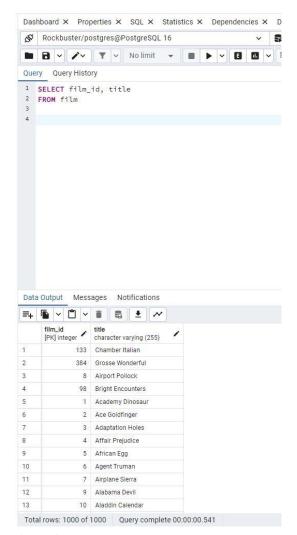# Data Immersion 3.4

## 1.Refining my query



## Cost Comparison

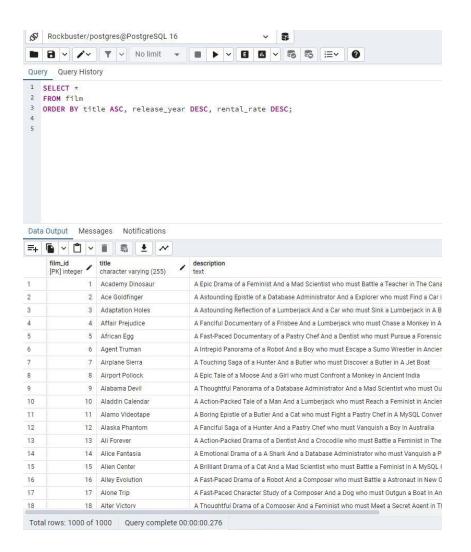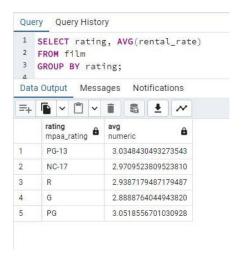**Original Query**                                    **Revised Query**

The cost associated with both queries is identical because each one performs a sequential scan over the entire film table, processing the same total number of records (1000 rows). However, the 'width'—or the average amount of data retrieved per row—differs due to the first query fetching all columns with **SELECT \***, while the second query only fetches two columns. Although the width is different, this doesn't affect the total cost because it's the number of rows scanned that primarily determines it. To make the queries faster, you can add a **LIMIT** clause to get just a few records instead of the whole table. This way, the database doesn't have to go through every single row, and you can get the results you need more quickly.
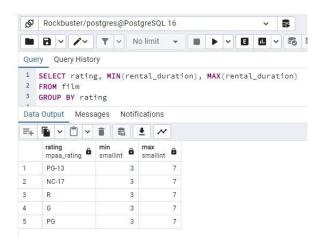
**2.**

**3.**

**Average rental rate for each rental category**



**Minimum and maximum rental durations for each rating category**



4. Database Migration

1. Extraction: Collecting the data from the source, which in this case is the new Rockbuster Android app.

2. Transformation: Converting the raw data into a usable format, cleaning it, dealing with missing or duplicate values, ensuring consistency, and potentially enriching the data by adding new calculated fields.

3. Loading: Inserting the cleaned and transformed data into the data warehouse where it can be accessed for analysis.

If analysis is attempted before the data is properly integrated into the data warehouse, the following issues may arise:

**Inconsistent Values:** Data from the app may not match the format or standards of the existing data, leading to unreliable analysis results.

**Missing Values:** Important pieces of data may not be captured correctly during extraction, which could lead to gaps in the analysis.

**Duplicate Values:** Without proper deduplication during the transformation phase, the analysis might count things twice, skewing any metrics or insights.

**Data Integrity:** Analysis based on raw data can lead to incorrect conclusions due to outliers or errors that have not been cleaned.