQL™

There are several different “dialects” of SQL

We’re going to be using Postgres for our session today - it’s widely used, an object oriented relational database management system (DBMS), it’s powerful, and it’s FREE!

Let's Talk about Tables

This is an entity relationship diagram (ERD) - it shows how tables connect to each other using a 'primary key'



SQL and Table Anatomy

SELECT (returns columns)

FROM
(entire table)

id	last_name	marital_status	gender	annual_income	industry	zip_code
A02601	Barrett	Married	M	\$ 115,613.00	Consumer Discretionary	80202
A00128	Boyd	Married	M	\$ 102,874.00	Consumer Staples	80202
A00102	Hudson	Married	M	\$ 87,772.00	Utilities	80202
A02479	Green	Single	M	\$ 81,574.00	Consumer Staples	80202
A02434	Parker	Married	F	\$ 90,410.00	Real Estate	80202
A01133	Woods	Married	F	\$ 48,462.00	Energy	80202
A02148	Page	Married	M	\$ 90,171.00	Consumer Staples	80202
A04898	Gonzalez	Married	M	\$ 111,065.00	Health Care	80202
A04686	Jacobs	Married	M	\$ 114,428.00	Consumer Discretionary	80202
A06081	Tyler	Single	M	\$ 69,769.00	Telecommunication Services	80202
A04755	Adkins	Married	M	\$ 103,336.00	Telecommunication Services	80202
A00936	Brock	Single	M	\$ 115,486.00	Health Care	80202
A04326	Rowe	Married	M	\$ 103,926.00	Consumer Staples	80202
A04236	Cohen	Married	M	\$ 104,903.00	Materials	80202
A02151	Maxwell	Single	M	\$ 64,316.00	Health Care	80202
A04291	Allen	Married	M	\$ 79,918.00	Materials	80202
A00773	Gregory	Married	M	\$ 101,686.00	Financials	80202

WHERE (filters
on rows)



SELECT last_name, annual_income
FROM membership
WHERE marital_status = 'Married'

```
SELECT last_name, annual_income FROM membership WHERE marital_status = 'Married'
```

last_name	annual_income
23121	77
23730	18
32207	18
23122	18
23114	24

SQL Order of Execution



What Query looks like:

1. SELECT
2. FROM
3. WHERE
4. GROUP BY
5. HAVING
6. ORDER BY
7. LIMIT

How it's Executed:

1. FROM
2. WHERE
3. GROUP BY
4. HAVING
5. SELECT
6. ORDER BY
7. LIMIT


Live Coding Activity

jupyter applayout_example Last Checkpoint: 20 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
11         icon='backward',
12         layout=Layout(width='80%',
13                        height='30%'))
14 next_button = Button(description="Next",
15                       icon='forward',
16                       layout=Layout(width='80%',
17                                    height='30%'))
18 footer = HTML("Filename: {}".format(image_file))
19
20 AppLayout(header=header,
21           left_sidebar=prev_button,
22           center=image,
23           right_sidebar=next_button,
24           footer=footer,
25           grid_gap='20px',
26           justify_items='center',
27           align_items='center')
```

Simple Image Viewer



Filename: images/cat.jpg



!! Requires a gmail account to edit !!

The background features several teal-colored geometric shapes. In the top-left corner, there is a large, rounded, organic shape. In the top-right, there is a sharp-pointed triangle with a white diagonal line cutting through it. In the bottom-left, there is another sharp-pointed triangle with a white diagonal line. In the bottom-right, there is a large, rounded, organic shape. The text 'Session Notebook' is centered in the middle of the page.

Session Notebook

Recap and Closing Notes

What Did We Learn Today?



Basic Queries

- SELECT, FROM, ORDER BY, LIMIT

Filtering Data

- WHERE, LIKE, ILIKE, IS, BETWEEN, IIN
- >, <, =, >=, <=

Aggregation

- SUM, COUNT
- AVERAGE, MAX, MIN
- GROUP BY, HAVING

Other Useful Functions

- ROUND, DISTINCT
- Aliasing



Remember - Stay Organized!

Things to Remember When Writing Your Queries



Use inline commenting to “brainstorm”

Frame your query as a business question first

Ask yourself, “What do I need for this query?”

You will always need a SELECT and FROM statement - but what else?

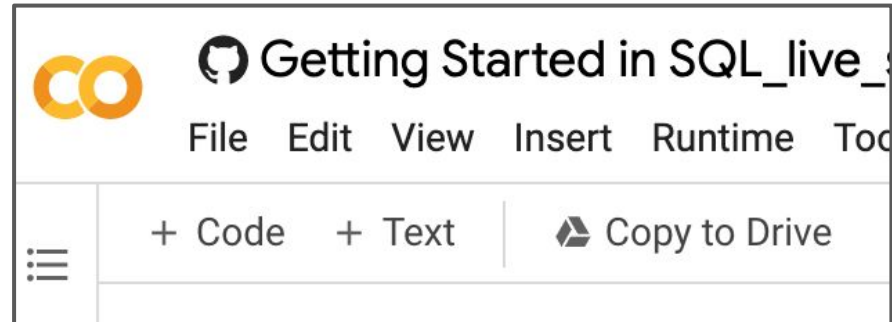
Test out each step along the way

Build your query one line at a time and run it frequently to check that it works

Further Practice



- **Today's Notebook is yours to keep**
 - Save a copy to your Google Drive
- **Remember:**
 - You will need to reload the Setting Up Postgres cells every time you reopen the notebook
 - If you see the error message “sql magic not found”, that means that Postgres was not loaded into the notebook
- **To add a new text cell, click “code”**
 - Start the cell with the code `%%sql` before starting a SQL query



Resources



Recap of Statements Used

Statement	Description
SELECT	Returns indicated columns
FROM	Indicates which table to query data from
WHERE	Filters data on rows
LIKE	Used in the WHERE statement to filter data
GROUP BY	Aggregates data as indicated in SELECT statement
HAVING	Filters on aggregated data
ORDER BY	Orders the rows returned by column indicated, can be ASC or DESC
LIMIT	Limits the number of rows returned

Recap of Filtering Functions Used

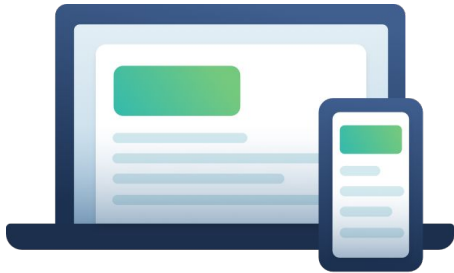
Function	Description
>, <, =	Greater than, less than, equal to
BETWEEN	Indicates a range of values, includes both values provided BETWEEN 'X' and 'Y'
IN	Filters for set of exact, known values. IN('term1', term2', 'term3')
LIKE	Can be used to search boolean terms, is case sensitive. LIKE 'X%' (where % stands for the unknown values)
ILIKE	Similar to LIKE, but is case indifferent
NOT	Used to negate the following filter function, as in, return values NOT LIKE 'x'

Recap of Aggregation /Other Functions Used

Function	Description
SUM	Adds values in column
COUNT	Returns count of values in column
AVG	Returns average of numerical values in column
MAX	Returns highest numerical value in column
MIN	Returns lowest numerical value in column

Function	Description
DISTINCT	Used in the SELECT statement to indicate unique values should be returned
ROUND	Used in the SELECT statement to round a number to indicated number of decimals

Additional Resources



Here are some other resources to help you get started on your journey to becoming a SQL Superstar!

- [DataCamp SQL Fundamentals Skill Track](#)
- [DataCamp Resources](#)
- [Upcoming Webinars](#)
- [SQL Style Guide](#)
- [StackOverflow](#)

Coming Soon!



Don't miss these upcoming webinars and live training sessions!

- Data Analysis in Excel (5/28)
- String Manipulation in SQL (6/4)
- Data Analysis in SQL (6/11)

NEW COURSE

Data Analysis in Excel

Learn how to use time-saving keyboard shortcuts, convert data types, and build impressive logic functions and conditional aggregations.

DataCamp

Exercise

Changing to text

By now, you know that there are four data types. A type something in a cell and hit enter, Excel determines which of these types you intend. If you might need to convert the data to the appropriate one.

Let's take a look at our data. Value is not accidentally used for calculation, and is readable as non-numeric characters long.

Instructions:

- Go to cell B2.
- Use `TEXT()` to convert

How is the id data aligned after the conversion?

Instructions 50XP

Possible Answers

A	B
1	id id_text
2	12846160 Hex Tile content Ad
3	13547406 Epic Stuff Library vol
4	20761861 Our House: The Lore
5	30059044 Zombicide: 2nd Editi
6	49752892 Longest Number Eve
7	52444976 =text DEAD+GONE: A Com
8	66675 the oracle of heaven
9	877 Dwarf board game
10	914 Traveller: Fifth Edition
11	024 TEXT Converts a value
12	159 TEXTJOIN th's Wake - Zi
13	006 BAHTTEXT THE HILLS 2ND
14	545 FORMULATEXT e: The Busines
15	441 ISNONTEXT /PATH TAROT:
16	599 ITEXT ts et Zombie
17	868 g Cards: The Je
18	899 Azote playing cards
19	132 Monster Menagerie:
20	8760 Hit the Streets: Defe
21	726009 Dice Palace: Display
22	265338582 peculiarity oracle, sa
23	272156229 DoubleSix Dice: Gen
24	279753888 The World of the Los
25	283822315 Super Dude Jump
26	302537271 Sloosh Cards
27	310730322 Dice Friend - Innovat
28	310781283 Dungeon Craft: Build
29	228751672 THE MANA BOOL CO

Sheet1

NEW COURSE

Introduction to Power BI

Unlock more value from your Microsoft plan. This interactive course empowers everyone with a 360-degree overview of how to analyze data and build impactful reports.

The screenshot displays the Power BI Desktop interface. At the top, the DataCamp logo and 'Course Outline' are visible. The ribbon includes 'File', 'Home', 'Insert', 'Modeling', 'View', and 'Help'. The 'Home' ribbon is active, showing options like 'Paste', 'Copy', 'Format painter', 'Get data', 'Excel', 'Power BI datasets', 'SQL Server', 'Enter data', 'Recent sources', 'Transform data', 'Refresh', 'New visual', 'Text box', 'More visuals', 'New measure', and 'Quick measure'. A notification states: 'Auto recovery contains some recovered files that haven't been opened.' The main workspace shows a report with a table and a chart. The table has columns: 'SalesOrderName', 'ProductCategoryName', 'AccountName', 'Amount', and 'DateKey'. The chart is a line graph titled 'Amount by Short Month' with a tooltip for 'Dec' showing 'Amount 5,020,903,994.80'. The status bar at the bottom indicates 'PAGE 1 OF 1'.

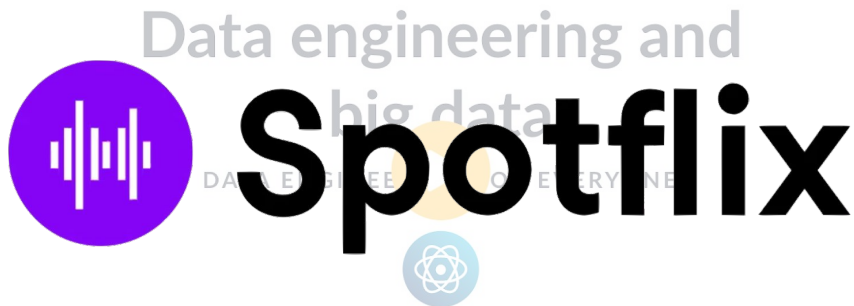
SalesOrderName	ProductCategoryName	AccountName	Amount	DateKey
Actual	Audio	Advertisement Expense	67,199.61	1/1/2017 12,000 Atd
Actual	Audio	Advertisement Expense	74,286.96	2/1/2017 12,000 Atd
Actual	Audio	Advertisement Expense	73,821.81	2/1/2017 12,000 Atd
Actual	Audio	Advertisement Expense	100,169.93	4/1/2017 12,000 Atd
Actual	Audio	Advertisement Expense	104,642.95	6/1/2017 12,000 Atd
Actual	Audio	Advertisement Expense	101,821.75	6/1/2017 12,000 Atd
Actual	Audio	Advertisement Expense	95,518.40	7/1/2017 12,000 Atd
Actual	Audio	Advertisement Expense	93,230.26	8/1/2017 12,000 Atd
Actual	Audio	Advertisement Expense	94,226.24	9/1/2017 12,000 Atd
Actual	Audio	Advertisement Expense	104,547.87	10/1/2017 12,000 Atd
Actual	Audio	Advertisement Expense	112,592.76	11/1/2017 12,000 Atd
Total			33,642,226,537.68	

☰ Course Outline



Data engineering and big data

50 X



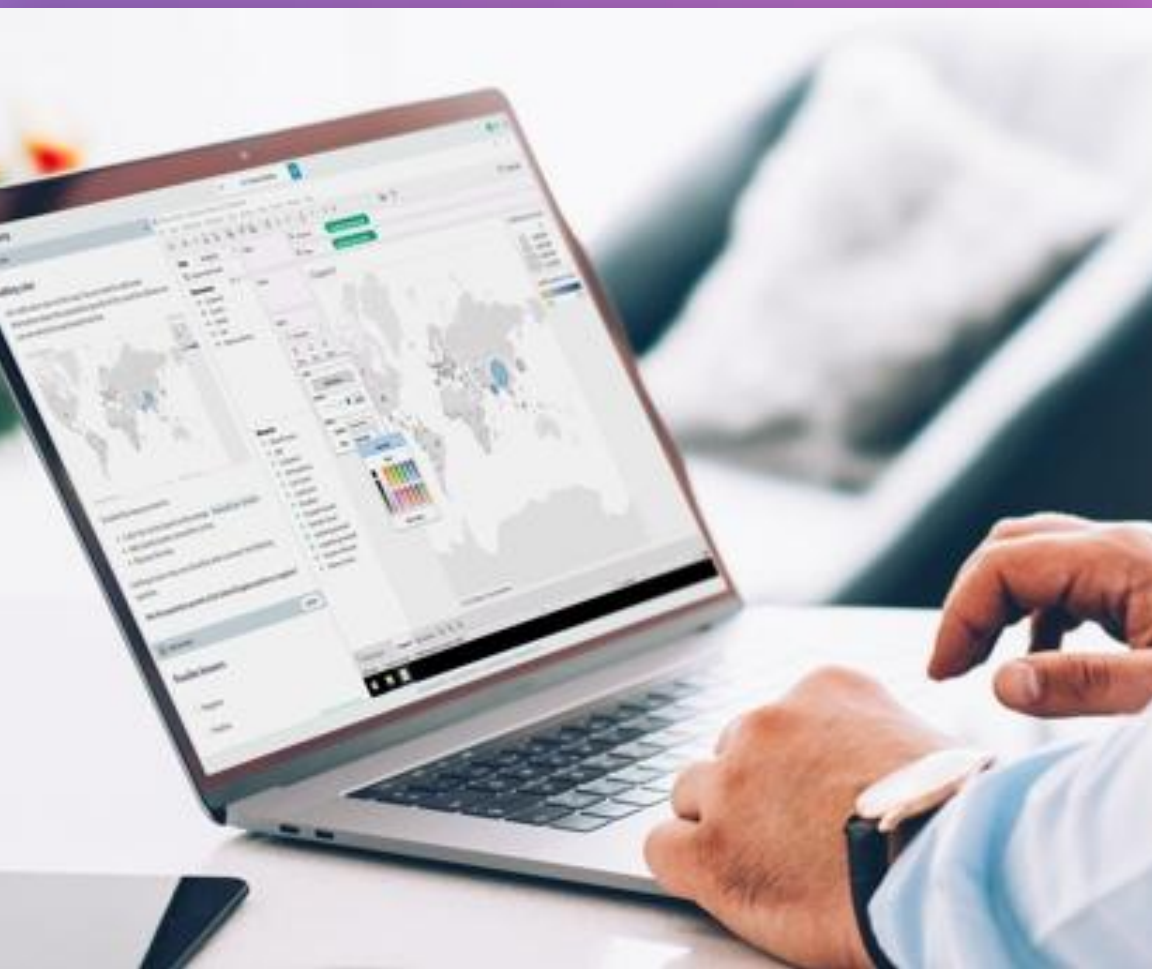
DataCamp

Got It

NEW COURSE

Data Engineering for Everyone

Learn about data engineering and why demand for them is at an all time high. Grow your knowledge in this area or take your first step towards becoming a data engineer.



NEW COURSE

Introduction to Tableau

Learn how to navigate, analyze, and build awesome dashboards that bring your data to life—all within your browser.

Thank you

Kelsey McNeillie

DataCamp Instructor Recruiter

kelsey@datacamp.com

<https://www.linkedin.com/in/kelsey-mcneillie/>

