**Streaming Platform Movie Analysis: Netflix vs Amazon Prime Video**

**Project Overview:**

This project analyzes a comprehensive dataset of nearly 25,000 movies available on **Netflix** and **Amazon Prime Video**. The dataset contains detailed data including Movie Title, IMDb and Rotten Tomatoes ratings, genres, year released, minimum age Rating, and platform availability.

My goal is to clean and preprocess the raw data, then perform exploratory and comparative analyses to uncover insights about movie ratings, genres, audience age groups, and platform-specific trends.

**Dataset Description:**

* Single raw table with movie fields:
  + Title
  + Year released
  + Minimum recommended age (MinAge)
  + IMDb score, Rotten Tomatoes score
  + Genre(s) (multi-label, comma separated)
  + Platform availability flags (Netflix, Amazon Prime Video)

**Data Cleaning & Preparation:**

* Cleaned textual data: trimmed whitespace, standardized genre names
* Handled missing or null values in key rating and age columns
* Used SQL Server’s STRING\_SPLIT function to normalize multi-genre data into individual genres for flexible analysis
* Cast rating columns to FLOAT to ensure safe numerical calculations
* Created views and CTEs to support complex queries without modifying original data

--creating a database

create database Moviesdb;

--using the database

use Moviesdb;

--table has been imported via import wizard table name-> netflixAmazon

-- I will create a duplicate of the original table to make sure i have the original data in case if i perform any mistakes

select \* into netflixAmazon2 from netflixAmazon;

--i will use and perform operations on table netflixAmazon2

--exploring the table

select top 10 \* from netflixAmazon2;

--i dont require the column column1

alter table netflixAmazon2 drop column column1;

--datatypes of columns

EXEC sp\_help 'dbo.netflixAmazon2';

-- i want to delete data where the movies are not streamed on either platforms

select \* from netflixAmazon2 where Netflix=0 and Amazon\_Prime\_Video=0;

/\*hopefully there are no rows satisfying this condition

if there were any rows i would delete them \*/

--exploring the IMDB column

select distinct IMDB from netflixAmazon2 order by 1;

--i have values like d;} nan

--exploring Rottenm\_Tomatoes column

select distinct Rotten\_Tomatoes from netflixAmazon2 order by 1;

-- i found value like na

--i want to delete data where i dont have any scores from both

select \* from netflixAmazon2

where (IMDB='d;}' or IMDB='nan') and Rotten\_Tomatoes='na';

--delete

delete from netflixAmazon2 where (IMDB='d;}' or IMDB='nan') and Rotten\_Tomatoes='na';

--changing the values d;} nan and na to -1

update netflixAmazon2

set IMDB='-1' where IMDB='d;}' or IMDB='nan';

update netflixAmazon2

set Rotten\_Tomatoes='-1' where Rotten\_Tomatoes='na';

--changing the datatypes

alter table netflixAmazon2 alter column Rotten\_Tomatoes int;

--searching for duplicates

select Title, COUNT(\*) as Count

from netflixAmazon2

group by Title

having COUNT(\*) > 1

order by Count desc;

--

select \* from netflixAmazon2 order by title,genre;

select Title,genre,COUNT(\*) as DuplicateCount

from netflixAmazon2

group by Title, genre

having COUNT(\*) > 1;

alter table netflixAmazon2 add MinAge int;

update netflixAmazon2

set MinAge = TRY\_CasT(REPLACE(Rating, '+', '') as int);

alter table netflixAmazon2 drop column Rating;

--exploring title column

select distinct title from netflixAmazon2;

--removing

update netflixAmazon2

set Title = LTRIM(RTRIM(Title));

--creating a view where genres are merged

create view MergedGenresview as

select Title, Year,MinAge,IMDb,Rotten\_Tomatoes,

STRING\_AGG(Genre, ', ') as CombinedGenres,

MAX(Netflix) as Netflix,

MAX(Amazon\_Prime\_Video) as Amazon\_Prime\_Video

from netflixAmazon2

group by Title, Year, MinAge, IMDb, Rotten\_Tomatoes,Netflix,Amazon\_Prime\_Video;

select \* from MergedGenresView

drop view if exists MergedGenresView;

create view MergedGenresview as

select Title,Year,MinAge,IMDb,Rotten\_Tomatoes,

REPLACE(STRING\_AGG(Genre, ', '), '&', ',') as CombinedGenres,

MAX(Netflix) as Netflix,

MAX(Amazon\_Prime\_Video) as Amazon\_Prime\_Video

from netflixAmazon2

group by Title, Year, MinAge, IMDb, Rotten\_Tomatoes;

select \* from MergedGenresView where IMDB<0 or Rotten\_Tomatoes<0;

delete from netflixAmazon2 where IMDb<0 or Rotten\_Tomatoes<0;

--Data analysis

--1. counting the total number of movies

select COUNT(\*) as total\_movies from MergedGenresView;

A screenshot of a phone number

AI-generated content may be incorrect.

--2. counting number of movies in each platform and their percentage

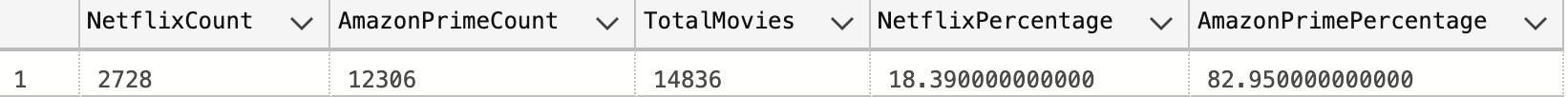
select SUM(Netflix) as NetflixCount, SUM(Amazon\_Prime\_Video) as AmazonPrimeCount,

COUNT(\*) as TotalMovies,

ROUND(100.0 \* SUM(Netflix) / COUNT(\*), 2) as NetflixPercentage,

ROUND(100.0 \* SUM(Amazon\_Prime\_Video) / COUNT(\*), 2) as AmazonPrimePercentage

from MergedGenresView;



--3. Average Rating per platform

select 'Netflix' as Platform, AVG(IMDb) as Avg\_IMDb, AVG(Rotten\_tomatoes) as AVG\_RottenTomatoes

from MergedGenresView where Netflix = 1

union all

select 'Amazon Prime Video' as Platform, AVG(IMDb) as Avg\_IMDb,AVG(Rotten\_tomatoes) as AVG\_RottenTomatoes

from MergedGenresView where Amazon\_Prime\_Video = 1;

A screenshot of a number

AI-generated content may be incorrect.

**Insights**

* Movies on Netflix tend to get better ratings overall. On average, Netflix titles score around 6.2 on IMDb and about 51% on Rotten Tomatoes.
* In comparison, movies on Amazon Prime Video have a slightly lower average, with IMDb ratings near 5.8 and Rotten Tomatoes scores around 42%.
* This means Netflix’s movie library generally receives better reviews from both audiences and critics.

**Recommendations**

* Amazon Prime Video might want to focus on adding higher-quality movies or promoting the best-rated titles more prominently to improve viewers’ perception.
* Netflix seems to be doing well with curating movies that get good ratings, so continuing this approach could help keep their audience happy.
* Both platforms can use these rating trends to feature top-rated movies in recommendations and marketing, helping users discover great content and stay engaged.

--4. Movies per year

select [Year] as year ,count(\*) as num\_movies from MergedGenresView group by [Year] order by 1;

--5 top 5 highest rated titles by IMDB

select top 5 title,IMDB

from MergedGenresView order by 2 desc;

A screenshot of a video game

AI-generated content may be incorrect.

--top 5 highest rated titles by rotten tomato

select top 5 title,Rotten\_Tomatoes

from MergedGenresView order by 2 desc;

A screenshot of a computer

AI-generated content may be incorrect.

--lowest rated movies of all time

select top 5 title,IMDB, Rotten\_Tomatoes

from MergedGenresView order by IMDB,Rotten\_Tomatoes;

A screenshot of a computer

AI-generated content may be incorrect.

--6 top genres by average IMDB rating

select TRIM(value) as Genre, COUNT(\*) as num\_titles, AVG(IMDb) as Avg\_IMDb

from MergedGenresView

cross apply STRING\_SPLIT(CombinedGenres, ',')

group by TRIM(value) having COUNT(\*) > 5

order by Avg\_IMDb desc;

A screenshot of a computer

AI-generated content may be incorrect.

-- 7 Title Count by Platform and Age Category

select Platform, MinAge, COUNT(\*) as NumTitles

from MergedGenresView

cross apply (

VALUES

('Netflix', Netflix),

('Amazon Prime Video', Amazon\_Prime\_Video)

) as Platforms(Platform, IsAvailable)

where IsAvailable = 1 and MinAge is not null

group by Platform, MinAge

order by Platform, MinAge;

A screenshot of a computer

AI-generated content may be incorrect.

-- 8 Average Ratings Per Genre and Platform

select TRIM(value) as Genre,Platform,COUNT(\*) as NumTitles,AVG(IMDb) as Avg\_IMDb,

AVG(Rotten\_Tomatoes) as Avg\_RT

from MergedGenresView

cross apply STRING\_SPLIT(CombinedGenres, ',')

cross apply (

VALUES

('Netflix', Netflix),

('Amazon Prime Video', Amazon\_Prime\_Video)

) as Platforms(Platform, IsAvailable)

where IsAvailable = 1

group by TRIM(value), Platform

order by Genre, Platform;

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a video

AI-generated content may be incorrect.

**Insights**

* Netflix movies generally have higher IMDb and Rotten Tomatoes ratings than Amazon Prime Video across most genres.
* Amazon Prime Video offers a much larger number of movies, especially in Action and Drama genres.
* Documentary, Biography, and Drama genres have the highest average ratings on both platforms.
* Horror, Mystery, and Thriller genres tend to have lower ratings, particularly on Amazon Prime Video.
* Netflix seems to focus on fewer but higher-rated titles, while Amazon Prime prioritizes quantity.

**Recommendations**

* Choose Netflix for better-rated movies and Amazon Prime for more variety.
* Amazon Prime could improve viewer satisfaction by adding higher-quality content in lower-rated genres like Horror and Thriller.
* Use these insights to tailor personalized movie recommendations based on user preferences for quality vs quantity.

-- 9 Year-on-Year Trends in Average IMDb Rating

select Year,COUNT(\*) as NumMovies,AVG(IMDb) as Avg\_IMDb

from MergedGenresView

group by Year

order by Year;

-- 10 Under-Rated movies (High IMDb, Low Rotten Tomatoes)

select Title,Year,IMDb,Rotten\_Tomatoes,CombinedGenres

from MergedGenresView

where IMDb >= 8 AND Rotten\_Tomatoes < 35

order by IMDb desc;

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Insights**

* There are quite a few movies that viewers really love—they have high IMDb ratings (above 8.0)—but critics don’t seem to rate them as highly, with Rotten Tomatoes scores around 30.
* Most of these underrated gems fall into the Documentary, Drama, and Comedy genres, showing that people connect strongly with these kinds of stories.
* This difference suggests that critics and audiences sometimes see things differently, especially when it comes to documentaries or more niche films.
* So, just because a movie doesn’t score high with critics doesn’t mean it’s not worth watching!

**Recommendations**

* If you’re looking for great movies, don’t just rely on critic scores—check out what other viewers are saying too.
* Streaming platforms could do a better job of showcasing these audience favorites that might fly under the radar.
* When recommending movies, it’s a good idea to balance critic reviews with audience ratings to give a fuller picture.
* Digging deeper into why these genres get different reactions could help platforms bring more of these hidden gems to light.

-- 11 Overhyped Movies(High RT, Low IMDb)

select Title,Year,IMDb,Rotten\_Tomatoes,CombinedGenres

from MergedGenresView

where Rotten\_Tomatoes > 70 AND IMDb < 6

order by Rotten\_Tomatoes desc;

A screenshot of a computer

AI-generated content may be incorrect.

**Insights**

* These overhyped movies mostly belong to the Action and Adventure genres, with some Drama mixed in.
* Critics give them high scores (Rotten Tomatoes 70+), but audiences rate them lower (IMDb around 5.4–5.9).
* This shows a clear gap where critics appreciate the production quality and technical aspects, but viewers feel less satisfied, possibly due to story or entertainment value.
* Action and Adventure movies often rely on spectacle, which may appeal more to critics than to general audiences.

**Recommendations**

* Viewers should consider both critic and audience ratings, especially for Action and Adventure films, before choosing to watch.
* Streaming platforms could highlight these rating differences to help viewers make better-informed choices.
* Filmmakers in these genres might focus more on storytelling and audience engagement to reduce the disconnect.
* Understanding why Action and Adventure movies get overhyped can improve future content strategies.

-- 12 Highest Rated Title Per Genre

WITH GenreRanks as (

select TRIM(value) as Genre,Title,Year,IMDb,

ROW\_NUMBER() OVER (PARTITION BY TRIM(value) order by IMDb desc) as rn

from MergedGenresView

cross apply STRING\_SPLIT(CombinedGenres, ',')

)

select \* from GenreRanks

where rn = 1

order by Genre;

-- 13 IMDb Rating Distribution

select FLOOR(IMDb) as RatingBucket,COUNT(\*) as NumTitles

from MergedGenresView

group by FLOOR(IMDb)

order by RatingBucket;

**A screenshot of a computer

AI-generated content may be incorrect.**

**Insights**

* Most movies are rated somewhere in the middle — the biggest group has a rating of 6.
* Ratings around 5 and 6 make up a large chunk of all movies, so a lot of films get average to slightly above-average scores.
* Very few movies have really low or really high ratings, which is pretty normal.
* This shows that most movies tend to be “middle of the road” when it comes to how people rate them.

**Recommendations**

* It’s a good idea to focus on these mid-range movies to figure out what keeps viewers satisfied.
* Don’t forget to highlight the really well-rated movies (7 and above) as those are likely crowd favorites or hidden gems.
* Also, keep an eye on the low-rated ones to understand what doesn’t work and help people avoid movies they might not enjoy.
* Using this mix of ratings can help build better recommendations that balance popular average picks with standout films.

-- 14 Find genres with high ratings but few titles (untapped opportunity)

WITH GenreStats as (

select TRIM(value) as Genre,COUNT(\*) as NumTitles,AVG(IMDb) as Avg\_IMDb

from MergedGenresView

cross apply STRING\_SPLIT(CombinedGenres, ',')

group by TRIM(value)

)

select \* from GenreStats

where NumTitles < 500 AND Avg\_IMDb > 6

order by Avg\_IMDb desc;

**A screenshot of a computer

AI-generated content may be incorrect.**

**Insights**

* Some genres have high average IMDb ratings but relatively few titles, indicating they might be underserved markets.
* For example, Biography (480 titles, avg IMDb ~6.82) and Anime (16 titles, avg IMDb ~6.67) stand out as well-rated but with limited content.
* These genres show strong audience appreciation despite their smaller catalogs.

**Recommendations**

* Streaming platforms could invest more in expanding content for these high-rated but underrepresented genres to attract dedicated audiences.
* Curators and content creators might explore adding more titles in Biography and Anime to meet viewer demand.
* Highlighting these genres in recommendations can help users discover quality content they might otherwise miss.
* Further analysis can identify other niche genres with similar patterns to capitalize on emerging interests.

-- 15 Calculate IMDb rating volatility within genres

select

TRIM(value) as Genre,

COUNT(\*) as NumTitles,

AVG(IMDb) as Avg\_IMDb,

STDEV(IMDb) as StdDev\_IMDb

from MergedGenresView

cross apply STRING\_SPLIT(CombinedGenres, ',')

group by TRIM(value)

having COUNT(\*) > 5

order by StdDev\_IMDb desc;

## A screenshot of a computer AI-generated content may be incorrect.

## **Insights**

* Some genres show high variability in IMDb ratings, meaning the quality of movies within those genres varies widely.
* Fantasy, Action, and Adventure have the highest rating volatility (std dev around 1.33–1.37), suggesting a mixed bag of very good and not-so-good movies.
* Genres like Documentary and Biography have lower volatility (around 1.0), indicating more consistent audience ratings.
* High volatility genres can be riskier for viewers but may also offer standout hits alongside weaker titles.

**Recommendations**

* For genres with high rating volatility (like Fantasy and Action), viewers may want to rely more on individual movie reviews rather than genre alone.
* Streaming services could improve user experience by highlighting top-rated movies within these volatile genres to help viewers find the best content.
* For genres with lower volatility (like Documentary and Biography), general recommendations by genre may be more reliable.
* Content creators might focus on raising the overall quality in high volatility genres to reduce the gap between hits and misses.

### Analysis of Genre Popularity Over Time

* **Overall Trend**:  
  The dataset shows a steady growth in the number of movies produced each year across almost all genres, especially starting from the 1920s and peaking around the 2000s and 2010s. Drama, Comedy, Action, and Adventure consistently dominate in production volume.
* **Dominant Genres by Era**:
  + **Early 1900s (1910s-1920s):** Drama, Comedy, Action, and Adventure were already present but in smaller numbers. Drama appears most frequently in early years.
  + **1930s to 1950s:** Action, Adventure, and Drama show significant growth; Comedy also becomes very prominent.
  + **1960s to 1980s:** The number of titles in Drama, Comedy, Action, and Adventure continues to increase substantially. Documentaries and Animation start appearing more frequently.
  + **1990s to 2010s:** High production volume for Drama and Comedy, with Action and Adventure closely behind. Documentary films see a noticeable rise, and genres like Thriller and Horror also increase steadily.
  + **Recent years (2010-2020):** Drama remains the most produced genre by a wide margin, followed by Comedy, Action, and Adventure. Some decline in production volume in 2020 and 2021, possibly due to external factors (e.g., pandemic).
* **Niche and Emerging Genres**:
  + Genres like Fantasy, Anime, Documentary, and Biography show smaller but steady numbers, with Documentary rising notably in later years.
  + Animation increases after the 1980s, becoming more mainstream.
  + Horror and Thriller grow steadily but remain behind the top genres.
* **Observations on Popularity Fluctuations**:
  + Peaks in production volumes can correspond to technological advancements, cultural trends, or industry shifts (e.g., rise of blockbusters in Action/Adventure).
  + Documentaries have a surge starting mid-20th century, growing consistently.
  + Some genres show fluctuating interest over decades (e.g., Romance peaks mid-20th century and declines slightly afterward).

-- 17 Most Increasing Genre (YOY Growth)

WITH GenreYearCounts as (

select Year,TRIM(value) as Genre,COUNT(\*) as NumTitles

from MergedGenresView

cross apply STRING\_SPLIT(CombinedGenres, ',')

group by Year, TRIM(value)

),

GenreWithLag as (

select \*,

LAG(NumTitles) OVER (PARTITION BY Genre order by Year) as PrevYearCount,

NumTitles - LAG(NumTitles) OVER (PARTITION BY Genre order by Year) as Growth

from GenreYearCounts

)

select top 5 \* from GenreWithLag

where Growth is not null

order by Growth desc;

A screenshot of a data

AI-generated content may be incorrect.

**Insights**

* **Comedy and Drama are the genres with the highest YOY growth spikes** in recent years:
  + **2017 Comedy:** 419 titles, up by 81 titles from 2016.
  + **2015 Drama:** 413 titles, up by 80 titles from 2014.
  + **2013 Drama:** 322 titles, up by 75 titles from 2012.
  + **2014 Comedy:** 281 titles, up by 60 titles from 2013.
  + **2017 Drama:** 515 titles, up by 54 titles from 2016.
* These spikes suggest periods of significant increase in popularity or production volume for these genres.
* The **Drama genre consistently shows strong growth** across multiple years, indicating steady audience interest and industry investment.
* Comedy shows notable surges particularly in 2014 and 2017, potentially linked to market trends, social factors, or popular releases driving more production.

### **Recommendations**

* For content producers or streaming services: Prioritize acquiring or creating **Drama and Comedy content**, especially around years when growth peaks.
* Investigate what specific factors or titles contributed to these growth spikes for strategic planning.
* Track these growth trends further to predict future shifts in genre popularity.

-- 18 Find hidden gems with high ratings in less popular genres

WITH GenrePopularity as (

select TRIM(value) as Genre,count(\*) as GenreCount

from MergedGenresView

cross apply STRING\_SPLIT(CombinedGenres, ',')

group by TRIM(value)

)

select m.Title, m.Year, m.IMDb, m.Rotten\_Tomatoes, TRIM(g.value) as Genre,gp.GenreCount

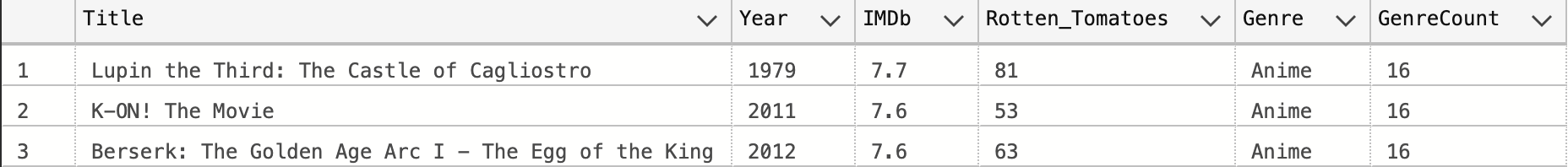
from MergedGenresView m

cross apply STRING\_SPLIT(m.CombinedGenres, ',') g

join GenrePopularity gp ON TRIM(g.value) = gp.Genre

where m.IMDb > 7.5 AND gp.GenreCount < 100

order by m.IMDb desc;



**Insights**

* Some less popular genres, like Anime, have highly rated movies with strong audience approval despite having a small number of titles.
* These "hidden gems" demonstrate quality content exists beyond mainstream genres and can appeal to niche viewers.

**Recommendations**

* Streaming services should spotlight these high-rating hidden gems to diversify viewing options and attract niche audiences.
* Investing in acquiring or producing more content in these underserved genres can boost engagement.
* Additionally, personalized recommendations can help users discover these quality titles, increasing overall satisfaction.

-- 19 Detect highly polarizing titles

-- Big differences between IMDb and Rotten Tomatoes, regardless of which is higher

select Title,Year,IMDb,Rotten\_Tomatoes,ABS((IMDb \* 10) - Rotten\_Tomatoes) as RatingGap,CombinedGenres

from MergedGenresView

where ABS((IMDb \* 10) - Rotten\_Tomatoes) >= 50

order by RatingGap desc;

-- 20 Platform similarity score by genre

-- How similar are Netflix and Prime's genre libraries?

WITH PlatformTotals as (

select

SUM(Netflix) as TotalNetflix,

SUM(Amazon\_Prime\_Video) as TotalPrime

from MergedGenresView

),

GenrePlatformCounts as (

select LTRIM(RTRIM(value)) as Genre,

SUM(Netflix) as Netflix\_Count,

SUM(Amazon\_Prime\_Video) as Prime\_Count

from MergedGenresView

cross apply STRING\_SPLIT(CombinedGenres, ',')

group by LTRIM(RTRIM(value))

),

GenreSimilarity as (

select g.Genre,g.Netflix\_Count,g.Prime\_Count,

case

when pt.TotalNetflix = 0 then 0

else CasT(g.Netflix\_Count as float) / pt.TotalNetflix

end as Netflix\_Share,

case

when pt.TotalPrime = 0 then 0

else CasT(g.Prime\_Count as float) / pt.TotalPrime

end as Prime\_Share

from GenrePlatformCounts g

cross join PlatformTotals pt

)

select Genre, Netflix\_Count, Prime\_Count, Netflix\_Share, Prime\_Share,

1.0 - ABS(Netflix\_Share - Prime\_Share) as SimilarityScore

from GenreSimilarity

order by SimilarityScore desc;

-- 21 IMDb volatility over time

-- Do certain years produce wildly inconsistent ratings?

select Year, COUNT(\*) as NumTitles, AVG(IMDb) as Avg\_IMDb, STDEV(IMDb) as StdDev\_IMDb

from MergedGenresView

group by Year

order by StdDev\_IMDb desc;

-- 22 Genre affinity mapping

-- Identify genres that co-occur frequently

WITH GenrePairs as (

select distinct a.Title, TRIM(x.value) as Genre1,TRIM(y.value) as Genre2

from MergedGenresView a

cross apply STRING\_SPLIT(CombinedGenres, ',') x

cross apply STRING\_SPLIT(CombinedGenres, ',') y

where TRIM(x.value) < TRIM(y.value)

)

select top 5 Genre1, Genre2, COUNT(\*) as PairCount

from GenrePairs

group by Genre1, Genre2

having COUNT(\*) > 5

order by PairCount desc;

-- 23 Cohort analysis: average IMDb by release decade

-- Helps see how ratings differ across eras

select (Year / 10) \* 10 as Decade,COUNT(\*) as NumTitles,AVG(IMDb) as Avg\_IMDb

from MergedGenresView

group by (Year / 10) \* 10

order by Decade;

--24 genres where Netflix has higher average IMDb than Amazon Prime Video

WITH GenrePlatformAvg as (

select LTRIM(RTRIM(g.value)) as Genre,m.Netflix,m.Amazon\_Prime\_Video,m.IMDb

from MergedGenresView m

cross apply STRING\_SPLIT(m.CombinedGenres, ',') g

where m.IMDb is not null

),

AvgIMDbByPlatform as (

select Genre,'Netflix' as Platform,AVG(IMDb) as AvgIMDb

from GenrePlatformAvg

where Netflix = 1

group by Genre

union all

select Genre, 'Amazon Prime Video' as Platform,AVG(IMDb) as AvgIMDb

from GenrePlatformAvg

where Amazon\_Prime\_Video = 1

group by Genre

),

AvgIMDbComparison as (

select n.Genre,n.AvgIMDb as NetflixAvgIMDb,p.AvgIMDb as PrimeAvgIMDb,n.AvgIMDb - p.AvgIMDb as IMDbDifference

from

(select Genre, AvgIMDb from AvgIMDbByPlatform where Platform = 'Netflix') n

inner join

(select Genre, AvgIMDb from AvgIMDbByPlatform where Platform = 'Amazon Prime Video') p

ON n.Genre = p.Genre

)

select \* from AvgIMDbComparison

where IMDbDifference > 0

order by IMDbDifference desc;

A screenshot of a computer

AI-generated content may be incorrect.

--25 corelation betweeen IMDB and Rotten tomatoes

with stats as (

select

COUNT(\*) as n,

SUM(CasT(IMDb as float)) as sum\_x,

SUM(CasT(Rotten\_Tomatoes as float)) as sum\_y,

SUM(CasT(IMDb as float) \* CasT(Rotten\_Tomatoes as float)) as sum\_xy,

SUM(CasT(IMDb as float) \* CasT(IMDb as float)) as sum\_xx,

SUM(CasT(Rotten\_Tomatoes as float) \* CasT(Rotten\_Tomatoes as float)) as sum\_yy

from MergedGenresview

where IMDb is not null AND Rotten\_Tomatoes is not null

)

select

case

when (n \* sum\_xx - sum\_x \* sum\_x) = 0 OR (n \* sum\_yy - sum\_y \* sum\_y) = 0 then null

else

(n \* sum\_xy - sum\_x \* sum\_y)

/

SQRT( (n \* sum\_xx - sum\_x \* sum\_x) \*(n \* sum\_yy - sum\_y \* sum\_y))

end as Correlation

from stats;

**Insights**

* Netflix outperforms Amazon Prime Video in average IMDb ratings across multiple genres, notably Horror (+0.87), Action (+0.57), and Adventure (+0.57).
* Genres like Comedy, Drama, and Crime also show a consistent but smaller edge for Netflix.
* This suggests Netflix’s catalog in these genres is perceived as higher quality by viewers.

**Recommendations**

* Amazon Prime Video could focus on improving or expanding content in genres where it trails Netflix, especially high-engagement genres like Horror, Action, and Comedy.
* Netflix should continue to highlight these genre strengths in marketing to maintain its competitive edge.
* Both platforms can analyze why certain genres perform better and apply these insights to content acquisition and production strategies.

**Consolidated Insights**

1. **Platform Ratings**

* **Netflix generally has higher average ratings** than Amazon Prime Video:
  + Netflix: Avg IMDb ≈ 6.2, Rotten Tomatoes ≈ 51%
  + Amazon Prime Video: Avg IMDb ≈ 5.8, Rotten Tomatoes ≈ 42%
* Netflix’s content is perceived as higher quality, while Amazon Prime offers **more titles overall**, especially in genres like Action and Drama.

1. **Genre-Specific Findings**

* **Drama, Comedy, and Documentary** consistently have higher average ratings on both platforms.
* **Horror, Mystery, and Thriller** genres tend to have **lower average ratings**, especially on Amazon Prime Video.
* Some genres, like **Biography (≈6.82 avg IMDb)** and **Anime (≈6.67 avg IMDb)**, are well-rated but **underrepresented** in terms of the number of titles.
* **High volatility genres** (Fantasy, Action, Adventure) have widely varying ratings, suggesting quality is hit-or-miss.
* **Documentary and Biography** genres have **low volatility**, showing more consistent quality.

1. **Audience vs. Critic Perception**

* Some movies are **hidden gems**—high IMDb ratings (8+) but low Rotten Tomatoes scores (~30), especially in Documentary, Drama, and Comedy.
* Conversely, some **overhyped titles** (mainly Action and Adventure) have high critic scores (RT 70+) but low audience scores (IMDb ~5.4–5.9).
* Most movies cluster around **mid-level ratings** (IMDb 5–6), showing that the majority are “average” in audience perception.

1. **Time Trends**

* Overall movie production **increases steadily across decades**, peaking in the 2000s and 2010s.
* **Drama and Comedy** show strong growth, with notable spikes in years like 2015 and 2017.
* **Documentary, Anime, and Biography** genres remain smaller but steadily grow over time.

**Consolidated Recommendations**

1. **For Streaming Platforms**

**Netflix**

* Continue emphasizing **quality over quantity**—this is resonating well with viewers.
* Promote strengths in genres like **Horror, Action, and Adventure** where ratings lead over Prime.
* Use high ratings as a marketing tool to differentiate from competitors.

**Amazon Prime Video**

* Improve content **quality in lower-rated genres** like Horror, Thriller, and Mystery.
* Promote well-rated titles more effectively to enhance viewer perception.
* Consider reducing reliance on sheer volume and focus on **curated, higher-quality offerings.**

1. **For Both Platforms**

* **Invest in underrepresented but high-rated genres** like Biography and Anime to capture niche, loyal audiences.
* Leverage insights from **audience vs. critic gaps**:
  + Promote hidden gems that audiences love but critics underrate.
  + Highlight rating differences for overhyped movies to help users make informed choices.
* In high-volatility genres (Fantasy, Action, Adventure), **spotlight top-rated titles** rather than recommending genre broadly.
* Keep a balance of **mid-range movies** with crowd-pleasers and standout hits in recommendations.
* Monitor **genre production trends** to anticipate shifts in viewer interests and tailor content acquisition.

1. **For Viewers**

* Use **both critic and audience scores** to decide what to watch—each offers valuable perspective.
* Explore **hidden gems** in niche genres for fresh viewing experiences.
* Be cautious of overhyped titles, especially in Action and Adventure, and check audience ratings before watching.

**Technologies Used**

* Microsoft SQL Server for data cleaning and analysis
* Azure data studio, docker
* Use of advanced SQL features like STRING\_SPLIT, CTEs, joins and aggregate functions

**Files Included**

- `SQQuery\_1.sql` - All SQL queries used

- `moviesdb.csv` - Cleaned dataset used for analysis

- `README.md` - Project overview and insights

**How to Use**

1. Load your raw movie dataset into SQL Server
2. Run the provided cleaning and transformation SQL scripts
3. Execute analysis queries to reproduce insights or customize for your needs