Week 17

Using SQL to clean data.

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.

Learn basic SQL queries

SELECT

INSERT INTO

LEN(name)

SUBSTR(name, start, end)

TRIM()

Typecasting: changing data typr from one to the other.

CAST(data AS type)

CONCAT(data\_a,data\_b) AS data\_c

COALESCE(data\_a, data\_b) AS data\_c (when data\_a is null, data\_c would be data\_b)