LEARNING OBJECTIVES

- Review previously learned Database/SQL concepts
- Effectively use advanced SQL functions
- Wrangle data within a SQL database

REVIEW

What is a database?

TEXT

- Databases are computer systems that manage the storage, querying and processing of data
- Databases provide a way to organize data along with efficient methods to retrieve specific information
- Databases also allow users to create rules that ensure proper data management and verification

REVIEW

What is a relational database?

TEXT

A relational database is a database based on tabular data and links between data entities or concepts. Typically, a relational database is organized into tables. Each table should correspond to one entity or concept. Each table is similar to a single CSV file or Pandas dataframe.

REVIEW

What's the main difference between Postgres and SQLite?

TEXT

- SQLite is a file-based database, which means it's a database that we interact with/update directly but in essence we just read/write to a file directly

- Postgres is a server-based database, which means that there is a server process running (bind to some port), that accepts requests (i.e. SQL queries). Your process connects to the server and sends queries to the server which itself is responsible to update the database files.

REVIEW

What is SQL?

REVIEW

SQL is a language built specifically for querying and interacting with databases and won't have much of the functionality you can expect in traditional programming languages

REVIEW

What are the different methods we learned to query a database?

TEXT

SQLite

- console connection
- python sqlite3 package
- pandas
- firefox browser extension SQLite Manager

Postgres

- console connection
- python sqlalchemy package
- pandas

REVIEW

What is CRUD?

REVIEW

- The four basic functions of persistent data storage implemented in relational database applications. The corresponding commands in SQL are INSERT (Create) SELECT (READ) UPDATE(Update) DELETE (DELETE).

REVIEW

What are some SQL operators you learned so far?

REVIEW

- SELECT
- WHERE
- ORDER BY
- AS
- LIKE
- DISTINCT
- LIMIT
- GROUP BY
- HAVING
- Joins
- Subqueries

REVIEW

What is psql?

TEXT

- psql is a terminal-based frontend to PostgreSQL
- psql stands for Postgres interactive terminal

TEXT

Let’s learn a few more

TEXT

CONNECT TO THE NORTHWIND DATABASE & EXPLORE (3 MINUTES)

psql -h dsi.c20gkj5cvu3l.us-east-1.rds.amazonaws.com -p 5432 -U dsi_student northwind

password: gastudents

- Make sure Postgres is running!
- Explore: view list of all tables, view the contents of specific ones

THE CASE STATEMENT

- More or less the SQL equivalent of an if/then statement
- Try with northwind (3 minutes)

```
SELECT column_name,
       CASE WHEN foo = '1' THEN 'yes'
            ELSE NULL END AS foo_new_column
  FROM table_name
```

CASE EXAMPLE

```
SELECT "City",
        CASE WHEN "City" = 'Berlin' THEN 'one'
        ELSE NULL END AS "New City"
FROM customers;
```

CONCAT

- The CONCAT function is a useful way to join two columns of strings within a SQL statement.

```
SELECT CONCAT(first_column, second_column)
    FROM table;
```

CONCAT EXAMPLE

```
SELECT  CONCAT("City", "Country")

FROM customers;
```

SCALAR FUNCTIONS

- Scalar functions are useful tools for text and number formatting within a SQL statement. Instead of formatting and wrangling data within python, SQL scalar functions can help us take care of these tasks in the preprocessing stage.

- Scalar functions range from text formatting such as LOWER(), which converts a string to lower case letters and MID() which extracts characters from a string, to ROUND() which rounds a number to a specific amount of decimals.

- SELECT LOWER("column_name") FROM table;

SCALAR FUNCTION EXAMPLE

```
SELECT LOWER("City") FROM customers;
```