Find the right table

APPLYING SQL TO REAL-WORLD PROBLEMS

Dmitriy (Dima) Gorenshteyn
Lead Data Scientist, Memorial Sloan Kettering Cancer Center

What table should I use?

- What columns are in your tables?
- What is the content in these columns?

```
SELECT *
FROM payment;
```

LIMIT your results

```
SELECT *
FROM payment;
```

```
rental_id rental_date         inventory_id customer_id return_date
2         2005-05-24 22:54:33    1525    459          2005-05-28 19:40:33
3         2005-05-24 23:03:39    1711    408          2005-06-01 22:12:39
4         2005-05-24 23:04:41    2452    333          2005-06-03 01:43:41
5         2005-05-24 23:05:21    2079    222          2005-06-02 04:33:21
6         2005-05-24 23:08:07    2792    549          2005-05-27 01:32:07
7         2005-05-24 23:11:53    3995    269          2005-05-29 20:34:53
8         2005-05-24 23:31:46    2346    239          2005-05-27 23:33:46
.....................16035 MORE ROWS.....................................
```

-

LIMIT your results

```
SELECT *
FROM payment
LIMIT 5;
```

```
rental_id rental_date        inventory_id customer_id return_date
2         2005-05-24 22:54:33    1525    459         2005-05-28 19:40:33
3         2005-05-24 23:03:39    1711    408         2005-06-01 22:12:39
4         2005-05-24 23:04:41    2452    333         2005-06-03 01:43:41
5         2005-05-24 23:05:21    2079    222         2005-06-02 04:33:21
```

What tables are in my database?

PostgreSQL:

```
SELECT *
FROM pg_catalog.pg_tables
;
```

```
schemaname      tablename      tableowner
public          address        postgres
public          actor          postgres
public          film_actor     postgres
public          language       postgres
...             ...            ...
```

What tables are in my database?

PostgreSQL:

```
SELECT *
FROM pg_catalog.pg_tables
WHERE schema_name = 'public;
```

```
schemaname      tablename      tableowner
public          address        postgres
public          actor          postgres
public          film_actor     postgres
public          language       postgres
...             ...            ...
```

What tables are in my database?

PostgreSQL:

```
SELECT * FROM pg_catalog.pg_tables;
```

SQL Server - TSQL:

```
SELECT * FROM INFORMATION_SCHEMA.TABLES;
```

MySQL:

```
SHOW TABLES;
```

...

Find the tables you need!

APPLYING SQL TO REAL-WORLD PROBLEMS

Join the correct tables

APPLYING SQL TO REAL-WORLD PROBLEMS

Dmitriy (Dima) Gorenshteyn
Lead Data Scientist, Memorial Sloan Kettering Cancer Center

All tables & columns

PostgreSQL:

```
SELECT * FROM information_schema.columns;
```

SQL Server - TSQL:

```
SELECT * FROM information_schema.columns;
```

MySQL:

```
SELECT * FROM information_schema.columns;
```

...

All tables & columns

PostgreSQL:

```
SELECT *
FROM information_schema.columns
;
```

table_catalog	table_schema	table_name	column_name
pagilla	pg_catalog	pg_proc	proname
pagilla	pg_catalog	pg_proc	pronamespace
pagilla	pg_catalog	pg_proc	proowner
pagilla	pg_catalog	pg_proc	prolang
...	...	...	...

All tables & columns

PostgreSQL:

```
SELECT *
FROM information_schema.columns
WHERE table_schema = 'public';
```

```
table_catalog    table_schema    table_name    column_name
pagilla          public          address       address_id
pagilla          public          address       address
pagilla          public          address       district
pagilla          public          address       city
...              ...             ...           ...
```

Aggregate the columns

```
SELECT table_name,
       STRING_AGG(column_name, ', ') AS columns



;
```

Aggregate the columns

```
SELECT table_name,
       STRING_AGG(column_name, ', ') AS columns
FROM information_schema.columns


;
```

Aggregate the columns

```
SELECT table_name,
       STRING_AGG(column_name, ', ') AS columns
FROM information_schema.columns
WHERE table_schema = 'public'
GROUP BY table_name;
```

```
table          columns
rental         rental_id, rental_date, inventory_id, customer_id, return_date
film_actor     actor_id, film_id
film           film_id, title, description, release_year, language_id, ...
customer       customer_id, first_name, last_name, email, address_id, active
...            ...
```

A VIEW of tables and columns

A **VIEW** is a virtual table.

```
CREATE VIEW name_of_view AS
...
```

```
CREATE VIEW table_columns AS
SELECT table_name,
       STRING_AGG(column_name, ', ') AS columns
FROM information_schema.columns
WHERE table_schema = 'public'
GROUP BY table_name;
```

table_columns

```
SELECT *
FROM table_columns;
```

```
table           columns
rental          rental_id, rental_date, inventory_id, customer_id, return_date
film_actor      actor_id, film_id
film            film_id, title, description, release_year, language_id, rental_duration,
customer        customer_id, first_name, last_name, email, address_id, active
actor           actor_id, first_name, last_name
...             ...
```

Let's find some data!

APPLYING SQL TO REAL-WORLD PROBLEMS

Complex joins

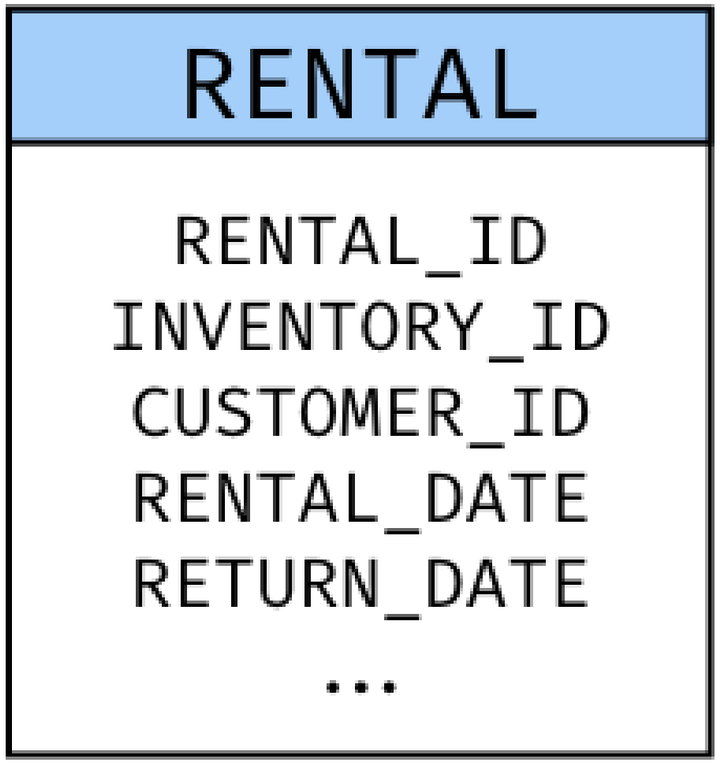
APPLYING SQL TO REAL-WORLD PROBLEMS

Dmitriy (Dima) Gorenshteyn
Lead Data Scientist, Memorial Sloan Kettering Cancer Center

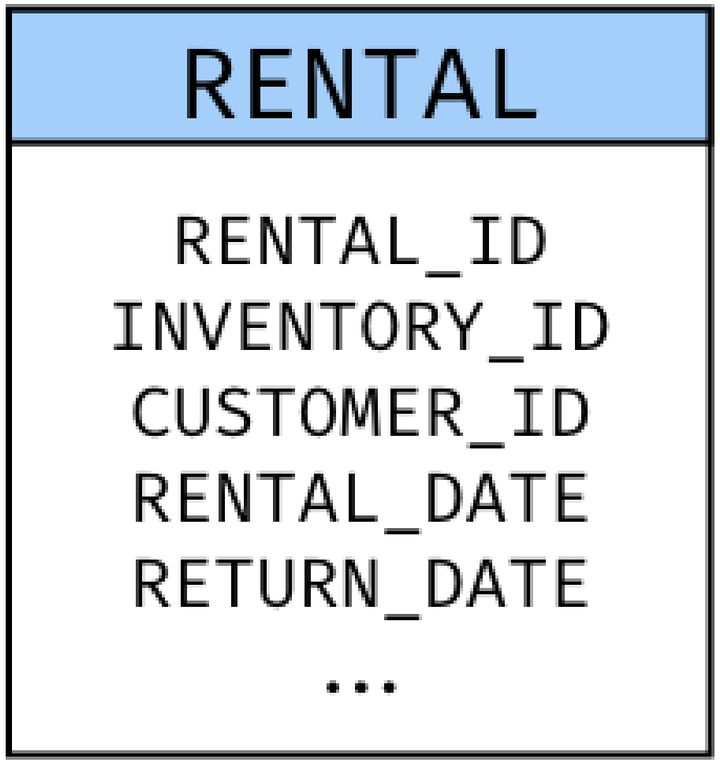
A complex question

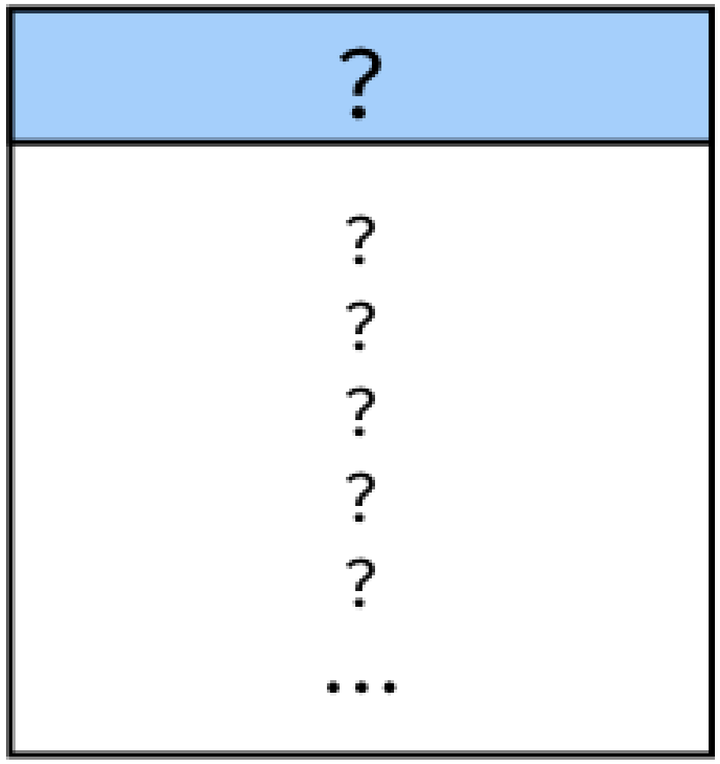
How many videos were rented in each city?

Connecting the data

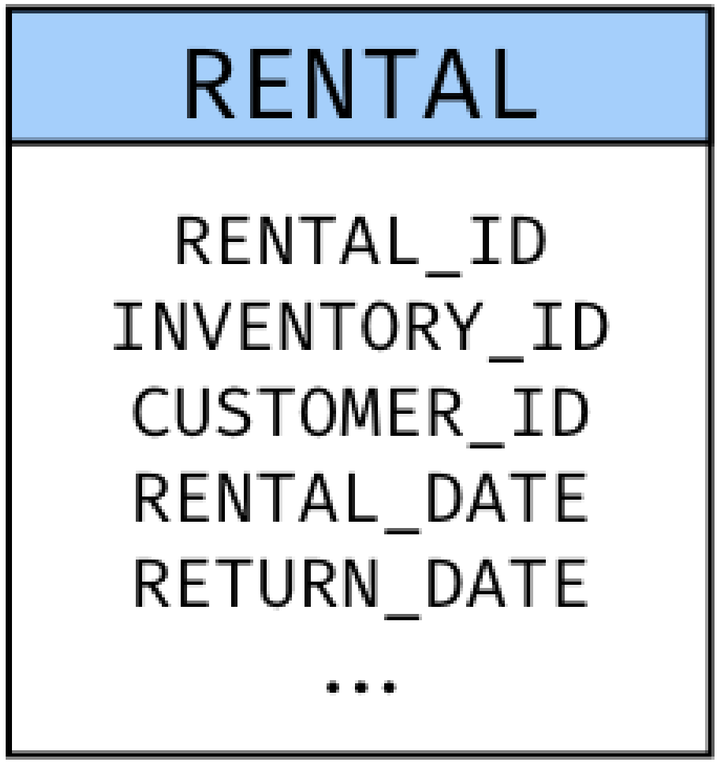


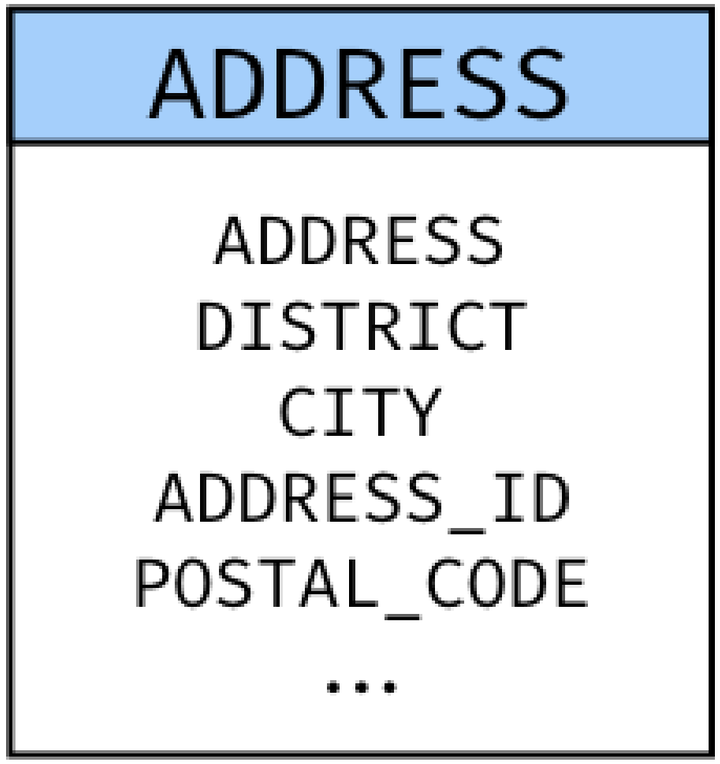
Connecting the data



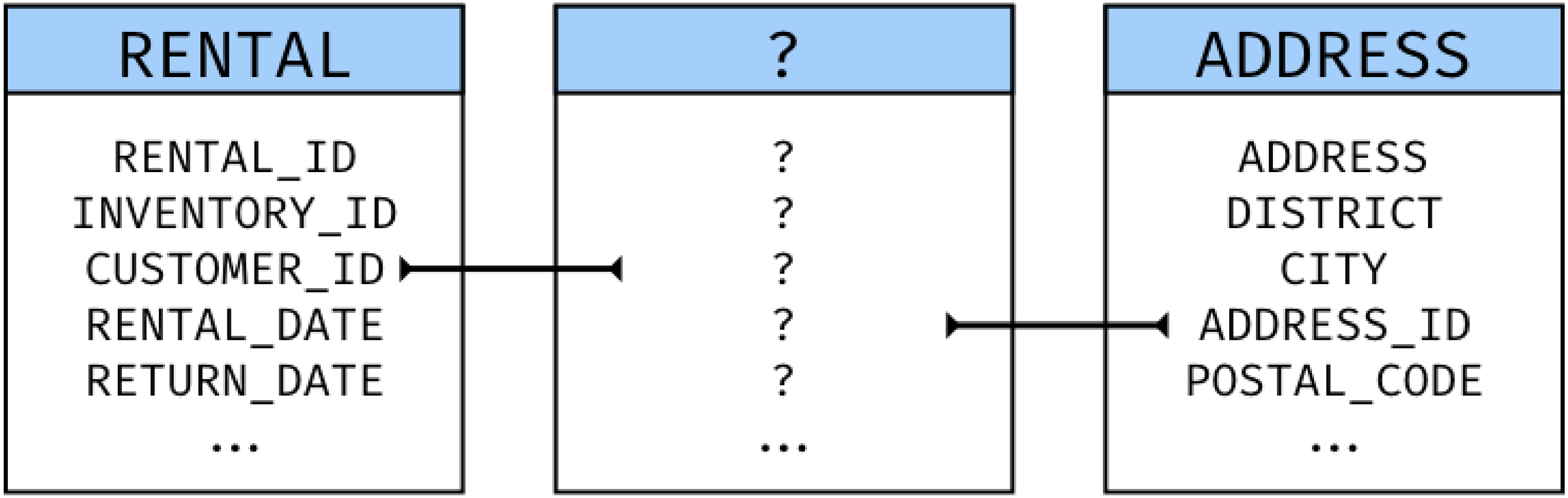


Connecting the data

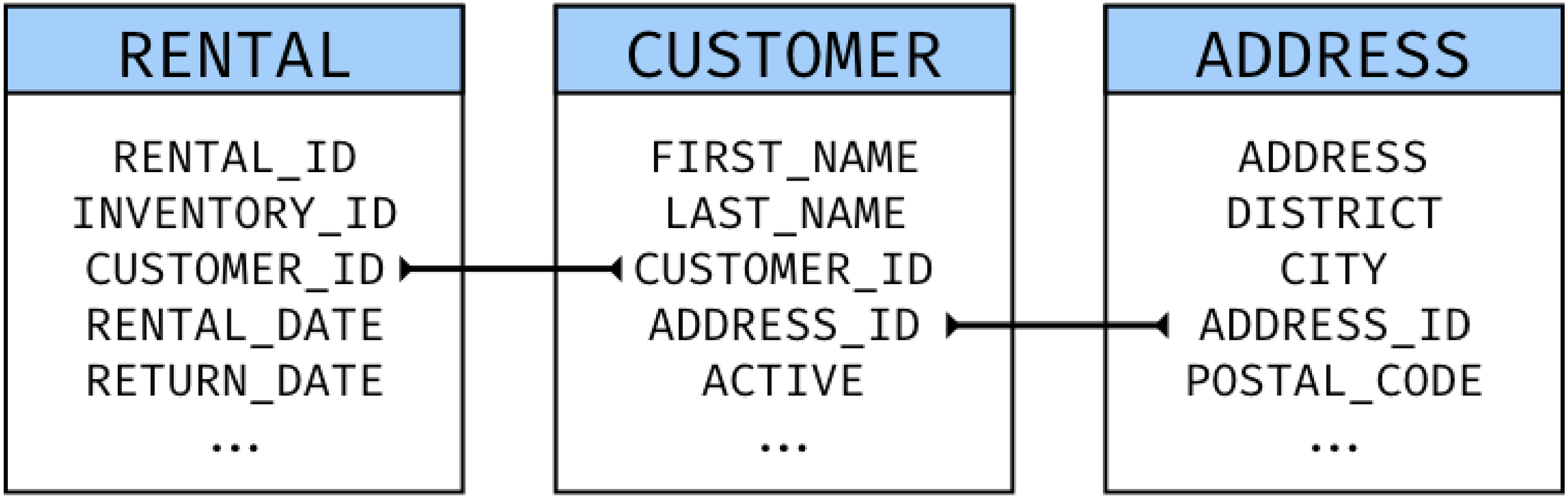




Connecting the data



Entity Relationship Diagram (ERD)



Tools for finding your data

```sql
-- LIMIT your results
SELECT *
FROM ___
LIMIT 10;
```

```sql
-- List the tables you have
SELECT *
FROM pg_catalog.pg_tables
WHERE schemaname = 'public';
```

```sql
-- Explore tables & columns using your new VIEW
SELECT * FROM table_columns;
```

Your turn!

APPLYING SQL TO REAL-WORLD PROBLEMS