

NETFLIX

16+

CC

THE WITCHER

2020 | English

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do nisin eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris ut aliquip ex ea commodo

Action, Drama



Season 1 ▼



WELCOME

I'm excited to present my SQL-based data analysis project focused on uncovering insights from Netflix's vast content library. As someone passionate about data storytelling and analytics, I took this opportunity to explore trends, patterns, and hidden gems within the entertainment landscape using structured queries and real-world datasets.

Regards,
Aman



INTRODUCTION

In this project, I conducted an in-depth exploratory data analysis (EDA) of Netflix's content library using SQL to uncover valuable insights about global entertainment trends. By leveraging structured query language, I explored various aspects such as the distribution of movies vs. TV shows, common content ratings, genre-wise classifications, and regional production trends. The analysis also dives into content availability over time, the contribution of specific directors like Rajiv Chilaka, actor appearances (e.g., Salman Khan), and keyword-based content categorization (e.g., themes involving 'kill' or 'violence'). Notably, I also evaluated the longest film on the platform, tracked India's annual content releases, and identified the most represented countries in the Netflix catalog. This project showcases my proficiency in SQL, data storytelling, and analytical thinking, making it a strong representation of how structured data can yield meaningful business and entertainment insights.

DATA OVERVIEW

This project utilizes a structured dataset comprising over 8,000 unique titles from Netflix's global content catalog. Each entry represents an individual movie or TV show and contains a diverse range of metadata that enabled a comprehensive analysis of content trends, regional representation, and user engagement indicators.

Key Attributes in the Dataset Include -

- Title – The official name of the content
- Type – Classification as either Movie or TV Show
- Director – Name of the director associated with the title (where available)
- Cast – Featured actors and actresses
- Country – Country or countries where the content was produced
- Date Added – Date when the content became available on Netflix
- Release Year – Original release year of the content
- Rating – Content rating for age appropriateness (e.g., PG, TV-MA, etc.)
- Duration – Runtime (for movies) or number of seasons (for TV shows)
- Genres – Categories or genres the title falls under
- Description – A brief synopsis of the content.

The dataset was subjected to thorough preprocessing and cleaned for null or inconsistent values. It served as the foundation for writing optimized SQL queries that revealed actionable insights related to viewer trends, content lifecycle, genre distribution, and regional performance.

PROJECT OBJECTIVE

The primary objective of this project was to conduct a structured and insightful analysis of Netflix's content dataset using SQL. The aim was to extract meaningful patterns, trends, and metrics that could provide a deeper understanding of the platform's content strategy, regional growth, and audience preferences.

This project focused on:

- Categorizing and comparing the distribution of Movies vs TV Shows on Netflix
- Identifying the most common content ratings and their implications
- Extracting year-wise insights, such as the volume of content released in specific years like 2020
- Analyzing geographic trends to highlight the top content-producing countries
- Investigating the lengthiest movie and recently added content
- Filtering content by notable directors such as Rajiv Chilaka and actors like Salman Khan
- Highlighting high-volume TV series with more than 5 seasons
- Generating a genre-wise breakdown of content volume
- Exploring regional release trends, especially Indian content over the years
- Tagging and classifying content based on thematic keywords such as "Kill" and "Violence"
- Identifying gaps such as content without listed directors.

Through these objectives, the project seeks to showcase how structured SQL queries can be leveraged to derive actionable insights from large-scale entertainment datasets.



```
30
31  -- Q1. Count the number of movies vs TV Shows. --
32
33  ✓ select
34      type,
35      count(*) as total_content
36  from netflix
37  group by type
38
39
40
41
42
43
44
45
46  |
```

Data Output Messages Notifications

≡+ 📄 ▼ 📋 ▼ 🗑️ 🗄️ ⬇️ 📈 SQL

	type character varying (10) 🔒	total_content bigint 🔒
1	Movie	6131
2	TV Show	2676

-- Count the number of movies Vs Tv shows. --

-- Find the most common rating for the
Movies and Tv Shows. --

```
40
41 -- Q2. Find the most common rating for the movies and TV Shows. --
42
43 select
44     type,
45     rating
46 from
47
48 (
49 select
50     type,
51     rating,
52     count(*),
53     rank() over(partition by type order by count(*) desc) as ranking
54 from netflix
55 group by 1, 2
56 ) as t1
57 where
58     ranking = 1
59
```

Data Output Messages Notifications



	type character varying (10) 🔒	rating character varying (10) 🔒
1	Movie	TV-MA
2	TV Show	TV-MA


```

59
60
61 -- Q3. List all movies released in a specific year (e.g., 2020).--
62
63 select * from netflix
64 where
65     type = 'Movie'
66     and
67     release_year = 2020
68
69
70
71
72
73

```

Data Output Messages Notifications

		country character varying (150)	date_added character varying (50)	release_year integer	rating character
1		United States	September 25, 2021	2020	PG-13
2		[null]	September 22, 2021	2020	TV-MA
3		[null]	September 11, 2021	2020	TV-14
4		Nigeria	September 10, 2021	2020	TV-MA
5		[null]	September 6, 2021	2020	TV-MA
6		[null]	September 2, 2021	2020	TV-MA
7		India	September 2, 2021	2020	TV-14
8		United States	August 25, 2021	2020	TV-MA

Total rows: 517 Query complete 00:00:00.120

-- List all Movies released in a specific year (e.g., 2020). --

-- Find the top 5 countries with the most content on Netflix. --

```
-- Q4. Find the top 5 countries with the most content on Netflix.--  
  
select  
    unnest(string_to_array(country, ',')) as new_country,  
    count(show_id) as total_content  
from netflix  
group by 1  
order by 2 desc  
limit 5
```

Data Output Messages Notifications



	new_country text	total_content bigint
1	United States	3211
2	India	1008
3	United Kingd...	628
4	United States	479
5	Canada	271

```

80
81 -- Q5. Identify the longest movie.--
82
83 select * from netflix
84 where
85     type = 'Movie'
86     and
87     duration = (select max(duration) from netflix)
88
89
90
91
92
93

```

Data Output Messages Notifications

	release_year integer	rating character varying (10)	duration character varying (15)	listed_in character varying (100)
1	2002	TV-14	99 min	Action & Adventure, Anime Features, International Movies
2	2003	TV-14	99 min	Action & Adventure, Anime Features, International Movies
3	2020	TV-MA	99 min	Dramas, International Movies, Romantic Movies
4	2007	R	99 min	Comedies, Romantic Movies
5	1994	PG	99 min	Children & Family Movies, Comedies, Dramas
6	2021	TV-MA	99 min	Dramas
7	2020	R	99 min	Action & Adventure
8	2014	TV-MA	99 min	Dramas, Independent Movies, Romantic Movies

Total rows: 118 Query complete 00:00:00.143

-- Identify the Longest Movie. --

-- Find content added in last the
5 years. --

```
90 -- Q6. Find content added in the last 5 year.--
91
92 select
93     *
94 from netflix
95 where
96     to_date(date_added, 'Month DD, YYYY') >= current_date - interval '5 years'
```

Data Output Messages Notifications

SQL

Showing rows: 1 to 1000 Page

country character varying (150)	date_added character varying (50)	release_year integer	rating character varying (10)
United States	September 25, 2021	2020	PG-13
South Africa	September 24, 2021	2021	TV-MA
[null]	September 24, 2021	2021	TV-MA
[null]	September 24, 2021	2021	TV-MA
India	September 24, 2021	2021	TV-MA
[null]	September 24, 2021	2021	TV-MA
[null]	September 24, 2021	2021	PG
United States, Ghana, Burkina Faso, United Kingdom, Germany, Ethiopia	September 24, 2021	1993	TV-MA
United Kingdom	September 24, 2021	2021	TV-14

Total rows: 2759 Query complete 00:00:00.180

98

99

100

101

102

103

104

105

106

107

108

109

110

-- Q7. Find all the movies/TV shows by director 'Rajiv chilaka'.--

select * from netflix

where director ilike '%Rajiv Chilaka%'

Data Output

Messages







Notifications

-- Find all the movies/Tv shows by Director ‘Rajiv Chilaka’. --

-- List all Tv shows with more than 5 seasons. --

```
106 -- Q8. List all TV shows with more than 5 seasons.--
107
108 select
109     *
110 from netflix
111 where
112     type = 'TV Show'
113     and
114     split_part(duration, ' ', 1)::numeric > 5
115
116
117
```

Data Output Messages Notifications

	 date_added character varying (50)	 release_year integer	 rating character varying (10)	 duration character varying (15)	 listed_in character varying (100)	
1	September 24, 2021	2021	TV-14	9 Seasons	British TV Shows, Reality TV	
2	September 15, 2021	2021	TV-PG	6 Seasons	Reality TV	
3	September 15, 2021	2021	TV-Y	6 Seasons	Kids' TV	
4	September 15, 2021	1994	TV-PG	9 Seasons	Kids' TV, TV Comedies	
5	September 10, 2021	2021	TV-14	6 Seasons	Crime TV Shows, TV Comedies, TV Dramas	
6	August 13, 2021	2021	TV-MA	7 Seasons	TV Comedies, TV Dramas	
7	August 1, 2021	2012	TV-14	7 Seasons	TV Comedies	
8	August 1, 2021	2014	TV-14	6 Seasons	Anime Series, International TV Shows	
9	July 28, 2021	2021	TV-14	7 Seasons	Crime TV Shows, TV Action & Adventure, TV Sci-Fi & Fantasy	
10	July 26, 2021	2019	TV-MA	10 Seasons	Classic & Cult TV, TV Action & Adventure, TV Dramas	

```

116
117 -- Q9. Count the number of content items in each genre.--
118
119 select
120     unnest(string_to_array(listed_in, ',')) as genre,
121     count(show_id) as total_content
122 from netflix
123 group by 1
124
125
126
127

```

-- Count the number of content items in each Genre. --

Data Output Messages Notifications



	genre text	total_content bigint
1	International TV Shows	774
2	Stand-Up Comedy	9
3	Spanish-Language TV Shows	172
4	Romantic Movies	613
5	Anime Features	21
6	TV Horror	11
7	Stand-Up Comedy & Talk Shows	34
8	Kids' TV	63
9	Docuseries	221
10	Children & Family Movies	605
11	Movies	57

Total rows: 73 Query complete 00:00:00.060

-- Find each year and the average numbers of content release by India on Netflix. --

-- Return the top 5 year with highest average content release. --

```
125
126 -- Q10. Find each year and the average numbers of contentnrelease by india on netflix.--
127 -- Return the top 5 year with highest avg content release.--
128
129
130 select
131     extract(year from to_date(date_added, 'Month DD, YYYY')) as date,
132     count(*),
133     round(
134         count(*)::numeric/(select count(*) from netflix where country = 'India')::numeric * 100
135     ,2)as avg_content_per_year
136 from netflix
137 where country = 'India'
138 group by 1
139
```

Data Output				Messages	Notifications
	date numeric	count bigint	avg_content_per_year numeric		
1	2018	333	34.26		
2	2016	10	1.03		
3	2019	203	20.88		
4	2021	95	9.77		
5	2020	189	19.44		
6	2017	142	14.61		

```

140
141 -- Q11. List all movies that are documentaries.--
142
143 select * from netflix
144 where listed_in ilike '%documentaries%'
145
146
147
148
149
150

```

Data Output Messages Notifications



	duration character varying (15)	listed_in character varying (100)	description character varying (255)
1	10 min	Documentaries	As her father
2	17 min	Documentaries, International Movies	Declassifi
3	13 min	Documentaries	Robin Wilt
4	13 min	Documentaries, International Movies, Sports Movies	Through e
5	15 min	Documentaries, Sports Movies	From a ch
6	12 min	Documentaries, International Movies	This docu
7	17 min	Documentaries, Music & Musicals	Using inte
8	10 min	Documentaries, Sports Movies	Under pres
9	14 min	Documentaries	This docu
10	10 min	Documentaries, Music & Musicals	The world
11	13 min	Documentaries, LGBTQ Movies	The inspir

Total rows: 869 Query complete 00:00:00.166

-- List all movies that are documentaries. --

-- Find all content without Director. --

146
147
148
149
150
151
152
153

-- Q12. Find all content without a director.--

select * from netflix
where director is null

Data OutputMessagesNotifications

≡+

SQL

		director character varying (208)	casts character varying (1000)
9		[null]	Blanca Suárez, Iván Marcos, Óscar
10		[null]	Engin Altan Düzyatan, Serdar Gö
11		[null]	Brooke Satchwell
12		[null]	Lauren Ash, Rory O'Malley, RuPa
13		[null]	Asa Butterfield, Gillian Anderson
14		[null]	Lee Jung-jae, Park Hae-soo, Wi K
15		[null]	Dami Lee, Jason Lee, Bommie C
16		[null]	Antti Pääkkönen, Heljä Heikkine
17		[null]	Vatsal Dubey, Julie Tejwani, Rup
18		[null]	Yuri Lowenthal, Kimberly Brooks
19		[null]	Richard Mofe-Damijo, Dakore Ak

Total rows: 2634Query complete 00:00:00.162

```
-- Q13. Find how many movies actor 'salman khan' appeared in last 10 years.--
```

```
select * from netflix
where
  casts ilike '%salman khan%'
  and
  release_year > extract(year from current_date) - 10
```

-- Find In how many movies, actor
'Salman Khan' appeared in last 10
years. --

Data Output Messages Notifications



	varying (208) 🔒	casts character varying (1000) 🔒
1	an Kumar	Lorena Franco, Bijesh Jayarajan, Neet Chowdhary, Rajeev Gaursingh, Karran Jeet, Rajesh Sharma, Gautam Kurup, Poonam Mathur, Salman Khan, Manish Devkush

-- Find the top 10 actors who have appeared in the highest number of movies produced in India. --

```
162 -- Q14. Find the top 10 actors who have appeared in the highest number of movies produced in India.--
163
164 select
165     unnest(string_to_array(casts, ',')) as actors,
166     count(*) as total_content
167 from netflix
168 where country ilike '%india%'
169 group by 1
170 order by 2 desc
171 limit 10
172
173
```

Data Output Messages Notifications

SQL

	actors text	total_content bigint
1	Anupam Kher	36
2	Om Puri	26
3	Boman Irani	25
4	Paresh Rawal	25
5	Shah Rukh Khan	25
6	Akshay Kumar	23
7	Naseeruddin Sh...	20
8	Amitabh Bachch...	20
9	Kareena Kapoor	20
10	Asrani	17

Total rows: 10 Query complete 00:00:00.070

```

174 -- Q15. Categorize the content based on the presence of the keywords 'kill' and 'violence' in the
175 description field. Label content containing these keywords as 'bad' and all other content as 'good' .
176 count how many items fall into each category.--
177
178
179 with new_table
180 as
181 (
182 select
183 *,
184     case
185     when description ilike '%kill%' or
186         description ilike '%violence%' then 'bad_content'
187     else 'good_content'
188     end category
189 from netflix
190 )
191 select
192     category,
193     count(*) as total_content
194 from new_table
195 group by 1

```

Data Output Messages Notifications



	category text	total_content bigint
1	bad_content	342
2	good_content	8465

-- Categorize the content based on the presence of the keywords 'kill' and 'violence' in the description field. Label content containing these keywords as 'Bad' & all other content as 'Good'. Count how many items fall into each category. --



16+

CC

KEY INSIGHTS & FINDINGS

The analysis of Netflix's content reveals several key patterns and trends:

- Netflix offers a larger number of movies compared to TV shows, indicating a stronger focus on feature-length films.
- The most common rating for both movies and TV shows is 'TV-MA,' which suggests a preference for mature content.
- The year 2020 saw the highest number of content releases, likely due to the increased streaming demand during the pandemic.
- The top five countries with the most content on Netflix are the USA, India, the UK, Canada, and France.
- The longest movie available on Netflix has a runtime of over 300 minutes.
- Over the last five years, Netflix has shown consistent growth in the volume of content added.
- Director Rajiv Chilaka has contributed several animated titles to the platform.
- Very few TV shows exceed five seasons, pointing to a preference for shorter series.
- The most common genres on Netflix include documentaries, dramas, and comedies.
- The volume of Indian content on Netflix has steadily increased over recent years.
- Documentaries make up a significant portion of Netflix's catalog.
- Several pieces of content do not have a listed director, possibly due to collaborative or regional productions.
- Salman Khan has appeared in only a few Netflix films in the past decade.
- Certain actors are featured prominently in Indian productions on the platform.
- The analysis of keywords like "kill" and "violence" reveals thematic patterns in Netflix's content.

CONCLUSION

In conclusion, the analysis of Netflix's content reveals key patterns that highlight its strategic focus on movies, mature content, and a diverse range of genres. The platform has significantly expanded its library in recent years, with a marked increase in content from countries like India. While there is a trend towards shorter TV series, documentaries remain a dominant genre. Additionally, content featuring prominent directors and actors continues to draw attention, with regional productions and thematic explorations like violence and kill-related keywords offering deeper insights into audience preferences. Overall, Netflix's evolving content strategy is designed to cater to global and regional tastes, ensuring sustained growth and user engagement.

TOOLS & TECHNOLOGIES USED

- SQL: Used for querying and analyzing the Netflix dataset, extracting meaningful insights such as content counts, ratings, and genre distributions.
- Python: Employed for data processing, automation of queries, and performing advanced analytics, including content categorization and keyword analysis.
- Power BI: Utilized for data visualization, creating interactive dashboards to present key findings such as top countries, longest movies, and genre distributions.
- Excel: Used for initial data cleaning, handling smaller datasets, and performing quick analysis.
- Google Cloud Platform (GCP): Enabled cloud storage and management of large datasets, as well as scalability for analysis.
- Jupyter Notebooks: Applied for running Python scripts in an organized and reproducible manner during data analysis.
- AWS: Used for cloud computing to run complex analysis and large-scale queries on large Netflix datasets.

LEARNING & OUTCOMES

- Data Analysis and Querying: Gained hands-on experience in using SQL and Python for querying large datasets and extracting meaningful insights, enhancing data manipulation and analysis skills.
- Data Visualization: Developed proficiency in using tools like Power BI and Tableau to create interactive dashboards and reports, enabling better data-driven decision-making.
- Cloud Computing: Improved understanding of cloud platforms such as AWS and Google Cloud for managing and processing large-scale datasets.
- Content Categorization: Learned how to categorize content based on various attributes like genre, country, director, and keywords, which provided deeper insights into the structure of Netflix's offerings.
- Trend Identification: Identified and analyzed key trends in Netflix's content release patterns, including the rise of certain genres, content release frequencies, and country-specific content growth.
- Time Management: Developed time management skills by managing the project from data collection to analysis and final report presentation, ensuring timely completion of tasks.
- Problem-Solving and Critical Thinking: Developed stronger problem-solving skills by addressing challenges such as missing data, data cleaning, and handling large datasets efficiently.
- Technical Skills Enhancement: Improved technical skills, particularly in using Python libraries for data manipulation and analysis, including pandas and numpy.
- Communication and Presentation: Enhanced communication skills by presenting complex data insights in a clear and visually appealing manner using Power BI and Tableau.

CONTACT ME



THANKS

I would like to express my sincere gratitude to everyone who supported me throughout this project. Special thanks to my mentors and colleagues for their valuable guidance, encouragement, and feedback. Their insights helped refine the analysis and improved the overall quality of this work. I also appreciate the opportunity to explore and learn new tools and technologies, which significantly contributed to my personal and professional growth .

Thank you for your time and consideration.

Regards,

Aman

aamansharma027@gmail.com