Relational databases

This lesson moves away from Excel and explores relational databases and why they're used.

So far, you've worked with data in Excel, which is a popular way to work with reasonably sized datasets. But Excel lacks the ability to work with datasets that have well over a million rows. So, for big data, relational databases are used. Relational databases hold massive amounts of data and require the use of a programming language to pull desired data.

Tabular database

A type of database that stores data in a table structure, where data is viewed by rows and columns.

Relational database

A type of database in which the relationships between the data points matter, and that requires a programming language to pull desired data.

Relational database management system

Also known as RDBMS, a class of programs that can be used to create, update, and administer a relational database.

Relational databases versus tabular databases

As you now know, Excel is a tabular database. It holds data in a table structure that the user can see and is organized by rows and columns. A tabular database lets a user scroll through and view the full data. But a major limitation of tabular databases is that they can't hold much more than a million rows of data. In this data-driven world, datasets can be millions of rows.

A relational database is a database in which the relationships between data matter. Imagine that you have a database with two tables: table A, which has customer demographic data, and table B, which has customer order data. Each order in table B points to a customer whose information is stored in table A. Because these tables are related to each other, the database is relational.

A relational database can house millions of rows of data without slowing down, so it is the preferred database for working with big data. In fact, relational databases can hold hundreds of data tables, each with millions of rows! But because these databases hold so much data, the user can't simply view or scroll through the data, as they can with Excel. So, programming languages are used to pull or see desired data from a relational database. A programming language is a code that instructs the database management system on what data to pull or how to manipulate the data.

Relational database types

A relational database is a general term for these large databases that can store millions of rows of data. There are many relational database management systems (RDBMS) that are used to store data. Some popular ones are PostgreSQL, MySQL, SQL Server, and Oracle. In this course, you'll use PostgreSQL as the primary relational database, paired with pgAdmin as the platform to write programming language code.