# Ramon Magallan

Data Immersion 3.4

SQL for Data Analysts

**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.**

**A screenshot of a computer

AI-generated content may be incorrect.**

* + **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?**

When comparing these 2 queries, I get the same cost of 0-98. With the intention of seeing all film\_id and titles, I can not think of a way to optimize this query further. If one intends to produce a more concise table, one could employ a LIMIT or GROUP BY function.

**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.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**3. Grouping Data:**

* **What is the average rental rate for each rating category?**

**A screenshot of a computer

AI-generated content may be incorrect.**

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

**A screenshot of a computer

AI-generated content may be incorrect.**

**3. Database Migration:**

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

The Rockbuster Android app data will go through the standard ETL (**Extract**, **Transform**, **Load**) process. First, the data will be **extracted** from the source systems where it is initially collected. Next, it will go through the **transformation** phase, where it will be reformatted, cleaned, and enriched — this may include converting data types, standardizing fields, and adding calculated metrics that are important for analysis. Finally, the data will be **loaded** into the Rockbuster Data Warehouse, where it can be accessed and analyzed alongside other business data.

The Data Engineering team will be primarily responsible for managing and executing this ETL process, ensuring that the data is properly structured, reliable, and ready for use by analysts and stakeholders.

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

If we started to analyze the data before it’s been loaded in the data warehouse, many issues may come to fruition. For one we would not have a full scope of the data as it applies to our entire database. This would limit our insights received while exploring questions like, “Which films have the highest rental rate?” Without all relevant data, patterns and trends could be misinterpreted or completely missed. Additionally, incomplete or inconsistent data might lead to errors in reports or dashboards, which could misguide business decisions. Overall, jumping into analysis too early compromises the integrity and reliability of the insights we hope to gain.