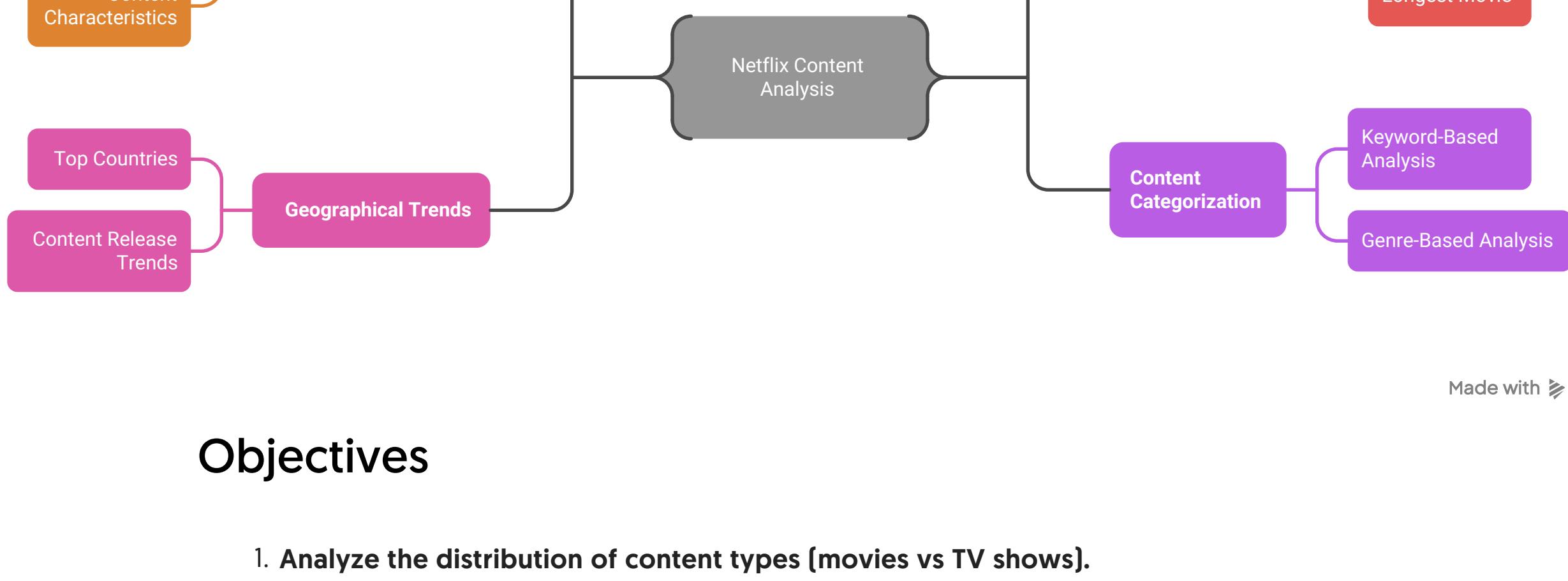


Netflix Movies and TV Shows Analysis using SQL

This document outlines a comprehensive SQL-based analysis of Netflix's movie and TV show data. The project aims to extract valuable insights by addressing specific business problems related to content distribution, ratings, geographical trends, and content categorization. The analysis provides a detailed understanding of the Netflix content landscape, including identifying popular content types, common ratings, top countries, and content characteristics based on keywords. The findings and conclusions summarize the key insights derived from the SQL queries, offering a clear picture of the content available on Netflix.

SQL-Based Analysis of Netflix Content



Made with Napkin

Objectives

1. **Analyze the distribution of content types [movies vs TV shows].**
2. **Identify the most common ratings for movies and TV shows.**
3. **List and analyze content based on release years, countries, and durations.**
4. **Explore and categorize content based on specific criteria and keywords.**

Business Problems and SQL Solutions

1. Count the Number of Movies vs TV Shows

Problem: Determine the total count of movies and TV shows available on Netflix.

SQL Solution:

```
SELECT type, COUNT(*) AS count
FROM netflix_titles
GROUP BY type;
```

Expected Outcome: This query will return two rows, one for "Movie" and one for "TV Show," along with the respective counts.

2. Find the Most Common Rating for Movies and TV Shows

Problem: Identify the most frequent rating given to movies and TV shows.

SQL Solution:

```
SELECT rating, COUNT(*) AS count
FROM netflix_titles
GROUP BY rating
ORDER BY count DESC
LIMIT 1;
```

Expected Outcome: This query will return the rating that appears most often in the dataset, indicating the most common rating.

3. List All Movies Released in a Specific Year [e.g., 2020]

Problem: Retrieve a list of all movies released in the year 2020.

SQL Solution:

```
SELECT title
FROM netflix_titles
WHERE type = 'Movie' AND strftime('%Y', date(release_year)) = '2020';
```

Expected Outcome: This query will return a list of movie titles that were released in the year 2020.

4. Find the Top 5 Countries with the Most Content on Netflix

Problem: Identify the top 5 countries that have the highest number of content items [movies and TV shows] on Netflix.

SQL Solution:

```
SELECT country, COUNT(*) AS count
FROM netflix_titles
GROUP BY country
ORDER BY count DESC
LIMIT 5;
```

Expected Outcome: This query will return the top 5 countries along with the number of content items from each country.

5. Identify the Longest Movie

Problem: Find the movie with the longest duration.

SQL Solution:

```
SELECT title, duration
FROM netflix_titles
WHERE type = 'Movie'
ORDER BY CAST(REPLACE(duration, ' min', '' ) AS INTEGER) DESC
LIMIT 1;
```

Expected Outcome: This query will return the title and duration of the longest movie in the dataset.

6. Find Content Added in the Last 5 Years

Problem: List all content [movies and TV shows] that were added to Netflix in the last 5 years.

SQL Solution:

```
SELECT title, date_added
FROM netflix_titles
WHERE strftime('%Y', date(date_added)) >= strftime('%Y', date('now', '-5 years'));
```

Expected Outcome: This query will return a list of titles and their addition dates for content added in the last 5 years.

7. Find All Movies/TV Shows by Director 'Rajiv Chilaka'

Problem: Retrieve all movies and TV shows directed by 'Rajiv Chilaka'.

SQL Solution:

```
SELECT title, type
FROM netflix_titles
WHERE director LIKE '%Rajiv Chilaka%';
```

Expected Outcome: This query will return a list of titles and their types [Movie or TV Show] directed by Rajiv Chilaka.

8. List All TV Shows with More Than 5 Seasons

Problem: List all TV shows that have more than 5 seasons.

SQL Solution:

```
SELECT title, duration
FROM netflix_titles
WHERE type = 'TV Show' AND CAST(REPLACE(duration, ' Seasons', '' ) AS INTEGER) > 5;
```

Expected Outcome: This query will return a list of TV show titles and their durations for shows with more than 5 seasons.

9. Count the Number of Content Items in Each Genre

Problem: Determine the number of content items [movies and TV shows] in each genre.

SQL Solution:

```
SELECT listed_in, COUNT(*) AS count
FROM netflix_titles
GROUP BY listed_in
ORDER BY count DESC;
```

Expected Outcome: This query will return a list of genres and the number of content items in each genre, sorted in descending order.

10. Find each year and the average numbers of content release in India on netflix.

Problem: Determine the average number of content releases in India for each year on Netflix.

SQL Solution:

```
SELECT strftime('%Y', date(release_year)) AS release_year, AVG(CASE WHEN country = 'India' THEN 1 ELSE 0 END) AS avg_releases
FROM netflix_titles
GROUP BY release_year;
```

Expected Outcome: This query will return each year and the average number of content releases in India for that year.

11. List All Movies that are Documentaries

Problem: Retrieve a list of all movies that are classified as documentaries.

SQL Solution:

```
SELECT title
FROM netflix_titles
WHERE type = 'Movie' AND listed_in LIKE '%Documentaries%';
```

Expected Outcome: This query will return a list of movie titles that are classified as documentaries.

12. Find All Content Without a Director

Problem: Identify all content items [movies and TV shows] that do not have a director listed.

SQL Solution:

```
SELECT title, type
FROM netflix_titles
WHERE director IS NULL OR director = '';
```

Expected Outcome: This query will return a list of titles and their types for content items without a director listed.

13. Find How Many Movies Actor 'Salman Khan' Appeared in the Last 10 Years

Problem: Determine the number of movies in which the actor 'Salman Khan' appeared in the last 10 years.

SQL Solution:

```
SELECT COUNT(*) AS movie_count
FROM netflix_titles
WHERE type = 'Movie'
AND strftime('%Y', date(release_year)) >= strftime('%Y', date('now', '-10 years'))
AND cast LIKE '%Salman Khan%';
```

Expected Outcome: This query will return the number of movies featuring Salman Khan that were released in the last 10 years.

14. Find the Top 10 Actors Who Have Appeared in the Highest Number of Movies Produced in India

Problem: Identify the top 10 actors who have appeared in the highest number of movies produced in India.

SQL Solution:

```
SELECT cast, COUNT(*) AS movie_count
FROM netflix_titles
WHERE type = 'Movie' AND country = 'India'
GROUP BY cast
ORDER BY movie_count DESC
LIMIT 10;
```

Expected Outcome: This query will return the top 10 actors and the number of movies they have appeared in that were produced in India.

15. Categorize Content Based on the Presence of 'Kill' and 'Violence' Keywords

Problem: Categorize content items based on whether their descriptions contain the keywords 'kill' or 'violence'.

SQL Solution:

```
SELECT title,
CASE
    WHEN description LIKE '%kill%' THEN 'Kill'
    WHEN description LIKE '%violence%' THEN 'Violence'
    ELSE 'None'
END AS keyword_category
FROM netflix_titles
WHERE description LIKE '%kill%' OR description LIKE '%violence%';
```

Expected Outcome: This query will return a list of titles and their corresponding keyword categories ['Kill', 'Violence', or 'None'] based on the presence of the specified keywords in their descriptions.

Findings and Conclusion

The SQL analysis of the Netflix dataset provides several key insights:

- **Content Distribution:** The dataset contains a diverse range of movies and TV shows, with a significant number of both types available on the platform.
- **Common Ratings:** The analysis of ratings provides an understanding of the content's target audience, with certain ratings being more prevalent than others.
- **Geographical Insights:** The identification of top countries and the average content releases by India highlight regional content distribution and preferences.
- **Content Categorization:** Categorizing content based on specific keywords helps in understanding the nature of content available on Netflix, allowing for targeted content discovery and recommendations.

In conclusion, this project successfully utilized SQL to extract valuable insights