Step 1:

EXPLAIN SELECT * FROM film

Seq Scan on film (cost=0.00..98.00 rows=1000 width=384)

EXPLAIN SELECT film_id, title FROM film

Seq Scan on film (cost=0.00..98.00 rows=1000 width=19)

The only difference between the two is the width. Which makes sense considering that the second query only wants 2 columns, not all of the columns withing the table. Also, without knowing the exact thing I'm trying to look for, I can't optimize this query anymore.

Step 2:

SELECT film_id, title FROM film ORDER BY title

SELECT film_id, release_year FROM film ORDER BY release_year DESC

SELECT film_id, rental_rate FROM film ORDER BY rental_rate DESC

Step 3:

SELECT rating, AVG(rental_rate) AS Average_rental_rate FROM film GROUP BY rating

SELECT rating, MIN(rental_duration) AS Min_rental_duration FROM film GROUP BY rating

SELECT rating, MAX(rental_duration) AS Max_rental_duration FROM film GROUP BY rating

Step 4:

The data engineer's are responsible for extracting the data from an outside source, then the data is transformed into a new format so that it can be more easily integrated into our internal database.

If I were to analyze the data before it's integrated into our systems, there would be a lot of empty gaps, and few, if any, keys to the rest of the data. This could cause problems for the quality of the data.