



NETFLIX MOVIES AND TV SHOWS DATA ANALYSIS USING SQL

NETFLIX

TICKET \$15

NOVEMBER 14, 2024

PROJECT DONE BY
ARPITA CHAKROBORTY



OVERVIEW :

THIS PROJECT INVOLVES A COMPREHENSIVE ANALYSIS OF NETFLIX'S MOVIES AND TV SHOWS DATA USING SQL. THE GOAL IS TO EXTRACT VALUABLE INSIGHTS AND ANSWER VARIOUS BUSINESS QUESTIONS BASED ON THE DATASET. THE FOLLOWING README PROVIDES A DETAILED ACCOUNT OF THE PROJECT'S OBJECTIVES, BUSINESS PROBLEMS, SOLUTIONS, FINDINGS, AND CONCLUSIONS

OBJECTIVES :

- ANALYZE THE DISTRIBUTION OF CONTENT TYPES (MOVIES VS TV SHOWS).
- IDENTIFY THE MOST COMMON RATINGS FOR MOVIES AND TV SHOWS.
- LIST AND ANALYZE CONTENT BASED ON RELEASE YEARS, COUNTRIES, AND DURATIONS.
- EXPLORE AND CATEGORIZE CONTENT BASED ON SPECIFIC CRITERIA AND KEYWORDS.



NETFLIX

- DATA IMPORT USING PANDAS



```
from sqlalchemy import create_engine
conn_string = 'mysql+pymysql://root:myNewPass123!@localhost/netflix_project'
db= create_engine(conn_string)
conn=db.connect()

df= pd.read_csv("E:\\netflix_titles.csv")

print(df.info)
df.describe
```

	show_id	type	title	director	release_date	imdb_score	popularity
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	2017-01-01	7.0	99.0
1	s2	TV Show	Blood & Water		2018-01-01	7.0	99.0
2	s3	TV Show	Ganglands	Julien Leclercq	2018-01-01	7.0	99.0
3	s4	TV Show	Jailbirds New Orleans		2018-01-01	7.0	99.0
4	s5	TV Show	Kota Factory		2018-01-01	7.0	99.0
...	7.0	99.0
8802	s8803	Movie	Zodiac	David Fincher	2013-01-01	8.2	99.0
8803	s8804	TV Show	Zombie Dumb		2018-01-01	7.0	99.0
8804	s8805	Movie	Zombieland	Ruben Fleischer	2009-01-01	8.2	99.0
8805	s8806	Movie	Zoom	Peter Hewitt	2012-01-01	7.0	99.0
8806	s8807	Movie	Zubaan	Mozez Singh	2018-01-01	7.0	99.0



BUSINESS PROBLEMS AND SOLUTIONS

1. COUNT THE NUMBER OF MOVIES VS TV SHOWS

```
select type ,count(*) as count_of_type  
from netflix_titles  
group by type;
```

Result Grid	
type	count_of_type
Movie	6131
TV Show	2676

2. LIST ALL MOVIES RELEASED IN A SPECIFIC YEAR (E.G., 2020)

```
select type,title, release_year  
from netflix_titles  
where release_year='2020' and type='movie';
```

type	title	release_year
Movie	Dick Johnson Is Dead	2020
Movie	Europe's Most Dangerous Man: Otto Skorzeny i...	2020
Movie	Tughlaq Durbar	2020
Movie	Omo Ghetto: the Saga	2020
Movie	Shadow Parties	2020





3. FIND THE MOST COMMON RATING FOR MOVIES AND TV SHOWS



```
with rating_count as
(select type, rating, count(*) as rating_count
from netflix_titles
group by 1,2),
rnk as (
select type, rating, rating_count
,rank() over (partition by type order by rating_count desc) as rnk
from rating_count
)
select type, rating as most_common_rating
from rnk
where rnk=1;
```

Result Grid		Filter Rows:
	type	most_common_rating
▶	Movie	TV-MA
	TV Show	TV-MA





4. FIND THE TOP 5 COUNTRIES WITH THE MOST CONTENT ON NETFLIX

```
SELECT country, COUNT(*) AS total_content
FROM (
    SELECT
        SUBSTRING_INDEX(SUBSTRING_INDEX(country, ',', numbers.n), ',', -1) AS country
    FROM netflix_titles
    JOIN (
        SELECT 1 AS n UNION ALL SELECT 2 UNION ALL SELECT 3 UNION ALL SELECT 4 UNION ALL SELECT 5
        UNION ALL SELECT 6 UNION ALL SELECT 7 UNION ALL SELECT 8 UNION ALL SELECT 9 UNION ALL SELECT 10
    ) AS numbers ON CHAR_LENGTH(country) - CHAR_LENGTH(REPLACE(country, ',', '')) >= numbers.n - 1
) AS t1
WHERE country IS NOT NULL
GROUP BY country
ORDER BY total_content DESC
LIMIT 5;
```

Result Grid		Filter Rows:
	country	total_content
▶	United States	3211
	India	1008
	United Kingdom	628
	United States	478
	Canada	271





5. IDENTIFY THE LONGEST MOVIE

```
SELECT
    title,duration
FROM netflix_titles
WHERE type = 'Movie'
ORDER BY CAST(SUBSTRING_INDEX(duration, ' ', 1) AS UNSIGNED) DESC
limit 1;
```

Result Grid		Filter Rows:
	title	duration
▶	Black Mirror: Bandersnatch	312 min

6. FIND CONTENT ADDED IN THE LAST 5 YEARS

```
select *
from netflix_titles
where str_to_date(date_added,'%M %D %Y') >= curdate() - interval 5 year;
```

	show_id	type	title	director	cast	country	date_added	release_y
▶	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	HULL	United States	September 25, 2021	2020
	s2	TV Show	Blood & Water	HULL	Ama Qamata, Khosi Ngema, Gail Mabalane, Thabang Molaba, ...	South Africa	September 24, 2021	2021
	s3	TV Show	Ganglands	Julien Ledercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabiba Akkari, Sofi...	HULL	September 24, 2021	2021
	s4	TV Show	Jailbirds New Orleans	HULL	HULL	HULL	September 24, 2021	2021



NETFLIX

7. FIND ALL MOVIES/TV SHOWS BY DIRECTOR 'RAJIV CHILAKA'

```
SELECT *
FROM NETFLIX_TITLES
WHERE TYPE = 'MOVIE' AND DIRECTOR='RAJIV CHILAKA';
```

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
s407	Movie	Chhota Bheem - Neeli Pahaadi	Rajiv Chilaka	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jigna Bhardwaj, R...	HULL	July 22, 2021	2013	TV-Y7	64 min	Children & Family Movies
s408	Movie	Chhota Bheem & Ganesh	Rajiv Chilaka	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jigna Bhardwaj, R...	HULL	July 22, 2021	2009	TV-Y7	68 min	Children & Family Movies
s409	Movie	Chhota Bheem & Krishna: Mayanagar	Rajiv Chilaka	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jigna Bhardwaj, R...	HULL	July 22, 2021	2011	TV-Y7	67 min	Children & Family Movies
s410	Movie	Chhota Bheem & Krishna: Pataliputra- City of th...	Rajiv Chilaka	Vatsal Dubey, Julie Tejwani, Rupa Bhimani, Jigna Bhardwaj, R...	HULL	July 22, 2021	2010	TV-Y7	66 min	Children & Family Movies

8. LIST ALL TV SHOWS WITH MORE THAN 5 SEASONS

```
SELECT *
FROM NETFLIX_TITLES
WHERE TYPE= 'TV SHOW'
AND CAST(SUBSTRING_INDEX(DURATION, ' ',1) AS UNSIGNED ) > 5;
```

	type	title
▶	TV Show	The Great British Baking Show
	TV Show	Nailed It
	TV Show	Numberblocks
	TV Show	Saved by the Bell
	TV Show	Lucifer



9. COUNT THE NUMBER OF CONTENT ITEMS IN EACH GENRE



```
SELECT
    genre,
    COUNT(*) AS total_content
FROM (
    SELECT
        TRIM(SUBSTRING_INDEX(SUBSTRING_INDEX(listed_in, ',', numbers.n), ',', -1)) AS genre
    FROM netflix_TITLES
    JOIN (
        SELECT 1 AS n UNION ALL SELECT 2 UNION ALL SELECT 3 UNION ALL SELECT 4 UNION ALL SELECT 5
        UNION ALL SELECT 6 UNION ALL SELECT 7 UNION ALL SELECT 8 UNION ALL SELECT 9 UNION ALL SELECT 10
    ) numbers ON CHAR_LENGTH(listed_in) - CHAR_LENGTH(REPLACE(listed_in, ',', '')) >= numbers.n - 1
) AS genres
WHERE genre IS NOT NULL
GROUP BY genre;
```

Result Grid		Filter Rows:
genre	total_content	
Documentaries	869	
TV Mysteries	98	
TV Dramas	763	
International TV Shows	1351	
TV Action & Adventure	168	



10. FIND EACH YEAR AND THE AVERAGE NUMBERS OF CONTENT RELEASE IN INDIA ON NETFLIX

```
SELECT
    country,
    release_year,
    COUNT(show_id) AS total_release,
    ROUND(
        COUNT(show_id) /
        (SELECT COUNT(show_id) FROM netflix_TITLES WHERE country = 'India') * 100, 2
    ) AS avg_release
FROM netflix_TITLES
WHERE country = 'India'
GROUP BY country, release_year
ORDER BY avg_release DESC
LIMIT 5;
```

country	release_year	total_release	avg_release
India	2017	101	10.39
India	2018	94	9.67
India	2019	87	8.95
India	2020	75	7.72
India	2016	73	7.51

11. LIST ALL MOVIES THAT ARE DOCUMENTARIES

```
SELECT
    type, title, listed_in
FROM netflix_titles
WHERE type = 'Movie' and listed_in= 'documentaries'
ORDER BY CAST(SUBSTRING_INDEX(listed_in, ' ', 1) AS UNSIGNED) DESC
;
```

type	title	listed_in
Movie	Dick Johnson Is Dead	Documentaries
Movie	My Heroes Were Cowboys	Documentaries
Movie	Final Account	Documentaries
Movie	Bob Ross: Happy Accidents, Betrayal & Greed	Documentaries
Movie	Lady Boss: The Jackie Collins Story	Documentaries





12. FIND ALL CONTENT WITHOUT A DIRECTOR

```
select * from netflix_titles  
where director is null;
```

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
s2	TV Show	Blood & Water	NULL	Ama Qamata, Khosi Ngema, Gail Mabalane, Thabang Molaba, ...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows,
s4	TV Show	Jailbirds New Orleans	NULL	NULL	NULL	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV
s5	TV Show	Kota Factory	NULL	Mayur More, Jitendra Kumar, Ranjan Raj, Alam Khan, Ahsaas ...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows,
s11	TV Show	Vendetta: Truth, Lies and The Mafia	NULL	NULL	NULL	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, Docuse





13. FIND HOW MANY MOVIES ACTOR 'SALMAN KHAN' APPEARED IN THE LAST 10 YEARS

```
select *
from netflix_titles
where cast like '%salman khan%' and str_to_date(date_added,'%M %D %Y') >= curdate() - interval 10 year
```

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
s1442	Movie	Hello Brother	Sohail Khan	Salman Khan, Arbaaz Khan, Rani Mukerji, Johnny Lever, Neeraj...	India	January 4, 2021	1999	TV-14	126 min	Action & Adventure, Comedies, In
s1675	Movie	Bodyguard	Siddique	Salman Khan, Kareena Kapoor, Raj Babbar, Asrani, Vidya Sinh...	India	November 19, 2020	2011	TV-14	130 min	Action & Adventure, Comedies, In
s2340	Movie	Prem Ratan Dhan Payo	Sooraj R. Barjatya	Salman Khan, Sonam Kapoor, Anupam Kher, Neil Nitin Mukesh,...	India	June 25, 2020	2015	TV-14	164 min	Action & Adventure, Dramas, Inte
s2546	Movie	Andaz Apna Apna	Rajkumar Santoshi	Aamir Khan, Salman Khan, Raveena Tandon, Karisma Kapoor, ...	India	May 11, 2020	1994	TV-14	160 min	Action & Adventure, Comedies, Cl





14. FIND THE TOP 10 ACTORS WHO HAVE APPEARED IN THE HIGHEST NUMBER OF MOVIES PRODUCED IN INDIA

```
• SELECT
    actor,
    COUNT(*) AS actor_count
  FROM (
    SELECT
      TRIM(SUBSTRING_INDEX(SUBSTRING_INDEX(cast, ',', numbers.n), ',', -1)) AS actor
    FROM netflix_titles
    JOIN (
      SELECT 1 AS n UNION ALL SELECT 2 UNION ALL SELECT 3 UNION ALL SELECT 4 UNION ALL SELECT 5
      UNION ALL SELECT 6 UNION ALL SELECT 7 UNION ALL SELECT 8 UNION ALL SELECT 9 UNION ALL SELECT 10
    ) numbers ON CHAR_LENGTH(cast) - CHAR_LENGTH(REPLACE(cast, ',', '')) >= numbers.n - 1
    WHERE country = 'India'
  ) AS actor_list
  WHERE actor IS NOT NULL
  GROUP BY actor
  ORDER BY actor_count DESC
  LIMIT 10;
```

Result Grid		Filter Rows:
	actor	actor_count
▶	Anupam Kher	39
	Shah Rukh Khan	32
	Naseeruddin Shah	29
	Amitabh Bachchan	28
	Akshay Kumar	28
	Paresh Rawal	27



15. CATEGORIZE CONTENT BASED ON THE PRESENCE OF 'KILL' AND 'VIOLENCE' KEYWORDS



```
• SELECT
    category,
    COUNT(*) AS content_count
  FROM (
    SELECT
      CASE
        WHEN description LIKE '%kill%' OR description LIKE '%violence%' THEN 'Bad'
        ELSE 'Good'
      END AS category
    FROM netflix_titles
  ) AS categorized_content
 GROUP BY category;
```

Result Grid		
	category	content_count
▶	Good	8465
	Bad	342

FINDINGS AND CONCLUSION



- **CONTENT DISTRIBUTION:** THE DATASET CONTAINS A DIVERSE RANGE OF MOVIES AND TV SHOWS WITH VARYING RATINGS AND GENRES.
- **COMMON RATINGS:** INSIGHTS INTO THE MOST COMMON RATINGS PROVIDE AN UNDERSTANDING OF THE CONTENT'S TARGET AUDIENCE.
- **GEOGRAPHICAL INSIGHTS:** THE TOP COUNTRIES AND THE AVERAGE CONTENT RELEASES BY INDIA HIGHLIGHT REGIONAL CONTENT DISTRIBUTION.
- **CONTENT CATEGORIZATION:** CATEGORIZING CONTENT BASED ON SPECIFIC KEYWORDS HELPS IN UNDERSTANDING THE NATURE OF CONTENT AVAILABLE ON NETFLIX.

