Task 3.3 SQL for Data Analysts

By Lee Heng Chuah

Your line manager has told you they'd like to improve the movie search filters on the Rockbuster app and website. After thinking about the best way to do this, you decide to add more film categories so users can browse more genres they are interested in.

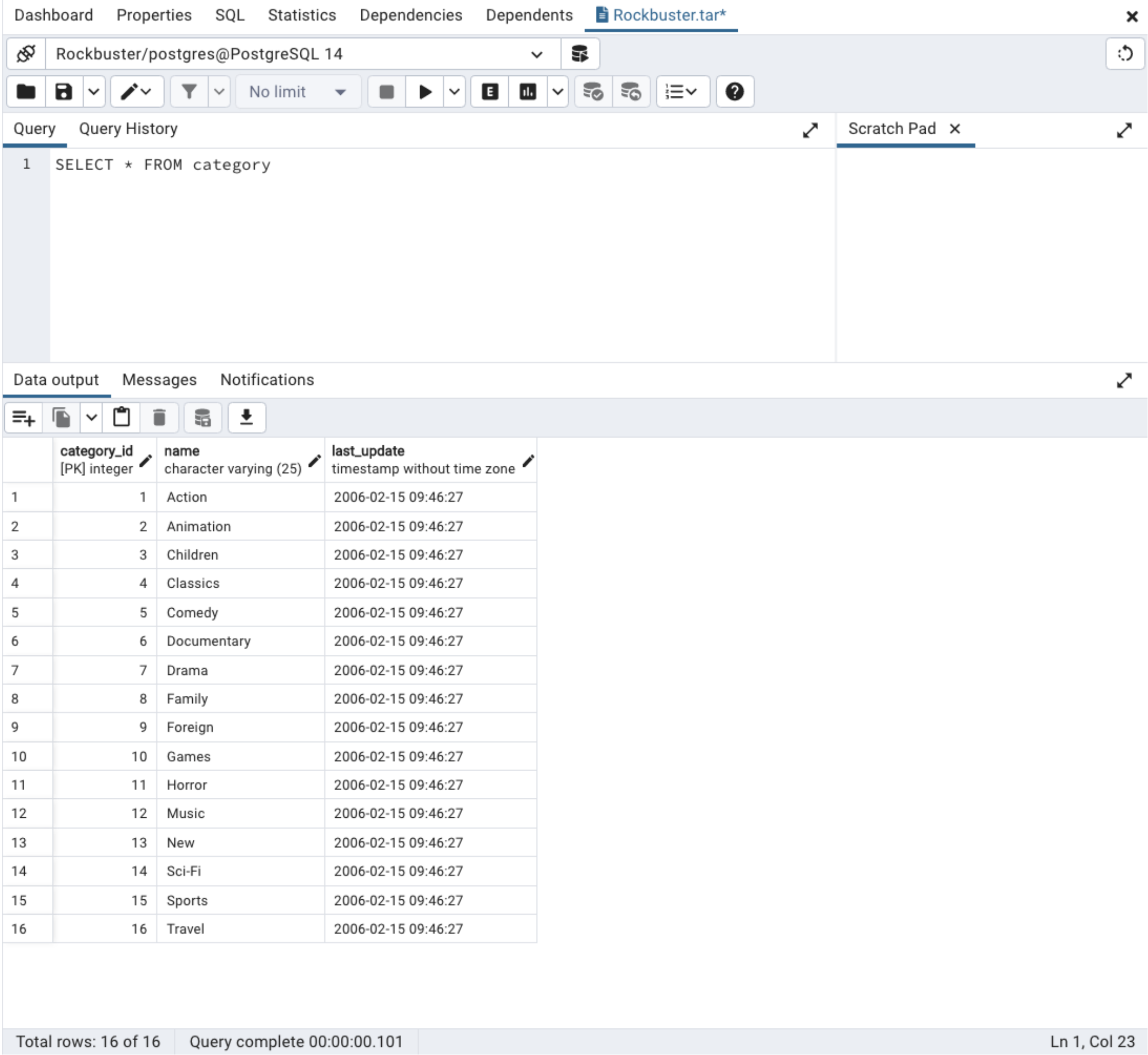
Directions:

Create a new text document and call it "Answers 3.3". You will submit your queries, outputs and written answers in this document at the end of the task. - Completed

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.
- Copy-paste the output into your answers document or write the answers out – it is up to you. Make sure to include the category ID for each genre.



Rockbuster/postgres@PostgreSQL 14

No limit

```
SELECT * FROM category
```

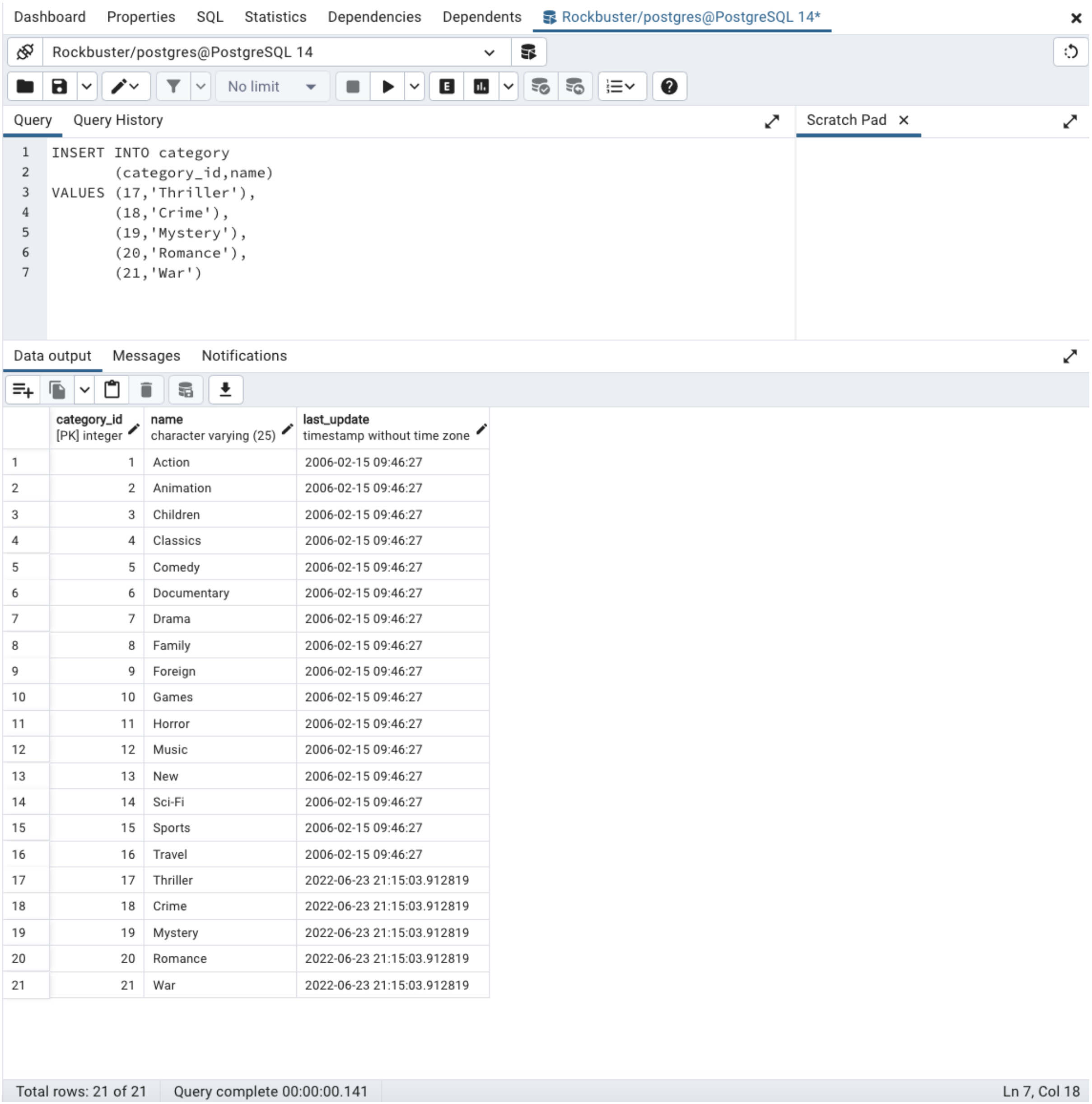
	category_id [PK] integer	name character varying (25)	last_update timestamp without time zone
1	1	Action	2006-02-15 09:46:27
2	2	Animation	2006-02-15 09:46:27
3	3	Children	2006-02-15 09:46:27
4	4	Classics	2006-02-15 09:46:27
5	5	Comedy	2006-02-15 09:46:27
6	6	Documentary	2006-02-15 09:46:27
7	7	Drama	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

Total rows: 16 of 16 Query complete 00:00:00.101 Ln 1, Col 23

Step 2

You are 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.



```sql
INSERT INTO category
       (category_id,name)
VALUES (17,'Thriller'),
       (18,'Crime'),
       (19,'Mystery'),
       (20,'Romance'),
       (21,'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	2022-06-23 21:15:03.912819
18	18	Crime	2022-06-23 21:15:03.912819
19	19	Mystery	2022-06-23 21:15:03.912819
20	20	Romance	2022-06-23 21:15:03.912819
21	21	War	2022-06-23 21:15:03.912819

Total rows: 21 of 21   Query complete 00:00:00.141   Ln 7, Col 18

- 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?

```sql
  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)
);
```

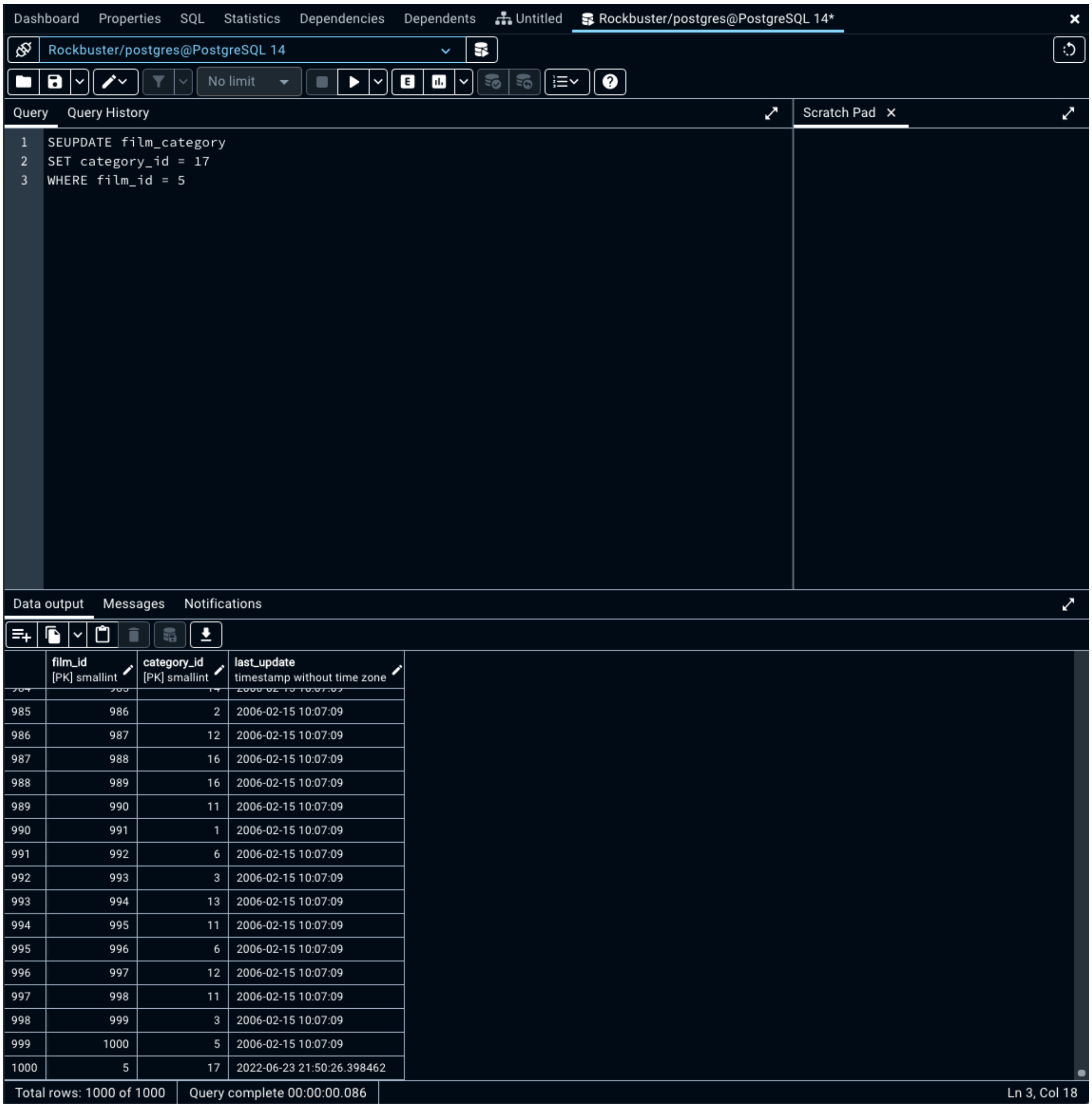
NOT NULL Constraint: This ensures that the category_id, an integer, that can't have any empty or missing values. This record value use the DEFAULT command to assign a default value to the affected field. Similar to the name (text) and last_update (timestamp with time zone) that both columns can't have any empty or missing values.

PRIMARY KEY Constraint: The primary key gives each record in a table a unique ID. The primary key column can't contain any null or duplicate values. PRIMARY KEY constraints set the category_id column as the primary key.

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.



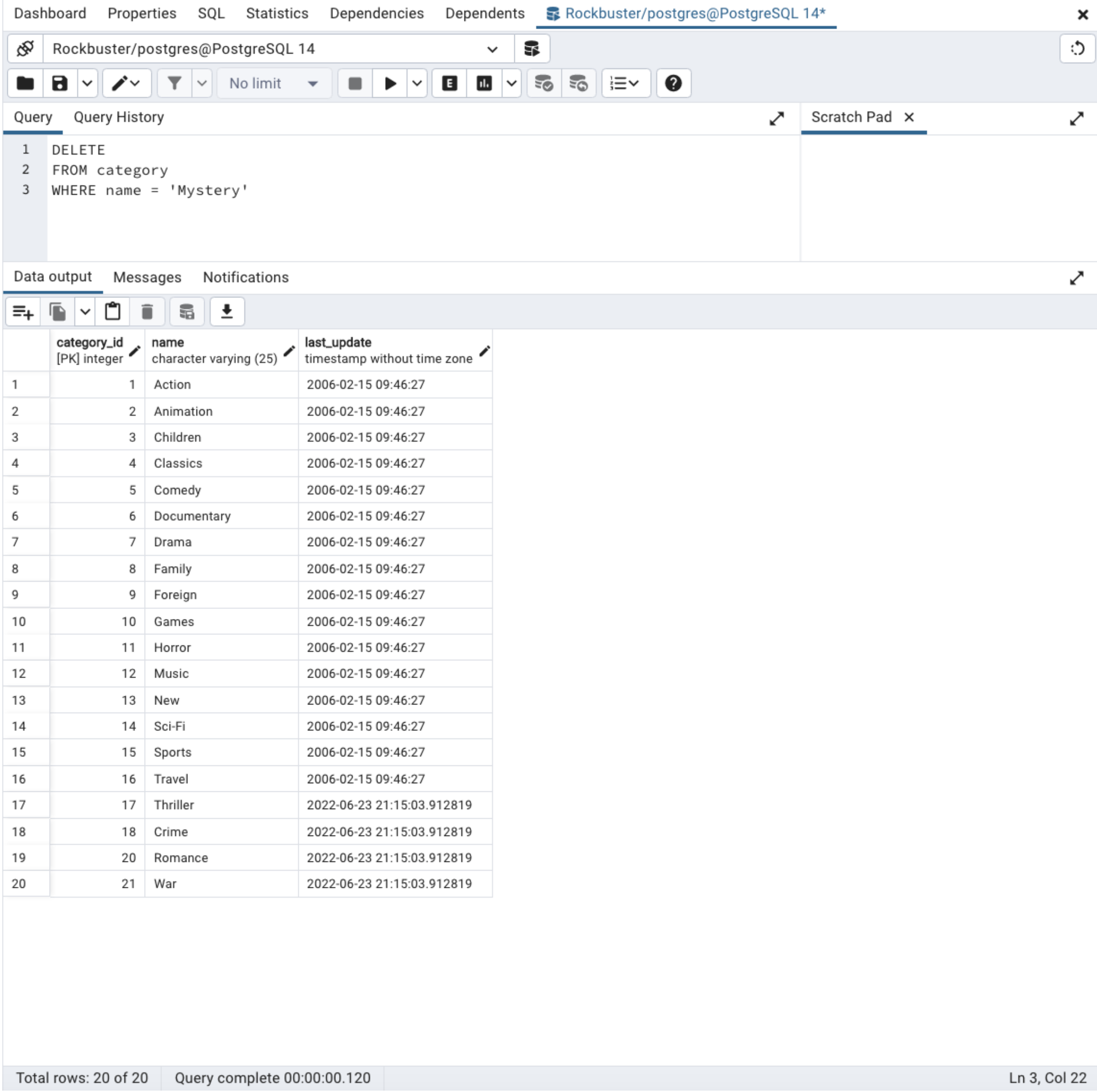
```
SEUPDATE film_category
SET category_id = 17
WHERE film_id = 5
```

	film_id [PK] smallint	category_id [PK] smallint	last_update timestamp without time zone
984	985	14	2006-02-15 10:07:09
985	986	2	2006-02-15 10:07:09
986	987	12	2006-02-15 10:07:09
987	988	16	2006-02-15 10:07:09
988	989	16	2006-02-15 10:07:09
989	990	11	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	13	2006-02-15 10:07:09
994	995	11	2006-02-15 10:07:09
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	2022-06-23 21:50:26.398462

```
Total rows: 1000 of 1000    Query complete 00:00:00.086    Ln 3, Col 18
```

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'
```

	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	2022-06-23 21:15:03.912819
18	18	Crime	2022-06-23 21:15:03.912819
19	20	Romance	2022-06-23 21:15:03.912819
20	21	War	2022-06-23 21:15:03.912819

Total rows: 20 of 20 Query complete 00:00:00.120 Ln 3, Col 22

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.

SQL is faster, easier, and safer than Excel. It can take minutes in SQL to find the data, but Excel could take an hour. Excel can technically handle one million rows, but that's before the pivot tables, multiple tabs and functions we are probably using. SQL stores the data separately from our analysis. In summary, Excel is excellent, but it can be inadequate for tasks requiring heavier lifting. SQL can complete most of those tasks faster and is more user-friendly from start to finish.

Step 6

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