3/28/2025

Netflix Viewer Trend Analysis

Jeremiah Mendoza

Introduction: Netflix is a streaming company that deals with providing customers with movies, shows and exclusives to subscribers of the service. I am using the data that was provided to solve data problems and task given to me.

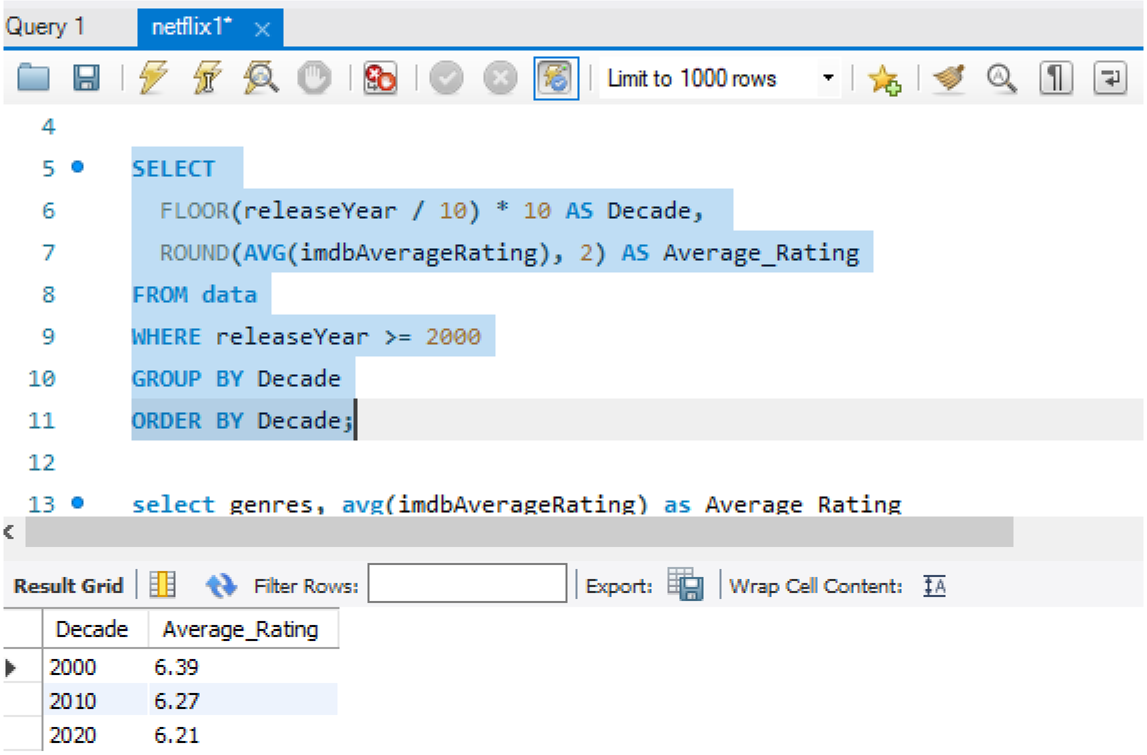
Task:

"Analyze how audience interests and content performance have changed over time on Netflix using IMDB ratings, genres, and engagement data from the 2000s and above. The goal is to uncover trends in viewer preferences and genre popularity to inform content strategy and investment decisions. "

Netflix has provided a set of 9 data questions. In addition, you are tasked with developing 3 original data questions of your own to further explore the dataset. Your objective is to analyze these questions and present your insights into a final presentation focused on those 3 self formulated questions that tackle the task given.

Data Questions:

1. How have average ratings changed by decade since 2000?



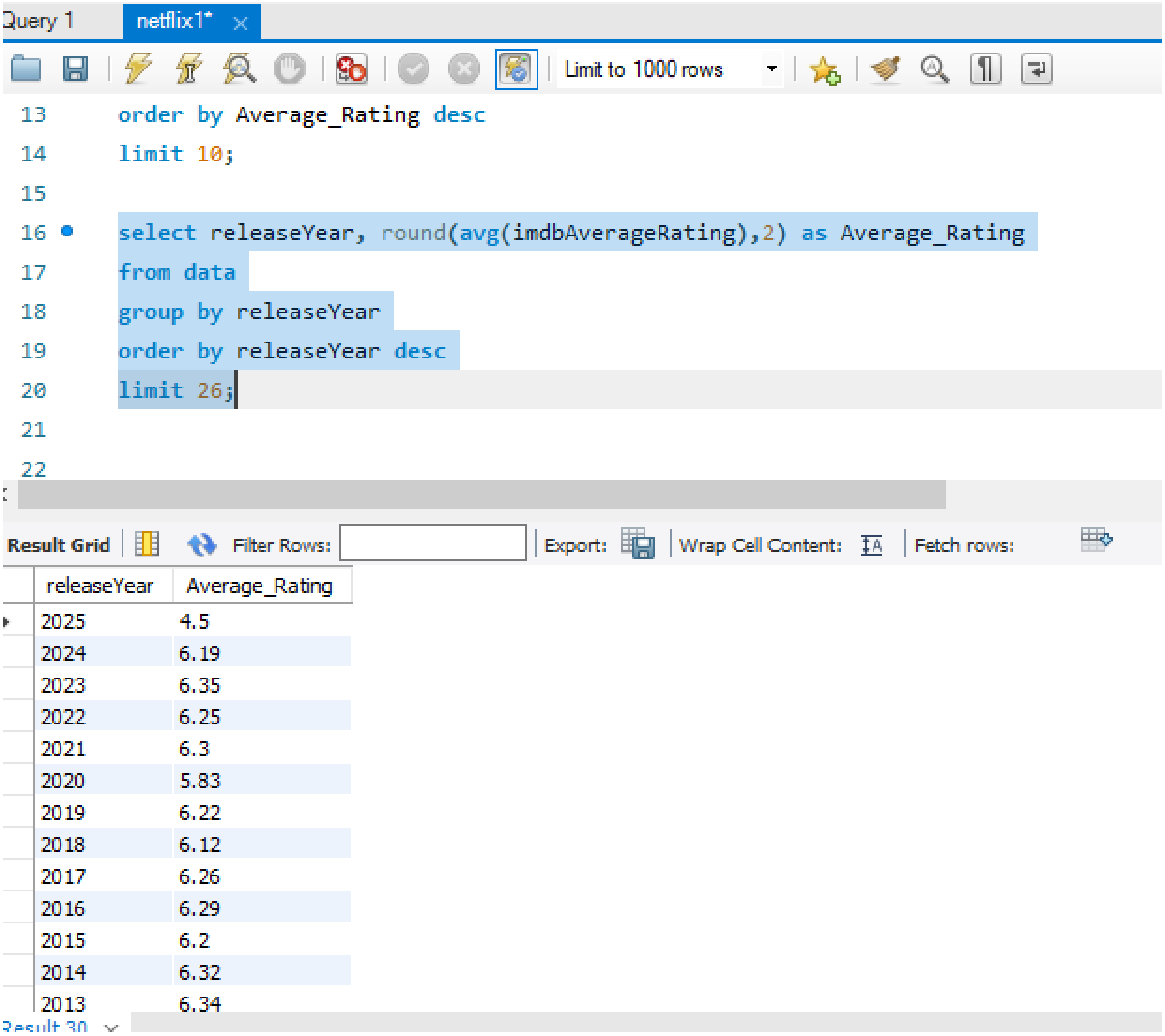
```sql

SELECT
  FLOOR(releaseYear / 10) * 10 AS Decade,
  ROUND(AVG(imdbAverageRating), 2) AS Average_Rating
FROM data
WHERE releaseYear >= 2000
GROUP BY Decade
ORDER BY Decade;

select genres, avg(imdbAverageRating) as Average_Rating
```

Decade	Average_Rating
2000	6.39
2010	6.27
2020	6.21

Explanation:

This query calculates the average IMDb rating for each decade starting from the year 2000. It uses the FLOOR(releaseYear / 10) * 10 expression to group movies by decade (e.g., 2000s, 2010s, 2020s). The AVG() function is applied to find the mean rating for each decade, and ROUND() limits the result to two decimal places. This allows us to analyze how movie ratings have shifted over long periods and detect trends across generations of film content.

2. How have average ratings shifted year-by-year?



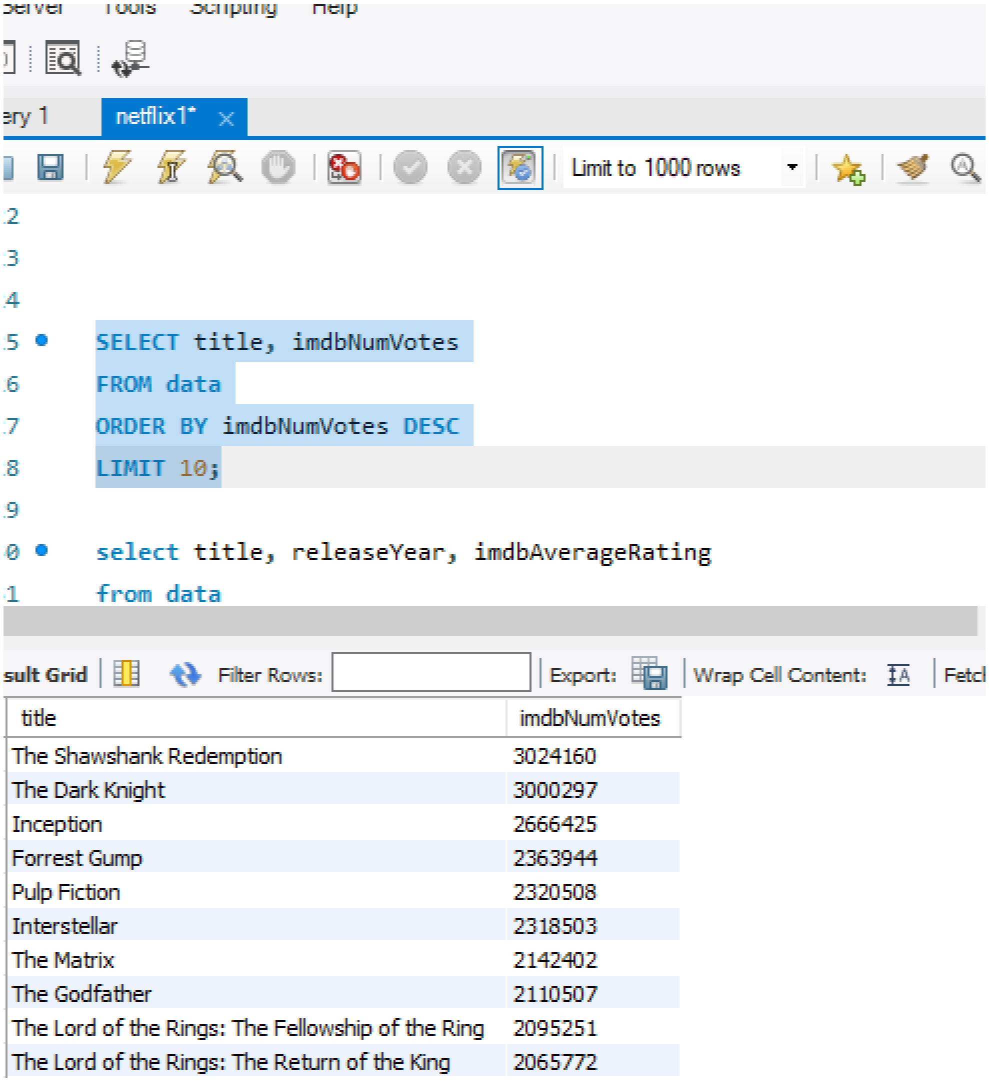
```sql
order by Average_Rating desc
limit 10;

select releaseYear, round(avg(imdbAverageRating),2) as Average_Rating
from data
group by releaseYear
order by releaseYear desc
limit 26;
```

releaseYear	Average_Rating
2025	4.5
2024	6.19
2023	6.35
2022	6.25
2021	6.3
2020	5.83
2019	6.22
2018	6.12
2017	6.26
2016	6.29
2015	6.2
2014	6.32
2013	6.34

Explanation:

This query focuses on identifying average IMDb ratings for each of the last 26 years by grouping records by releaseYear. Using AVG() to calculate mean ratings and ROUND() to improve readability, it highlights how ratings have changed on a yearly basis. By ordering the results in descending order, it provides a top down view of recent trends in movie quality.

3. What are the most voted (engaged-with) Netflix movies?

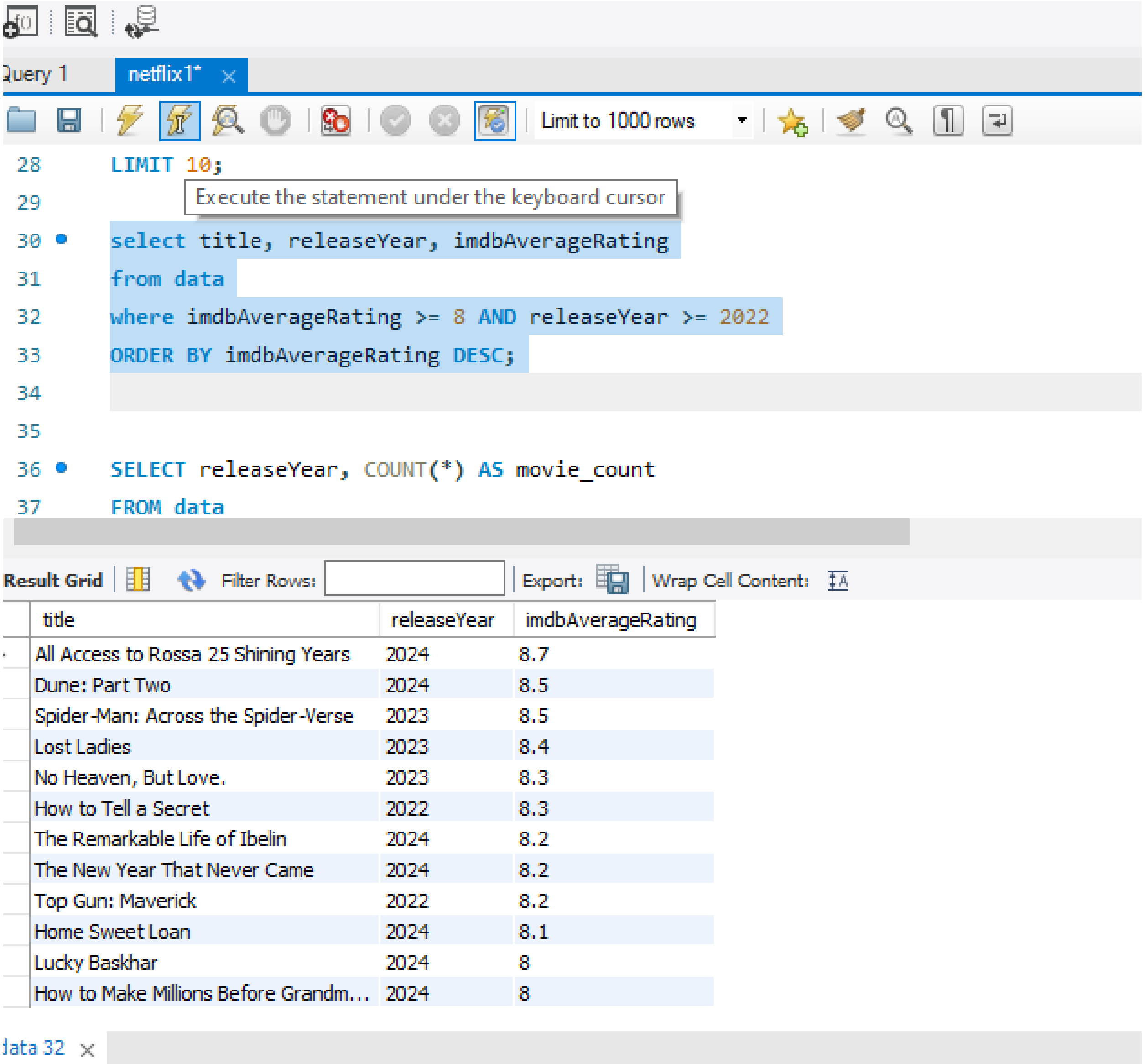


```
SELECT title, imdbNumVotes
FROM data
ORDER BY imdbNumVotes DESC
LIMIT 10;

select title, releaseYear, imdbAverageRating
from data
```

title	imdbNumVotes
The Shawshank Redemption	3024160
The Dark Knight	3000297
Inception	2666425
Forrest Gump	2363944
Pulp Fiction	2320508
Interstellar	2318503
The Matrix	2142402
The Godfather	2110507
The Lord of the Rings: The Fellowship of the Ring	2095251
The Lord of the Rings: The Return of the King	2065772

Explanation:
This query finds the top 10 movies that received the most IMDb votes, which serves as a strong proxy for engagement and popularity. It selects each movie's title and vote count and sorts them in descending order of imdbNumVotes. This is useful for identifying blockbuster hits or cult favorites that generated significant audience interaction.

4. What are the best-rated Netflix movies released since 2022?

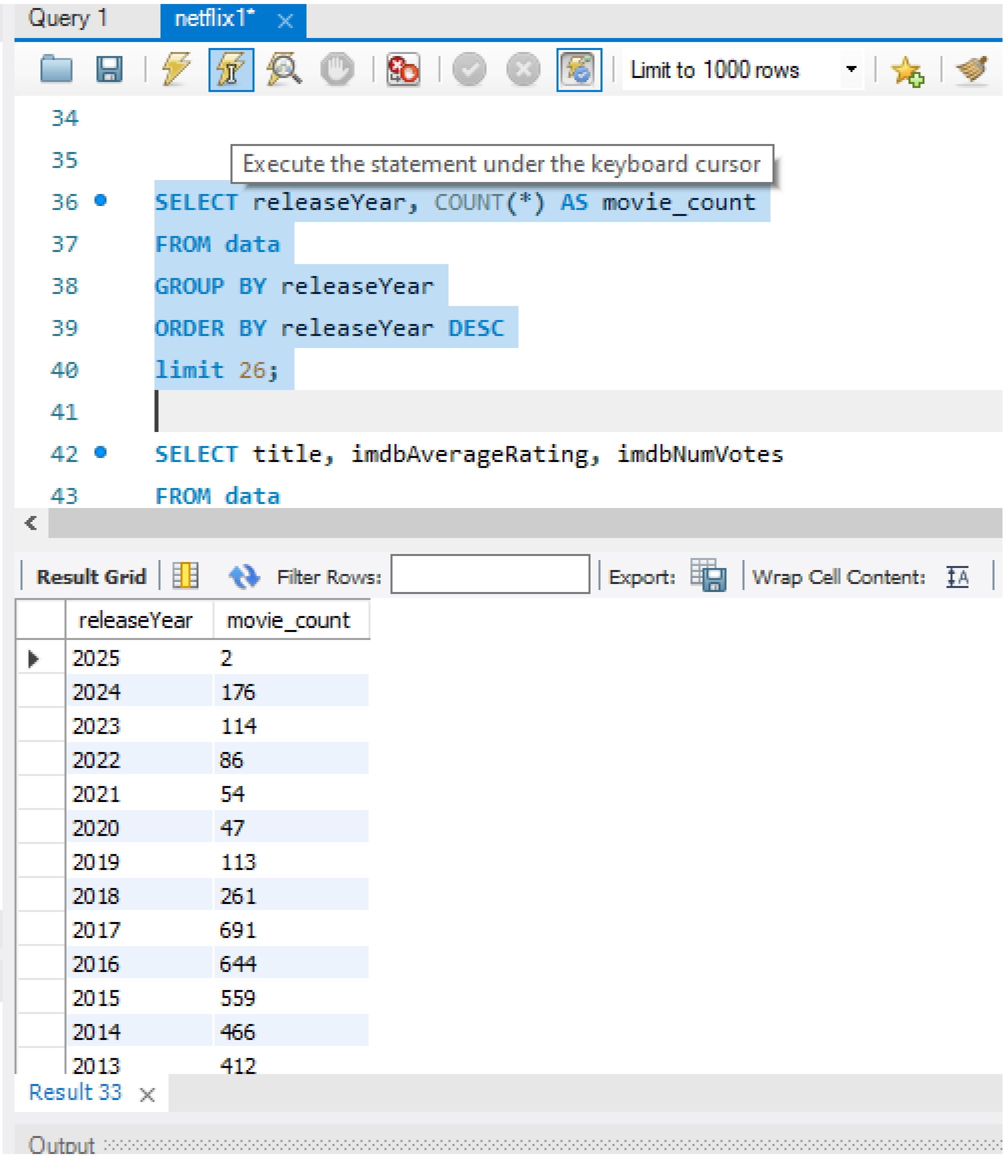


```
LIMIT 10;

select title, releaseYear, imdbAverageRating
from data
where imdbAverageRating >= 8 AND releaseYear >= 2022
ORDER BY imdbAverageRating DESC;


SELECT releaseYear, COUNT(*) AS movie_count
FROM data
```

title	releaseYear	imdbAverageRating
All Access to Rossa 25 Shining Years	2024	8.7
Dune: Part Two	2024	8.5
Spider-Man: Across the Spider-Verse	2023	8.5
Lost Ladies	2023	8.4
No Heaven, But Love.	2023	8.3
How to Tell a Secret	2022	8.3
The Remarkable Life of Ibelin	2024	8.2
The New Year That Never Came	2024	8.2
Top Gun: Maverick	2022	8.2
Home Sweet Loan	2024	8.1
Lucky Baskhar	2024	8
How to Make Millions Before Grandm...	2024	8

Explanation:
This query targets high-performing recent movies by filtering for those released after 2022 with IMDb ratings of 8.0 or higher. It provides insight into which new releases are resonating best with viewers and could guide strategic decisions on which genres, styles, or stories Netflix should continue investing in.

5. How many movies has Netflix released each year?



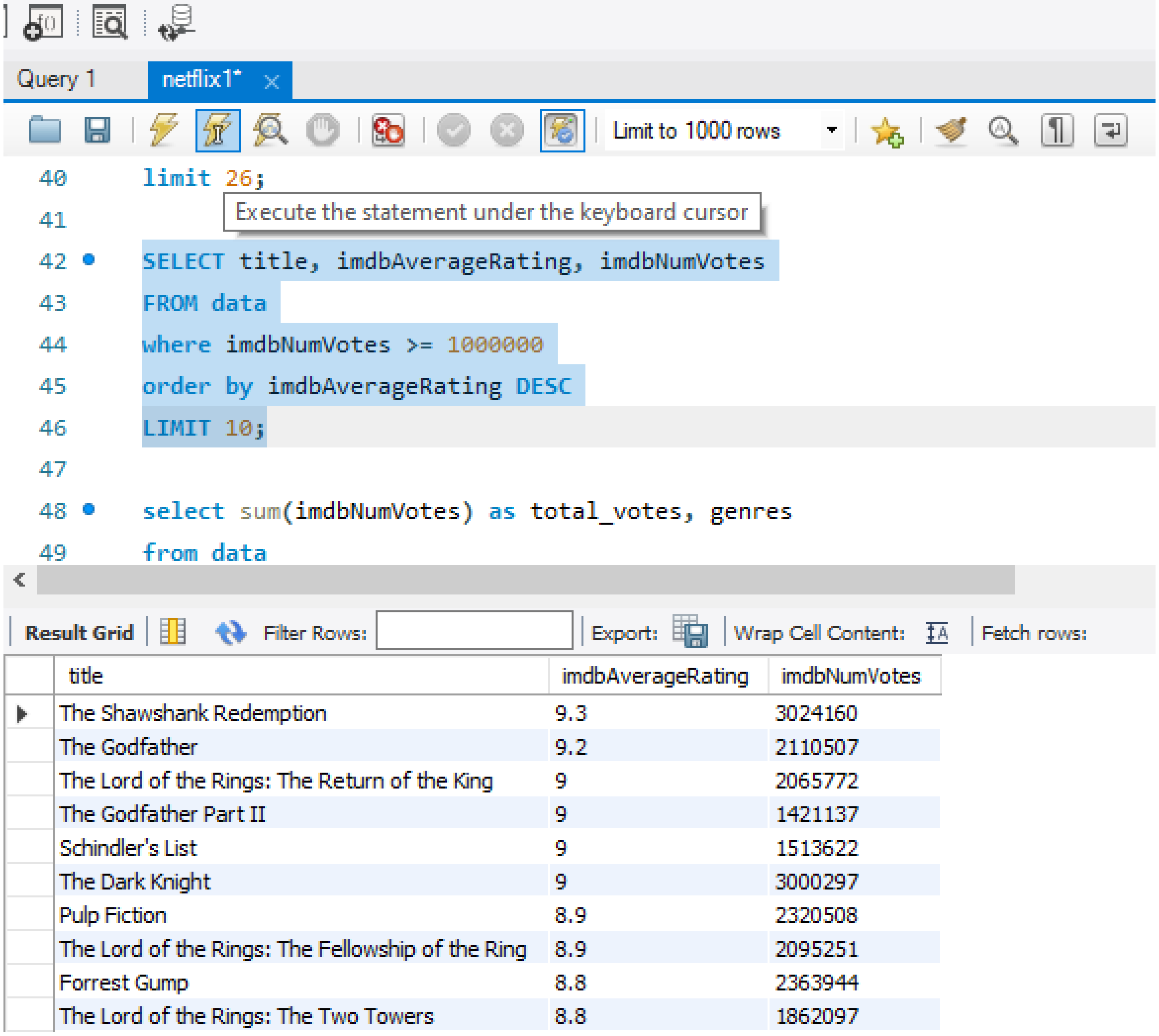
```
SELECT releaseYear, COUNT(*) AS movie_count
FROM data
GROUP BY releaseYear
ORDER BY releaseYear DESC
limit 26;

SELECT title, imdbAverageRating, imdbNumVotes
FROM data
```

releaseYear	movie_count
2025	2
2024	176
2023	114
2022	86
2021	54
2020	47
2019	113
2018	261
2017	691
2016	644
2015	559
2014	466
2013	412

Explanation:

This query counts how many movies were released each year using the COUNT(*) function grouped by releaseYear. It shows how Netflix's content production volume has evolved, helping analysts evaluate whether the platform is ramping up or reducing movie releases over time. It's especially useful when compared with viewer ratings to assess both quality and quantity.

6. What Netflix movies have over 1 million votes, limit to 10.



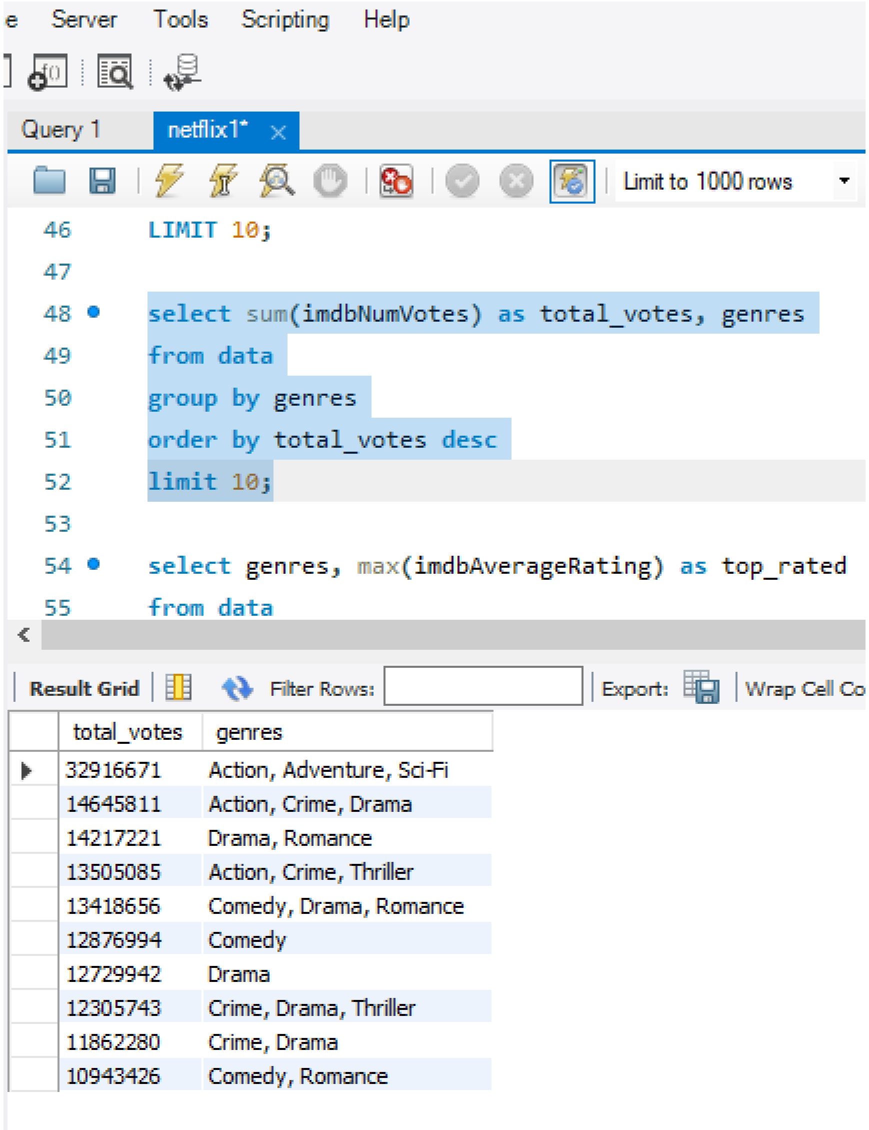
```
limit 26;

SELECT title, imdbAverageRating, imdbNumVotes
FROM data
where imdbNumVotes >= 1000000
order by imdbAverageRating DESC
LIMIT 10;

select sum(imdbNumVotes) as total_votes, genres
from data
```

title	imdbAverageRating	imdbNumVotes
The Shawshank Redemption	9.3	3024160
The Godfather	9.2	2110507
The Lord of the Rings: The Return of the King	9	2065772
The Godfather Part II	9	1421137
Schindler's List	9	1513622
The Dark Knight	9	3000297
Pulp Fiction	8.9	2320508
The Lord of the Rings: The Fellowship of the Ring	8.9	2095251
Forrest Gump	8.8	2363944
The Lord of the Rings: The Two Towers	8.8	1862097

Explanation:

This query retrieves the top 10 highest rated movies that have received over 1,000,000 IMDb votes. By applying a WHERE clause to filter movies with high vote counts (implying significant viewer engagement) and ordering the results by imdbAverageRating in descending order, the query surfaces highly rated movies that are also very popular. This gives insight into which widelywatched movies are also the most well received, making them ideal references for future content investments.

7. Which genres have the highest total audience engagement (votes)?



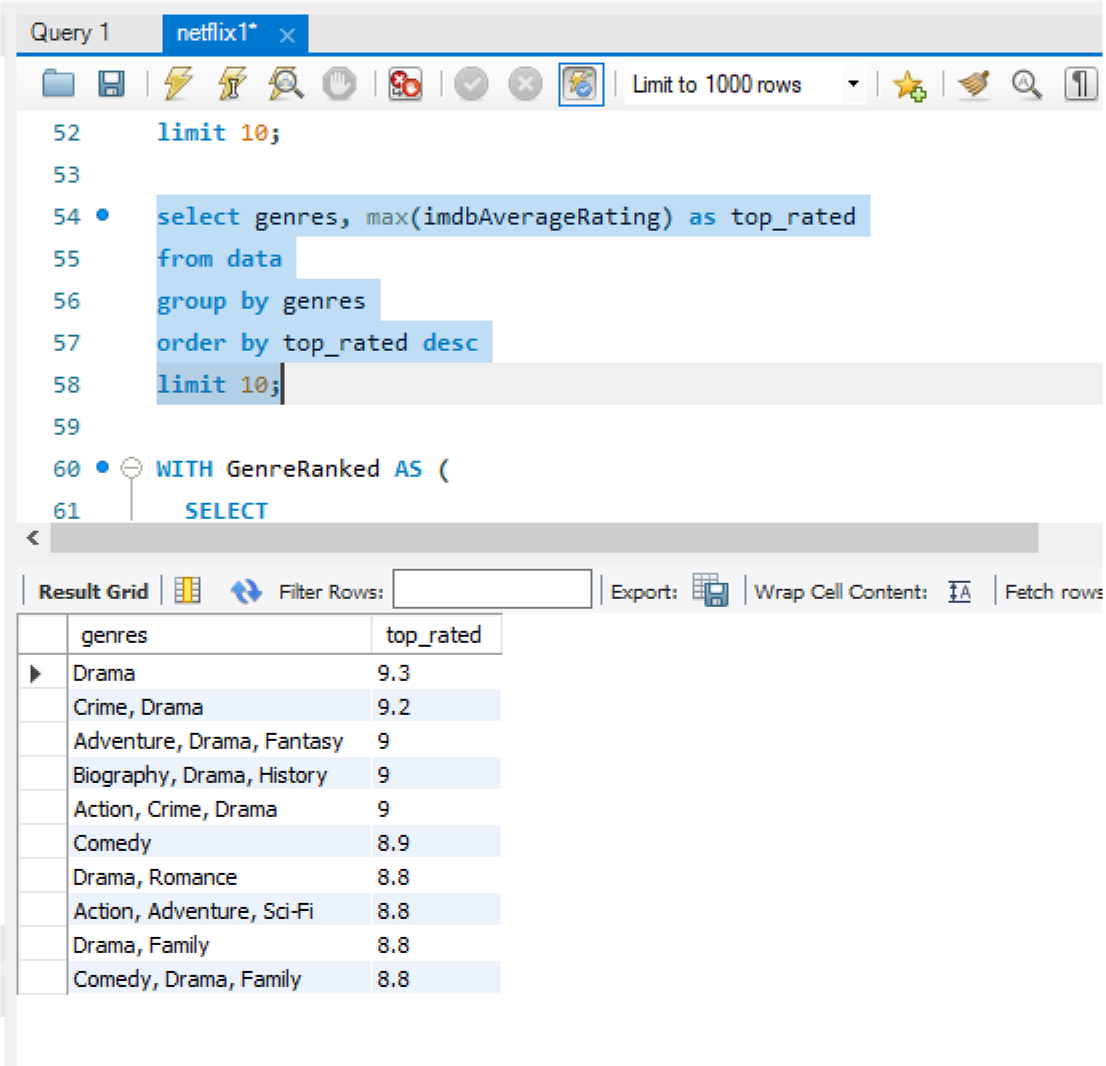
```
LIMIT 10;

select sum(imdbNumVotes) as total_votes, genres
from data
group by genres
order by total_votes desc
limit 10;

select genres, max(imdbAverageRating) as top_rated
from data
```

total_votes	genres
32916671	Action, Adventure, Sci-Fi
14645811	Action, Crime, Drama
14217221	Drama, Romance
13505085	Action, Crime, Thriller
13418656	Comedy, Drama, Romance
12876994	Comedy
12729942	Drama
12305743	Crime, Drama, Thriller
11862280	Crime, Drama
10943426	Comedy, Romance

Explanation:

This query calculates the total number of votes per genre by summing imdbNumVotes and grouping the results by genres. It then orders the results in descending order of total votes and limits the output to the top 10 genres. This query identifies which genres have received the most engagement overall, offering insight into what types of content attract the most audience attention.

8. What are the 10 highest rated genres



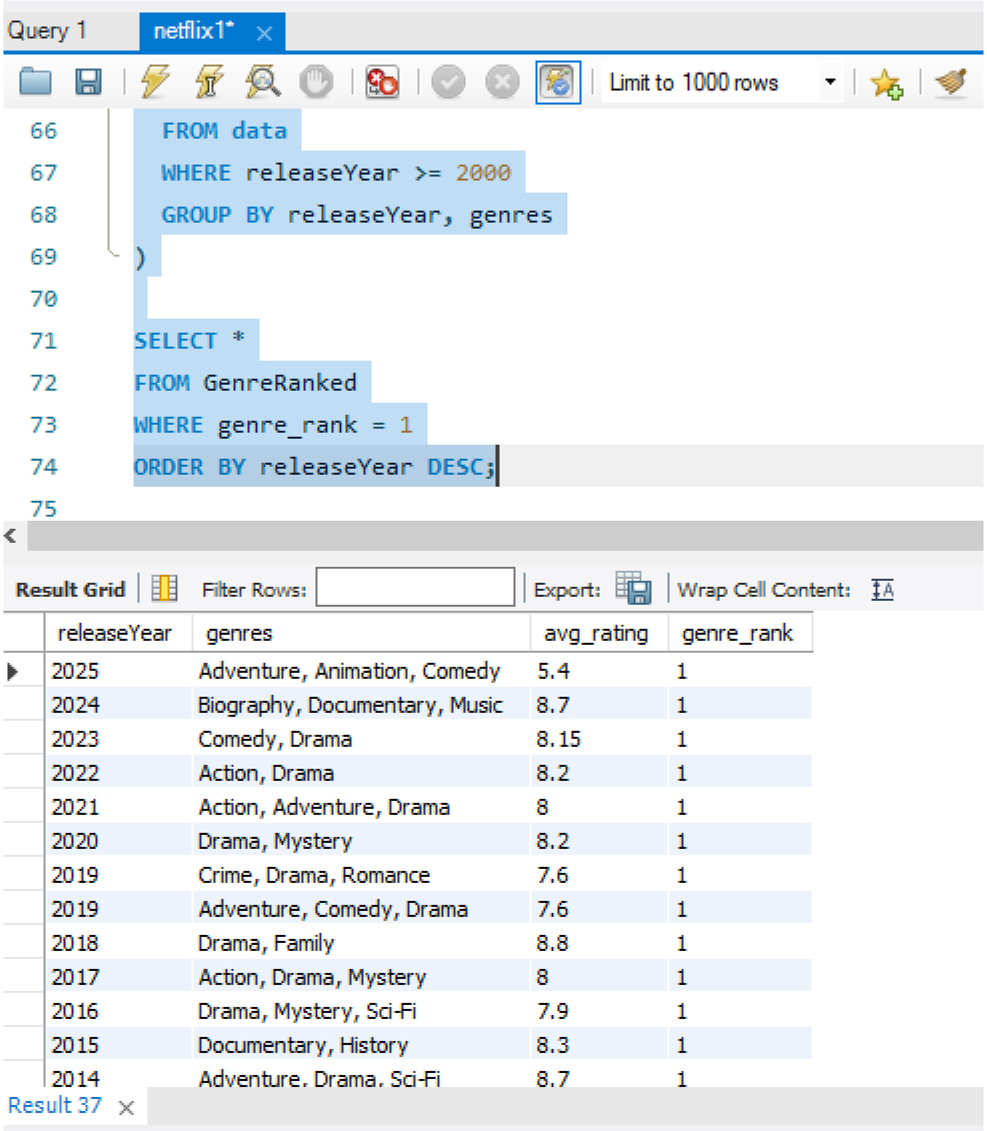
```
limit 10;

select genres, max(imdbAverageRating) as top_rated
from data
group by genres
order by top_rated desc
limit 10;

WITH GenreRanked AS (
  SELECT
```

genres	top_rated
Drama	9.3
Crime, Drama	9.2
Adventure, Drama, Fantasy	9
Biography, Drama, History	9
Action, Crime, Drama	9
Comedy	8.9
Drama, Romance	8.8
Action, Adventure, Sci-Fi	8.8
Drama, Family	8.8
Comedy, Drama, Family	8.8

Explanation:

This query identifies the highest rated movie within each genre by using the MAX function on imdbAverageRating and grouping by genre. The result shows the best rated content for each genre and helps stakeholders understand the upper limits of quality and audience satisfaction by genre. It also helps identify standout genres where exceptional quality content exists.

9. What is the top-rated genre each year after 2000?



```
    FROM data
    WHERE releaseYear >= 2000
    GROUP BY releaseYear, genres
)

SELECT *
FROM GenreRanked
WHERE genre_rank = 1
ORDER BY releaseYear DESC;
```

releaseYear	genres	avg_rating	genre_rank
2025	Adventure, Animation, Comedy	5.4	1
2024	Biography, Documentary, Music	8.7	1
2023	Comedy, Drama	8.15	1
2022	Action, Drama	8.2	1
2021	Action, Adventure, Drama	8	1
2020	Drama, Mystery	8.2	1
2019	Crime, Drama, Romance	7.6	1
2019	Adventure, Comedy, Drama	7.6	1
2018	Drama, Family	8.8	1
2017	Action, Drama, Mystery	8	1
2016	Drama, Mystery, Sci-Fi	7.9	1
2015	Documentary, History	8.3	1
2014	Adventure, Drama, Sci-Fi	8.7	1

Explanation:

This Common Table Expression (CTE) ranks genres by their average IMDb rating for each year starting from 2000. The RANK() window function assigns a rank to each genre per year based on descending average rating. The final query filters for only the top-ranked genre per year (genre_rank = 1) and orders the results by releaseYear. This allows analysis of the most highly rated genres year-over-year, helping track shifts in viewer preferences and top performing genres over time.

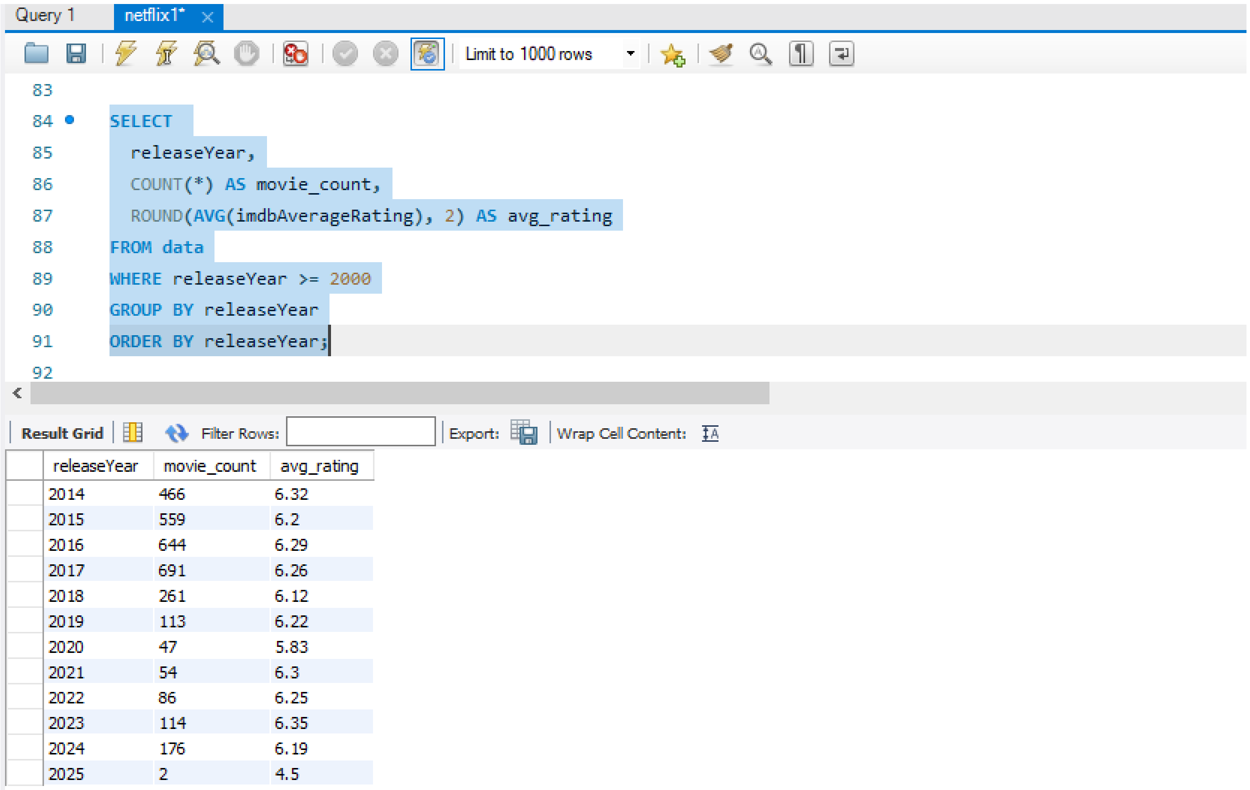
Data Questions I used to solve the given task

1. Average Rating and Total Votes for Retrospective Years(2000-2025)

By knowing the average rating and total votes for each year we can gain a basic understanding of how movies are generally viewed by the public during certain years.

Through the data we can see a gradual increase after 2024 which means we should start to invest more into the platforms movies as it is gaining more interest.

Side Note: 2025 only consists of 2 movies which makes the data skewed.



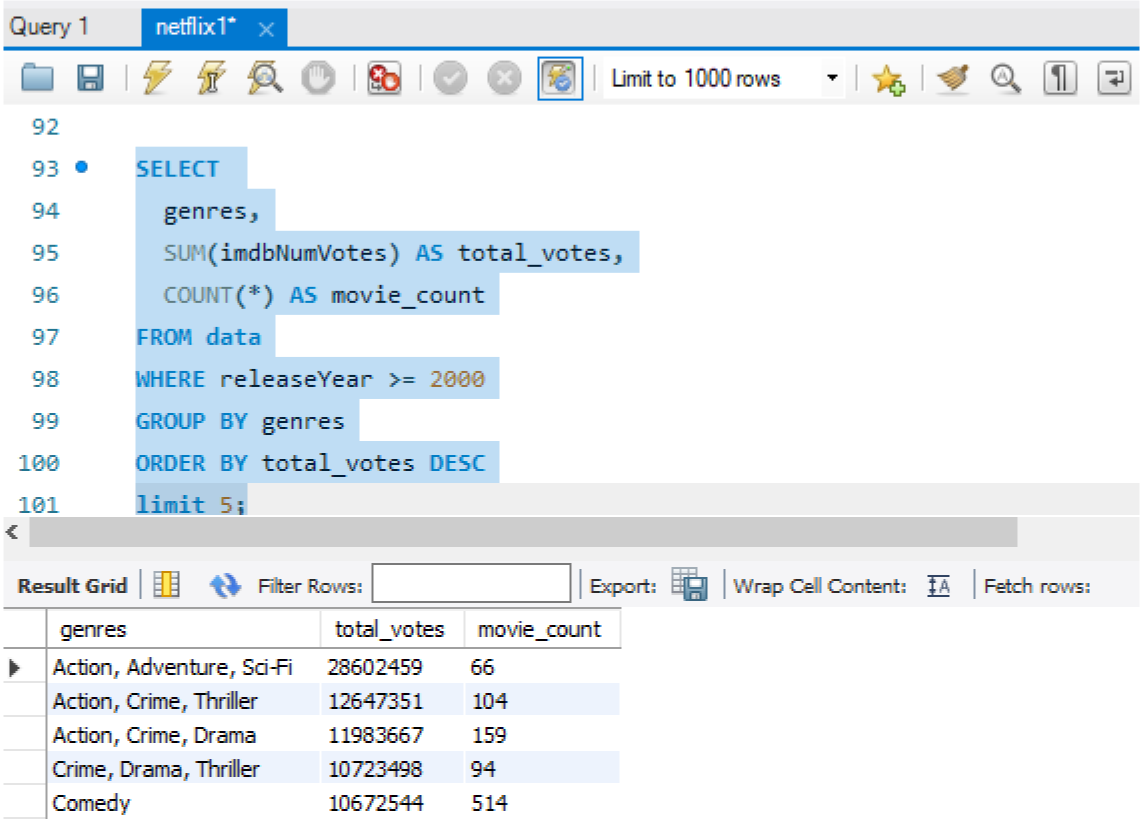
```sql

SELECT
  releaseYear,
  COUNT(*) AS movie_count,
  ROUND(AVG(imdbAverageRating), 2) AS avg_rating
FROM data
WHERE releaseYear >= 2000
GROUP BY releaseYear
ORDER BY releaseYear;

```

releaseYear	movie_count	avg_rating
2014	466	6.32
2015	559	6.2
2016	644	6.29
2017	691	6.26
2018	261	6.12
2019	113	6.22
2020	47	5.83
2021	54	6.3
2022	86	6.25
2023	114	6.35
2024	176	6.19
2025	2	4.5

2. The Most Voted for Movies and their Numbered Releases (Top 5)

Despite having the least movie releases, Action Adventure, Sci-Fi is still the most engaged in genre, because of this we can

- Capitalize on an already popular genre
- Take advantage of there not being a lot of releases of the genre

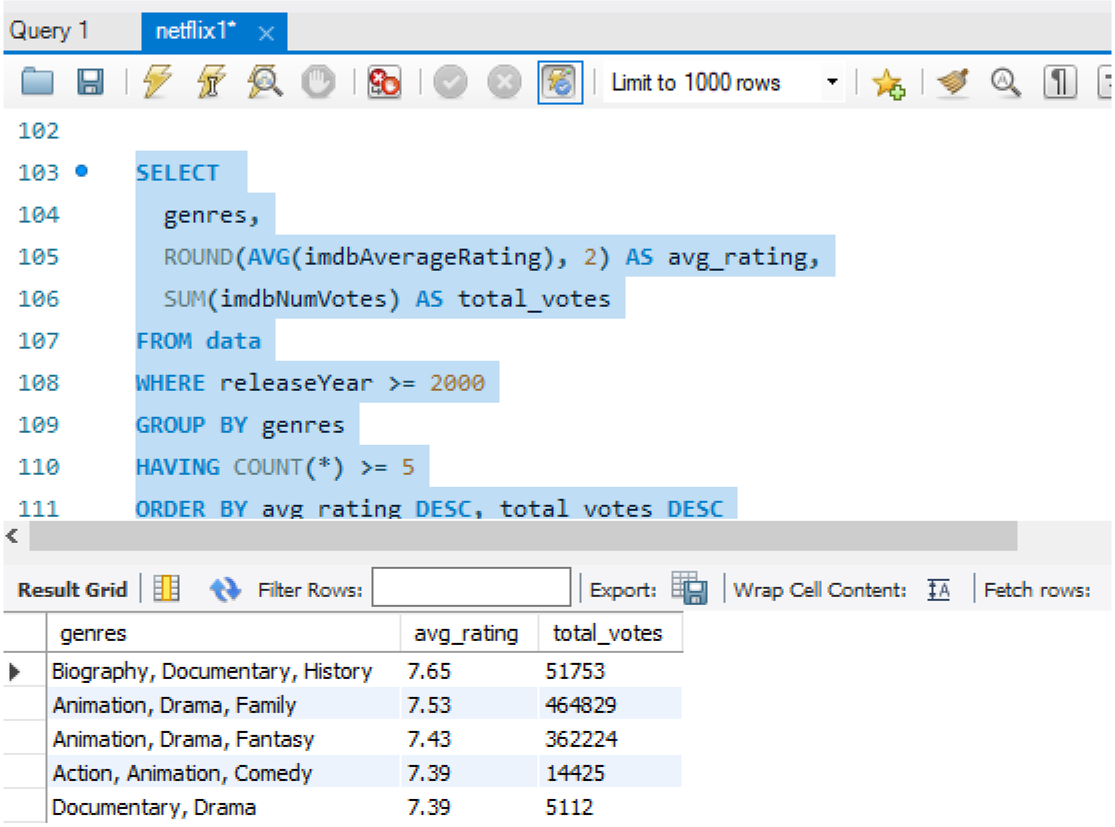


```
SELECT
  genres,
  SUM(imdbNumVotes) AS total_votes,
  COUNT(*) AS movie_count
FROM data
WHERE releaseYear >= 2000
GROUP BY genres
ORDER BY total_votes DESC
limit 5;
```

genres	total_votes	movie_count
Action, Adventure, Sci-Fi	28602459	66
Action, Crime, Thriller	12647351	104
Action, Crime, Drama	11983667	159
Crime, Drama, Thriller	10723498	94
Comedy	10672544	514

3. Most Consistent Performing Genres in Votes and Average Rating (Top 5)

We want to find a genre that has a good amount of engagement along with a good amount of average rating. Focusing on movies from the Animation, Drama, Family genres will yield the most engagement and satisfaction.



```sql

SELECT
  genres,
  ROUND(AVG(imdbAverageRating), 2) AS avg_rating,
  SUM(imdbNumVotes) AS total_votes
FROM data
WHERE releaseYear >= 2000
GROUP BY genres
HAVING COUNT(*) >= 5
ORDER BY avg_rating DESC, total_votes DESC
```

genres	avg_rating	total_votes
Biography, Documentary, History	7.65	51753
Animation, Drama, Family	7.53	464829
Animation, Drama, Fantasy	7.43	362224
Action, Animation, Comedy	7.39	14425
Documentary, Drama	7.39	5112