**Memproses Data dari Kotor ke Bersih**

**Menggunakan SQL untuk membersihkan data**

SQL adalah structured query language untuk berinteraksi dengan database. Analis data menggunakan SQL untuk menangani dataset yang sangat besar karena SQL dapat menangani data dalam jumlah yang besar.

**Menggunakan SQL sebagai analis data junior**

**Tugas dan konteks bisnis**

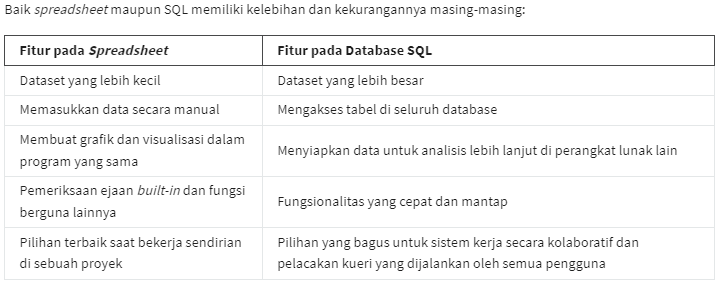
Pada contoh berikut ini, seorang analis data junior bekerja di perusahaan media sosial. Sebuah model bisnis baru telah diterapkan pada tanggal 15 Februari 2020 dan perusahaan ingin memahami bagaimana pertumbuhan pengguna mereka dibandingkan dengan tahun sebelumnya. Secara khusus, analis data diminta untuk mencari tahu berapa banyak pengguna yang telah bergabung sejak 15 Februari 2020.

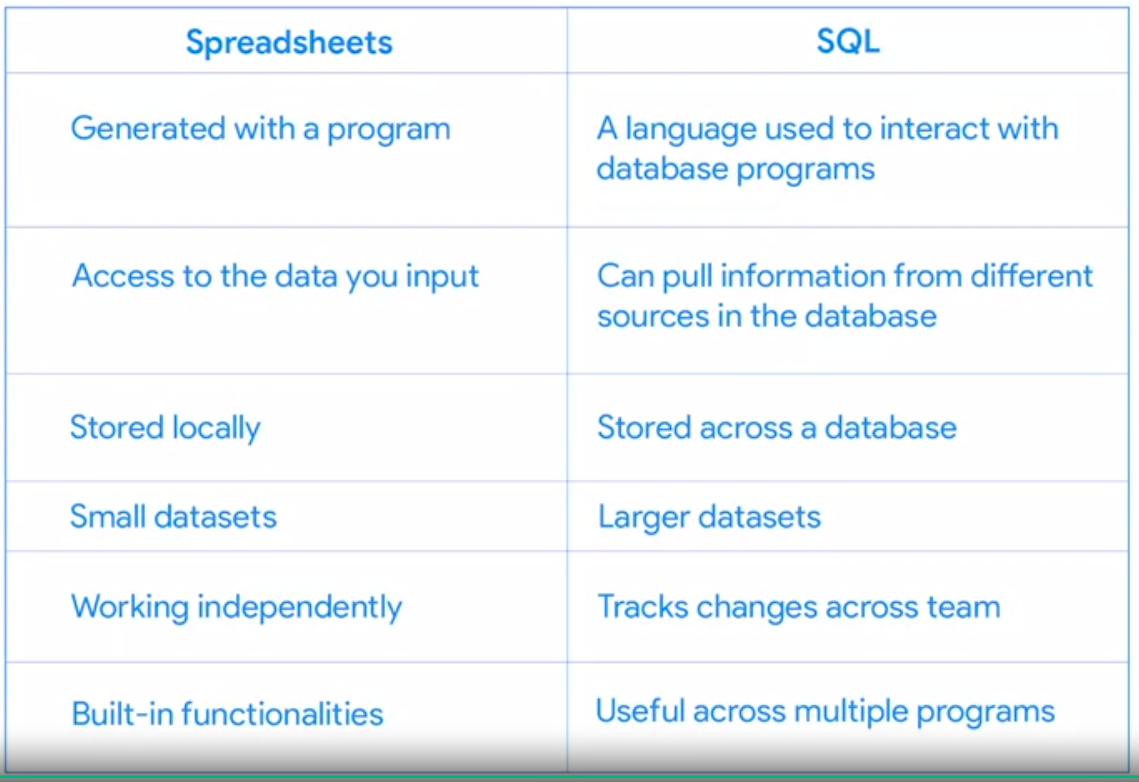
**Alat apa yang akan digunakan, fungsi dan rumus spreadsheet atau kueri SQL?**

Sebelum dapat menjawab pertanyaan ini, analis data perlu memilih alat apa yang akan digunakan. Pertama, analis data harus memikirkan letak data tersebut. Jika data tersebut disimpan di dalam database, maka SQL adalah alat terbaik untuk menyelesaikan pekerjaan itu. Tetapi jika disimpan di dalam spreadsheet, maka dia harus melakukan analisis di dalam spreadsheet tersebut. Dalam skenario tersebut, mereka dapat membuat tabel pivot dan kemudian menerapkan rumus dan filter tertentu ke data mereka sampai mendapatkan jumlah pengguna yang bergabung setelah 15 Februari. Ini bukan proses yang sangat rumit, tetapi memiliki langkah-langkah yang banyak.

Jika data disimpan di dalam database, maka analis data harus bekerja dengan SQL. Dan untuk mendapatkan hasil yang sama, analis data cukup menjalankan satu kueri SQL:







**NOTE!!!**

* Blog LearnSQL, [What Is a SQL Dialect, and Which One Should You Learn?](https://learnsql.com/blog/what-sql-dialect-to-learn/)
* Artikel Software Testing Help, [Differences Between SQL Vs MySQL vs SQL Server](https://www.softwaretestinghelp.com/sql-vs-mysql-vs-sql-server/)
* Blog Datacamp, [SQL Server, PostgreSQL, MySQL... what’s the difference Di mana saya memulai?](https://www.datacamp.com/blog/sql-server-postgresql-mysql-whats-the-difference-where-do-i-start) Perhatikan bahwa ada kesalahan dalam artikel blog ini. Tabel perbandingan salah menyatakan bahwa SQlite menggunakan subkueri alih-alih window functions. Untuk klarifikasi yang tepat, lihat dokumentasi tentang [SQLite Window Functions](https://sqlite.org/windowfunctions.html).
* SQL Tutorial, [What is SQL](https://www.sqltutorial.org/what-is-sql/)

