data science
student society

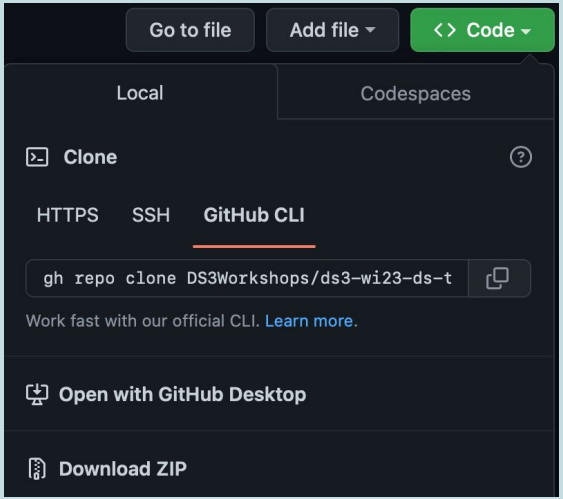
Introduction to SQL

Baraa Zekeria and Ojas Vashishtha

Download Code

data science student society

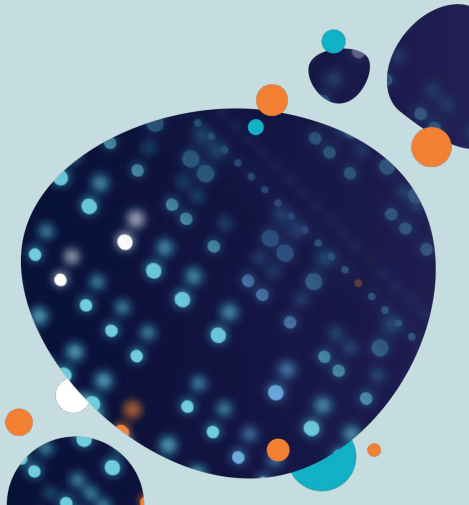
- Go to [bit.ly/ds3sql](bit.ly/ds3sql)
- Two ways
  - Clone repository (ideal for DataHub users)
    - `git clone https://github.com/DS3Workshops/ds3-wi23-ds-tools-workshops`
  - Download as a ZIP



What is SQL?

data science
student society

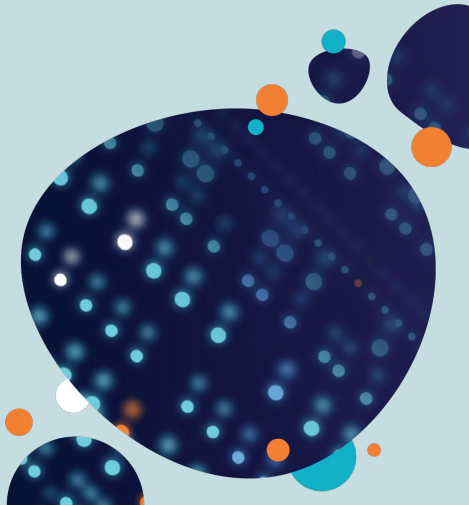
- Structured Query Language
- Not just gathering information, but structuring, cleaning, formatting, setting up data for future use
- Works in relation with databases/raw data
- What’s going on BEFORE we analyze the data?
- Key programming language for data modeling



SQL Data Types

data science
student society

- Used to define the values that a column can contain
- Types
  - Numeric (INT, BOOL, FLOAT, DECIMAL, etc.)
  - String (CHAR, BINARY, LONGTEXT, SET, etc.)
  - Date and Time (DATE, TIMESTAMP, YEAR, etc.)



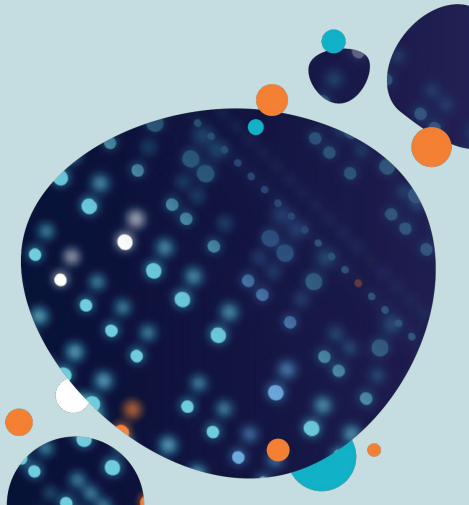
Comparison

- +
- -
- *
- /
- % (returns remainder)

Arithmetic

- =
- != or <>
- >
- <
- >=
- <=
- !<
- !>

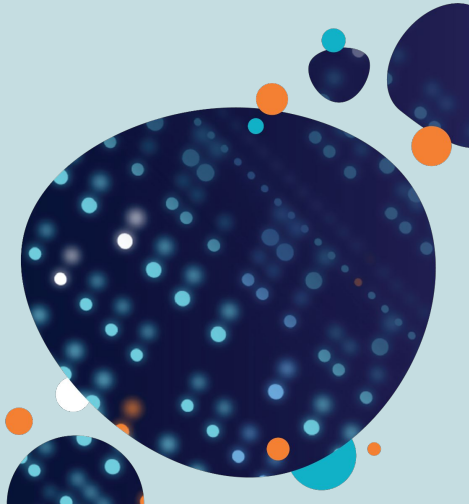
data science
student society



Logical Operators

data science
student society

- ALL (compares a value to all values in another value set)
- AND
- ANY (compares the values in the list according to the condition)
- BETWEEN
- IN
- NOT
- OR
- EXIST (search for a row)
- LIKE (compares a value to similar values using a wildcard operator)



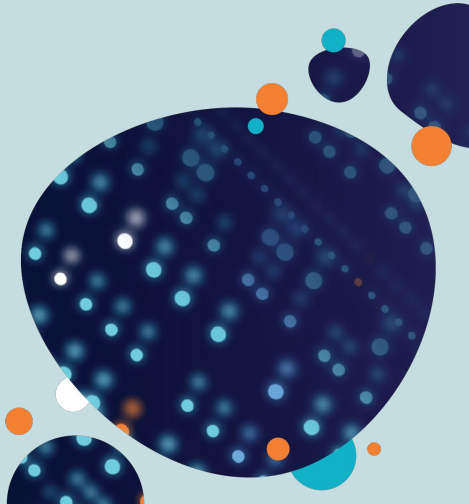
SQL

- Data Management
- Large Datasets
- “Communicating” with data
- Data Engineering
- Data automation

Pandas

- Data Analysis
- Slower for bigger data
- Generating insights from data
- Much better for visualizations
- Complicated functions/ML

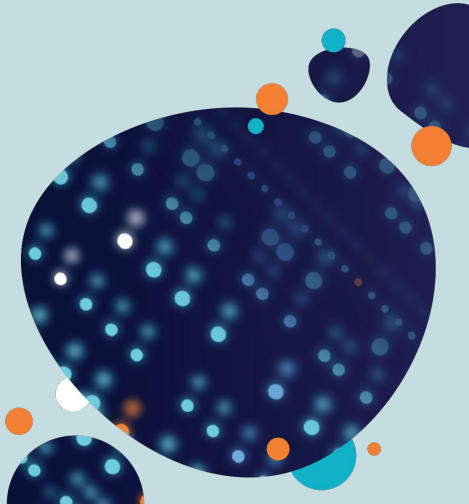
data science
student society



Some Basic Commands

data science student society

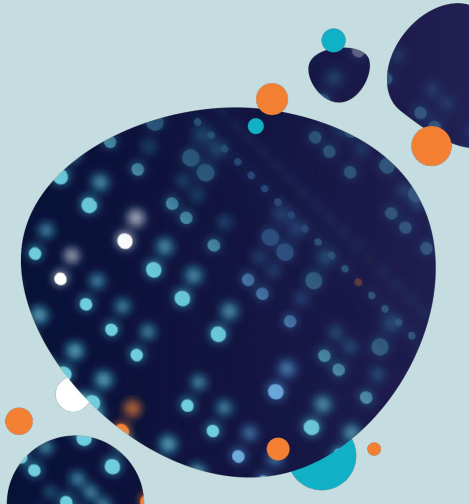
- `SELECT *`
- `WHERE` clause
  - Add a condition in this clause to filter
- `ORDER BY`
- `MIN, MAX, AVG, SUM, COUNT` (aggregates)
- JOINS
  - Also a very important clause
  - How we connect different data tables
  - Important for creating future tables for analysts to use
  - Must have a match between tables



Interpreting a Query

data science student society

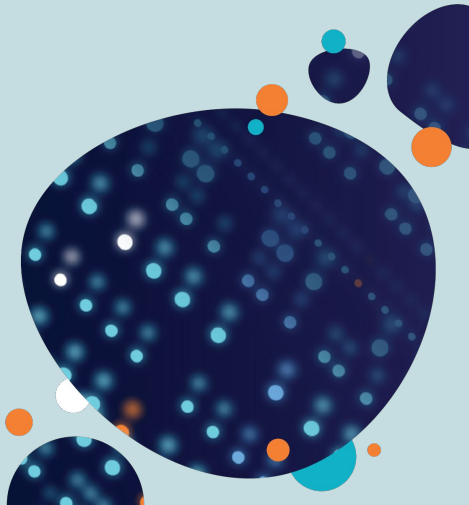
- Always start from the **`FROM`** clause to create the data (includes joins)
- Using the **`WHERE`** clause, the data is then filtered
- The data is then grouped (**`GROUP BY`**) by the columns specified
- **`SELECT`** is ran (i.e columns to output)
- Data is then sorted from the **`ORDER BY`** clause



Query

data science student society

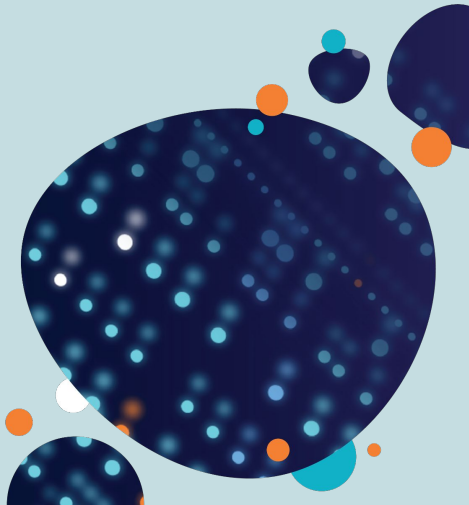
- A request for data from one or more tables
- Example: Output a given student's grade
  - ```
    SELECT *
    FROM grades
    WHERE student_id =  “A012”
    ```


    ```

Subqueries

data science student society

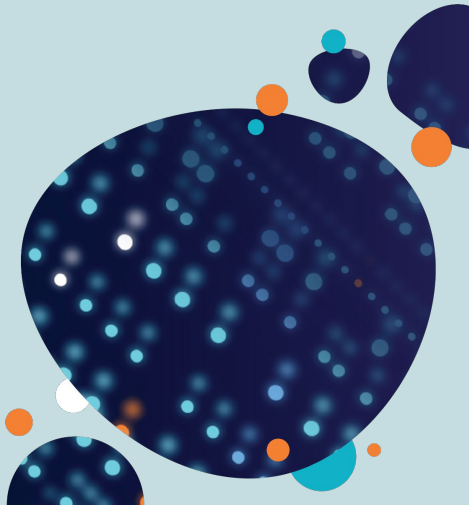
- A query nested insider a larger query
- Usually added within the WHERE clause
- Use cases:
  - Compare an expression to the result of the query
  - Determine if an expression is included in the results of the query
  - Check whether the query selects any rows
- Example: Identify students who has a better grade than student “A012”
  - ```
    SELECT a.student, a.name, b.total_grade
    FROM student a, grades b
    WHERE a.student_id = b.student_id AND
    b.total_grade >(SELECT total_grade
                    FROM grades
                     WHERE stuent_id =
    “A012”)
    ```


    ```

SQL Injection

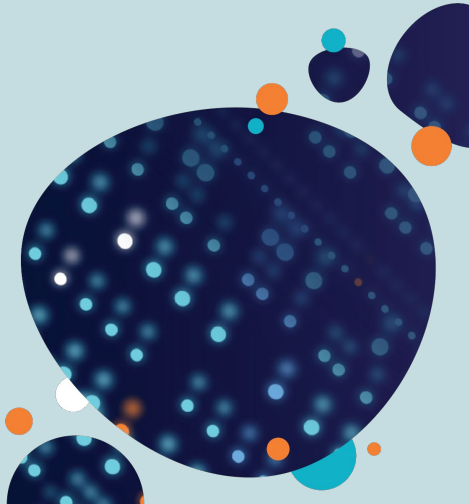
data science student society

- A code that can harm/steal data from your database
- Placement of malicious SQL code via web page input
- Example:
  - `user_id`: 105 or 1 = 1
  - ```
    SELECT *
    FROM users
    WHERE user_id = 105 OR 1 = 1;
    ```
  - Will always return rows from the `users` table to the user
- Use SQL parameters to remedy the hack


    ```

data science
student society

DEMO



Resources

- Practice Questions
  - DataLemur
  - StrataScratch (includes R and Python (pandas) as well!)
- Tutorials
  - W3Schools
- Books
  - *SQL Queries for Mere Mortals* by John L. Viescas
- UCSD Courses
  - DSC 100, DSC 102
  - CSE 132A-C
- Many more online!

data science student society

Feedback

data science
student society