Gwyn Reniers

Exercise 3.4

1. **Refining Your Query:** You need to get some data from the “film” table and decide to use the query SELECT \* FROM film.
   * You realize that only the “film\_id” and “title” columns are needed. Write a new query that selects only those 2 columns.

Graphical user interface, application

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

Graphical user interface, text

Description automatically generated

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?

*In this instance it appears that both queries returned the same cost, although there were less rows in the more specific query so it was returned faster.. Some suggestions to optimize even more would be to consider the data you need, and tailor the query around that. Do you need certain titles that start with certain letters, do a BETWEEN query, or how many titles start with N, limit it that way. The more specific the query, the less cost to the company and the less time it will take.*

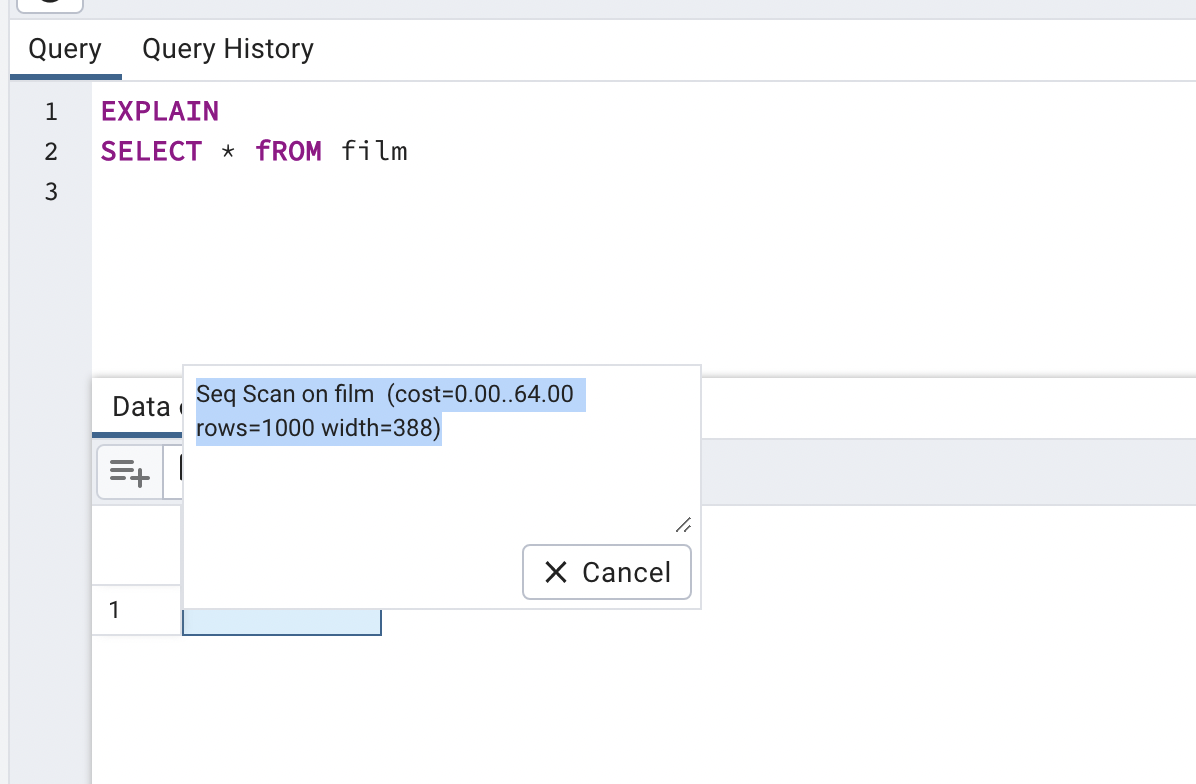
1. **Ordering the Data:**
   * In the pgAdmin Query Tool, run a query that selects every film from the “film” table, with the movies sorted by title from A to Z, then by most recent release year, and then by highest to lowest rental rate.
   * Graphical user interface

     Description automatically generated

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.)

<https://drive.google.com/file/d/1uUsfPGg-L5MIYz4Kul3l2Xtc4BRq8Aj-/view?usp=sharing>

*The min/max and average are on the last few pages of this doc- since it uploaded all the film info beginning page 35 I believe.*



Graphical user interface, application, Word

Description automatically generated

1. **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?

|  |  |
| --- | --- |
| **rating** | **avg** |
| R | 4.774358974 |
| NC-17 | 5.142857143 |
| G | 4.837078652 |
| PG | 5.082474227 |
| PG-13 | 5.053811659 |
|  |  |

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

|  |  |
| --- | --- |
| **Rating** | **min** |
| R | 3 |
| NC-17 | 3 |
| G | 3 |
| PG | 3 |
| PG-13 | 3 |

|  |  |
| --- | --- |
| **rating** | **max** |
| R | 7 |
| NC-17 | 7 |
| G | 7 |
| PG | 7 |
| PG-13 | 7 |

1. **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?

*Step 1: Extract the data, connect to the source, select and collect the necessary data needed.*

*Step 2: Transform the data- apply rules and conditions to the extracted data to convert it to the standard format of the target database*

*Step 3: Load the data: Involves importing extracted and transformed data into the target database/data warehouse.*

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

*Data may not be in usable form, or it may be incomplete leading to inaccuracies.*

**Bonus Task**

You’ve not yet covered custom sorting; however, let’s imagine you’ve found the two resources below that explain it. Read each one, then try to write a query to answer the following question: 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?

Graphical user interface

Description automatically generated

*WOOOOHOOOO!!! I did it!! ☺ lol*