NETFLIX

INTRODUCTION

The objective of this analysis is to explore the content available on Netflix, including its distribution by type, genre, and other key metrics such as duration, rating, and country. The analysis also involves predictive modeling using machine learning to classify content types (Movies vs. TV Shows) based on several features such as duration, number of seasons, and release year. The combination of Power BI visualizations and machine learning insights offers a deeper understanding of trends and patterns in Netflix content over time.

Objective:

1) To analyze the distribution and trends of Netflix content, focusing on key features like duration, genre, rating, and content type.
2) To use machine learning models to predict content type (Movie vs. TV Show) and identify the most significant features that influence these predictions.
3) To provide recommendations for future Netflix content strategies based on the insights gained.

Insights

1. Duration Insights by Year

- **Finding**: The highest **total duration_in_minutes** was in **2018**, followed by **2017** and **2016**.
- **ML Insight**: **Duration_in_minutes** was the most important feature in predicting whether content is a **Movie or TV Show** (with **58.96% importance** in the Random Forest model).

2. Growth of Movies and TV Shows Over Time (2008 - 2021)

- **Movies** saw a **99,200.00% increase** in availability from 2008 to 2021.
- **TV Shows** saw a **50,400.00% increase**, especially since 2015, with a **1,842.31% rise** in six years.

3. Content Type Distribution

- **Movies** made up **69.69%** of the **total GenreCount**, whereas **TV Shows** comprised the remaining **30.31%**.
- **TV Show Not Specified in Type** accounted for **90.65%** of all TV shows, indicating incomplete classification in the dataset.

4. Correlation Between Movie and TV Show Content

- A **negative correlation** was observed between the **count of movies** and **count of TV shows**, suggesting a shift in focus from one content type to another over time.

5. Popular Content Ratings

- **TV-MA** was the most common rating, accounting for **3,205** titles, followed by **TV-14** and **TV-PG**.
- In **2018**, **TV-MA** made up **6.25%** of the total content type, reflecting a trend towards mature content on Netflix.

6. Genre and Duration Distribution

- **Top 10 genres** by movie count revealed that **action**, **drama**, and **comedy** were the most popular genres.
- The average count of **duration_in_minutes** for **movies** and **TV Shows** ranged from **1 to 2 minutes**.

7. Type Distribution Over Time

- **Movies** dominated Netflix in earlier years but **TV Shows** have shown a sharp rise since **2015**.

8. Country Insights

- **The United States** contributed the most content to Netflix, with **39.10%** of the total content originating from the US.
- Across **79 countries**, the total count of content ranged from **1 to 2,395**, indicating global representation.

9. Predictive Modeling

- **Machine Learning Models** (Logistic Regression and Random Forest) achieved **100% accuracy** in predicting whether a piece of content is a **Movie or TV Show**.
- **Key Features**:
  - **Duration_in_minutes** (58.96% importance)
  - **Number of Seasons** (38.91% importance)
  - **Release Year** and **Director Count** had minimal impact.

10. Feature Importance and Model Validation

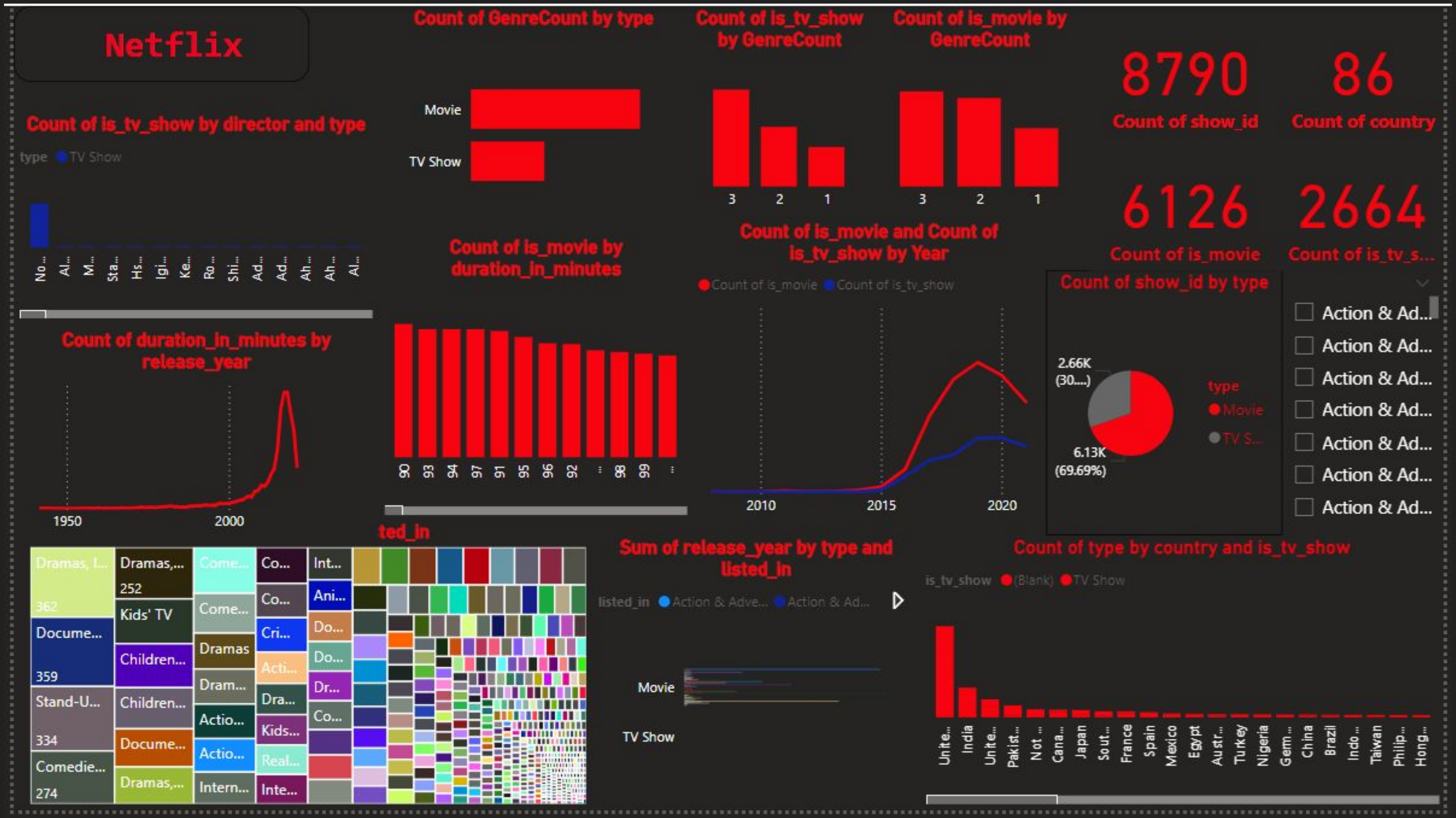
- **Cross-validation** confirmed model robustness with both models maintaining **100% accuracy**.
- **Hyperparameter tuning** further optimized the **Random Forest model** performance, finding the best parameters and ensuring no overfitting.

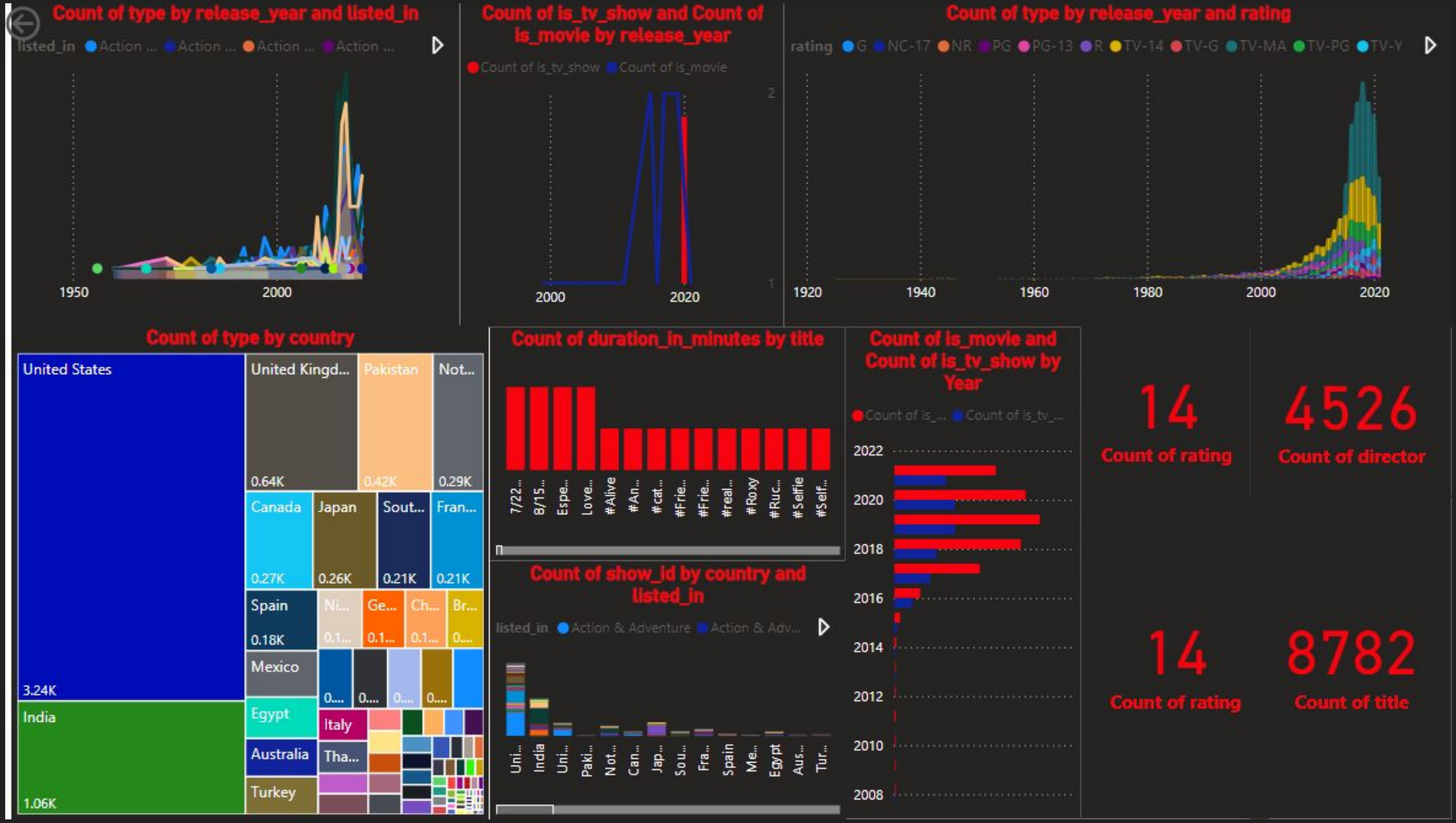
Recommendations

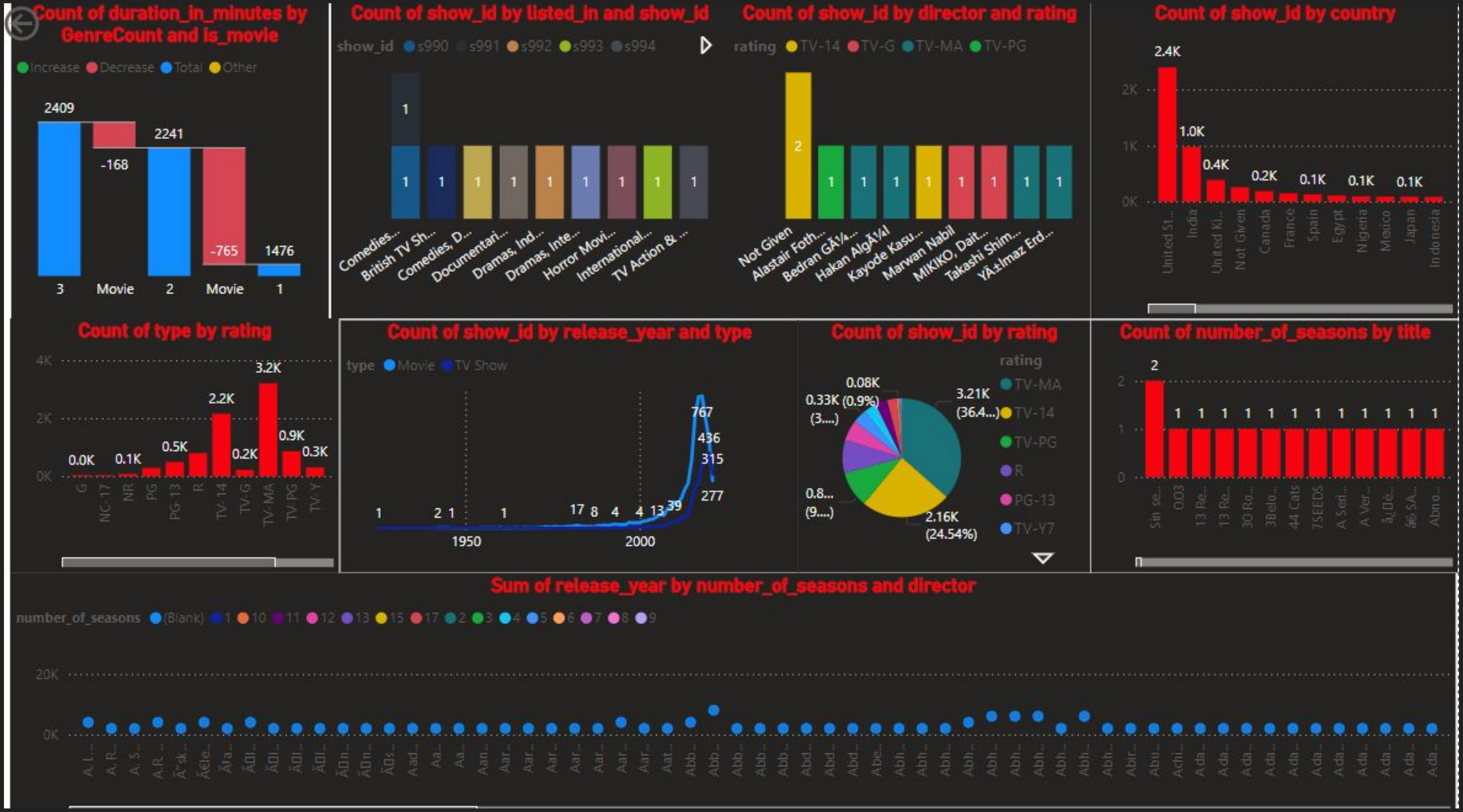
1. **Content Strategy Based on Duration**:
    - Since **duration_in_minutes** is highly predictive of whether content is a movie or TV show, Netflix should explore more variations in content length, especially for genres where content length is critical.
2. **Increase in TV Shows**:
    - Given the rapid rise in **TV Shows** after **2015**, Netflix should continue to invest in this content type, especially in **drama**, **crime**, and **thriller** genres which have been most popular.
3. **Targeted Mature Content**:
    - With **TV-MA** rated content being the most popular, Netflix should consider producing more mature content, especially for **thriller**, **crime**, and **mystery** genres.
4. **Country-Specific Content**:
    - **The US** leads Netflix content, but there’s an opportunity to expand localized content in emerging markets like **India**, **Brazil**, and **South Korea**, where demand for regional content is increasing.
5. **Global Content Expansion**:
    - The diversity in content origin suggests Netflix could further explore underrepresented countries, producing content that resonates with different cultural contexts.
6. **Improve Content Classification**:
    - **Not Given** classification in **TV shows** (90.65% in type) indicates a need for better metadata handling. Netflix should refine its data classification process to improve searchability and content recommendations.

CONCLUSION:

The combined insights from **Power BI visualizations** and **Machine Learning models** provide a comprehensive overview of Netflix's content landscape. Key factors like **duration**, **number of seasons**, and **genre** significantly influence the type of content (Movie or TV Show). The predictive models achieved perfect accuracy, highlighting the robustness of the features used. Recommendations based on content trends and feature importance suggest areas for further content development, especially in mature-rated TV shows and country-specific content production. Netflix can leverage these insights to make data-driven decisions that align with audience preferences and global market trends.


Netflix
Count of is_tv_show by director and type
type TV Show
Count of GenreCount by type
Movie
TV Show
Count of is_tv_show by GenreCount
3 2 1
Count of is_movie by GenreCount
3 2 1
8790
Count of show_id
86
Count of country
6126
Count of is_movie
2664
Count of is_tv_s...
Count of is_movie by duration_in_minutes
90 93 94 97 91 95 96 92 98 99
Count of is_movie and Count of is_tv_show by Year
Count of is_movie Count of is_tv_show
2010 2015 2020
Count of show_id by type
2.66K (30....)
6.13K (69.69%)
type
Movie
TV S...
Action & Ad...
Count of duration_in_minutes by release_year
1950 2000
Dramas, I... 362
Dramas,... 252
Kids' TV
Docume... 359
Children...
Stand-U... 334
Comedie... 274
Sum of release_year by type and listed_in
listed_in Action & Adve... Action & Ad...
Movie
TV Show
Count of type by country and is_tv_show
is_tv_show (Blank) TV Show
Unite... India Unite... Pakist... Not ... Cana... Japan Sout... France Spain Mexico Egypt Austr... Turkey Nigeria Germ... China Brazil Indo... Taiwan Philip... Hong...


Count of type by release_year and listed_in
listed_in Action ... Action ... Action ... Action ...
1950
2000
Count of is_tv_show and Count of is_movie by release_year
Count of is_tv_show Count of is_movie
2000
2020
2
1
Count of type by release_year and rating
rating G NC-17 NR PG PG-13 R TV-14 TV-G TV-MA TV-PG TV-Y
1920
1940
1960
1980
2000
2020
Count of type by country
United States
3.24K
India
1.06K
United Kingd...
0.64K
Pakistan
0.42K
Not...
0.29K
Canada
0.27K
Japan
0.26K
Sout...
0.21K
Fran...
0.21K
Spain
0.18K
Ni...
0.1...
Ge...
0.1...
Ch...
0.1...
Br...
0...
Mexico
Egypt
Italy
Australia
Tha...
Turkey
Count of duration_in_minutes by title
7/22...
8/15...
Espe...
Love...
#Alive
#An...
#cat...
#Frie...
#Frie...
#real...
#Roxy
#Ruc...
#Selfie
#Self...
Count of show_id by country and listed_in
listed_in Action & Adventure Action & Adv...
Uni...
India
Uni...
Paki...
Not...
Can...
Jap...
Sou...
Fra...
Spain
Me...
Egypt
Aus...
Tur...
Count of is_movie and Count of is_tv_show by Year
Count of is_... Count of is_tv_...
2022
2020
2018
2016
2014
2012
2010
2008
14
Count of rating
4526
Count of director
14
Count of rating
8782
Count of title


Count of duration_in_minutes by GenreCount and is_movie
Increase Decrease Total Other
2409
-168
2241
-765
1476
3 Movie 2 Movie 1
Count of show_id by listed_in and show_id
show_id s990 s991 s992 s993 s994
1
1 1 1 1 1 1 1 1 1
Comedies... British TV Sh... Comedies, D... Documentari... Dramas, Ind... Dramas, Inte... Horror Movi... International... TV Action & ...
Count of show_id by director and rating
rating TV-14 TV-G TV-MA TV-PG
2
1 1 1 1 1 1 1 1 1
Not Given Alastair Foth... Bedran GÃ¼... Hakan AlgÃ¼l Kayode Kasu... Marwan Nabil MIKIKO, Dait... Takashi Shim... YÃ±lmaz Erd...
Count of show_id by country
2.4K
1.0K
0.4K
0.2K
0.1K 0.1K 0.1K
2K
1K
0K
United St... India United Ki... Not Given Canada France Spain Egypt Nigeria Mexico Japan Indonesia
Count of type by rating
4K
2K
0K
3.2K
2.2K
0.9K
0.5K
0.2K
0.3K
0.0K
0.1K
G NC-17 NR PG PG-13 R TV-14 TV-G TV-MA TV-PG TV-Y
Count of show_id by release_year and type
type Movie TV Show
767
436
315
277
1 2 1 1 17 8 4 4 13 39
1950 2000
Count of show_id by rating
rating
TV-MA
TV-14
TV-PG
R
PG-13
TV-Y7
0.08K
0.33K (0.9%)
(3....)
3.21K
(36.4...)
0.8...
(9....)
2.16K
(24.54%)
Count of number_of_seasons by title
2
1 1 1 1 1 1 1 1 1 1 1 1
2
1
0
Sin se... 003 13 Re... 13 Re... 30 Ro... 3Belo... 44 Cats 7SEEDS A Seri... A Ver...
Sum of release_year by number_of_seasons and director
number_of_seasons (Blank) 1 10 11 12 13 15 17 2 3 4 5 6 7 8 9
20K
0K
A. L... A. R... A. S... A.R. ... Ã"sk... Ãele... Ãla... ÃDl... ÃDl... ÃDl... ÃDn... ÃDn... ÃDs... Aad... Aa... Aa... Aan... Aar... Aar... Aar... Aar... Aar... Aar... Aar... Aat... Abb... Abb... Abb... Abb... Abb... Abd... Abd... Abd... Abe... Abh... Abh... Abh... Abh... Abh... Abh... Abh... Abh... Abh... Abir... Abu... Achi... Ada... Ada... Ada... Ada... Ada... Ada... Ada... Ada... Ada... Ada...

THANKS