**3.4: Database Querying in SQL**
**Prepared by: Rob Rowland**

**Question 1:**

SELECT * FROM film

| | QUERY PLAN text |
|---|---|
| 1 | Seq Scan on film (cost=0.00..64.00 rows=1000 width=384) |

SELECT film_id,
title
FROM film

| | QUERY PLAN text |
|---|---|
| 1 | Seq Scan on film (cost=0.00..64.00 rows=1000 width=19) |

Both queries have a cost of 64 for all rows, meaning that there is not a large difference. The second query eliminates unneeded columns, which helps with readability and unneeded information, but does not appear to reduce the cost of the query. Further optimization is not needed at this point, until we have more information on the purpose of the query.

**Question 2:**

**Query by Title A to Z** (also the file when using ORDER BY for all three variables in the same SQL)
**Query by Release Date**
**Query by Rental Rate**

**Question 3:**

**Query by Avg Rental Rate by Rating Category**
**Query by Min and Max Rental Duration by Rating Category**`

**Question 4:**

Rockbuster will use the Extract, Transform, and Load procedure to pull data from the Rockbuster Android application and make it usable data.

- Extract: Data will be pulled from the app on user behavior (login frequency, duration, time of day, etc.) by data scientists.
- Transform: The data is converted into the appropriate format based on the database it will eventually live in by data scientists.

- Load: The transferred data is then inserted into the new database by data scientists, though a data analyst may be asked to perform this step (or others) based on the size and capabilities of the company.

Analyzing data before it has been loaded into the data warehouse may cause inconsistencies in the way data is recorded in the same variable. This will then cause inaccuracies in the data reporting and risk poor decisions being made off undependable data. For example, a data scientist likely would want to differentiate between activate user minutes and idle user minutes. A user may login into the app, and then leave their phone open but unused for several minutes as opposed to actively using the app. Without calculating "activate minutes" by eliminating the idle minutes, or at least creating a new variable to see the difference, we run the risk of miscalculating assumptions on user behavior.

**Bonus Task:**

```
1  SELECT rating, MIN(replacement_cost), MAX(replacement_cost)
2  FROM film
3  GROUP BY rating
4  ORDER BY CASE WHEN rating = 'G' THEN '1'
5  WHEN rating = 'PG' THEN '2'
6  WHEN rating = 'PG-13' THEN '3'
7  WHEN rating = 'R' THEN '4'
8  WHEN rating = 'NC-17' THEN '5'
9  ELSE '6'
10 end asc
```

Data Output | Explain | Messages | Notifications

| | rating mpaa_rating | min numeric | max numeric |
|---|---|---|---|
| 1 | G | 9.99 | 29.99 |
| 2 | PG | 9.99 | 29.99 |
| 3 | PG-13 | 9.99 | 29.99 |
| 4 | R | 9.99 | 29.99 |
| 5 | NC-17 | 9.99 | 29.99 |