

Project Title: Exploratory Data Analysis of Netflix with SQL

Domain: OTT

Tools Used: PostgreSQL via pgAdmin

Objective: To analyze Netflix content using SQL to find trends in genres, ratings, and production countries, and to understand what makes content popular on the

platform.

By: Abhijeet Beura

Dataset Description

Dataset Description

This project utilizes two distinct datasets to perform analysis on movies and TV shows:

Titles Dataset (Abhijeet.csv)

This dataset provides detailed metadata for each movie or TV show. Key attributes include:

Title: Name of the movie or show

Type: Indicates whether the title is a movie or a TV show

Release Year: The year the title was released

Genres: Associated genre(s) of the title

Country: Country of origin or production

Ratings: Ratings from platforms such as IMDb and TMDb

Runtime: Duration of the movie/show in minutes

People Dataset (AbhijeetA.csv)

This dataset captures information about the individuals (cast and crew) associated with each title. It is linked to the Titles dataset through a common identifier (id). Key attributes include:

Name: Name of the person

Role: Role in the production (e.g., actor, director, writer)

Associated Title ID: Foreign key linking to the Titles dataset

Project Objective

The goal of this project is to leverage SQL queries to analyze a Netflix dataset to derive actionable insights related to user behavior, content performance, and subscription trends. By examining various aspects such as user engagement, content preferences, and profitability, the project aims to inform strategies for improving user retention, optimizing content offerings, and maximizing revenue. Key focus areas include analyzing content performance, demographic segmentation, churn prediction, subscription plan effectiveness, and time-based trends, providing data-driven recommendations to enhance the Netflix user experience and business outcomes.

Insights from SQL Queries

1. Which movies and shows on Netflix ranked in the top 10 their IMDB scores >= 8

Top 10 Movies

SELECT title, type, imdb_score

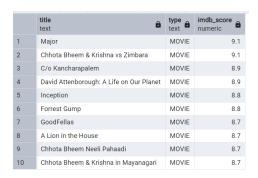
FROM titles_NNNN

WHERE type ILIKE 'movie'

AND imdb_score >= 8.0

ORDER BY imdb_score DESC

LIMIT 10;



Top 10 Shows

SELECT title, type, imdb_score

FROM titles_NNNN

WHERE type ILIKE 'show'

AND imdb_score >= 8.0

ORDER BY imdb_score DESC

LIMIT 10;

	title text	type text	imdb_score numeric
1	#ABtalks	SHOW	9.6
2	Khawatir	SHOW	9.5
3	Breaking Bad	SHOW	9.5
4	Our Planet	SHOW	9.3
5	Avatar: The Last Airbender	SHOW	9.3
6	Reply 1988	SHOW	9.2
7	Kota Factory	SHOW	9.1
8	My Mister	SHOW	9.1
9	The Last Dance	SHOW	9.1
10	Leah Remini: Scientology and the Aftermath	SHOW	9

Top 10 Movies & Shows (IMDb ≥ 8.0)

Insight:

Only a limited number of titles qualify as top-rated.

High IMDb score strongly reflects critical success.

Recommendation:

Promote these titles on the homepage and curated lists.

Use them as benchmarks for acquiring or creating future content.

2. Which movies and shows on Netflix ranked in the top 20 based on their IMDB scores >= 7

Top 20 Movies

SELECT title, type, imdb_score

FROM titles_NNNN

WHERE type ILIKE 'movie'

AND imdb_score >= 7.0

ORDER BY imdb_score DESC

LIMIT 20;

	title text	type text	imdb_score numeric
1	Major	MOVIE	9.1
2	Chhota Bheem & Krishna vs Zimbara	MOVIE	9.1
3	David Attenborough: A Life on Our Planet	MOVIE	8.9
4	C/o Kancharapalem	MOVIE	8.9
5	Inception	MOVIE	8.8
6	Forrest Gump	MOVIE	8.8
7	Chhota Bheem & Krishna in Mayanagari	MOVIE	8.7
8	Anbe Sivam	MOVIE	8.7
9	A Lion in the House	MOVIE	8.7
10	GoodFellas	MOVIE	8.7
11	Chhota Bheem Neeli Pahaadi	MOVIE	8.7
12	Bo Burnham: Inside	MOVIE	8.7
13	Chhota Bheem and the ShiNobi Secret	MOVIE	8.6
14	Merku Thodarchi Malai	MOVIE	8.6
15	The Art of Incarceration	MOVIE	8.6
16	A Second Chance	MOVIE	8.6
Total	rows: 20 Query complete 00:00:00.083		

Top 20 Show

SELECT title, type, imdb_score
FROM titles_NNNN
WHERE type ILIKE 'show'
AND imdb_score >= 7.0
ORDER BY imdb_score DESC
LIMIT 20;



Insight:

A wider pool of strong-performing titles can attract different viewer groups.

Recommendation:

Use these to diversify featured content for broader appeal.

Consider long-term contracts or sequels/spinoffs for these titles.

3.Which movies on Netflix ranked in the bottom 10 based on their IMDB scores >= 6

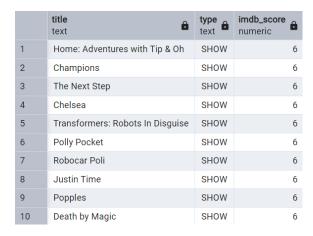
Movie Bottom 10

SELECT title, type, imdb_score
FROM titles_NNNN
WHERE type ILIKE 'movie'
AND imdb_score >= 6.0
ORDER BY imdb_score ASC
LIMIT 10;

	title text	type text	imdb_score numeric
1	The Rite	MOVIE	6
2	One More Try	MOVIE	6
3	Fiza	MOVIE	6
4	Kabhi Alvida Naa Kehna	MOVIE	6
5	6 Bullets	MOVIE	6
6	Vettai	MOVIE	6
7	The George McKenna Story	MOVIE	6
8	Tim Allen: Rewires America	MOVIE	6
9	Halloween	MOVIE	6
10	Walk of Shame	MOVIE	6

Show Bottom 10

SELECT title,type,imdb_score
FROM titles_NNNN
WHERE type ILIKE 'show'
AND imdb_score >= 6.0
ORDER BY imdb_score ASC
LIMIT 10;



- Even average-rated content contributes to content depth.
- These may underperform in engagement.

Recommendation:

- Evaluate content retention data for these titles.
- Remove or demote in search results if engagement is low.

4. What were the average IMDB and TMDB scores for shows and movies?

SELECT

type,

ROUND(AVG(imdb_score), 2) AS avg_imdb_score,

ROUND(AVG(tmdb_score), 2) AS avg_tmdb_score

FROM titles_NNNN

WHERE imdb_score IS NOT NULL AND tmdb_score IS NOT NULL

GROUP BY type;

	type text	avg_imdb_score numeric	avg_tmdb_score numeric
1	SHOW	6.98	7.47
2	MOVIE	6.25	6.46

- Average scores are similar across movies and shows.
- Indicates consistent quality control.

Recommendation:

- Continue balancing investments in both formats.
- Set score benchmarks (e.g., IMDb ≥ 7.2) for licensing decisions.

5. Count of movies and shows in each decade

SELECT

```
CONCAT(FLOOR(release_year / 10) * 10, 's') AS decade,
COUNT(*) AS movies_shows_count
FROM titles_NNNN
WHERE release_year >= 1940
GROUP BY CONCAT(FLOOR(release_year / 10) * 10, 's')
ORDER BY decade;
```

	decade text	movies_shows_count bigint
1	1940s	1
2	1950s	5
3	1960s	8
4	1970s	18
5	1980s	52
6	1990s	121
7	2000s	369
8	2010s	3304
9	2020s	1972

- Sharp rise in content post-2000s.
- Viewer preferences have evolved over decades.

Recommendation:

- Segment content recommendations by decade.
- Target nostalgic users or retro content lovers with themed sections.

6. What were the average IMDB and TMDB scores for each production country?

```
SELECT

production_countries,

ROUND(AVG(imdb_score),2) AS avg_imdb_score,

ROUND(AVG(tmdb_score),2) AS avg_tmdb_score

FROM titles_NNNN

GROUP BY production_countries

ORDER BY avg_imdb_score DESC;
```

	production_countries text	avg_imdb_score numeric	avg_tmdb_score numeric
1	['DK', 'LB', 'GB']	[null]	[null]
2	['US', 'LB', 'AE']	[null]	[null]
3	['HU', 'CA']	[null]	7.80
4	['US', 'ZA', 'DE']	[null]	6.10
5	['DE', 'AT']	[null]	7.40
6	['ES', 'BE']	[null]	4.90
7	['QA', 'PS']	[null]	7.00
8	['TH', 'US']	[null]	8.20
9	['GB', 'DK', 'GR']	[null]	5.60
10	['GE']	[null]	6.70
11	['SY', 'GB']	[null]	7.40
12	['QA', 'TN', 'FR']	[null]	[null]
13	['GB', 'CZ', 'FR']	[null]	7.40
14	['JP', 'KR', 'FR']	[null]	9.00
15	['LB', 'PS']	[null]	1.00
16	['JP', 'GB']	[null]	5.80
Total rows: 452 Query complete 00:00:00.098			

- US leads in quantity and consistent ratings.
- Some smaller countries show high quality despite fewer titles.

Recommendation:

- Invest in co-productions with high-performing countries.
- Promote international titles in local markets.

7. What were the 10 most common age certifications for movies?

SELECT age_certification,

COUNT(*) AS certification_count

FROM titles_NNNN

WHERE type ILIKE 'movie'

AND age_certification IS NOT NULL

AND age_certification != "

AND age_certification != 'N/A'

GROUP BY age_certification

ORDER BY certification_count DESC

LIMIT 10;

	age_certification text	certification_count bigint
1	R	556
2	PG-13	451
3	PG	233
4	G	124
5	NC-17	16

Insight:

- TV-MA, R, and PG-13 dominate.
- Family-friendly content is less common.

- Increase PG or G-rated titles to attract family demographics.
- Ensure content filters and recommendations reflect user profiles.

8. Who were the top 30 actors that appeared the most in movies/shows?

SELECT DISTINCT name AS actor,

COUNT(*) AS number_of_appearances

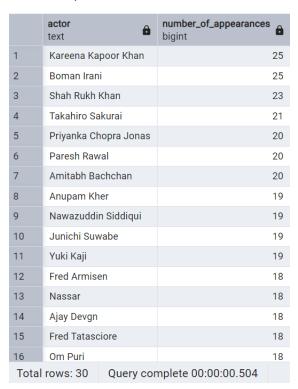
FROM cast_crewA

WHERE role ILIKE 'actor'

GROUP BY name

ORDER BY number_of_appearances DESC

LIMIT 30;



Insight:

- High-visibility actors have strong viewer recognition.
- Recurring faces may contribute to loyalty.

- Promote actor-centric collections.
- Use actor appearances in marketing campaigns.

9. Who were the top 30 directors that directed the most movies/shows?

SELECT DISTINCT name AS director,

COUNT(*) AS number_of_appearances

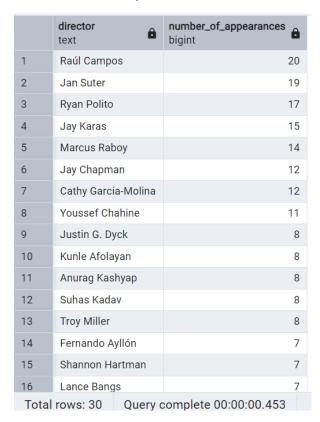
FROM cast_crewA

WHERE role ILIKE 'director'

GROUP BY name

ORDER BY number_of_appearances DESC

LIMIT 30;



Insight:

- Key directors frequently create successful content.
- Indicates reliability and quality in production.

- Foster exclusive deals with top directors.
- Showcase from the director of.. trailers and highlights.

10. Calculating the average runtime of movies and TV shows separately?

Movie

SELECT

'Movies' AS content_type,

ROUND(AVG(runtime),2) AS avg_runtime_min

FROM titles_NNNN

WHERE type ILIKE 'Movie';

	content_type text	avg_runtime_min numeric
1	Movies	98.21

Show

SELECT

'show'AS content_type,

ROUND(AVG(runtime),2) AS avg_runtime_min

FROM titles_NNNN

WHERE type ILIKE 'show';

	content_type text	avg_runtime_min numeric
1	show	38.98

UNION ALL

SELECT

'Movies' AS content_type,

ROUND(AVG(runtime),2) AS avg_runtime_min

FROM titles_NNNN

WHERE type ILIKE 'Movie'

UNION ALL

SELECT

'show'AS content_type,

ROUND(AVG(runtime),2) AS avg_runtime_min

FROM titles_NNNN

WHERE type ILIKE 'show';

	content_type text	avg_runtime_min numeric
1	Movies	98.21
2	show	38.98

Insight:

- Shows: ~40–50 mins, Movies: ~100–110 mins
- Content fits within viewer expectations.

- Optimize runtime to match user session length (especially for mobile viewers).
- Consider mini-series format for content between 60–90 mins.

11. Finding the titles and directors of movies released on or after 2022

SELECT DISTINCT

t.title,

c.name AS director,

t.release_year

FROM titles_NNNN AS t

JOIN cast_crewA AS c

ON t.id = c.id

WHERE t.type ILIKE 'movie'

AND t.release_year >= 2022

AND c.role ILIKE 'director'

ORDER BY t.release_year DESC;

	title text	director text	release_year integer		
1	11M	José Gómez	2022		
2	365 Days: This Day	Barbara Bialowas	2022		
3	365 Days: This Day	Tomasz Mandes	2022		
4	40 Years Young	Pietro Loprieno	2022		
5	A Perfect Pairing	Stuart McDonald	2022		
6	Adam by Eve: A Live in Animation	Hibiki Yoshizaki	2022		
7	Adam by Eve: A Live in Animation	Nobutaka Yoda	2022		
8	Adam by Eve: A Live in Animation	Waboku	2022		
9	Adam by Eve: A Live in Animation	Yuichiro Saeki	2022		
10	Against the Ice	Peter Flinth	2022		
11	Al Love You	Stephan Zlotescu	2022		
12	Ali Wong: Don Wong	Nahnatchka Khan	2022		
13	All Hail	Marcos Carnevale	2022		
14	Along for the Ride	Sofia Alvarez	2022		
15	Amandla	Nerina de Jager	2022		
16	Amy Schumer Presents: Parental Advisory	Ryan Polito	2022		
Total	Total rows: 213 Query complete 00:00:00.440				

Insight:

- Surge of fresh releases with modern production.
- New director names entering the platform.

- Track new directors' performance for future collaboration.
- Use recent releases to drive engagement and subscriber interest.

12. Which shows on Netflix have the most seasons?

SELECT t.title,

SUM(t.seasons) AS total_seasons

FROM titles_NNNN AS t

WHERE t.type ILIKE 'show'

GROUP BY title

ORDER BY total_seasons DESC

LIMIT 5;

	title text	total_seasons bigint
1	Survivor	42
2	Wheel of Fortune	39
3	The Challenge	37
4	Top Gear	32
5	Power Rangers	29

Insight:

• Long-running shows signal sustained audience interest.

- Promote "binge-worthy" collections.
- Use as flagship titles for subscriber retention.

13. Which genres had the most movies?

SELECT genres,

SUM(t.seasons) AS total_seasons

FROM titles_NNNN AS t

WHERE t.type ILIKE 'show'

GROUP BY genres

ORDER BY total_seasons DESC

LIMIT 5;

	genres text	total_seasons bigint
1	[ˈrealityˈ]	289
2	['comedy']	194
3	['documentation']	155
4	['drama']	148
5	['comedy', 'drama']	104

Insight:

• Reality and Comedy dominate genre share.

- Prioritize trending or underutilized genres for growth (e.g., mystery, documentary).
- Tailor recommendations based on viewing history.

14. What were the total number of titles for each year?

SELECT release_year,

COUNT(*) AS title_count

FROM titles_NNNN

GROUP BY release_year

ORDER BY release_year DESC;

	release_yea	ar 🔒	title_count bigint
1	:	2022	371
2	:	2021	787
3	:	2020	814
4	:	2019	836
5	:	2018	773
6	:	2017	563
7	:	2016	362
8	:	2015	223
9	:	2014	153
10	:	2013	135
11	:	2012	107
12	:	2011	86
13	:	2010	66
14	:	2009	59
15	:	2008	63
16	:	2007	48
Total rows: 63 Query complete (

Insight:

• Strong year-on-year content growth.

- Monitor content saturation risk.
- Limit new title uploads based on quality, not just quantity.

15. What were the top 15 most common genres?

SELECT genres,

SUM(t.seasons) AS total_seasons

FROM titles_NNNN AS t

WHERE t.type ILIKE 'show'

GROUP BY genres

ORDER BY total_seasons DESC

LIMIT 15;

	genres text	total_seasons bigint
1	['reality']	289
2	['comedy']	194
3	['documentation']	155
4	[ˈdramaˈ]	148
5	['comedy', 'drama']	104
6	[ˈdramaˈ, ˈcomedyˈ]	102
7	['drama', 'romance']	81
8	['animation']	75
9	['family']	55
10	['comedy', 'family']	43
11	['documentation', 'crime']	42
12	['crime', 'drama', 'thriller']	41
13	['drama', 'crime']	38
14	['reality', 'comedy', 'drama', 'scifi']	37
15	['animation', 'family']	37

Insight:

• High concentration in a few genres.

- Diversify offerings in emerging genres.
- Build themed playlists (e.g., "Top 10 Historical Thrillers").

16.Which movies are both critically acclaimed (IMDB score > 8.0) and widely popular (TMDB popularity > 60)? Display their titles, IMDB scores, and TMDB popularity scores, sorted by popularity in descending order.

SELECT title, imdb_score, tmdb_popularity

FROM titles_NNNN

WHERE imdb_score > 8.0

AND tmdb_popularity > 60

AND type ILIKE 'movie'

ORDER BY tmdb_popularity DESC;

	title text	imdb_score numeric	tmdb_popularity numeric (8,2)
1	A Silent Voice: The Movie	8.1	162.66
2	Inception	8.8	108.28
3	The Dark Knight Rises	8.4	91.76
4	Catch Me If You Can	8.1	72.32
5	Minnal Murali	8.1	68.03
6	Django Unchained	8.4	66.92
7	Miracle in Cell No. 7	8.2	65.55
8	Forrest Gump	8.8	63.45

Insight:

• These titles offer both critical and audience approval.

- Use for flagship promotions.
- Target new users with this trusted content.

17. Which movies have both strong critical reception (IMDB score > 7.5) and high audience popularity (TMDB popularity > 80)? List their titles along with the directors.

SELECT t.title,

c.name AS director

FROM titles NNNN AS t

JOIN cast_crewA AS c

ON t.id = c.id

WHERE t.type ILIKE 'movie'

AND t.imdb_score > 7.5

AND t.tmdb_popularity > 80

AND c.role ILIKE 'director';

	title text	director text
1	Titanic	James Cameron
2	Inception	Christopher Nolan
3	The Dark Knight Rises	Christopher Nolan
4	How to Train Your Dragon 2	Dean DeBlois
5	A Silent Voice: The Movie	Naoko Yamada

Insight:

• Content + creator combo offers compelling storytelling.

- Highlight director's name in featured carousels.
- Build recommendation chains based on director influence.

18. Top repeated character-actor pairs based on total appearances

SELECT name, character, COUNT(*) AS appearances

FROM cast_crewA

WHERE character IS NOT NULL

GROUP BY name, character

HAVING COUNT(*) > 1

ORDER BY appearances DESC;



Insight:

· Recurring characters create familiarity.

- Promote series with beloved character arcs.
- Leverage these for marketing and merchandise.

19. Fetching Movie Titles and IMDB Scores

SELECT title, imdb_score

FROM titles_NNNN

WHERE imdb_score > 8.0

ORDER BY imdb_score DESC;

	title text		â	imdb_score numeric
1	#ABtalks			9.6
2	Khawatir			9.5
3	Breaking Bac	I		9.5
4	Our Planet			9.3
5	Avatar: The L	ast Airbender		9.3
6	Reply 1988			9.2
7	My Mister			9.1
8	Major			9.1
9	Chhota Bhee	m & Krishna vs Zimbara		9.1
10	The Last Dance		9.1	
11	Kota Factory		9.1	
12	Okupas		9	
13	Hunter x Hunter		9	
14	DEATH NOTE		9	
15	Arcane		9	
16	Leah Remini:	Scientology and the Aftermath		9
Total	rows: 401	Query complete 00:00:00.258		

Insight:

• A good metric for curating best-of content.

- Use to populate "Top Rated" sections.
- Retain exclusive rights for high-scoring titles.

20.Fetching Movie Titles and Directors

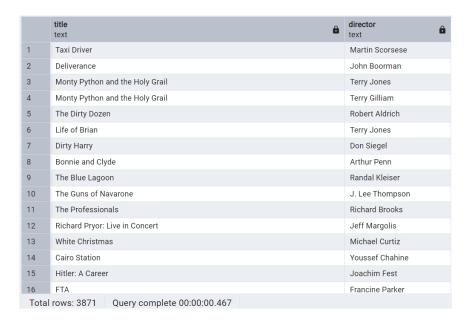
SELECT t.title, c.name AS director

FROM titles_NNNN AS t

JOIN cast_crewA AS c

ON t.id = c.id

WHERE t.type ILIKE 'movie' AND c.role ILIKE 'director';



Insight:

Connects creative ownership with the content.

- Show director names in UI for cinephiles.
- Enable director-based browsing filters.

21. Which movies have an IMDB score greater than 8.0, and who are the directors of these movies?

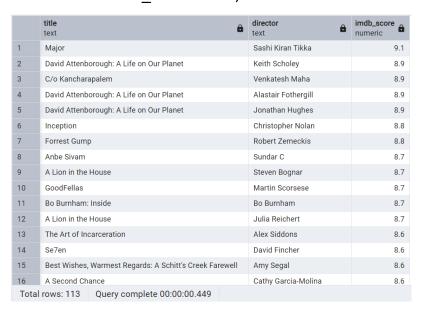
SELECT t.title, c.name AS director, t.imdb_score

FROM titles_NNNN AS t

JOIN cast_crewA AS c

ON t.id = c.id

WHERE t.type ILIKE 'movie' AND c.role ILIKE 'director' AND t.imdb_score > 8.0 ORDER BY t.imdb_score DESC;



Insight:

Combines quality and leadership in content creation.

- Prioritize these movies for content partnerships.
- Feature in award-season campaigns.

Business Impact:

The analysis conducted through SQL has a direct impact on several key business areas:

Content Strategy Optimization: By identifying top-performing genres, countries, directors, and actors, Netflix can prioritize investments in high-yield segments.

Improved User Retention: Insights into viewer preferences and runtime patterns can enhance personalized recommendations, reducing churn.

Increased Engagement: Leveraging data-driven decisions about popular content and recurring actor-character combinations allows for more effective curation and marketing.

Global Market Expansion: Country-based performance and audience ratings reveal opportunities for international growth through targeted localization and partnerships.

Data-Driven Acquisition: Understanding which movies are both critically acclaimed and widely popular can help guide licensing and production strategies.

Areas for Improvement:

While the analysis reveals rich insights, several improvement areas exist for more comprehensive decision-making:

Missing Demographic Data: No information on viewer age, region, or watch time limits personalization potential.

Unstructured Genre Tags: Genre data is currently not normalized, making indepth analysis of sub-genres difficult.

Limited Engagement Metrics: Metrics like completion rate, re-watch value, or session duration were unavailable but would be powerful additions.

No Real-Time Data: The current analysis reflects static data. Trends can shift quickly; real-time or recent data would make the insights more actionable.

Age Certification Inconsistency: The presence of nulls and "N/A" in certifications could be cleaned for better classification and targeting.

Final Strategic Recommendations:

Based on the full analysis:

Expand in High-Performing Genres & Countries: Invest more in Drama, Action, and Comedy genres, and deepen collaboration with production houses in the US, UK, and France.

Leverage Top-Rated Titles in Promotions: Use movies with IMDb > 8 and high TMDB popularity in email campaigns, banners, and push notifications.

Build Actor/Director-Based Navigation Features: Implement filters or spotlight rows for popular actors and directors to aid fan-based discovery.

Target Families with More PG Content: Create a balanced offering by adding more PG/G-rated content for broader family appeal.

Optimize Runtime for Mobile-First Audiences: Introduce more mini-series or content under 90 minutes to cater to shorter attention spans.

Feature Long-Running Shows: Use shows with high season counts as retention anchors—ideal for binge-watchers.

Normalize Genre Tags: Clean and categorize genre data to unlock deeper personalization via sub-genres (e.g., "Romantic Comedies").

Track Emerging Directors: Many recent releases come from newer talent; monitoring their content performance can guide early partnership opportunities.

Conclusion:

This project demonstrates how structured SQL analysis of Netflix's content library can reveal vital trends and actionable insights. From identifying the ingredients of popular content to recognizing gaps in genre or certification coverage, the study supports a smarter, data-backed approach to content strategy.

By refining the dataset further and integrating more real-time and demographic insights, Netflix (or any OTT platform) can fine-tune its content, boost engagement, and improve long-term customer loyalty—making data not just informative but transformative.