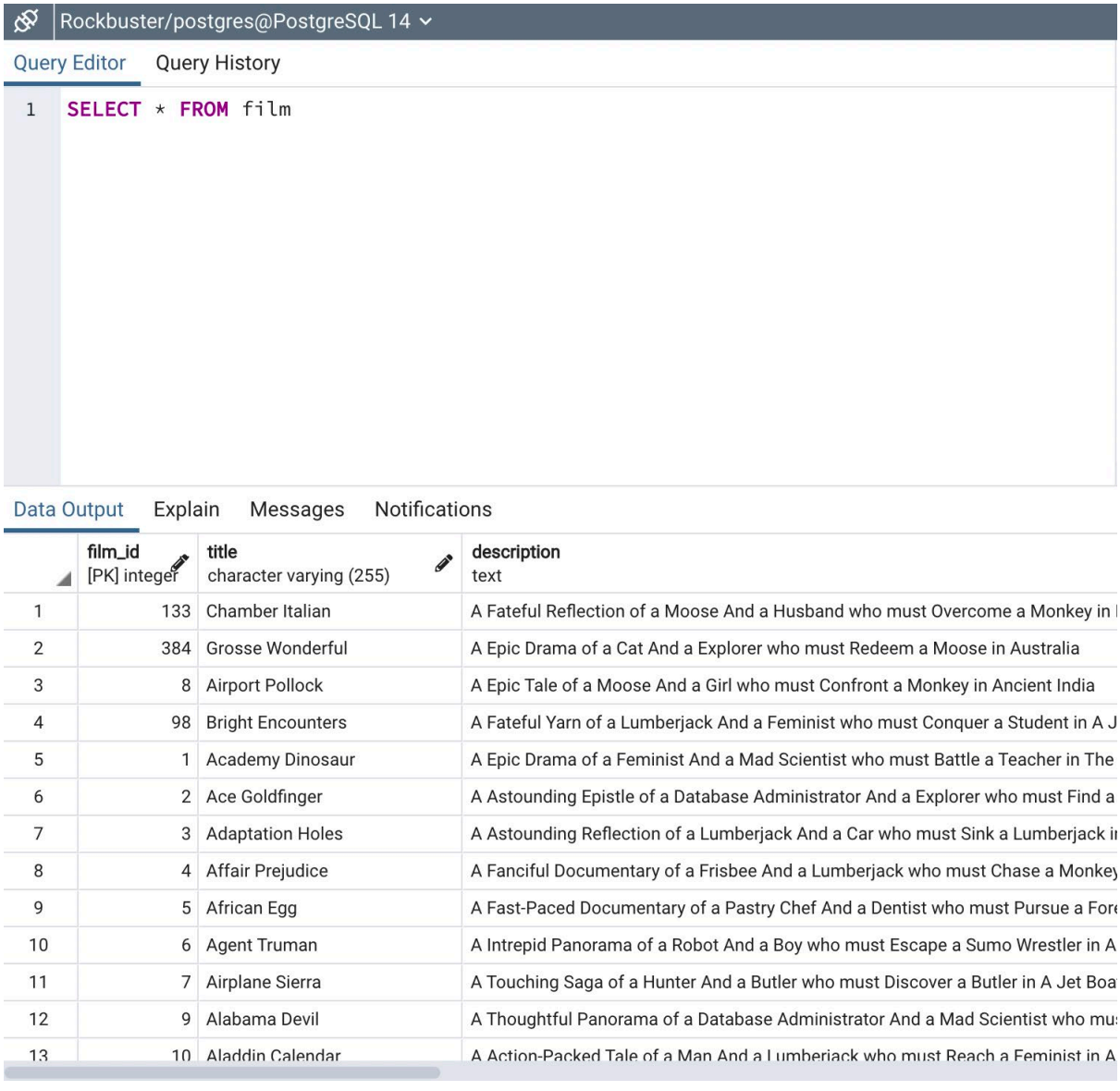
Juan Ignacio Galvalisi

Exercise 3.4: Database Querying in SQL

Refining Your Query

You need to get some data from the "film" table and decide to use the query SELECT * FROM film.



	film_id [PK] integer	title character varying (255)	description text
1	133	Chamber Italian	A Fateful Reflection of a Moose And a Husband who must Overcome a Monkey in
2	384	Grosse Wonderful	A Epic Drama of a Cat And a Explorer who must Redeem a Moose in Australia
3	8	Airport Pollock	A Epic Tale of a Moose And a Girl who must Confront a Monkey in Ancient India
4	98	Bright Encounters	A Fateful Yarn of a Lumberjack And a Feminist who must Conquer a Student in A J
5	1	Academy Dinosaur	A Epic Drama of a Feminist And a Mad Scientist who must Battle a Teacher in The
6	2	Ace Goldfinger	A Astounding Epistle of a Database Administrator And a Explorer who must Find a
7	3	Adaptation Holes	A Astounding Reflection of a Lumberjack And a Car who must Sink a Lumberjack i
8	4	Affair Prejudice	A Fanciful Documentary of a Frisbee And a Lumberjack who must Chase a Monkey
9	5	African Egg	A Fast-Paced Documentary of a Pastry Chef And a Dentist who must Pursue a Fore
10	6	Agent Truman	A Intrepid Panorama of a Robot And a Boy who must Escape a Sumo Wrestler in A
11	7	Airplane Sierra	A Touching Saga of a Hunter And a Butler who must Discover a Butler in A Jet Boa
12	9	Alabama Devil	A Thoughtful Panorama of a Database Administrator And a Mad Scientist who mu
13	10	Aladdin Calendar	A Action-Packed Tale of a Man And a Lumberjack who must Reach a Feminist in A

You realize that only the "film_id" and "title" columns are needed. Write a new query that selects only those 2 columns.

Rockbuster/postgres@PostgreSQL 14

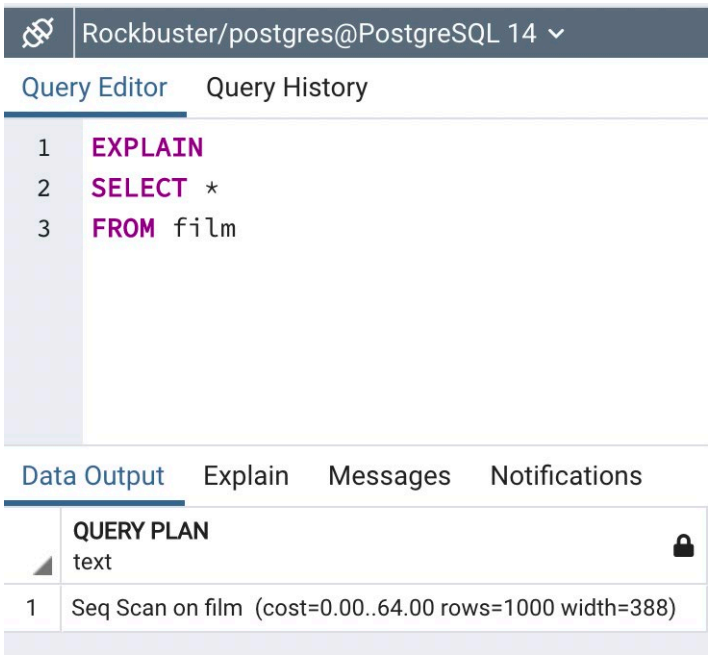
Query Editor | Query History

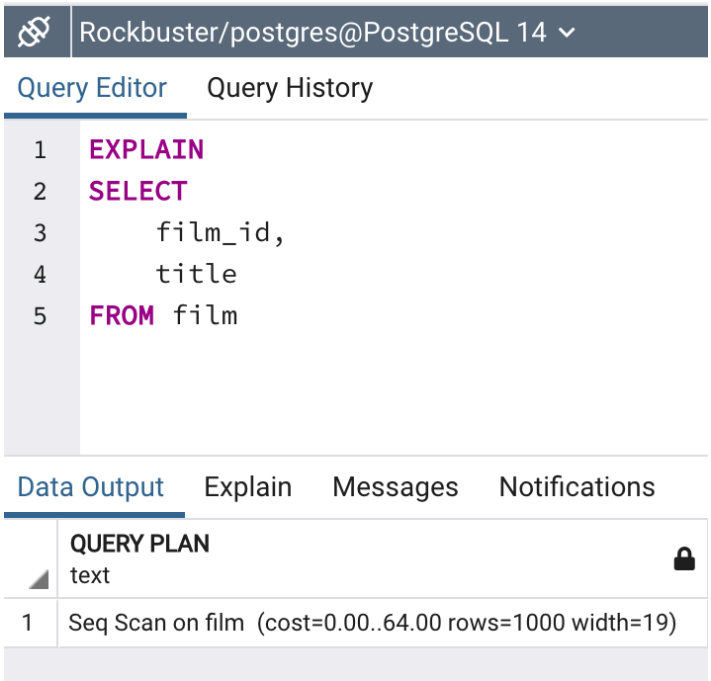
```
1  SELECT
2      film_id,
3      title
4  FROM film
```

Data Output | Explain | Messages | Notifications

	film_id [PK] integer	title character varying (255)
1	133	Chamber Italian
2	384	Grosse Wonderful
3	8	Airport Pollock
4	98	Bright Encounters
5	1	Academy Dinosaur
6	2	Ace Goldfinger
7	3	Adaptation Holes
8	4	Affair Prejudice
9	5	African Egg
10	6	Agent Truman
11	7	Airplane Sierra
12	9	Alabama Devil
13	10	Aladdin Calendar
14	11	Alamo Videotape
15	12	Alaska Phantom
16	213	Date Speed
17	13	Ali Forever

Compare the cost of the original query and the revised query, and write a few sentences explaining the comparison. Can you suggest any ways to optimize this query?





At a glance, both queries have the same cost of 0.00-64.00. Nevertheless, there is a difference when you want to retrieve each one. The second one was faster using the filters than not using them: the first query took 69 milliseconds while the second took 59 milliseconds. In other words, we are more efficient when using filters than not using them because you are not retrieving all the data.

On the other hand, using the LIMIT function could be helpful in order to optimize our query, being even more specific.

Ordering the Data

In the pgAdmin Query Tool, run a query that selects every film from the "film" table, with the movies sorted by title from A to Z, then by most recent release year, and then by highest to lowest rental rate.

Rockbuster/postgres@PostgreSQL 14

Query Editor | Query History

```
1  SELECT *
2  FROM film
3  ORDER BY
4      title,
5      release_year DESC,
6      rental_rate DESC
```

Data Output | Explain | Messages | Notifications

	film_id [PK] integer	title character varying (2	description text	release_year integer	language_id smallint	rental_duration smallint	rental_rate numeric (4,2)	length smallint
1	1	Academy Dinosaur	A Epic Dra...	2006	1	6	0.99	86
2	2	Ace Goldfinger	A Astoundi...	2006	1	3	4.99	48
3	3	Adaptation Holes	A Astoundi...	2006	1	7	2.99	50
4	4	Affair Prejudice	A Fanciful D...	2006	1	5	2.99	117
5	5	African Egg	A Fast-Pace...	2006	1	6	2.99	130
6	6	Agent Truman	A Intrepid P...	2006	1	3	2.99	169
7	7	Airplane Sierra	A Touching ...	2006	1	6	4.99	62
8	8	Airport Pollock	A Epic Tale ...	2006	1	6	4.99	54
9	9	Alabama Devil	A Thoughtf...	2006	1	3	2.99	114
10	10	Aladdin Calendar	A Action-Pa...	2006	1	6	4.99	63
11	11	Alamo Videotape	A Boring Ep...	2006	1	6	0.99	126
12	12	Alaska Phantom	A Fanciful S...	2006	1	6	0.99	136
13	13	Ali Forever	A Action-Pa...	2006	1	4	4.99	150
14	14	Alice Fantasia	A Emotiona...	2006	1	6	0.99	94
15	15	Alien Center	A Brilliant D...	2006	1	5	2.99	46
16	16	Alley Evolution	A Fast-Pace...	2006	1	6	2.99	180
17	17	Alone Trip	A Fast-Pace	2006	1	3	0.99	82

Extract the data output of your query into a csv file for the film collection department to analyze in Excel. (You may need to explore how to save your output as a csv file in the Query Tool.)

3.4 - Ordering the Data.cvs

Grouping Data

The strategy department has asked you the questions below. Write a SQL query to retrieve the correct answers, then extract your results as a csv file.

What is the average rental rate for each rating category?

Rockbuster/postgres@PostgreSQL 14

Query Editor | Query History

```
1  SELECT rating AS rating_category,
2      AVG(rental_rate) AS average_rental_rate
3  FROM film
4  GROUP BY rating
```

Data Output | Explain | Messages | Notifications

	rating_category mpaa_rating	average_rental_rate numeric
1	R	2.9387179487179487
2	NC-17	2.970952380952381
3	G	2.888876404494382
4	PG	3.0518556701030928
5	PG-13	3.034843049327354

What are the minimum and maximum rental durations for each rating category?

Rockbuster/postgres@PostgreSQL 14

Query Editor | Query History

```
1  SELECT rating AS rating_category,
2      MIN(rental_duration) AS minimum_rental_rate,
3      MAX(rental_duration) AS maximum_rental_rate
4  FROM film
5  GROUP BY rating
```

Data Output | Explain | Messages | Notifications

	rating_category mpaa_rating	minimum_rental_rate smallint	maximum_rental_rate smallint
1	R	3	7
2	NC-17	3	7
3	G	3	7
4	PG	3	7
5	PG-13	3	7

3.4 - Grouping Data.cvs

Database Migration

Your team has decided to use an external tool to collect data on user behavior in the new Rockbuster Android app. Data collected from this new source will need to be loaded into the data warehouse before you can analyze it.

Can you outline the procedure for migrating the data and who will be responsible for it?

The procedure for migrating the data from one database to another is the ETL process (Extract, Transform, and Load). In general terms, we part from a stage where our data

is not optimized for analytics to a central host optimized and standardized, called the Data Warehouse system.

The data engineering team mostly handles data migration, but analysts also need to know how the process functions in order to work together successfully.

What problems do you foresee if you start analyzing the data before it's been loaded into the data warehouse?

In that case, it could be the same situation when you are working with, for example, dirty data. Errors are much more likely if the data has not been properly transformed before the migration process. You could handle duplicates, different names, blank values, and misspellings, issues we want to address in the transformation process during data migration. Suppose we start analyzing the data before it's loaded into the data warehouse. In that case, the ETL process will have been in vain, as the main reason for data migration is to provide cleaner and more reliable data.

Bonus Task

What are the minimum and the maximum replacement costs for each rating category ordered by rating as follows: G, PG, PG-13, R, NC-17?

Rockbuster/postgres@PostgreSQL 14

Query Editor | Query History

```
1  SELECT rating AS rating_category,
2      MIN(replacement_cost) AS minimum_replacement_cost,
3      MAX(replacement_cost) AS maximum_replacement_cost
4  FROM film
5  GROUP BY rating
6  ORDER BY CASE WHEN rating = 'G' THEN '1'
7                WHEN rating = 'PG' THEN '2'
8                WHEN rating = 'PG-13' THEN '3'
9                WHEN rating = 'R' THEN '4'
10               WHEN rating = 'NC-17' THEN '5'
11          END
```

Data Output | Explain | Messages | Notifications

	rating_category mpaa_rating	minimum_replacement_cost numeric	maximum_replacement_cost numeric
1	G	9.99	29.99
2	PG	9.99	29.99
3	PG-13	9.99	29.99
4	R	9.99	29.99
5	NC-17	9.99	29.99