3.3 SQL for Data Analysts

Step 1

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

Data Output | Messages | Notifications

	name character varying (25)
1	Action
2	Animation
3	Children
4	Classics
5	Comedy
6	Documentary
7	Drama
8	Family
9	Foreign
10	Games
11	Horror
12	Music
13	New
14	Sci-Fi
15	Sports
16	Travel

Query | Query History

```
1  SELECT name
2  FROM category;
3
4
```

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

	category_id [PK] integer	name character varying (25)	last_update timestamp without time zone
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	2022-11-27 14:25:41.260276
18	18	Crime	2022-11-27 14:28:38.745929
19	19	Mystery	2022-11-27 14:28:38.745929
20	20	Romance	2022-11-27 14:28:38.745929
21	21	War	2022-11-27 14:28:38.745929

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)
);
```

category_id

NOT NULL DEFAULT: these constraints state that this variable cannot have null values and the results can be generated by default.

PRIMARY KEY: this constraint stablishes that the variable is the primary key (unique value that identifies each record of the table).

last_update
NOT NULL DEFAULT now (): again, these constraints indicate that the values cannot be null and the time has to be right in moment when the query is executed.

Step 3

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

Film ID

Query | Query History

```
1  SELECT *
2  FROM film
3  WHERE title = 'African Egg';
```

	film_id [PK] integer	title character varying (255)	description text
1	5	African Egg	A Fast-Paced Documentary of a Pastry Chef And a Dentist who must Pursue a Fo

Category ID

	category_id [PK] integer	name character varying (25)	last_up timest
14	14	Sci-Fi	2006-(
15	15	Sports	2006-(
16	16	Travel	2006-(
17	17	Thriller	2022-1

Query Query History

```
1  UPDATE film_category
2  SET category_id = 17
3  WHERE film_id = 5;
4
5  SELECT *
6  FROM film_category
7  WHERE film_id = 5;
8
```

	film_id [PK] smallint	category_id [PK] smallint	last_update timestamp without time zone
1	5	17	2022-11-27 15:06:04.085761

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.

1. **Identify mystery id**

	category_id [PK] integer	name character varying (25)	last_update timestamp without time zone
1	19	Mystery	2022-11-27 14:28:38.745929

Query Query History

```
1  DELETE FROM category
2  WHERE category_id =19;
```

	category_id [PK] integer	name character varying (25)	last_update timestamp without time zone
16	16	Travel	2006-02-15 09:46:27
17	17	Thriller	2022-11-27 14:25:41.260276
18	18	Crime	2022-11-27 14:28:38.745929
19	20	Romance	2022-11-27 14:28:38.745929
20	21	War	2022-11-27 14:28:38.745929

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.

Step 1.
To identify the genres, I would have opened the data table with the genre details.
Step 2. I would have had to add manually every genre, checking which is the following id number, add each genre and use the function now() for update the time.
Step 3. To identify the film_id in the fil table I would have opened the film table, filter by name and write down the film_id.
Later, I would have opened the film_category table and filter the film_id and finally write down manually new category_id for the film.
Step 4. Go to the category table, filter the name by Mystery and delete the row.

PROS

It is easier to interact with the tables, you don't have to be looking the values directly in the tables, just with code you can easily read, add or delete information.
Also, there are less chances to make mistakes such as data type, write in the wrong column and also you can see what you have done in the query editor.
It is faster to get to the result, sometimes excel very easily get stuck and with SQL you can get the result very fast.

CONST

It is more complex to do work wit code than just work directly in a spreadsheet. You need to understand how the functions work and structure your ideas to get to the expected result.