



Data Camp Live Training: Data Analysis in SQL



Michel Semaan

Data Scientist



Introduction

- SQL is the most widely-used database language
- Usually, data analysis is done in languages other than SQL
 - Data is exported out of SQL
 - Languages like R or Python are used to analyze the data
- But performing data analysis in SQL has several advantages



Power BI

Dashboards and reports

Many business intelligence (BI) tools connect directly to the database, and generate dashboards, reports, and self-service analytics tools through SQL queries.

Often, the only way to generate these dashboards and reports is to write SQL.

Data analysis in SQL can also be used with...



Triggers

Triggers in SQL are functions automatically that fire off when a certain condition in the database are fulfilled (e.g., if a user churns, send an email).

Access restrictions

Most databases have security rules that make downloading datasets slow, difficult, or borderline impossible.

Database optimizations

It's often much quicker to do certain operations in SQL, especially when indices are created, than in R or Python.

user_sessions:

- **session_date:** *When the user logged into the website*
- **user_id:** *Unique identifier for a user*
- **time_spent_in_mins:** *How much time the user spent on the website in that session*

user_data:

- **user_id:** *Unique identifier for a user*
- **country:** *The user's country*
- **age:** *The user's age*

Dataset Overview

We're going to study the user activity data of a social media platform, Facespace.

Though the data is simulated, it very closely follows real world data in both content and structure.

jupyter applayout_example Last Checkpoint: 20 hours ago (autosaved)


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

◀◀ Prev



▶▶ Next

Filename: images/cat.jpg

Coding Environment



!! Requires a Gmail account to edit !!

Session Agenda

- *Introduction*
- Data overview + Active users (DAU & MAU)
- Q&A
- Registrations + growth rate
- Q&A
- Retention rate
- Q&A
- Recap + Next Steps

The background features several abstract teal-colored shapes. In the top-left corner, there is a large, rounded, organic shape. In the top-right, there is a sharp triangle with a thin line extending from its base. In the bottom-left, there is another sharp triangle with a thin line extending from its base. In the bottom-right, there is a large, rounded, organic shape. The text "Session Notebook" is centered in the middle of the page, underlined.

Session Notebook

Recap

What Did We Learn Today?

Core business KPIs

Active users (DAU, MAU), registrations, growth rates, and retention rates are used everywhere to judge a platform's performance

Date and window functions

Date manipulation and window operations have a variety of common uses

CTEs and self-joins

With CTEs and self-joins, you can manipulate and link query results in more advanced ways

Relevant DataCamp Courses

[Analyzing Business Data in SQL](#) - Write SQL queries to calculate various common KPIs and build reports

Michel Semaan

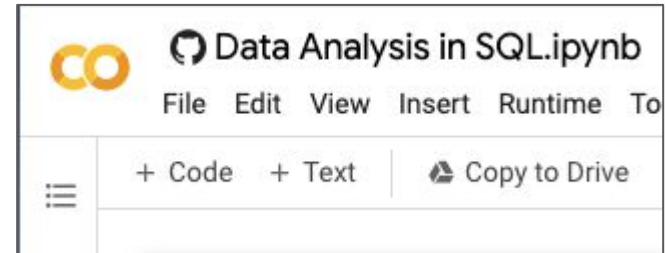
[PostgreSQL Summary Stats & Window Functions](#) - Dive deeper into window functions and their various applications

Fernando Gonzalez Prada & Michel Semaan

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



Coming Soon!



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

- [Cleaning Data in R](#) (June 16)
- [Cleaning Data with PySpark](#) (June 17)

Thank you

Michel Semaan

Data Scientist

michlsemn@gmail.com


[linkedin.com/in/michlsemn](https://www.linkedin.com/in/michlsemn)



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. It might need to convert the data to the appropriate one.

Let's take a look at our data. VLOOKUP is not accidentally used for card names that are readable as non-numeric characters long.

Instructions:


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

How is the id data aligned after the conversion?


Instructions

50XP

Possible Answers



DATA ANALYSIS IN EXCEL



Course Outline

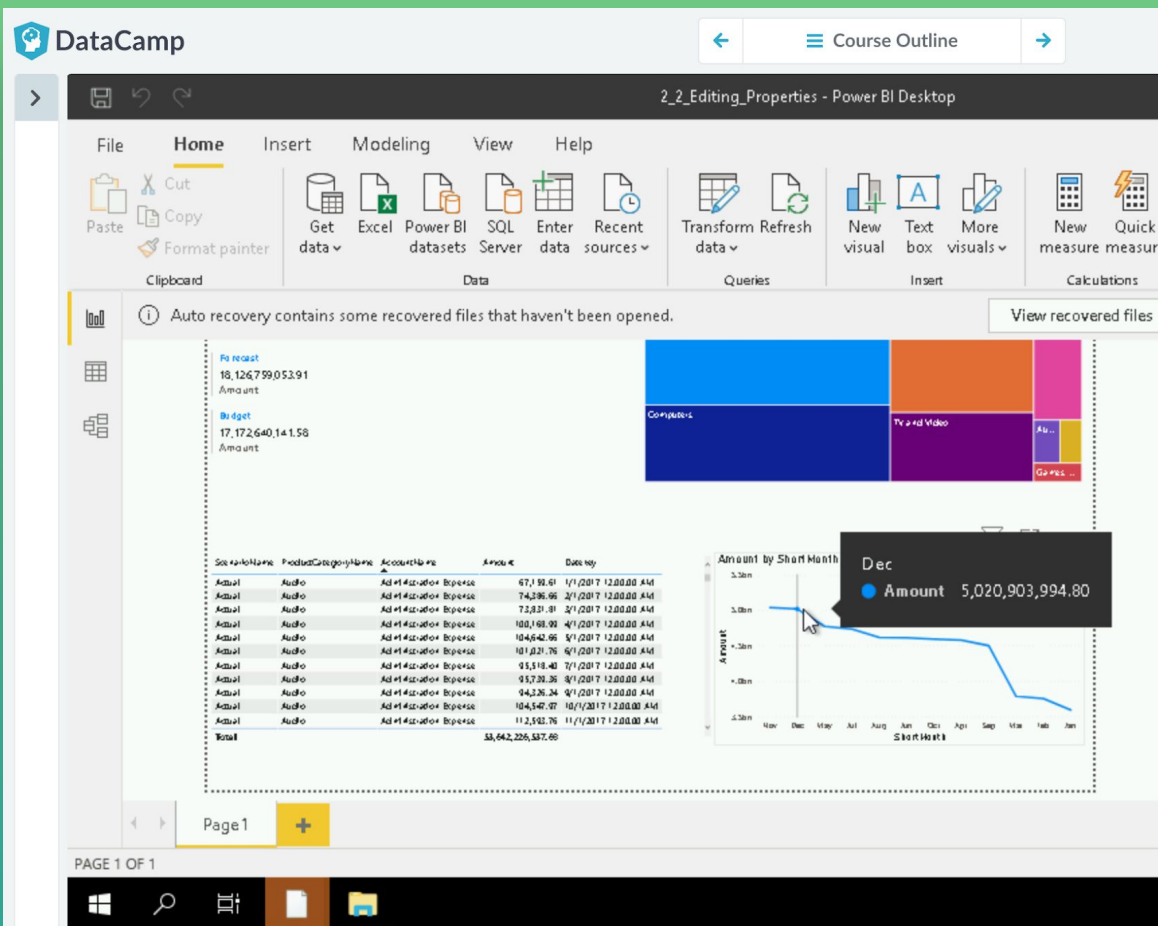
A	B
1	id id_text
2	12846160 Hex Tile content Ad
3	13547406 Epic Stuff Library vol
4	20761861 Our House: The Lore
5	30059044 Zombicide: 2nd Editi
	49752892 Longest Number Eve
	52444976 =text DEAD+GONE: A Com
	66675 the oracle of heaven
	877 Dwarf board game
	914 Traveller: Fifth Edition
	024 Converts a value
	159 TEXT
	006 TEXTJOIN th's Wake - Zir
	545 BAHTTEXT THE HILLS 2ND
	441 FORMULATEXT e: The Busines
	599 ISNONTEXT /PATH TAROT:
	868 ITEXT ts et Zombie
	899 g Cards: The Je
	132 Azote playing cards
	8760 Monster Menagerie
	7726009 Hit the Streets: Defe
	265338582 Dice Palace: Display
	272156229 peculiarity oracle, sa
24	279753888 DoubleSix Dice: Gen
25	283822315 The World of the Los
26	302537271 Super Dude Jump
27	310730322 Sloosh Cards
28	310781283 Dice Friend - Innovat
29	328751672 Dungeon Craft: Build
	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.

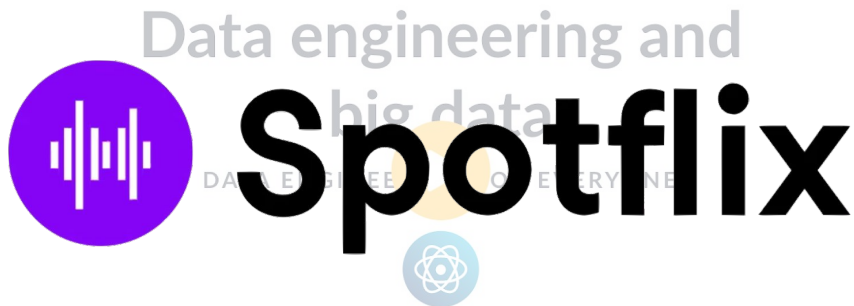


ta engineering and big data

Course Outline



50 X



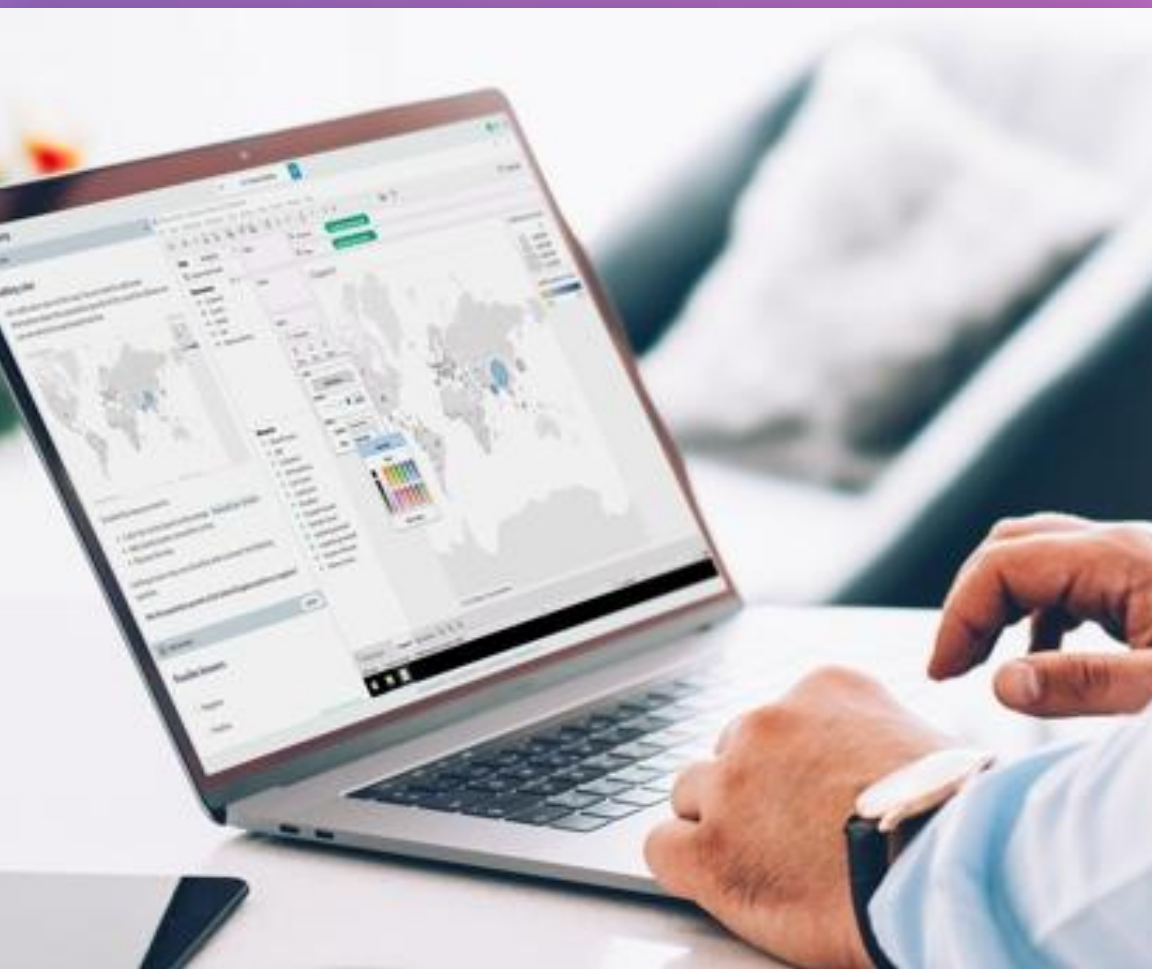
DataCamp

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.