William Su

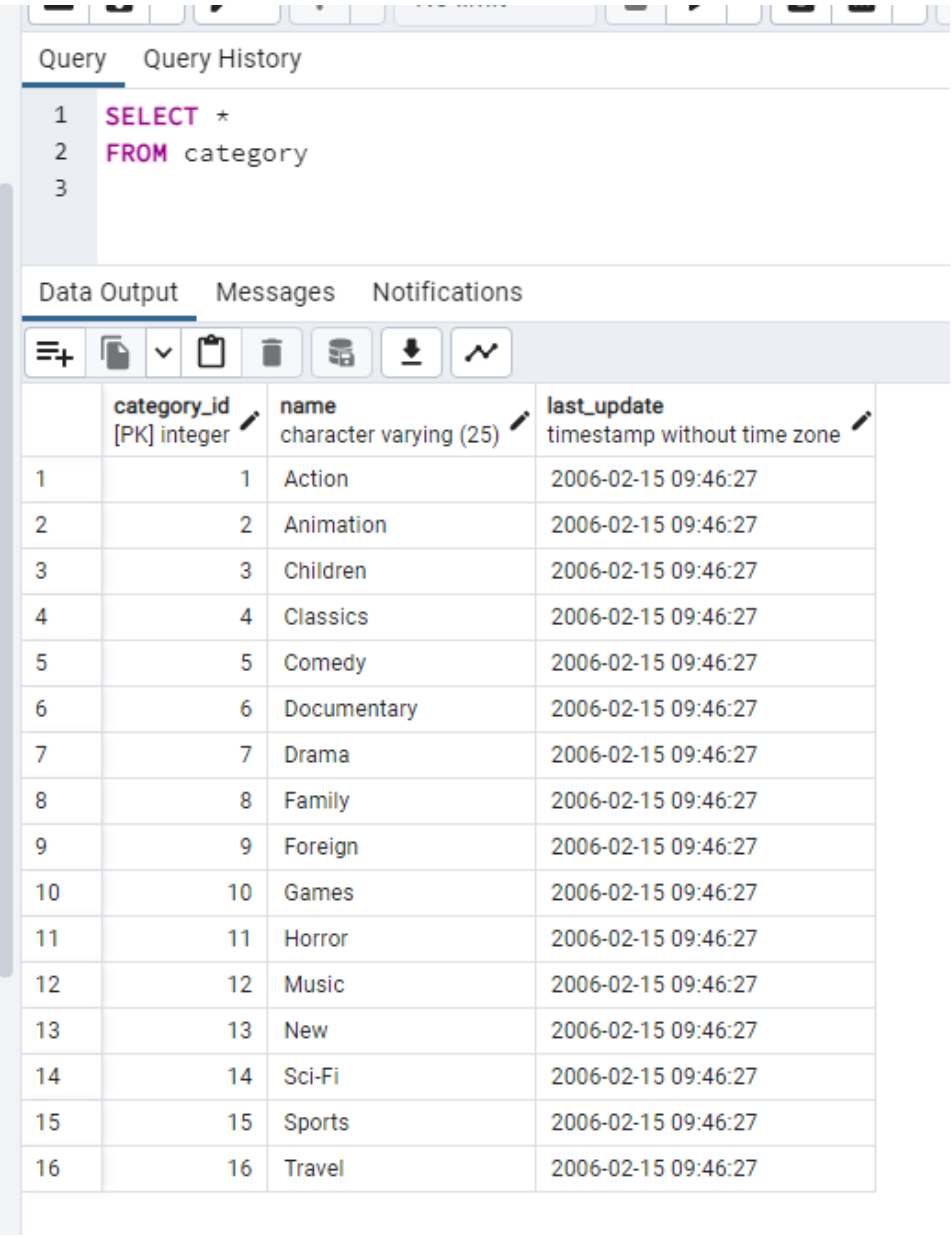
Task 3.3 – SQL for Data Analysts

Your first task is to find out what film genres already exist in the category table:

- Open pgAdmin 4, click the Rockbuster database, and open the Query Tool.
- Write a SELECT command to find out what film genres exist in the category table.
- Copy-paste the output into your answers document or write the answers out—it's up to you. Make sure to include the category ID for each genre.

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

Step 2:

You're ready to add some new genres! Write an INSERT statement to add the following genres to the category table: Thriller, Crime, Mystery, Romance, and War:

- Copy-paste your INSERT commands into your answers document.

INSERT INTO category(name)

VALUES ('Thriller'), ('Crime'), ('Mystery'), ('Romance'), ('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	2022-10-20 15:02:21.618802
18	18	Crime	2022-10-20 15:02:21.618802
19	19	Mystery	2022-10-20 15:02:21.618802
20	20	Romance	2022-10-20 15:02:21.618802
21	21	War	2022-10-20 15:02:21.618802

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

1) The three columns all have a NOT NULL constraint, so their value type cannot be empty.
2) The category_id's data type is integer, and its value cannot be null.
3) The name's data type is text, and its value cannot be null.
4) The last_update's data type is timestamp with a time zone and also cannot be null.
5) Category_id was also made the primary key of the table, forcing it to have no null or duplicate values.

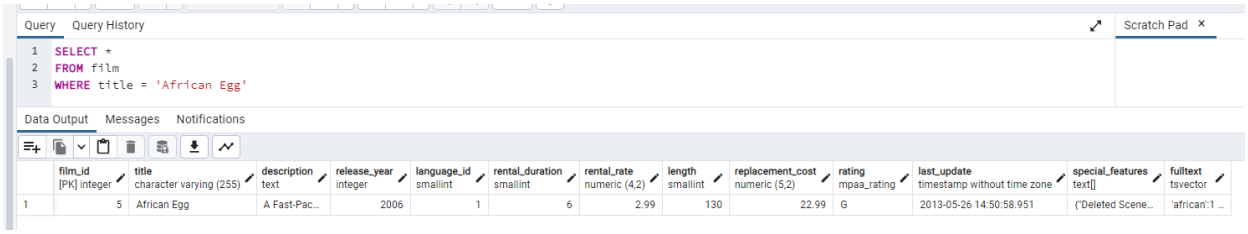
The use of these constraints is to force the formatting of the table to be uniform to ensure as few errors as possible as well as making querying the database quicker and easier to access.

Step 3:

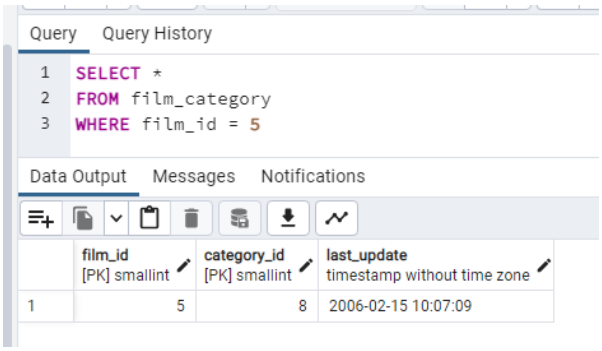
The genre for the movie *African Egg* needs to be updated to thriller. Work through the steps below to make this change:

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

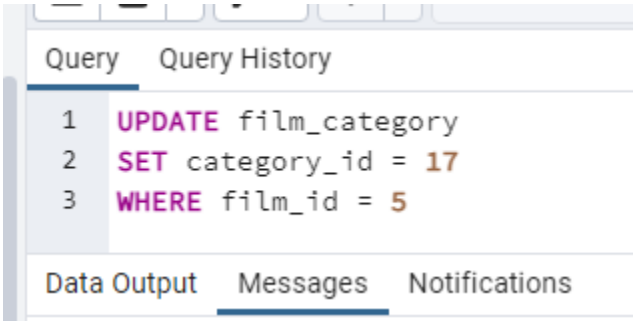
Finding Film_ID (the film_id is 5)

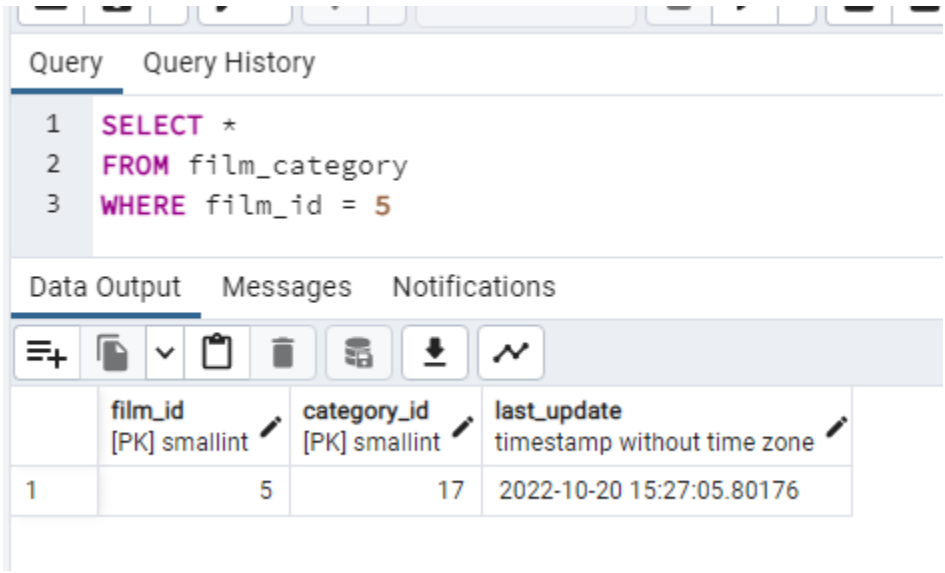


Finding category_id (the category_id was 8)



Updating





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.

Pros

1) SQL is much faster to work with, if you know what you're querying. Finding information by querying is a lot faster than manually digging through the information.
2) It's able to pinpoint whichever data you want extremely quickly, and show only that data so you won't get distracted.
3) Using constraints makes it so that some fields are automatically generated for you, like when updating the category table.

Cons

1) You need to have an understanding of the language already to be able to query properly.
2) It feels like it's slightly harder to update things or fix simple typos if they occur on SQL than Excel.