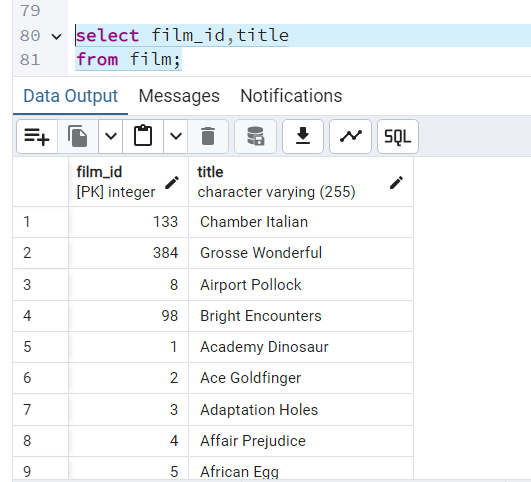
**Task 3.4**

**Refining Query**

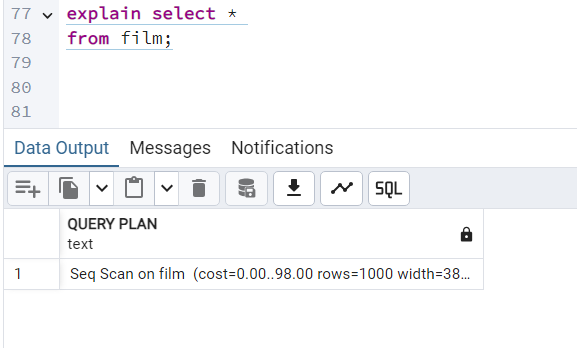
**1.a select film\_id,title**

**from film;**

****

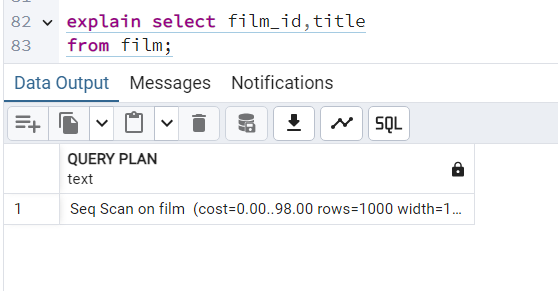
**explain select \***

**from film;**

****

**1.b explain select film\_id,title**

**from film;**

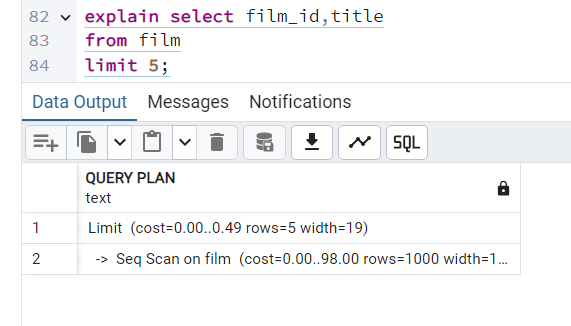
****

**Original query: Seq Scan on film (cost=0.00..98.00 rows=1000 width=384)**

**Revised query: Seq Scan on film (cost=0.00..98.00 rows=1000 width=19)**

**The revised query is 20 times faster because it focuses the command on the required data only. Optimization suggestion is to limit the results to the top 5, as expanding to the top 10 does not result in any improvements.**

**Optimized query: Limit (cost=0.00..0.49 rows=5 width=19)**

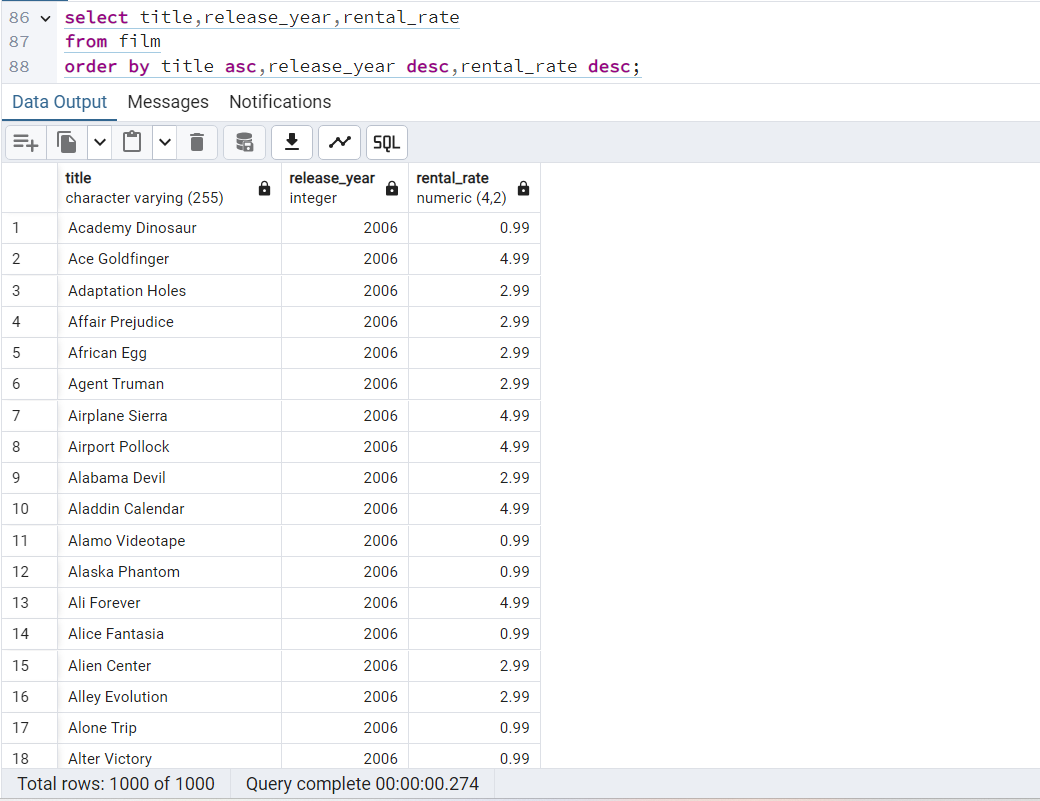
****

**Ordering the Data**

**2.a select title,release\_year,rental\_rate**

**from film**

**order by title asc,release\_year desc,rental\_rate desc;**

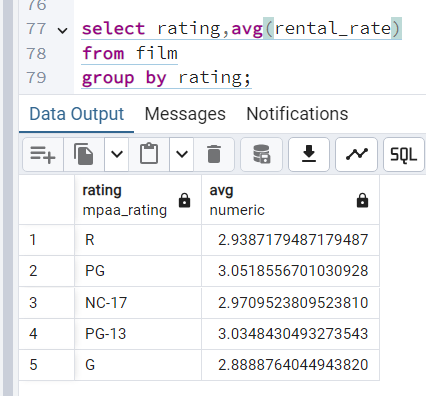
****

**Grouping Data**

**3.a select rating,avg(rental\_rate)**

**from film**

**group by rating;**

****

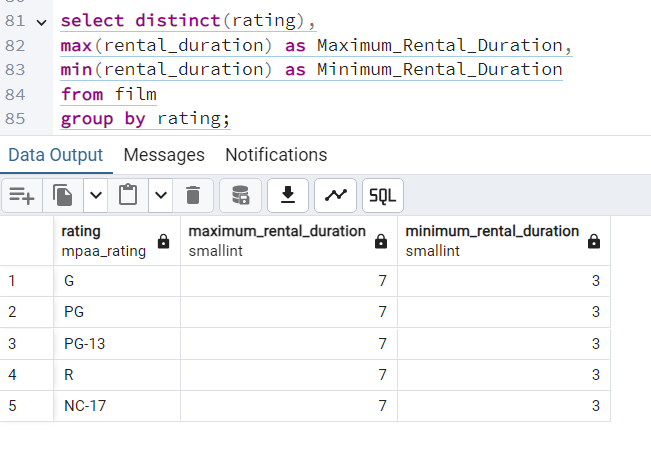
**3.b select distinct(rating),**

**max(rental\_duration) as Maximum\_Rental\_Duration,**

**min(rental\_duration) as Minimum\_Rental\_Duration**

**from film**

**group by rating;**

****

4. **Database Migration**

**Procedure for migrating the data and who will be responsible for it: Data Engineer**

**The procedure for the migration of data to the data warehouse is as follows:**

1.The new data on user behavior in the new Rockbuster Android app have to be extracted and collected.

2. The extracted data will then need to be transformed into a suitable format so that it is align with the format and requirement needed to load into the data warehouse.

3. The transformed data will then be loaded into the data warehouse. This ETL (Extract, Transform, Load) process is the job of data engineer. However, data analyst can also help with the coordination and handing of the timeline for the migration process as well as to check on the data completeness of the data at the different stages of the data migration process.

**Problems with analyzing data before it’s been loaded into the data warehouse:**

If we analyze the data before it is being loaded into the data warehouse, we might encounter many issues with the data such as inconsistency of the values, missing values, duplicate values. These will all cause inaccurate analysis of the data. More time and effort will end up being spent on cleaning the data before it can be usable for analysis, which might impact the project timeline. Another problem will be that analyzing the raw data will take more time then analyzing the data from a data warehouse where we can easily use a query to extract out the required data need for the analysis rather than go through the whole data set manually