Exercise 3. - SQL for Data Analysts

Step 1:

- Write a SELECT command to find out what film genres exist in the category table.

→SELECT * FROM category

	category_id [PK] integer	name character varying (25)	last_update timestamp without time zone
1	1	Action	2006-02-15 09:46:27
2	2	Animation	2006-02-15 09:46:27
3	3	Children	2006-02-15 09:46:27
4	4	Classics	2006-02-15 09:46:27
5	5	Comedy	2006-02-15 09:46:27
6	6	Documentary	2006-02-15 09:46:27
7	7	Drama	2006-02-15 09:46:27
8	8	Family	2006-02-15 09:46:27
9	9	Foreign	2006-02-15 09:46:27
10	10	Games	2006-02-15 09:46:27
11	11	Horror	2006-02-15 09:46:27
12	12	Music	2006-02-15 09:46:27
13	13	New	2006-02-15 09:46:27
14	14	Sci-Fi	2006-02-15 09:46:27
15	15	Sports	2006-02-15 09:46:27
16	16	Travel	2006-02-15 09:46:27

Total rows: 16 of 16 Query complete 00:00:01.178

Step 2:

- Write an INSERT statement to add the following genres to the category table: Thriller, Crime, Mystery, Romance, and War:

	category_id [PK] integer	name character varying (25)	last_update timestamp without time zone
1	1	Action	2006-02-15 09:46:27
2	2	Animation	2006-02-15 09:46:27
3	3	Children	2006-02-15 09:46:27
4	4	Classics	2006-02-15 09:46:27
5	5	Comedy	2006-02-15 09:46:27
6	6	Documentary	2006-02-15 09:46:27
7	7	Drama	2006-02-15 09:46:27
8	8	Family	2006-02-15 09:46:27
9	9	Foreign	2006-02-15 09:46:27
10	10	Games	2006-02-15 09:46:27
11	11	Horror	2006-02-15 09:46:27
12	12	Music	2006-02-15 09:46:27
13	13	New	2006-02-15 09:46:27
14	14	Sci-Fi	2006-02-15 09:46:27
15	15	Sports	2006-02-15 09:46:27
16	16	Travel	2006-02-15 09:46:27
17	17	Thriller	2022-06-27 15:21:06.048829
18	18	Crime	2022-06-27 15:22:39.06194
19	19	Mystery	2022-06-27 15:24:22.033428
20	20	Romance	2022-06-27 15:25:45.229813
21	21	War	2022-06-27 15:25:58.190657

- The CREATE statement below shows the constraints on the category table. Write a short paragraph explaining the various constraints that have been applied to the columns.

```
 CREATE TABLE category
(
 category_id integer NOT NULL DEFAULT nextval('category_category_id_seq'::regclass),
 name text COLLATE pg_catalog."default" NOT NULL,
 last_update timestamp with time zone NOT NULL DEFAULT now(),
 CONSTRAINT category_pkey PRIMARY KEY (category_id)
);
```

  - What do these constraints do exactly? Why are they important?

→ Constraints are important because they can help you make sure that the values in each column are consistently formatted. They can also help you make sure values in a column are unique, not null, or even check for values that don't belong. For the category table, the constraints for each column are:

- Category_id - The data type should be an integer and it cannot be null
- Name - The data type should be in text and it cannot be null
- Last_update - The data type should be the timestamp with time zone and cannot be null

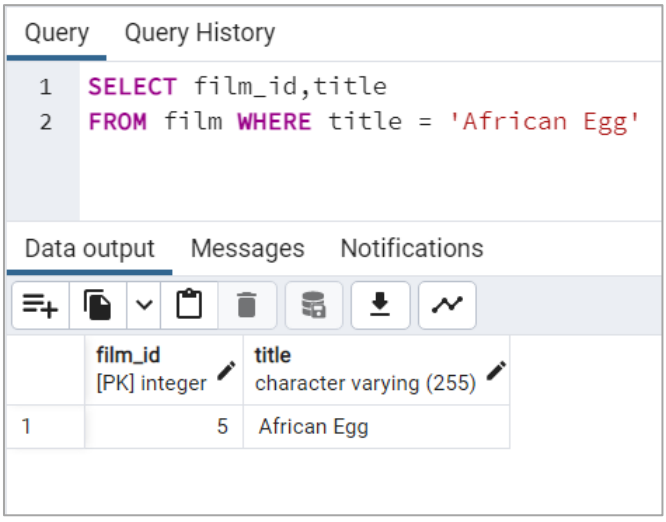
The primary key gives each record in a table a unique ID.

In this SQL command, category is the primary key.

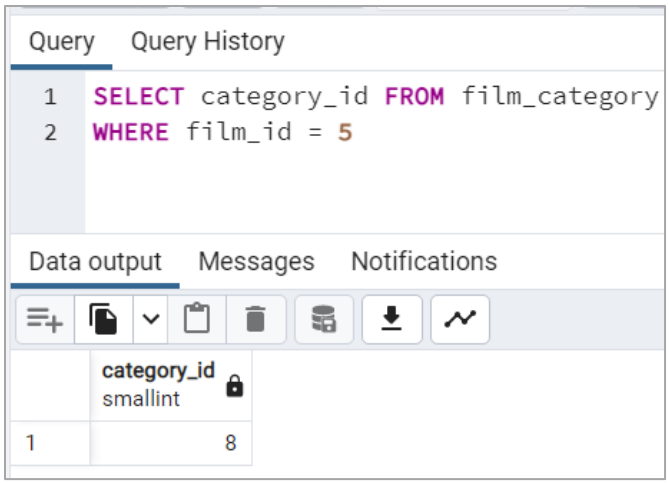
Constraints specify what type of data a table or column can accept, and they're typically set when a table is created. Constraints make querying the database quicker and easier and they also act as a data quality check in certain situations.

Step 3

  - Write the SELECT statement to find the film_id for the movie *African Egg*.

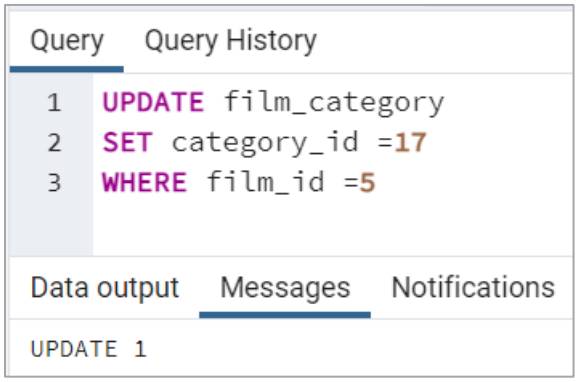


Once you have the film_ID then I need to find the category_ID :



- Once you have the film_ID and category_ID, write an UPDATE command to change the category in the film_category table (not the category table). Copy-paste this command into your answers document.

  ➔ Then I used this query to change the category_id:



Step 4:

Since there aren't many movies in the mystery category, you and your manager decide to remove it from the category table. Write a DELETE command to do so and copy-paste it into your answers document.

→
DELETE FROM category
WHERE name ='Mystery'

Step 5:

Based on what you've learned so far, think about what it would be like to complete steps 1 to 4 with Excel instead of SQL. Are there any pros and cons to using SQL? Write a paragraph explaining your answer.

→Using SQL to complete steps 1 to 4 is a lot easier than using Excel as with SQL finding tables and updating them on PgAdmin4 with just written queries and using commands makes it faster. Finding information on a particular table in excel will take longer especially when it has to do with large data and multiple data sets as this has to be done manually.

→The pro with excel is that some steps like updating the category-id can be done easily with the find and replace function Overall, it is easier with SQL if one knows how to write queries and use commands appropriately