Week 4

**Mastering spreadsheet basics**

key points about spreadsheets:

**Spreadsheets Basics:**

* Spreadsheets are essential tools for data analysts.
* They help organize and manipulate data effectively, making data analysis easier.
* Spreadsheets consist of cells, rows, and columns:
  + Cells: Basic units in a spreadsheet where you input data.
  + Rows: Horizontal organization of data, numbered.
  + Columns: Vertical organization of data, lettered.

**Data Entry and Labels:**

* You can input data into cells by selecting a cell and typing the desired information.
* Labels (column headers) are important for organizing and referencing data. They make your data clear and easy to find.
* You can format cells, rows, and columns by adjusting their size, using text wrapping, or applying bold formatting.

**Sorting Data:**

* You can sort data in a spreadsheet to organize it in various ways.
* To sort data, select the columns you want to sort and use the "Sort Range" option in the "Data" menu.
* Pay attention to the header row to ensure attribute labels remain in place.

**Formulas and Functions:**

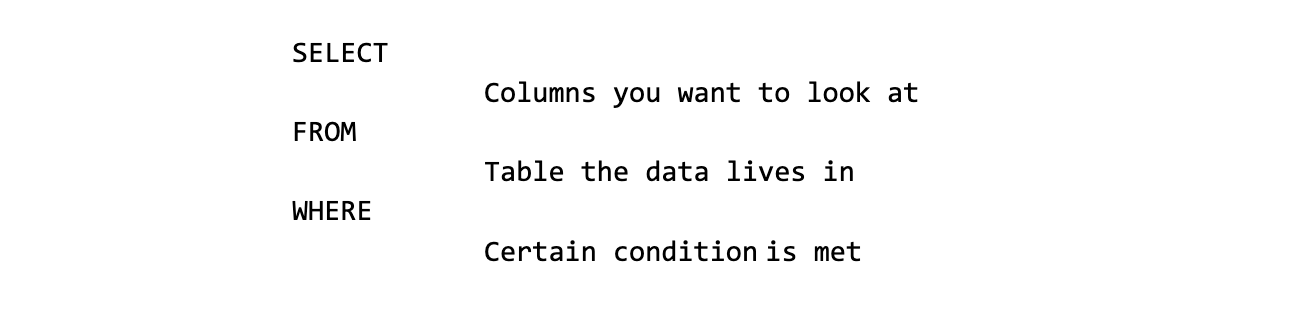
* Formulas are sets of instructions that perform specific calculations using data in a spreadsheet.
* Formulas start with an equal sign (=) and use cell references for calculations.
* Functions are predefined commands that perform specific tasks with data.
* Formulas and functions are powerful tools for manipulating and analyzing data.

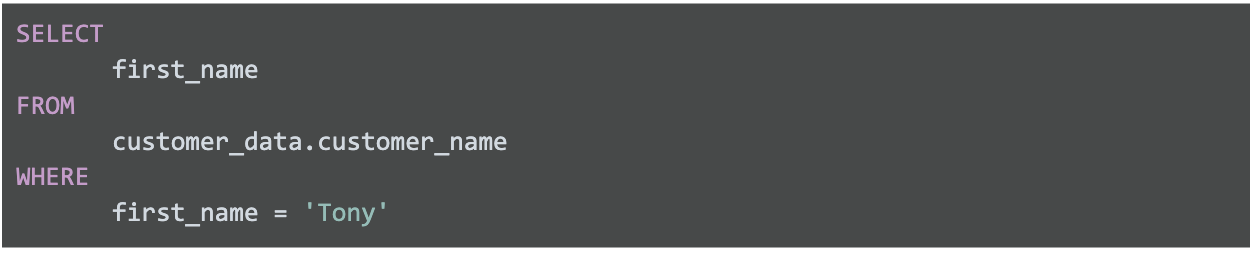
**Saving Spreadsheets:**

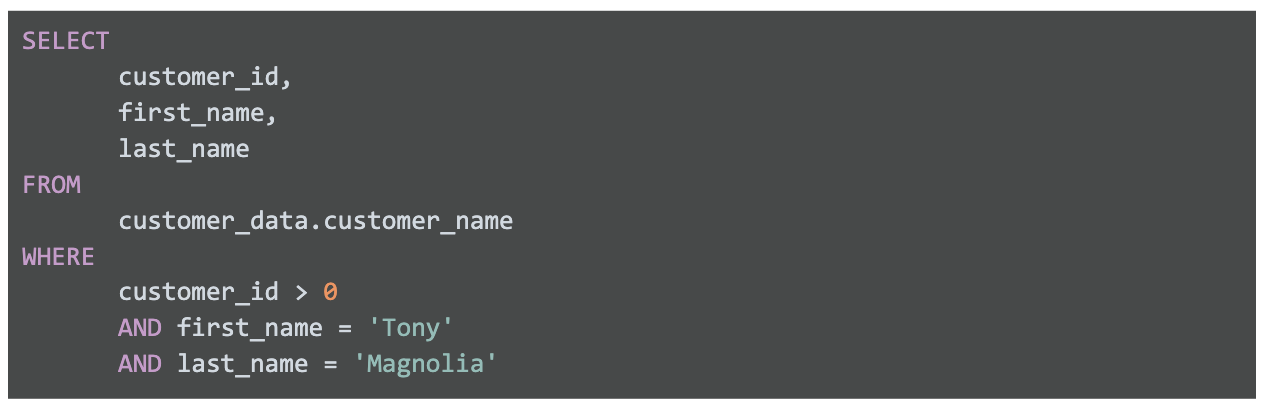
* Google Sheets automatically saves spreadsheets in your Google Drive.
* For other spreadsheet applications like Excel, you'll save spreadsheets as files.

## Structured Query Language (SQL)

* SQL (Structured Query Language) is a powerful tool for data analysis, similar to spreadsheets but designed for larger datasets.
* SQL requires a compatible database where it can communicate and perform operations.
* SQL queries have a structured format: SELECT (columns), FROM (table), and WHERE (conditions).
* Queries are used to request specific data or information from a database.
* SQL queries can filter data based on conditions, refining the results.
* SQL is universal and works similarly across various database management systems.
* It's particularly useful for handling extensive datasets that might overwhelm spreadsheets.
* As you progress in your data analytics journey, you'll learn to create more complex SQL queries for in-depth analysis.
* SQL (Structured Query Language) allows data analysts to communicate with databases and retrieve information efficiently.
* A query is a request for data or information from a database, and SQL is used to convey this request.
* SQL queries follow a specific syntax structure: SELECT (columns), FROM (table), and WHERE (conditions).
* Structuring SQL queries with SELECT, FROM, and WHERE in the mentioned order is a helpful practice.
* An example query: **SELECT first\_name FROM customer\_data.customer\_name WHERE first\_name = 'Tony'** selects the first name from a table named **customer\_name** in a dataset named **customer\_data** where the first name is 'Tony'.
* Multiple columns can be selected in a query by listing them with commas in the SELECT command.
* Conditions in the WHERE clause can use logical operators like AND to filter data based on multiple criteria.
* Efficient SQL queries only select the columns needed for analysis.
* As you gain experience, you can use more advanced operators like OR and NOT in your queries.

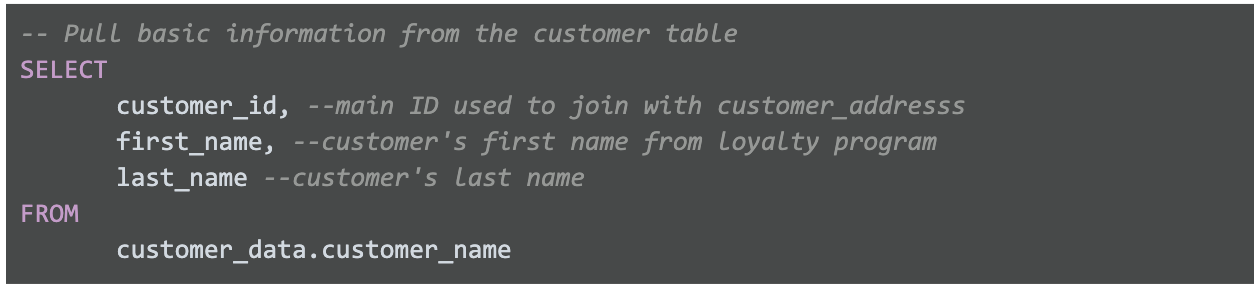






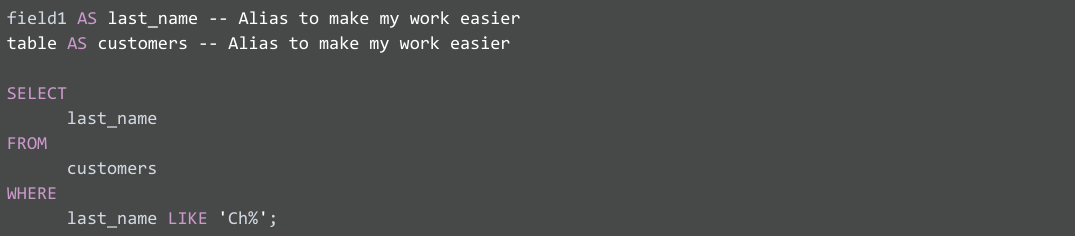
* SQL queries should be formatted neatly, even though SQL is not case-sensitive and doesn't require extra spaces.
* The semicolon (;) is used as a statement terminator in SQL, although not all databases enforce this, so you might encounter queries without semicolons.
* The WHERE clause in a SQL query is used to filter data based on specific conditions.
* You can use the LIKE clause with a wildcard (%) to search for patterns in data.
* Avoid using SELECT \* to select all columns as it can slow down queries and retrieve unnecessary data. It's better to specify the columns you need.
* Adding comments in SQL queries can provide clarity and make it easier to understand the purpose of the query.
* Aliases can be assigned to columns or tables using the AS clause to make queries more readable and manageable.

* An example scenario is given where SQL is used to retrieve employee data based on specific criteria, showcasing how SQL can provide valuable insights for analysis.
* Resources for learning SQL are recommended, including the W3Schools SQL Tutorial and an SQL Cheat Sheet for more advanced learners.



A close up of a grey surface

Description automatically generated



## Data visualization

* Data visualization involves representing data graphically to make it more understandable and insightful.
* Data analysts often find data visualization to be an exciting and rewarding aspect of their work.
* Florence Nightingale, known for her contributions to nursing, was also a data analyst who used visualizations to improve patient care during the Crimean War.
* Visualizations are powerful tools for conveying trends and patterns quickly and clearly to stakeholders.
* Data analysts should aim to create visualizations that are easy to understand and visually appealing to engage stakeholders effectively.
* Spreadsheets and data analysis tools like charts and graphs are essential for creating visualizations.
* The course introduces various types of visualizations, such as bar graphs and column charts, and provides tools to adjust and customize them.
* Learning data visualization, along with spreadsheet skills and SQL knowledge, equips data analysts to excel in their roles.
* Data visualization is the graphical representation of information and is a crucial aspect of data analysis.
* The data visualization process involves several steps:
  1. Explore the data for patterns and insights.
  2. Plan the visuals by defining what information you want to convey and to whom.
  3. Create the actual visualizations using appropriate tools.
* Spreadsheets like Microsoft Excel and Google Sheets offer built-in chart and graph capabilities, making it easy to create simple visualizations such as bar graphs and pie charts.
* More advanced data visualization tools like Tableau allow you to integrate data into dashboard-style visualizations and provide interactive features.
* Programming languages like R, combined with IDEs like RStudio, are also used for data visualization, especially by data analysts working with large datasets.
* Continuously exploring and experimenting with different data visualization tools and methods is essential for becoming a proficient data analyst.
* Staying curious, researching various options, and testing new programs and platforms will help you make the most of your data visualization skills

