# Answers 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.**

SELECT film\_id, title FROM film

* + **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?**

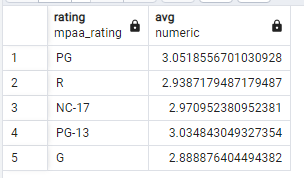
EXPLAIN SELECT film\_id, title FROM film = "Seq Scan on film (cost=0.00..64.00 rows=1000 width=19)"

EXPLAIN SELECT \* FROM film = "Seq Scan on film (cost=0.00..64.00 rows=1000 width=388)"

Both of these queries cost the same = 0

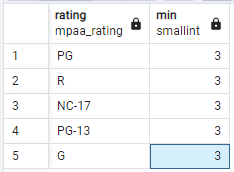
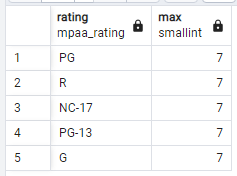
1. **Ordering the Data:**
   * **In the pgAdmin Query Tool, run a query that selects every film from the “film” table, with the movies sorted by title from A to Z, then by most recent release year, and then by highest to lowest rental rate.**
     + By Title from A to Z = SELECT title FROM film ORDER BY title ASC
     + By most recent year = SELECT title, release\_year FROM film ORDER BY release\_year ASC. (All movies were released in 2006)
     + By Highest to lowest rental rate = SELECT title, rental\_rate FROM film ORDER BY rental\_rate DESC
   * **Extract the data output of your query into a CSV file for the film collection department to analyze in Excel. To do this, click the button “Save results to file”:**
2. **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?**

SELECT rating, AVG(rental\_rate) FROM film GROUP BY rating



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

SELECT rating, MAX(rental\_duration) FROM film GROUP BY rating

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?**

First, the data needs to be extracted from the app, then it must be transformed into a generalized format to be used, and lastly it must be loaded into the new database for this data.   
This is the data engineer’s job.

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

The data may be in an improper format or there may be multiple formats making it difficult or impossible to analyze.