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.

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?

a.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

b. It is more expensive to query all the columns, rather than just the ones actually needed. 384 is a lot more than 19. I would recommend just querying the columns absolutely needed.

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.

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”:

Graphical user interface, text

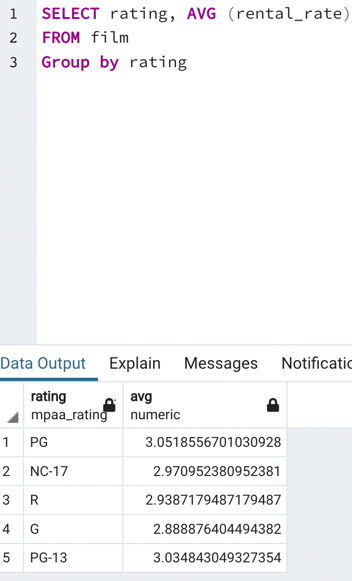
Description automatically generated

Table

Description automatically generated

b. link: <https://drive.google.com/file/d/1_K0MuehEdqVeZ6LodbcvKmPOzZZjiFR4/view?usp=sharing>

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.
   1. What is the average rental rate for each rating category?
   2. What are the minimum and maximum rental durations for each rating category?



<https://drive.google.com/file/d/1Be7FkTGAOkMdYXzr79dCgsTBfNT_admw/view?usp=sharing>

Table

Description automatically generated

<https://drive.google.com/file/d/1GYYZG8HUHKrBBSxxufJA5MzGaBiOt7yo/view?usp=sharing>

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.
   1. Can you outline the procedure for migrating the data and who will be responsible for it?
   2. What problems do you foresee if you start analyzing the data before it’s been loaded into the data warehouse?
2. The procedure for migrating the data from an outside source is called ETL (extract, transform, load). Usually ETL is a data engineer’s job to perform. ETL consists of extracting the data from the outside source, transforming the data to be compatible with the source you are seeking to combine it with (cleaning, organizing ect.), and loading it to the data source of choice to combine them successfully.

b. Problems can arise if data is analyzed before this process is complete. Some data may not be compatible or cleaned up properly and could give the wrong preemptive insights, this could lead to large errors in your final analysis. The data could also be structured incorrectly which could lead to errors in your ability to query the information correctly.