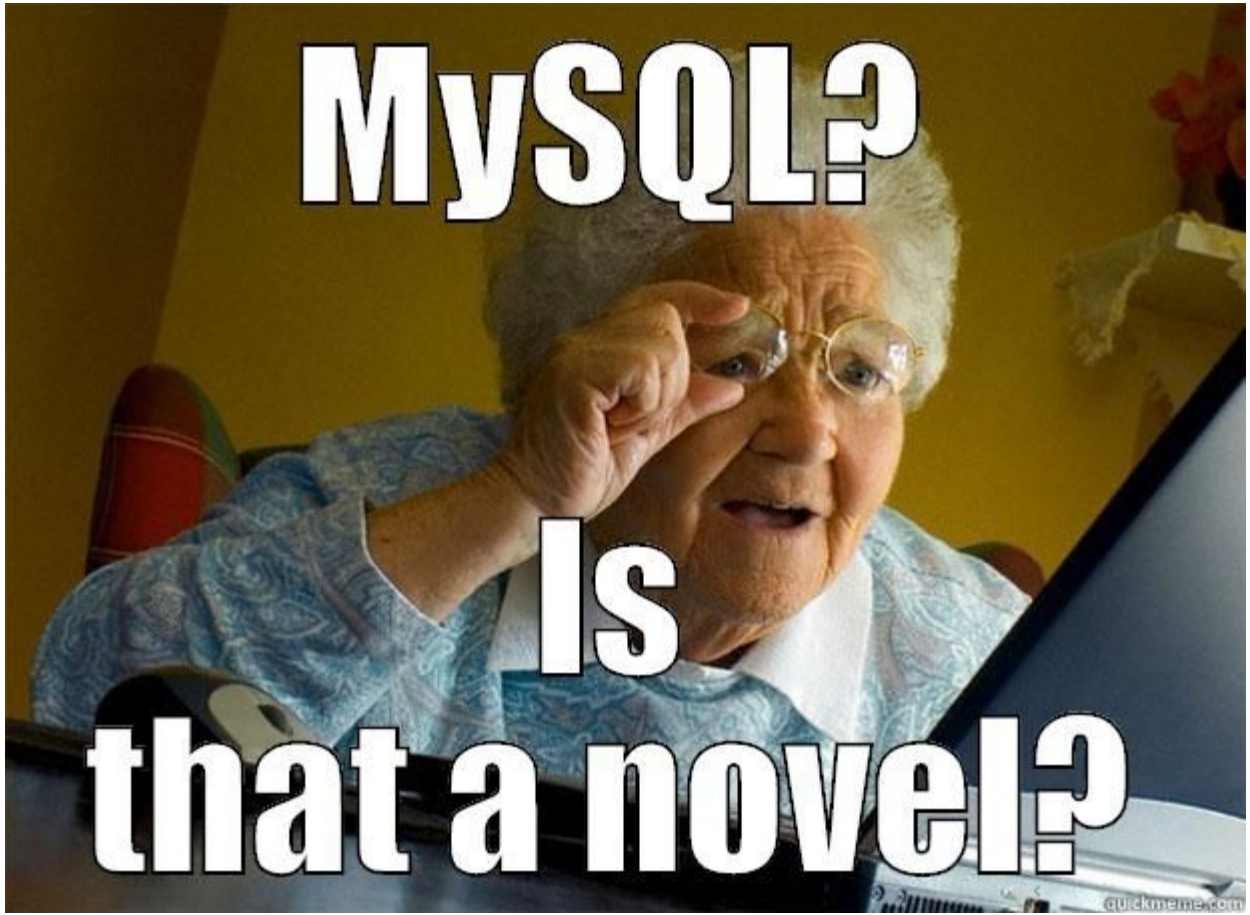
SQL

Structured Query Language

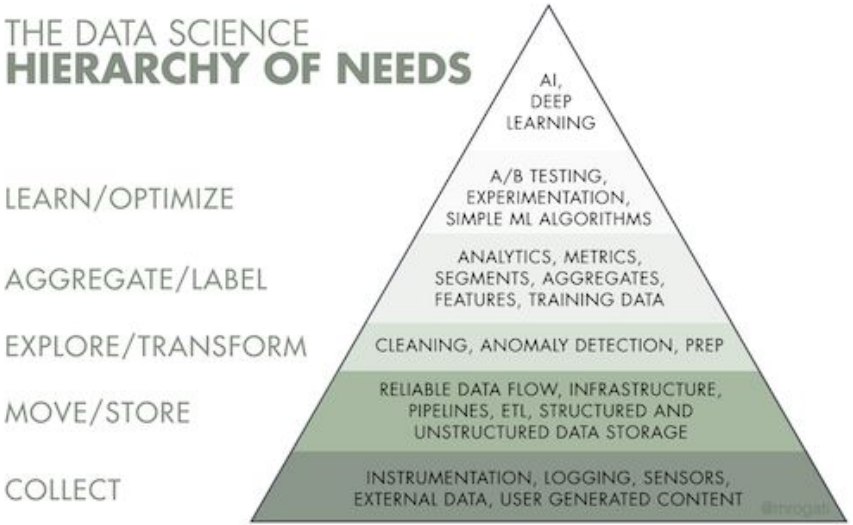
Adapted from the DC campus and instructors

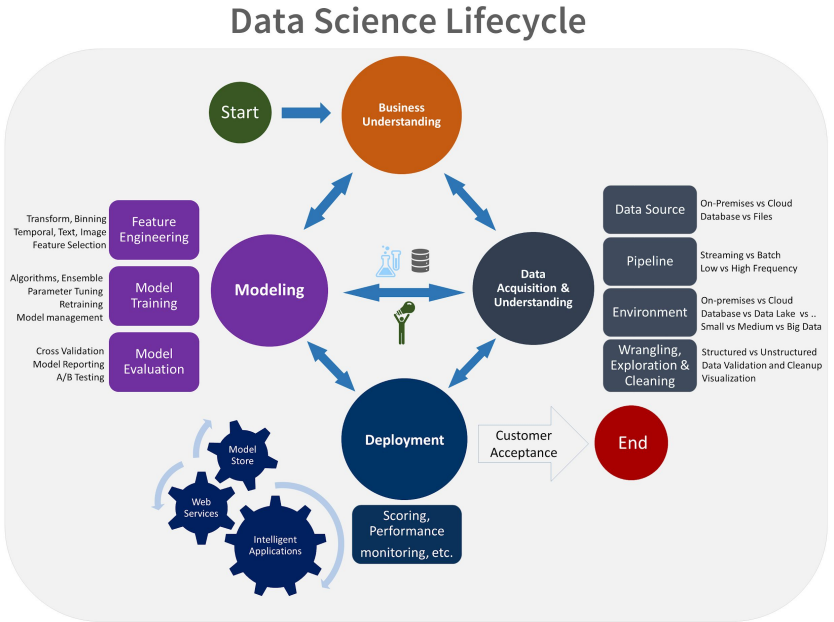

MySQL?
Is
that a novel?
quickmeme.com

Objectives:

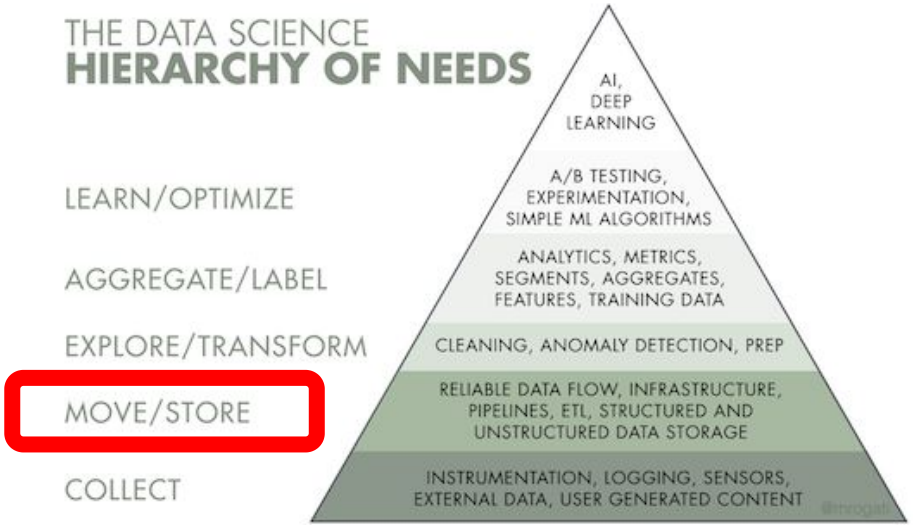
1. Summarize the use case for sql in the data science skill set
2. Define key sql terminology
3. Get information about DB schema and table structure
4. Use basic SQL commands and construct simple SQL queries
5. Convert SQL to pandas
6. Identify main JOIN options

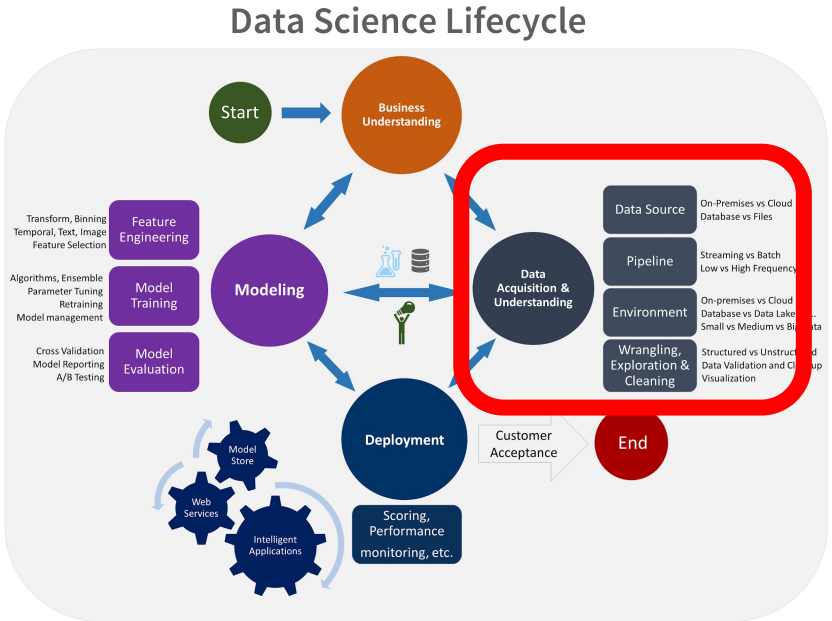
Data Science Use Case:



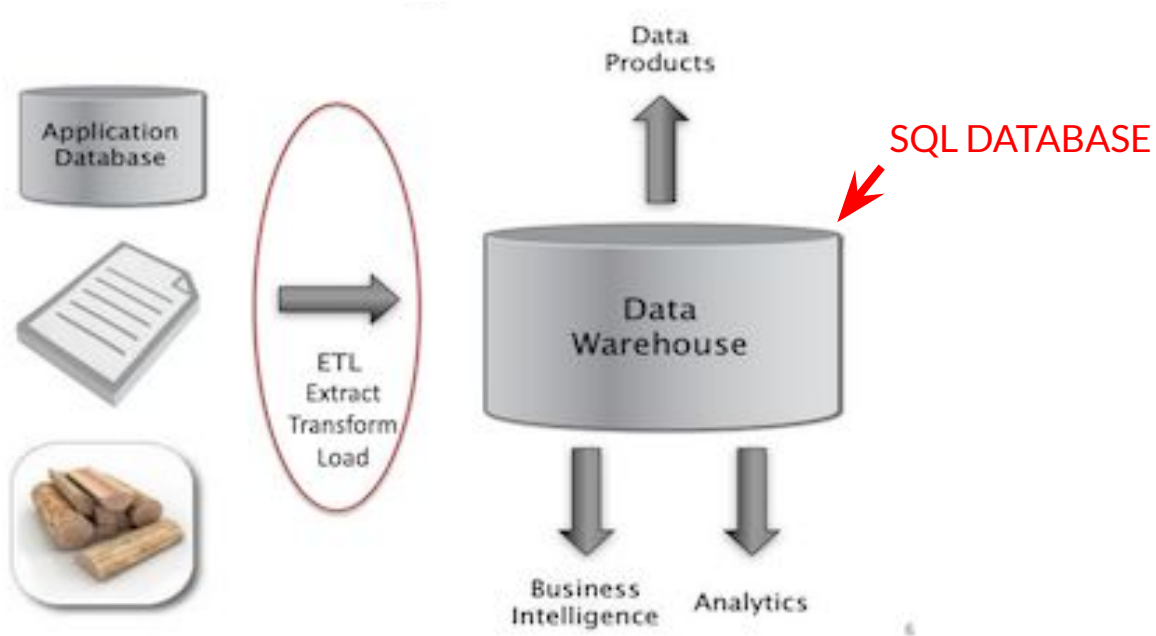


Data Science Use Case:





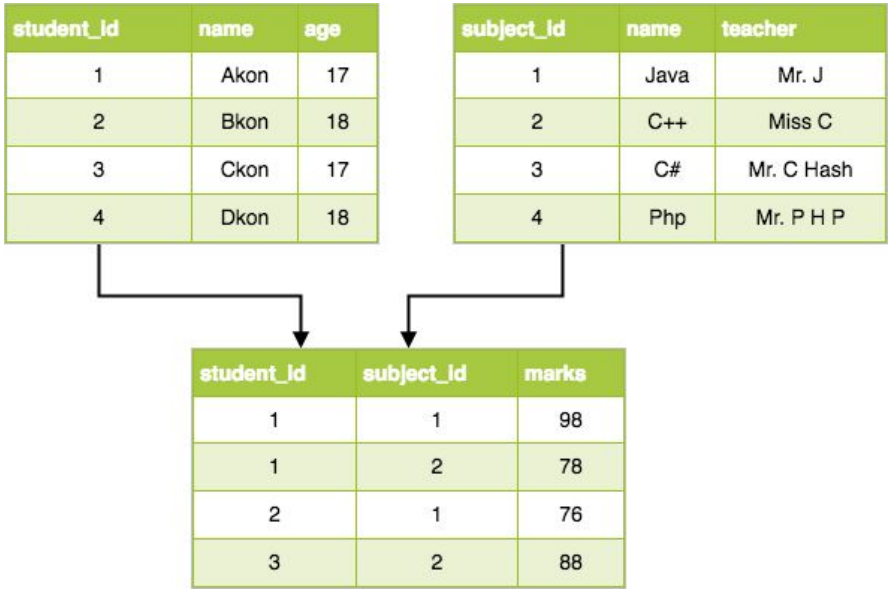
The Big Picture



Relational Databases

RDBMS:

- SQLite
- MySQL
- PostgreSql
- Oracle DB
- SQL Server

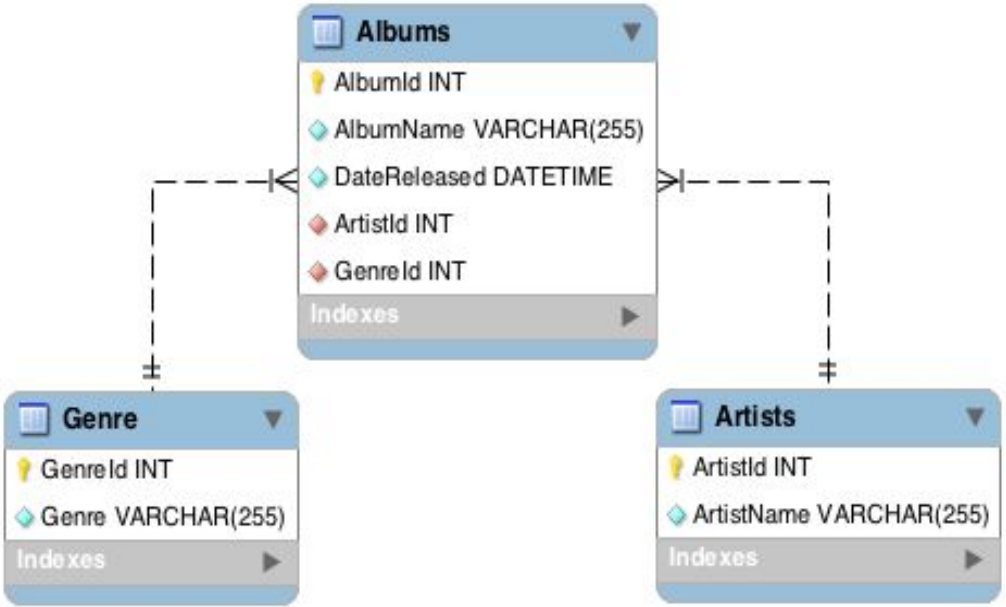


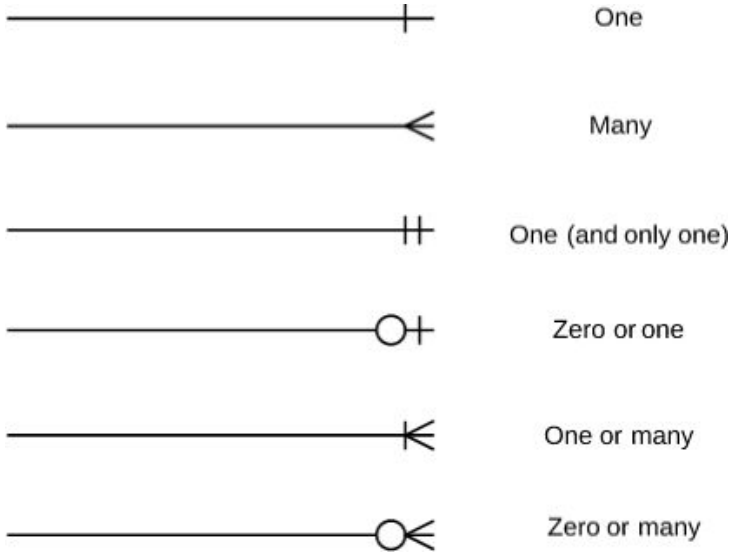
student_id	name	age
1	Akon	17
2	Bkon	18
3	Ckon	17
4	Dkon	18

subject_id	name	teacher
1	Java	Mr. J
2	C++	Miss C
3	C#	Mr. C Hash
4	Php	Mr. P H P

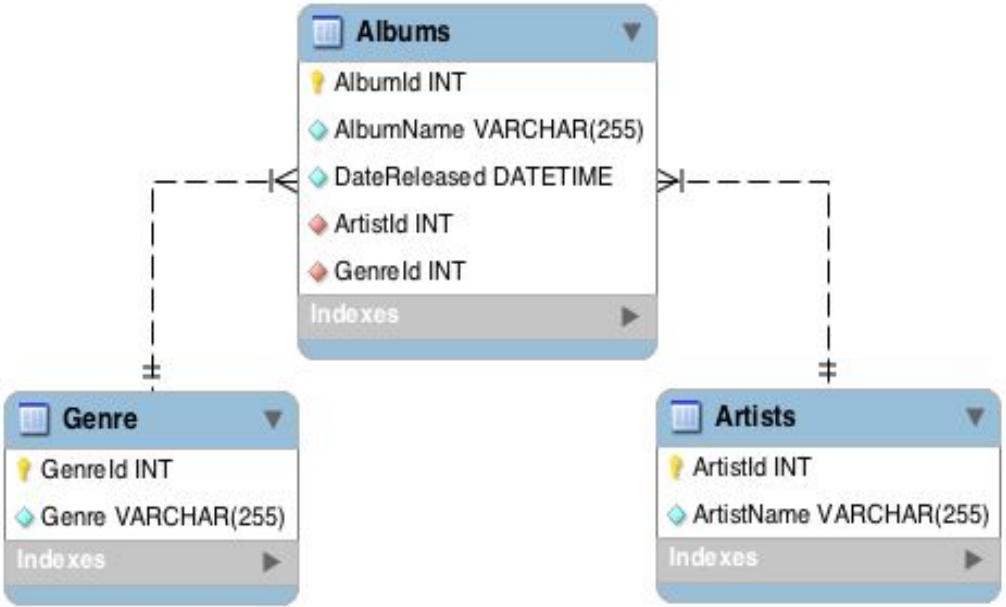
student_id	subject_id	marks
1	1	98
1	2	78
2	1	76
3	2	88

Database Schema





Database Schema



ALBUMS	DATA TYPE
Album ID	INT
AlbumName	VARCHAR
DateReleased	DATETIME
ArtistID	INT
GenreID	INT

Additional Terminology

- PRIMARY KEY: uniquely identifies a record (row) in the table.
- FOREIGN KEY: a field in the table that is primary key in another table. Links two tables.
- SCHEMA:
  - The structure/layout of the database (table, columns, relationships)
- STRUCTURED QUERIES
  - Our queries require us to follow a specific format
- VIEWS
  - Built on top of tables, almost like a sub-table
  - E.g. Classical Music Albums is View of the Albums Table

SQL Syntax: White Boarding Exercises

Students:

name	favorite_food	age
Christiaan	tofu	28
Gary	tacos	17
Princess	sushi	32
Patrick	tacos	28
Maia	tacos	19
Thoa	scallops	28
Sydney	Tacos	33

```
SELECT name

FROM students;
```

Christiaan
Gary
Princess
Patrick
Maia
Thoa
Sydney

Students:

name	favorite_food	age
Christiaan	tofu	28
Gary	tacos	17
Princess	sushi	32
Patrick	tacos	28
Maia	tacos	19
Thoa	scallops	28
Sydney	Tacos	33

```
SELECT name, favorite_food

FROM students

WHERE favorite_food = tacos ;
```

Gary	tacos
Patrick	tacos
Maia	tacos
Sydney	Tacos

Students:

name	favorite_food	age
Christiaan	tofu	28
Gary	tacos	17
Princess	sushi	32
Patrick	tacos	28
Maia	tacos	19
Thoa	scallops	28
Sydney	tacos	33

```
SELECT name, favorite_food, age

FROM students

WHERE favorite_food = tacos

Order by age;
```

Gary	tacos	17
Maia	tacos	19
Patrick	tacos	28
Sydney	tacos	33

Students:

name	favorite_food	age
Christiaan	tofu	28
Gary	tacos	17
Princess	sushi	32
Patrick	tacos	28
Maia	tacos	19
Thoa	scallops	28
Sydney	tacos	33

```
SELECT name, favorite_food, age
FROM students
WHERE favorite_food = tacos
Order by age
Limit 1;
```

Gary	tacos	17

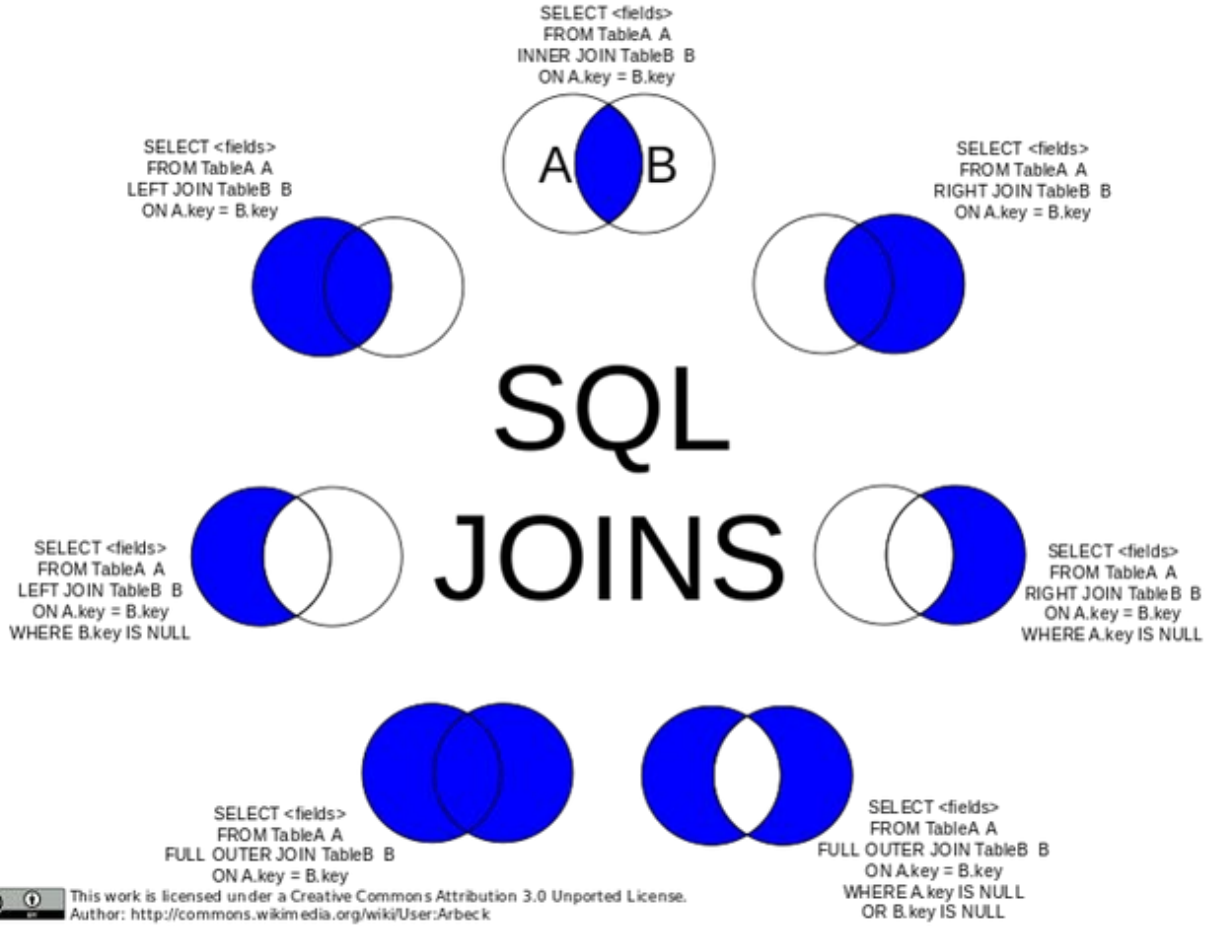
Let's try?

1. Get the names and ages of everyone who likes blue
2. Get the names of everyone who likes blue, sorted by age
3. Get the names, age and favorite colors of everyone who likes blue, sorted by name (alphabetically)

name	favorite_color	age
Christiaan	blue	28
Gary	red	17
Princess	blue	32
Patrick	orange	28
Maia	purple	19
Thoa	blue	28
Sydney	green	33

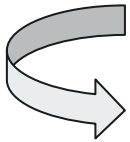
SQLite is a C library that provides a lightweight disk-based database that doesn’t require a separate server process and allows accessing the database using a nonstandard variant of the SQL query language. Some applications can use SQLite for internal data storage. It’s also possible to prototype an application using SQLite and then port the code to a larger database such as PostgreSQL or Oracle." - sqlite documentation

NOTEBOOK TIME!

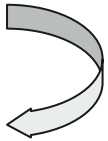


Student_id	name
1	Christiaan
2	Gary
3	Princess

food _id	student _id	food	calories
11	1	tofu	100
12	4	tacos	500
13	3	sushi	250



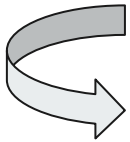
FULL OUTER JOIN



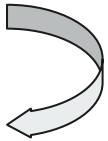
Student_id	name	food _id	student _id	food	calories
1	Christiaan	11	1	tofu	100
2	Gary	NULL	NULL	NULL	NILL
3	Princess	13	3	sushi	250
NULL	NULL	12	4	tacos	500

Student_id	name
1	Christiaan
2	Gary
3	Princess

food _id	student _id	food	calories
11	1	tofu	100
12	4	tacos	500
13	3	sushi	250



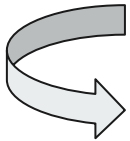
LEFT JOIN



Student_id	name	food _id	student _id	food	calories
1	Christiaan	11	1	tofu	100
2	Gary	NULL	NULL	NULL	NILL
3	Princess	13	3	sushi	250

Student_id	name
1	Christiaan
2	Gary
3	Princess

food _id	student _id	food	calories
11	1	tofu	100
12	4	tacos	500
13	3	sushi	250



INNER JOIN

Student_id	name	food _id	student _id	food	calories
1	Christiaan	11	1	tofu	100
3	Princess	13	3	sushi	250

Join Queries Syntax Example

SELECT column-names

FROM table-name1 ***JOIN*** table-name2

ON column-name1 = column-name2

WHERE condition

JOIN
LEFT JOIN
RIGHT JOIN
FULL OUTER JOIN
INNER JOIN