**DATA DIVE**

1. **Difficulties**

**(a)Categorical Data Ambiguities:**

* Listed\_in and Categories like 'TV-MA', 'PG-13', and genres such as 'Crime TV Shows' or 'International Movies' may overlap or be too broad, complicating granular analysis.

**(b)Special Characters and Encoding Issues:**

* Entries like 'My Little Pony: A New Generation' (s7) include characters like 'Ã©', 'â€™', which indicate encoding issues. These characters need to be cleaned for accurate text analysis

**(c)Missing Data:**

**Director and Cast Information**: Some entries, such as 'Blood & Water' (s2) and 'Jailbirds New Orleans' (s4), lack director and cast details. Missing data can hinder analyses related to popular directors, cast diversity, or the impact of specific directors or actors on show popularity.

* **Country Information**: Some shows, like 'Vendetta: Truth, Lies and The Mafia' (s11), lack country information, complicating geographical trend analysis.
* **Date Added**: Entries like 'Castle and Castle' (s50) lack the exact date added to Netflix, which is essential for temporal analysis of content addition.

**(d)Inconsistent Formats:**

* **Date Formats:** The date added is consistently in 'Month Day, Year' format, but inconsistencies or variations in formats can arise, leading to parsing errors.
* **Duration Format**: Movies list duration in minutes (e.g., '90 min'), while TV shows list seasons (e.g., '2 Seasons'), requiring different handling during data processing.

**(e)Historical Data Variability:**

* **Release Year vs. Date Added:** Shows like 'minsara kanavu' (s27) released in 1997 but added to Netflix in 2021, indicating a significant lag between release and availability on Netflix. This requires careful handling to distinguish between the release trend and Netflix acquisition trends.

**(2)Interest**

**International TV Shows/Movies:** Content produced outside the United States, often featuring international actors, settings, and cultural themes.

* Example: "Blood & Water" is an international TV show from South Africa.

**DATA FUN**

**Sql queries**

**(a)SELECT COUNT(\*) AS shows\_from\_us**

**FROM netflix\_titles**

**WHERE country LIKE '%United States%';**

**# shows\_from\_us**

**'25'**

**(b)SELECT COUNT(\*) AS shows\_added\_in\_sept\_2021**

**FROM netflix\_titles**

**WHERE date\_added LIKE 'September %, 2021';**

**# shows\_added\_in\_sept\_2021**

**'100'**

**(c)SELECT AVG(release\_year) AS avg\_release\_year**

**FROM netflix\_titles;**

**# avg\_release\_year**

**'2015.1200'**

**(d)SELECT COUNT(\*) AS total\_tv\_shows**

**FROM netflix\_titles**

**WHERE type = 'TV Show';**

**# total\_tv\_shows**

**'45'**

**(e)SELECT title**

**FROM netflix\_titles**

**WHERE cast LIKE '%Denzel Washington%';**

**# title**

**'Safe House'**

**'Training Day'**

**(f)SELECT COUNT(\*)**

**FROM netflix\_titles**

**WHERE rating = 'TV-MA';**

**# COUNT(\*)**

**'32'**

**(g)SELECT AVG(CAST(SUBSTRING\_INDEX(duration, ' ', 1) AS UNSIGNED)) AS average\_duration**

**FROM netflix\_titles**

**WHERE type = 'Movie';**

**# average\_duration**

**'100.7273'**

**(h)SELECT SUM(CAST(SUBSTRING(duration, 1, LOCATE(' ', duration) - 1) AS UNSIGNED)) AS total\_tv\_seasons**

**FROM netflix\_titles**

**WHERE type = 'TV Show' AND duration LIKE '%Season%';**

**# total\_tv\_seasons**

**'106'**

**Cool facts**

**(a)Diverse Origins:** The dataset showcases a notable diversity in the origin of the shows and movies, with productions coming from various countries like the United States, South Africa, India, Germany, France, Japan, Turkey, and more. This diversity reflects Netflix's strategy of offering a global content library that caters to a wide array of cultural tastes and preferences

**(b)Strong Presence of Documentaries and Docuseries:** A significant number of titles fall under the "Documentaries" and "Docuseries" categories, such as "Dick Johnson Is Dead," "Vendetta: Truth, Lies and The Mafia," and "Schumacher." This indicates that Netflix has invested heavily in non-fiction content, which has been growing in popularity among viewers seeking informative and real-life stories.

**3 ASK AWAY**

**(a)Questions**

**1:What are the top 5 longest movies from the United States?**

**SELECT title, duration,country**

**FROM netflix\_titles**

**WHERE type = 'Movie' AND country = 'United States'**

**ORDER BY CAST(SUBSTRING(duration, 1, LENGTH(duration)-4) AS UNSIGNED) DESC**

**LIMIT 5;**

**# title, duration, country**

**'Jaws', '124 min', 'United States'**

**'Training Day', '122 min', 'United States'**

**'Jaws 2', '116 min', 'United States'**

**'Kate', '106 min', 'United States'**

**'The Starling', '104 min', 'United States'**

***“I have learnt that this query show which movies are the oldest in the dataset, providing an understanding of the historical range and earliest content available from the United States.”***

**2: Which TV shows were released in 2021, and what are their titles, release years, and number of seasons?**

**SELECT title, release\_year, duration**

**FROM netflix\_titles**

**WHERE type = 'TV Show' AND release\_year = '2021'**

**ORDER BY title ASC**

**LIMIT 5;**

**# title, release\_year, duration**

**'Bangkok Breaking', '2021', '1 Season'**

**'Blood & Water', '2021', '2 Seasons'**

**'Castle and Castle', '2021', '2 Seasons'**

**'Chhota Bheem', '2021', '3 Seasons'**

**'Chicago Party Aunt', '2021', '1 Season'**

***I have learnt this query reveals which TV shows have the most seasons or episodes, providing an indication of the longest-running series in terms of seasons, along with their countries of origin.***