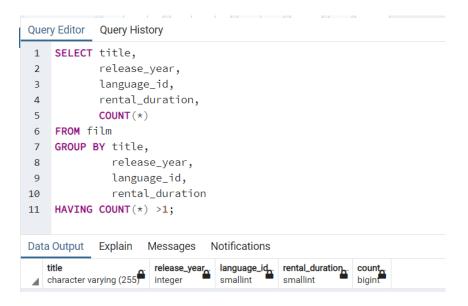
3.6: Summarizing & Cleaning Data in SQL Prepared by: Rob Rowland

Film Table Duplicate Data:

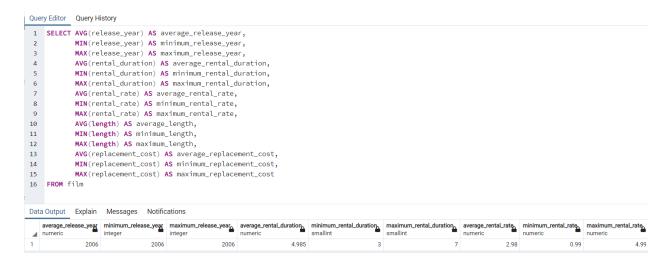


Customer Table Duplicate Data:

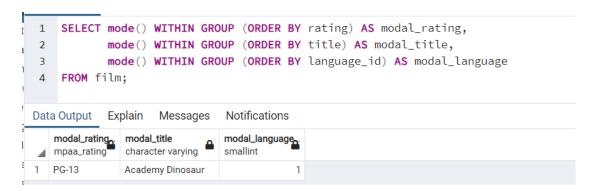


If either of these tables had duplicate data, I could either a) create a view with unique records using GROUP BY or DISTINCT or b) delete the duplicate record using DELETE. It would likely be most wise to create the view with unique records in case any other users needed to access the duplicate rows for any reason. Permanently deleting a record should be done with care.

Film Table, Numerical Columns (columns continue to right, not seen on picture):



Film Table, Non-Numerical Columns:



Customer Table, Numerical and Non=Numerical Columns (columns continue to right, not seen on picture):



I remain more familiar with Excel currently, so I can conduct the data profiling much quicker that way. As data sets get larger, I can certainly see the merit for using SQL. I think that, as I learn and become fluent in SQL, it will be the preferred option for quickly referencing data points from large data sets. However, when trying to build a larger profile involving multiple columns, Excel seems easier to use at this time.