

Using SQL to clean data

- Video:

Using SQL to clean data

45 sec
- Video:

Sally: For the love of SQL

3 min
- Video:

Understanding SQL capabilities

3 min
- Reading:

Using SQL as a junior data analyst

10 min
- Video:

Spreadsheets versus SQL

4 min
- Reading:

SQL dialects and their uses

10 min
- Practice Quiz:

Hands-On Activity: Processing time with SQL

2 questions
- Practice Quiz:

Test your knowledge on SQL

3 questions

Learn basic SQL queries

Transforming data

Weekly challenge 3

Using SQL as a junior data analyst

In this reading, you will learn more about how to decide when to use SQL, or Structured Query Language. As a data analyst, you will be tasked with handling a lot of data, and SQL is one of the tools that can help make your work a lot easier. SQL is the primary way data analysts extract data from databases. As a data analyst, you will work with databases all the time, which is why SQL is such a key skill. Let's follow along as a junior data analyst uses SQL to solve a business task.

The business task and context

The junior data analyst in this example works for a social media company. A new business model was implemented on February 15, 2020 and the company wants to understand how their user-growth compares to the previous year. Specifically, the data analyst was asked to find out how many users have joined since February 15, 2020.



Spreadsheets functions and formulas or SQL queries?

Before they can address this question, this data analyst needs to choose what tool to use. First, they have to think about where the data lives. If it is stored in a database, then SQL is the best tool for the job. But if it is stored in a spreadsheet, then they will have to perform their analysis in that spreadsheet. In that scenario, they could create a pivot table of the data and then apply specific formulas and filters to their data until they were given the number of users that joined after February 15th. It isn't a really complicated process, but it would involve a lot of steps.

In this case, the data is stored in a database, so they will have to work with SQL. And this data analyst knows they could get the same results with a single SQL query:

```
SELECT
    COUNT(DISTINCT user_id) AS count_of_unique_users
FROM
    table
WHERE
    join_date >= '2020-02-15'
```

Spreadsheets and SQL both have their advantages and disadvantages:

Features of Spreadsheets	Features of SQL Databases
Smaller data sets	Larger datasets
Enter data manually	Access tables across a database
Create graphs and visualizations in the same program	Prepare data for further analysis in another software
Built-in spell check and other useful functions	Fast and powerful functionality
Best when working solo on a project	Great for collaborative work and tracking queries run by all users

When it comes down to it, where the data lives will decide which tool you use. If you are working with data that is already in a spreadsheet, that is most likely where you will perform your analysis. And if you are working with data stored in a database, SQL will be the best tool for you to use for your analysis. You will learn more about SQL coming up, so that you will be ready to tackle any business problem with the best tool possible.

Mark as completed