Databases in data analytics

In this reading you will learn about what relational databases are and how databases can be organized. Databases let analysts manipulate, store, and process data.  This helps them search through data a lot more efficiently to get the best insights.

<https://d3c33hcgiwev3.cloudfront.net/imageAssetProxy.v1/PlkXAtKhQM6ZFwLSoSDOSA_994445d36d8b47c189cd78d88407778e_Screen-Shot-2021-03-04-at-10.26.19-PM.png?expiry=1625356800000&hmac=Qb8sTmvaQugiZdAPAn8yGf7_-NRFasTHjO00w0ikiRE>

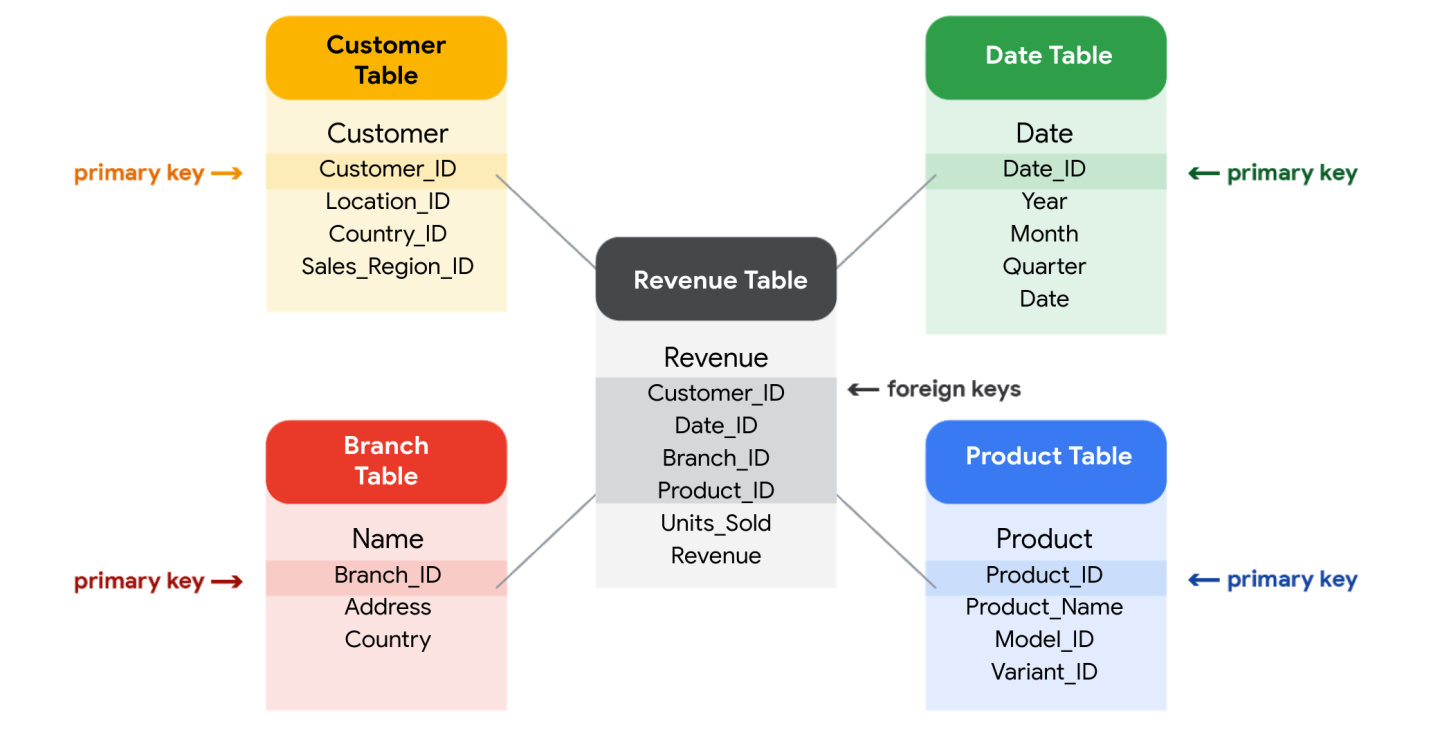
Relational databases

A **relational database** is a database that contains a series of tables that can be connected to show relationships. Basically, they allow data analysts to organize and link data based on what the data has in common.

In a non-relational table, you will find all of the possible variables you might be interested in analyzing all grouped together. This can make it really hard to sort through. This is one reason why relational databases are so common in data analysis: they simplify a lot of analysis processes and make data easier to find and use across an entire database.

**The key to relational databases**

Tables in a relational database are connected by the fields they have in common. You might remember learning about primary and foreign keys before. As a quick refresher, a **primary key** is an identifier that references a column in which each value is unique. Primary keys are unique, and the value assigned to that key for each row can only exist once within a given table. For example, if customer\_id is the primary key in the customer\_address table, no two customers will ever have the same customer\_id.  By contrast, a **foreign key** is a field within a table that is a primary key in another table. A table can only have one primary key, but it can have multiple foreign keys. These keys are what create the relationships between tables in a relational database, which helps organize and connect data across multiple tables in the database.



SQL? You’re speaking my language

Databases use a special language to communicate called a query language. **Structured Query Language** (SQL) is a type of query language that lets data analysts communicate with a database. So, a data analyst will use SQL to create a query to view the specific data that they want from within the larger set. In a relational database, data analysts can write queries to get data from the related tables. SQL is a powerful tool for working with databases — which is why you are going to learn more about it coming up!