ANSWER 3.3 -SQL for Data Analysts

Step 1

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

Step2

INSERT INTO category(name)
VALUES('Thriller'),('Crime'),('Romance'),('War')

category_id [PK] integer	name character varying (25)	last_update timestamp without time zone
1	Action	2006-02-15 09:46:27
2	Animation	2006-02-15 09:46:27
3	Children	2006-02-15 09:46:27
4	Classics	2006-02-15 09:46:27
5	Comedy	2006-02-15 09:46:27
6	Documentary	2006-02-15 09:46:27
7	Drama	2006-02-15 09:46:27
8	Family	2006-02-15 09:46:27
9	Foreign	2006-02-15 09:46:27
10	Games	2006-02-15 09:46:27
11	Horror	2006-02-15 09:46:27
12	Music	2006-02-15 09:46:27
13	New	2006-02-15 09:46:27
14	Sci-Fi	2006-02-15 09:46:27
15	Sports	2006-02-15 09:46:27
16	Travel	2006-02-15 09:46:27
17	Thriller	2022-08-05 10:26:34.528067
18	Crime	2022-08-05 10:26:34.528067
19	Romance	2022-08-05 10:26:34.528067
20	War	2022-08-05 10:26:34.528067

Step 2b

NOT NULL constraint- This help to ensure that no columns have any missing or empty value

category_id : (data type is integer)Value cannot be null

name: (data type is text) Values cannot be null

last_update: (data type is timestamp with time zone) value cannot be null

PRIMARY KEY constraint: is a unique identifier for each record in a table

Category_pkey which is the category_id is set as the primary key.

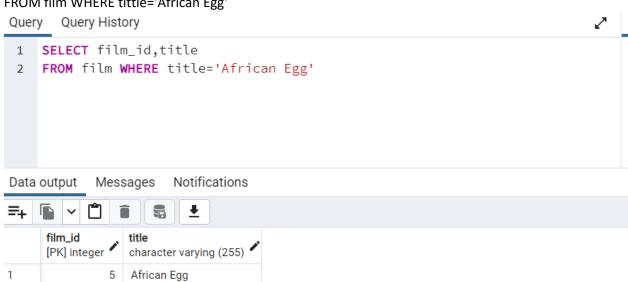**What do these constraints do exactly? Why are they important?**

Constraints are important as they ensure that values in columns are always formatted by preventing duplicates or values that are not required in a column. They can also help restrict values that don't meet a certain condition and many more.
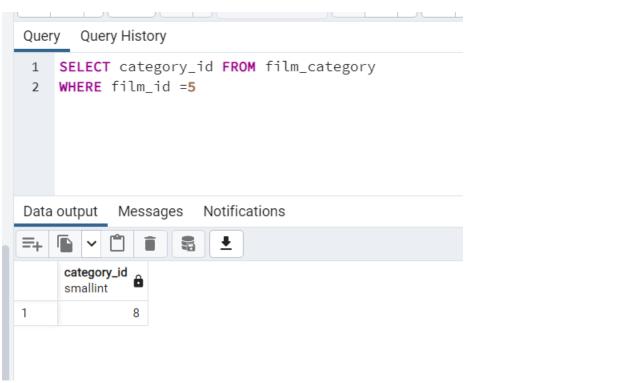
Step 3

--To find film_id

SELECT film_id, title

FROM film WHERE tittle='African Egg'

Query   Query History

```
1   SELECT film_id,title
2   FROM film WHERE title='African Egg'
```

Data output   Messages   Notifications

film_id [PK] integer	title character varying (255)
5	African Egg

--To find category_id

Query   Query History

```
1   SELECT category_id FROM film_category
2   WHERE film_id =5
```

Data output   Messages   Notifications

category_id smallint
8

--To update from category_id 8 to category_id 17 which is for the genre 'Thriller' I used the below query

UPDATE film_category
SET category_id =17
WHERE film_id =5

997	998	11	2006-02-15 10:07:09
998	999	3	2006-02-15 10:07:09
999	1000	5	2006-02-15 10:07:09
1000	5	17	2022-08-05 13:28:57.654893

Total rows: 1000 of 1000   Query complete 00:00:00 471

Step 4

DELETE FROM category
WHERE name ='Mystery'

Step5
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.

Bonus Task: Corrected

CREATE TABLE employees
 (
employee_id VARCHAR(30) NOT NULL,
name VARCHAR(50),
contact_number VARCHAR(30),
designation_id INT,
last_update TIMESTAMP NOT NULL DEFAULT now(),
CONSTRAINT employee_pkey PRIMARY KEY (employee_id)