3.4 Database querying

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?
        - The cost of both queries is the same, 0 for returning the first row and 64 to return all rows. Reducing the number of total rows to be returned by the query with LIMIT would help reduce the cost, this however depends on what the data is needed for.
2. **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.

```
SELECT * FROM film
ORDER BY title, release_year DESC, rental_rate DESC
```

    - Extract the data output of your query into a csv file for the film collection department to analyze in Excel.
3. **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
```

rating	avg
R	2.938718
NC-17	2.970952
G	2.888876
PG	3.051856
PG-13	3.034843

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

```
SELECT rating, MIN(rental_duration), MAX(rental_duration)
FROM film
GROUP BY rating
```

rating	min	max
R	3	7
NC-17	3	7
G	3	7
PG	3	7
PG-13	3	7

4. **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?
        - Data Engineers will collect data from multiple sources, transforming it where necessary with calculations before loading it to a new database.

- What problems do you foresee if you start analyzing the data before it’s been loaded into the data warehouse?
    - Data may not be correctly structured if used raw and will not be available to query.