

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











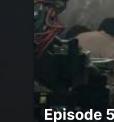
#### Season 1 ▼

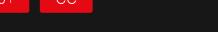












# WELCOME

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 v select
34
       type,
       count(*) as total_content
35
       from netflix
36
37
        group by type
38
39
40
41
42
43
44
45
46
Data Output
            Messages
                      Notifications
                                     SQL
=+
                         total_content
      type
     character varying (10)
                         bigint
      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. --

```
-- Q2. Find the most common rating for the movies and TV Shows. --
42
43
     select
44
        type,
        rating
45
46
        from
47
48
49
     select
50
         type,
         rating,
51
52
         count(*),
53
         rank() over(partition by type order by count(*) desc) as ranking
54
     from netflix
     group by 1, 2
56
     ) as t1
57
     where
         ranking = 1
58
59
Data Output Messages Notifications
     type
     character varying (10)
                        character varying (10)
                        TV-MA
     Movie
                         TV-MA
     TV Show
```

```
60
      -- Q3. List all movies released in a specific year (e.g., 2020).--
61
62
      select * from netflix
63
64
      where
          type = 'Movie'
65
66
          and
          release_year = 2020
67
68
69
70
71
72
73
Data Output
              Messages
                         Notifications
                                          5QL
=+
                                                                       release_year
                                                 date_added
                                                                                      rating
                                                 character varying (50)
              character varying (150)
                                                                                      character
              United States
                                                 September 25, 2021
                                                                                2020
                                                                                      PG-13
              [null]
                                                 September 22, 2021
                                                                                2020
                                                                                      TV-MA
3
                                                                                      TV-14
              [null]
                                                 September 11, 2021
                                                                                2020
              Nigeria
                                                 September 10, 2021
                                                                                2020
                                                                                     TV-MA
              [null]
                                                                                2020
                                                                                      TV-MA
                                                 September 6, 2021
              [null]
6
                                                 September 2, 2021
                                                                                2020
                                                                                     TV-MA
                                                                                2020
                                                                                      TV-14
              India
                                                 September 2, 2021
              United States
                                                 August 25, 2021
                                                                                2020
                                                                                     TV-MA
```

-- 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.--
71
72
     select
       unnest(string_to_array(country, ',')) as new_country,
73
74
         count(show_id) as total_content
75
     from netflix
     group by 1
     order by 2 desc
78
     limit 5
79
80
81
82
23
Data Output Messages Notifications
                                    SQL
     new_country total_content
     text
1
     United States
                          3211
2
                          1008
     India
3
     United Kingd...
                           628
```

United States

Canada

5

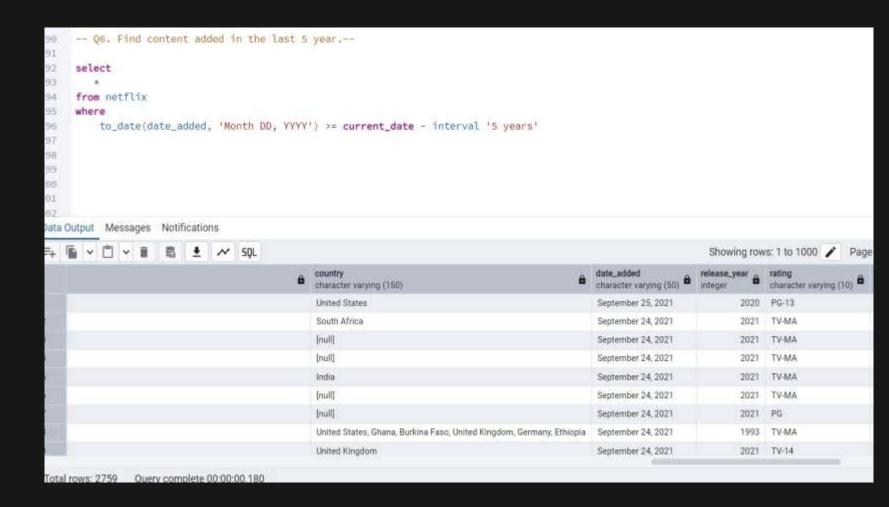
479

271

```
80
       -- Q5. Identify the longest movie.--
 81
 82
       select * from netflix
 83
       where
 84
            type = 'Movie'
 85
            and
 86
            duration = (select max(duration) from netflix)
 87
 88
 89
 90
 91
 92
 93
                         Notifications
             Messages
Data Output
=+
                          5
                                     N
                                          SQL
     elease_year
                                                                listed_in
                   character varying (10)
                                         character varying (15)
                                                               character varying (100)
     nteger
             2002 TV-14
                                          99 min
                                                                Action & Adventure, Anime Features, International Movies
             2003 TV-14
                                          99 min
                                                                Action & Adventure, Anime Features, International Movies
                                                                Dramas, International Movies, Romantic Movies
3
             2020
                   TV-MA
                                          99 min
             2007 R
                                          99 min
                                                                Comedies, Romantic Movies
5
                                                                Children & Family Movies, Comedies, Dramas
             1994 PG
                                          99 min
             2021
                   TV-MA
                                          99 min
                                                                Dramas
             2020 R
                                          99 min
                                                                Action & Adventure
             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. --



```
-- Q7. Find all the movies/TV shows by director 'Rajiv chilaka'.--
99
100
101
       select * from netflix
102
       where director ilike '%Rajiv Chilaka%'
103
104
105
106
107
108
109
110
Data Output Messages
                         Notifications
=+
      show_id
                                                                                                   director
                           character varying (10)
      character varying (6)
                                                   character varying (150)
                                                                                                   character varying (208)
      s407
                             Movie
                                                   Chhota Bheem - Neeli Pahaadi
                                                                                                   Rajiv Chilaka
                            Movie
                                                   Chhota Bheem & Ganesh
                                                                                                   Rajiv Chilaka
      s408
      s409
                            Movie
                                                   Chhota Bheem & Krishna: Mayanagari
                                                                                                   Rajiv Chilaka
                                                                                                   Rajiv Chilaka
      s410
                            Movie
                                                   Chhota Bheem & Krishna: Pataliputra- City of the Dead
                                                                                                   Rajiv Chilaka
      s411
                            Movie
                                                   Chhota Bheem And The Broken Amulet
      s412
                            Movie
                                                   Chhota Bheem And The Crown of Valhalla
                                                                                                   Rajiv Chilaka
                                                                                                   Rajiv Chilaka
      s413
                            Movie
                                                   Chhota Bheem and the Incan Adventure
                                                                                                   Rajiv Chilaka
      s414
                            Movie
                                                   Chhota Bheem and The ShiNobi Secret
                                                                                                   Rajiv Chilaka
      s415
                            Movie
                                                   Chhota Bheem Aur Hanuman
10
      s416
                            Movie
                                                   Chhota Bheem aur Krishna
                                                                                                   Rajiv Chilaka
                 Query complete 00:00:00.131
Total rows: 22
```

-- Find all the movies/Tv shows by Director 'Rajiv Chilaka'. -- -- List all Tv shows with more than 5 seasons. --

```
-- Q8. List all TV shows with more than 5 seasons,--
106
107
108
       select
109
       from netflix
110
       where
111
              type = 'TV Show'
112
              and
113
             split_part(duration, ' ', 1)::numeric > 5
114
115
116
117
Data Output Messages Notifications
                                     N
                                                                                                                                                Showing
                                              character varying (10)
                                                                    character varying (15)
         character varying (50)
                                                                                          character varying (100)
                                                                                           British TV Shows, Reality TV
          September 24, 2021
                                        2021 TV-14
                                                                     9 Seasons
2
          September 15, 2021
                                        2021
                                              TV-PG
                                                                     6 Seasons
                                                                                           Reality TV
3
          September 15, 2021
                                        2021 TV-Y
                                                                     6 Seasons
                                                                                           Kids'TV
                                        1994 TV-PG
          September 15, 2021
                                                                     9 Seasons
                                                                                           Kids' TV. TV Comedies
                                        2021 TV-14
                                                                                           Crime TV Shows, TV Comedies, TV Dramas
          September 10, 2021
                                                                     6 Seasons
                                              TV-MA
                                                                                           TV Comedies, TV Dramas
          August 13, 2021
                                        2021
                                                                     7 Seasons
          August 1, 2021
                                        2012 TV-14
                                                                     7 Seasons
                                                                                           TV Comedies
8
                                                                                           Anime Series, International TV Shows
          August 1, 2021
                                        2014 TV-14
                                                                     6 Seasons
9
                                                                                           Crime TV Shows, TV Action & Adventure, TV Sci-Fi & Fantasy
          July 28, 2021
                                        2021 TV-14
                                                                     7 Seasons
          July 26, 2021
10
                                        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
       unnest(string_to_array(listed_in, ',')) as genre,
120
121
          count(show_id) as total_content
      from netflix
122
123
      group by 1
124
125
126
127
```

#### Data Output Messages Notifications

=+		<b>● /</b> SQL
	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
10	Children & Family Movies	605 57

				in oach	
(C(0)UI 017 17 0	re mumb	er of con	tent items		Genre

- -- 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.--
      -- Return the top 5 year with highest avg content release.--
127
128
129
130
      select
         extract(year from to_date(date_added, 'Month DD, YYYY')) as date,
131
         count(*),
132
         round (
133
         count(*)::numeric/(select count(*) from netflix where country = 'India')::numeric * 100
134
135
         ,2)as avg_content_per_year
      from netflix
136
      where country = 'India'
137
138
      group by 1
139
Data Output Messages Notifications
                                    SQL
              count bigint avg_content_per_year numeric
     numeric **
          2018
                  333
                                    34.26
         2016
                   10
                                    1.03
          2019
                  203
                                    20.88
         2021
                   95
                                    9.77
          2020
                  189
                                    19.44
```

2017

142

14.61

```
140
        -- 011. List all movies that are documentaries.--
141
142
        select * from netflix
143
        where listed_in ilike '%documentaries%'
144
145
146
147
148
149
150
Data Output
              Messages
                           Notifications
                                             SQL
                            listed_in
      ration
                                                                                    description
      aracter varying (15)
                           character varying (100)
                                                                                    character
      ) min
                            Documentaries
                                                                                     As her fatl
                            Documentaries, International Movies
                                                                                    Declassifi
       min
3
                                                                                    Robin Wilt
      3 min
                            Documentaries
                            Documentaries, International Movies, Sports Movies
                                                                                    Through e
      3 min
                            Documentaries, Sports Movies
                                                                                    From a ch
      5 min
                            Documentaries, International Movies
                                                                                    This docu
      2 min
                            Documentaries, Music & Musicals
                                                                                    Using inte
      7 min
8
                            Documentaries, Sports Movies
                                                                                    Under pres
      ) min
                                                                                    This docu
      1 min
                            Documentaries
                            Documentaries, Music & Musicals
10
      min (
                                                                                     The world
     ₹ min
                            Documentaries I GRTO Movies
                                                                                     The insnir
```

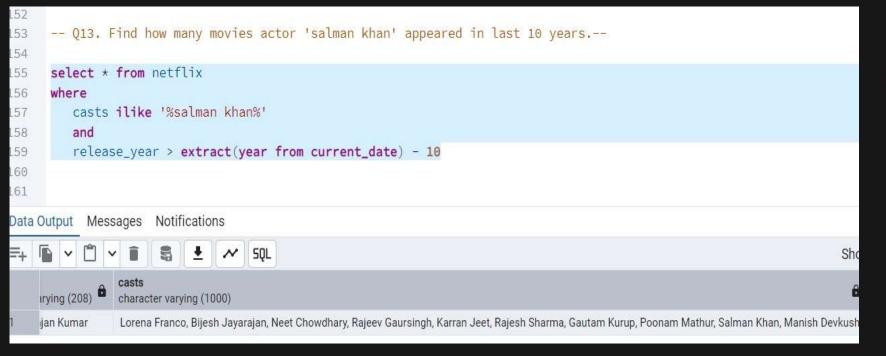
Query complete 00:00:00.166

Total rows: 869

-- List all movies that are documentaries. --

-- Find all content without Director. --

```
146
        -- 012. Find all content without a director.--
147
148
149
        select * from netflix
150
        where director is null
151
152
153
               Messages
                            Notifications
Data Output
                                                SQL
                      director
                                                casts
                      character varying (208)
                                                character varying (1000)
9
                      [null]
                                                Blanca Suárez, Iván Marcos, Óso
10
                      [null]
                                                Engin Altan Düzyatan, Serdar Gö
                                                Brooke Satchwell
11
                      [null]
12
                      [null]
                                                Lauren Ash, Rory O'Malley, RuPa
13
                      [null]
                                                Asa Butterfield, Gillian Andersor
14
                      [null]
                                                Lee Jung-jae, Park Hae-soo, Wi I
15
                      [null]
                                                Dami Lee, Jason Lee, Bommie C
16
                      [null]
                                                Antti Pääkkönen, Heljä Heikkine
17
                      [null]
                                                Vatsal Dubey, Julie Tejwani, Rup
                      [null]
                                                Yuri Lowenthal, Kimberly Brooks
18
19
                      Inuill
                                                 Richard Mofe-Damijo Dakore Ak
 Total rows: 2634
                      Query complete 00:00:00.162
```



-- Find In how many movies, actor 'Salman Khan' appeared in last 10 years. -- -- Find the top 10 actors who have appeared in the highest number of movies produced in India. --

```
-- Q14. Find the top 10 actors who have appeared in the highest number of movies produced in India.--
163
164
      select
          unnest(string_to_array(casts, ',')) as actors,
165
          count(*) as total_content
166
      from netflix
167
      where country ilike '%india%'
      group by 1
      order by 2 desc
      limit 10
172
Data Output Messages Notifications
                           . ✓ 5QL
     actors
     text
      Anupam Kher
                              36
                              26
      Om Puri
                              25
      Boman Irani
      Paresh Rawal
                              25
     Shah Rukh Khan
                              25
     Akshay Kumar
                              23
      Naseeruddin Sh...
                              20
      Amitabh Bachch...
                              20
                              20
      Kareena Kapoor
      Asrani
                              17
Total rows: 10 Query complete 00:00:00.070
```

```
-- Q15. Categorize the content based on the presence of the keywords 'kill' and 'violence' in the
      description field. Label content containing these keywords as 'bad' and all other content as 'good' .
175
      count how many items fall into each category.--
176
177
178
      with new_table
179
180
181
182
      select
183
184
         case
         when description ilike '%kill%' or
185
              description ilike '%violence%' then 'bad_content'
186
              else 'good_content'
187
              end category
188
      from netflix
190
191
      select
192
        category,
        count(*) as total_content
193
      from new table
194
      group by 1
195
Data Output Messages Notifications
                total_content
      bad_content
                        342
     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. --



#### 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