3.3 SQL for Data Analyst Task

Step 1:

Your first task is to find out what film genres already 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

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: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	2022-10-07 13:21:44.762264
18	18	Crime	2022-10-07 13:21:44.762264
19	19	Mystery	2022-10-07 13:21:44.762264
20	20	Romance	2022-10-07 13:21:44.762264
21	21	War	2022-10-07 13:21:44.762264

- 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?

```
  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)
);
```

The constraints used within the category table are NOT NULL constraint and the PRIMARY KEY constraint. The NOT NULL constrain pertains to the category_id, name, last_update columns which will ensure that the there are not any areas with missing values. The PRIMARY KEY constraint is related to the category_id column which ensures that the data listed as category_id will be unique.

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*.

SELECT film_id FROM film WHERE title = 'African Egg'

	film_id [PK] integer
1	5

- 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.

UPDATE film_category SET category_id=17 WHERE film_id=5

	film_id [PK] smallint	category_id [PK] smallint	last_update timestamp without time zone
995	996	6	2006-02-15 10:07:09
996	997	12	2006-02-15 10:07:09
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	2022-10-07 13:41:41.594874

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 –

- Ability to input commands to easily identify, replace, and delete information
- Information is easier to manage and look up
- PostGre does not require you to scroll through massive amounts of data to find applicable columns
- You can pull information down to the specific rows and columns you would like to see

Cons –

- You cannot physically see the changes occurring as they are executed. You have to view the tables after you have executed the command to see change
- Commands/SQL language has to be learned and memorized for mastery

Step 6:
Save your "Answers 3.3" document as a PDF and upload it here for your tutor to review.