**MANIPULATION**

**Generalizations**

Congratulations! You’ve learned six commands commonly used to manage data stored in a relational database and how to set constraints on such data. What can we generalize so far?

*SQL* is a programming language designed to manipulate and manage data stored in relational databases.

* A *relational database* is a database that organizes information into one or more tables.
* A *table* is a collection of data organized into rows and columns.

A *statement* is a string of characters that the database recognizes as a valid command.

* CREATE TABLE creates a new table.
* INSERT INTO adds a new row to a table.
* SELECT queries data from a table.
* ALTER TABLE changes an existing table.
* UPDATE edits a row in a table.
* DELETE FROM deletes rows from a table.

*Constraints* add information about how a column can be used.

* DROP
* ANALYZE

**QUERIES**

**Review**

Congratulations!

We just learned how to query data from a database using SQL. We also learned how to filter queries to make the information more specific and useful.

Let’s summarize:

* SELECT is the clause we use every time we want to query information from a database.
* AS renames a column or table.
* DISTINCT return unique values.
* WHERE is a popular command that lets you filter the results of the query based on conditions that you specify.
* LIKE and BETWEEN are special operators.
* AND and OR combines multiple conditions.
* ORDER BY sorts the result.
* LIMIT specifies the maximum number of rows that the query will return.
* CASE creates different outputs.

<https://discuss.codecademy.com/t/data-science-independent-project-1-watching-the-stock-market/419943>

**AGGREGATE FUNCTIONS**

**Review**

Congratulations!

You just learned how to use aggregate functions to perform calculations on your data. What can we generalize so far?

* COUNT(): count the number of rows
* SUM(): the sum of the values in a column
* MAX()/MIN(): the largest/smallest value
* AVG(): the average of the values in a column
* ROUND(): round the values in the column

*Aggregate functions* combine multiple rows together to form a single value of more meaningful information.

* GROUP BY is a clause used with aggregate functions to combine data from one or more columns.
* HAVING limit the results of a query based on an aggregate property.

<https://discuss.codecademy.com/t/data-science-independent-project-4-home-value-trends/419948>

<https://discuss.codecademy.com/t/data-science-independent-project-5-analyze-airfare-data/419949>

**MULTIPLE TABLES**

**Review**

In this lesson, we learned about relationships between tables in relational databases and how to query information from multiple tables using SQL.

Let’s summarize what we’ve learned so far:

* JOIN will combine rows from different tables if the join condition is true.
* LEFT JOIN will return every row in the *left* table, and if the join condition is not met, NULL values are used to fill in the columns from the *right* table.
* *Primary key* is a column that serves a unique identifier for the rows in the table.
* *Foreign key* is a column that contains the primary key to another table.
* CROSS JOIN lets us combine all rows of one table with all rows of another table.
* UNION stacks one dataset on top of another.
* WITH allows us to define one or more temporary tables that can be used in the final query.

<https://discuss.codecademy.com/t/data-science-independent-project-2-explore-a-sample-database/419945>

<https://discuss.codecademy.com/t/data-science-independent-project-3-education-census-data/419947>