3.3: SQL for Data Analysts

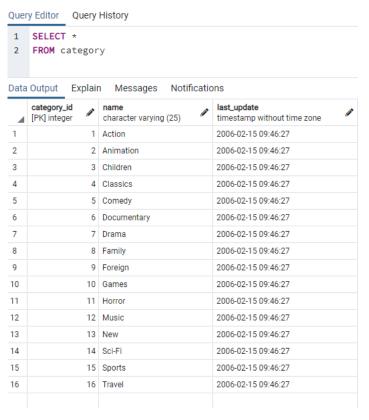
#### **DIRECTIONS:**

### STEP 1: 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.

SELECT \*
FROM category

> 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.



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')

4	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	2022-10-27 15:51:36.359744
18	18	Crime	2022-10-27 15:51:36.359744
19	19	Mystery	2022-10-27 15:51:36.359744
20	20	Romance	2022-10-27 15:51:36.359744
21	21	War	2022-10-27 15:51:36.359744

Grace Skelley on 10.27.2022

> 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)

);
```

**NOT NULL CONSTRAINTS** – This restriction prevents the value to be blank/empty.

PRIMARY KEY CONSTRAINTS - This prevents the value from being reproduced and duplicated. This is a unique identifier and each table must have one.

- The category\_id is an integer data type and the value can neither be null nor blank. This serves as a primary key for the category table.
- The name is a text data type that requires neither null nor blank value.
- The last\_update is a TIMESTAMP(6) WITH TIME ZONE data type that cannot be left null or empty in value.

### 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,title
FROM film
WHERE title='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.

SELECT category\_id FROM film\_category WHERE film\_id=5



UPDATE film\_category
SET category\_id=17
WHERE film\_id=5

Data O	utput Explain	Messages No	fications	
4	film_id [PK] smallint	category_id [PK] smallint	last_update timestamp without time zone	
981	982		1 2006-02-15 10:07:09	
982	983	1	2 2006-02-15 10:07:09	
983	984		9 2006-02-15 10:07:09	
984	985	1	4 2006-02-15 10:07:09	
985	986		2 2006-02-15 10:07:09	
986	987	1	2 2006-02-15 10:07:09	
987	988	1	6 2006-02-15 10:07:09	
988	989	1	6 2006-02-15 10:07:09	
989	990	1	1 2006-02-15 10:07:09	
990	991		1 2006-02-15 10:07:09	
991	992		6 2006-02-15 10:07:09	
992	993		3 2006-02-15 10:07:09	
993	994	1	3 2006-02-15 10:07:09	
994	995	1	1 2006-02-15 10:07:09	
995	996		6 2006-02-15 10:07:09	
996	997	1	2 2006-02-15 10:07:09	
997	998	1	1 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	1	7 2022-10-27 16:59:47.388349	

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'
Data Output Explain Messages Notifications

Data	Output Explair	n Messages Notificatio	ons	
4	category_id [PK] integer	name character varying (25)	last_update timestamp without time zone	
1	1	1 Action 2006-02-15 09:46		
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 Myst	tery = removed from the list
17	17	Thriller	2022-10-27 15:51:36.359744	
18	18	Crime	2022-10-27 15:51:36.359744	
19	20	Romance	2022-10-27 15:51:36.359744	
20	21	War	2022-10-27 15:51:36.359744	

3.3: SQL for Data Analysts Grace Skelley on 10.27.2022

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.

DATA TOOLS	PROS	CONS
SQL	<ul> <li>Easy database management (CRUD) for large-scale datasets</li> <li>Data Output can be generated in seconds</li> <li>Records of queries can be tracked through "Query History"</li> <li>Has high storage capacity for data updates, etc.</li> <li>Performs relational database management</li> </ul>	Must be familiar with SQL commands     System errors can occur and are hard to resolve at times
EXCEL	<ul> <li>User-friendly and doesn't require coding</li> <li>Search and data updates are easy to execute through the "Find and Replace" tabs</li> <li>File can be too large to maintain updates</li> </ul>	Database size is limited     Requires multiple steps to achieve the desired outcome (filtering, updates, etc.)     Susceptible to human error     Has visualization capabilities

3.3: SQL for Data Analysts

## **Bonus Task**

The SQL query below contains some typos. See if you can fix it based on what you've learned so far about SQL and data types; then try running it in pgAdmin 4. If the query works, copy it into your Answers 3.3 document.

If you get this you're a SQL champ!

```
CREATE TBL 3EMPLOYEES
{
employee_id VARINT(30) NOT EMPTY

name VARCHAR(50),
contact_number VARCHAR(30) ,
designation_id INT,
last_update TIMESTAMP NOT NULL DEF now()

CONSTRAIN employee_pkey PRIMARY KEY (employee_id)
}
```

# **CORRECT CODE**

```
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)

)
```