NETFLIX

NETFLIX DATA CLEANING, ANALYSIS & VISUALIZATION

PRESENTED BY :

SUBRATA DEY
DATA ANALYST INTERN
TOOLS USED : PYTHON, SQL, EXCEL, TABLEAU
UNIFIED MENTOR PRIVATE LIMITED
DATE :

Executive Summary

- Netflix hosts 19,294 titles across 86 countries, highlighting its strong global footprint.
- Movies dominate the platform, accounting for 68.34% (13,185 titles) compared to 31.66% TV Shows (6,109 titles).
- Content additions peaked around 2019, followed by a slight decline in recent years.
- TV-MA and TV-14 are the most common ratings, indicating strong demand for mature and teen-focused content.
- Drama and Comedy are the most dominant genres, guiding future content acquisition strategy.

Business Impact

- Supports strategic decisions on content investment, regional expansion, and genre prioritization.

Problem Statement & Objectives

Problem Statement

- Netflix needs to understand global content distribution, genre dominance, release trends, and audience targeting to optimize its content acquisition and production strategy.

Objectives

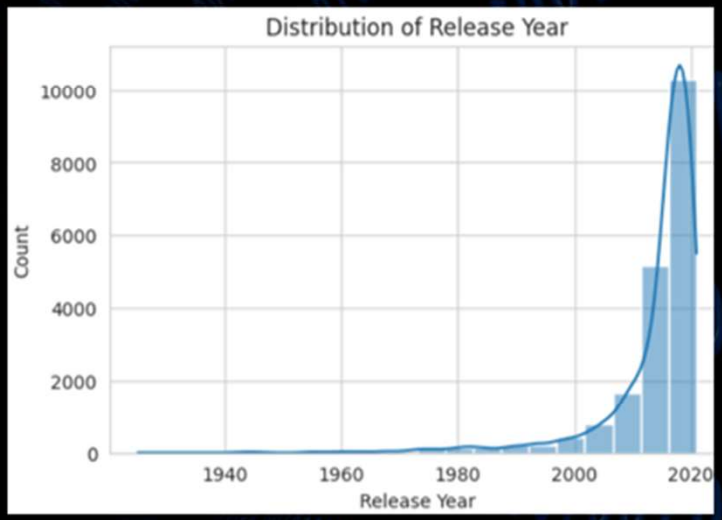
- Clean and prepare Netflix content data.
- Analyze content distribution by type, country, genre, rating, and time.
- Identify key growth trends and dominant categories.
- Build a dashboard to support executive decision-making.

Data Preparation and Cleaning

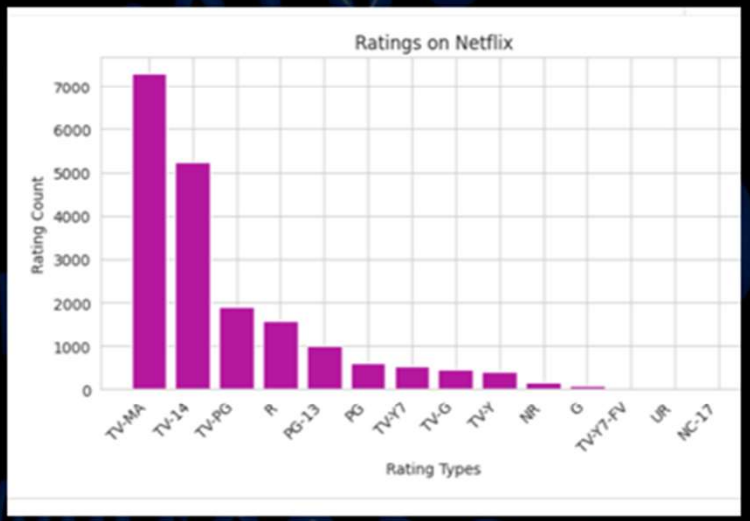
Data Loading & Cleaning

```
# Load the dataset:
df = pd.read_csv('/content/netflix1.csv')
# Looking for any Missing Values:
df.isnull().sum()
# Check number of duplicate data
print('Number of duplicate values =',df.duplicated().sum())
# Convert datatype of 'date_added' column to 'datetime' datatype
df2['date_added'] = pd.to_datetime(df2['date_added'])
# Show datatype to confirm changes
print(df2['date_added'].dtypes)
# Extract year and month columns from 'date_added' column
df2['year_added'] = df2['date_added'].dt.year
df2['month_added'] = df2['date_added'].dt.month
# Split the 'listed_in' column and count genres
df2['genre'] = df2['listed_in'].str.split(', ')
df2 = df2.explode('genre').reset_index(drop=True)
# Casting 'genre' column into 'category' datatype
df2['genre'] = df2['genre'].astype('category')
# Check datatype after changes
df2['genre'].dtypes
# number of unique values in dataframe df2
df2.nunique()
```

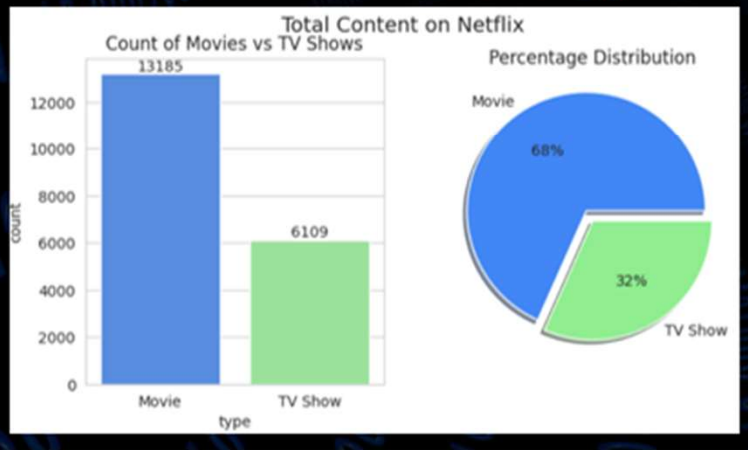
Movie & TV Show Release Year



Ratings on Netflix



Distribution of Content Type



Dataset Summary

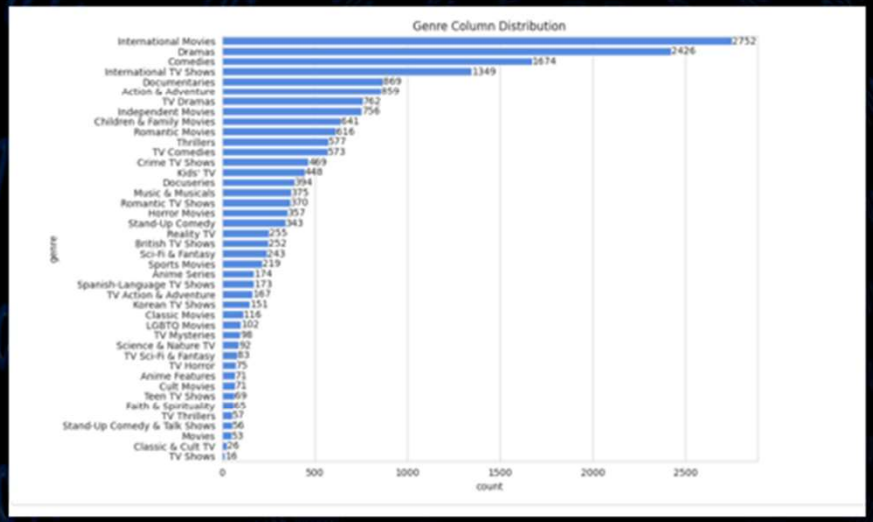
```
df2.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19294 entries, 0 to 19293
Data columns (total 13 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   show_id       19294 non-null  object
 1   type          19294 non-null  object
 2   title         19294 non-null  object
 3   director      19294 non-null  object
 4   country       19294 non-null  object
 5   date_added    19294 non-null  datetime64[ns]
 6   release_year  19294 non-null  int64
 7   rating        19294 non-null  object
 8   duration      19294 non-null  object
 9   listed_in     19294 non-null  object
 10  year_added    19294 non-null  int32
 11  month_added   19294 non-null  int32
 12  genre         19294 non-null  category
dtypes: category(1), datetime64[ns](1), int32(2), i
memory usage: 1.6+ MB
```

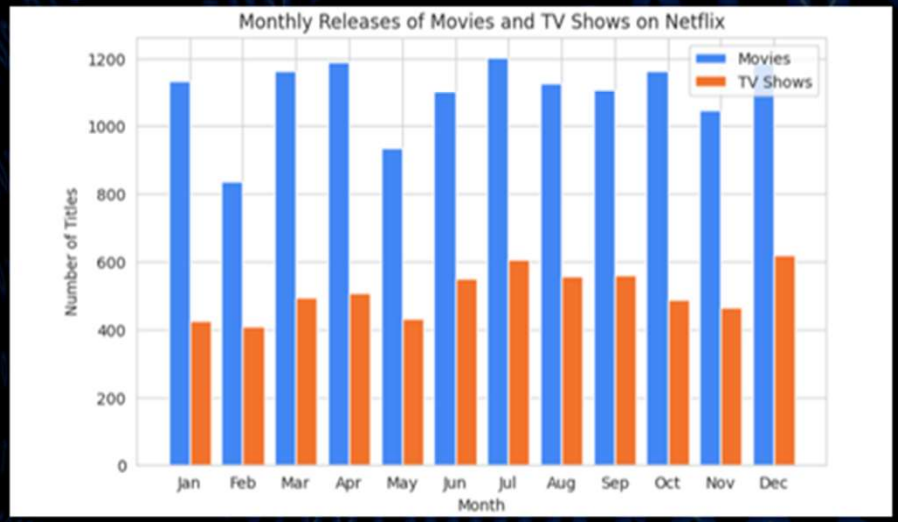
- Netflix content steadily increased after 2000, peaking around 2020, reflecting rapid expansion and aggressive growth.
- Movies dominate Netflix content, with TV Shows forming a substantial secondary library.
- Netflix primarily targets mature audiences while offering diverse ratings for all age groups.

Exploratory Data Analysis (EDA)

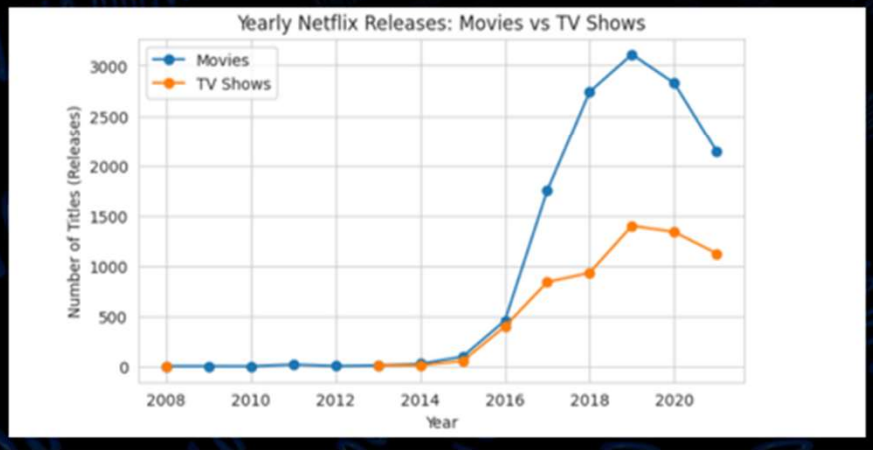
Genres released on Netflix



Monthly Releases: Movies vs TV Shows



Yearly Releases: Movies vs TV Shows



- Netflix emphasizes international movies, dramas, and comedies, offering diverse genres across its global content library.
- Netflix releases fluctuate monthly, peaking midyear and December, with slower February and strategic seasonal surges.
- Netflix content steadily grew, peaking around 2019–2020, accelerating post-2015, with movies consistently outnumbering TV shows.

SQL Analysis For Insights

1) Which directors often release contents across multiple years?

```
select director, count(distinct year_added) as active_years
from netflix_data
where director != 'Not Given'
group by director
having active_years >= 6
order by active_years desc;
```

director	active_years
Jay Karas	7
Alastair Fothergill	6
Lance Bangs	6

2) Top 10 countries producing the most Netflix content?

```
select country, count(*)  as total_titles
from netflix_data
where country != 'Not Given'
group by country
order by total_titles desc
limit 10;
```

country	total_titles
United States	5792
India	2816
United Kingdom	1431
Pakistan	901
Japan	619
South Korea	601
Canada	530
France	509
Spain	476
Mexico	340

3) Which ratings are most common for Movies vs TV Shows?

```
select type, rating, count(*) total_titles
from netflix_data
group by type, rating
order by type, total_titles desc;
```

Result Grid | Filter Rows: | Export:

type	rating	total_titles
Movie	TV-MA	4445
Movie	TV-14	3462
Movie	R	1578
Movie	TV-PG	1208
Movie	PG-13	998
Movie	PG	607
Movie	TV-G	263
Movie	TV-Y7	207
Movie	TV-Y	165
Movie	NR	154
Movie	G	73
Movie	TV-Y7-FV	10
Movie	UR	8
Movie	NC-17	7
TV Show	TV-MA	2861
TV Show	TV-14	1777
TV Show	TV-PG	697
TV Show	TV-Y7	320
TV Show	TV-Y	240
TV Show	TV-G	196
TV Show	NR	11
TV Show	R	4
TV Show	TV-Y7-FV	3

4) How many titles were added to Netflix each year?

```
select year_added, count(*) as total_titles
from netflix_data
group by year_added
order by year_added;
```

year_added	total_titles
2008	4
2009	3
2010	2
2011	19
2012	4
2013	23
2014	37
2015	152
2016	854
2017	2590
2018	3669
2019	4507
2020	4162
2021	3268

5) Top 10 most popular genres on Netflix?

```
select genre, count(*) as total_titles
from netflix_data
group by genre
order by total_titles desc
limit 10;
```

genre	total_titles
International Movies	2752
Dramas	2426
Comedies	1674
International TV Shows	1349
Documentaries	869
Action & Adventure	859
TV Dramas	762
Independent Movies	756
Children & Family Movies	641
Romantic Movies	616

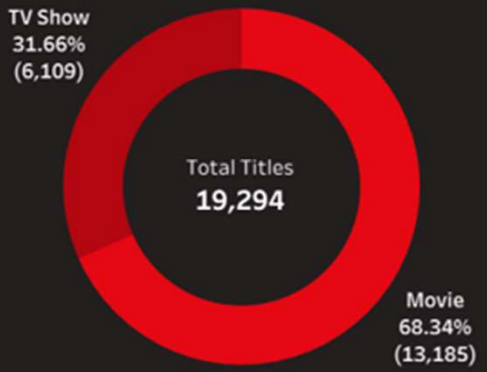
Tableau Dashboard Insights

NETFLIX DATA ANALYSIS DASHBOARD

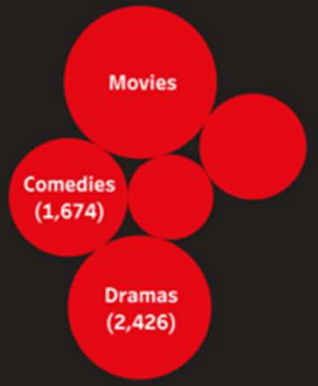
Title	Countries	Avg. Duration	Director
19,294	86	103.8	4,528

Country Map

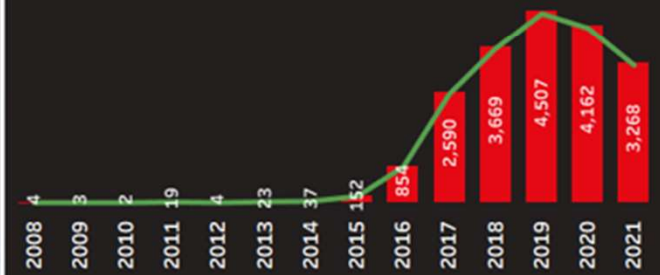
Percentage Ratio



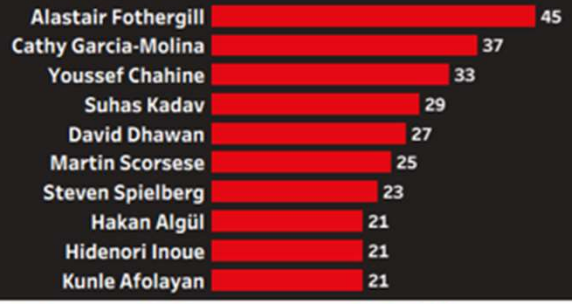
Top 5 Categories



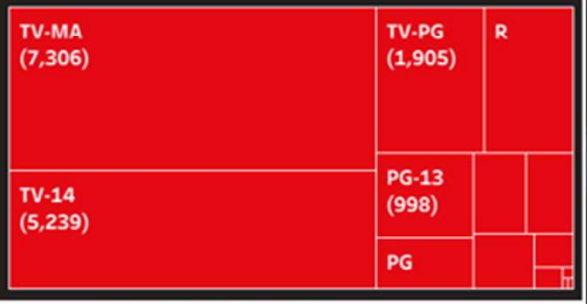
Yearly Progression



Top 10 Directors



Content by Rating



Recommendations

Recommendation 1 :

Increase investment in Drama and Comedy genres

- Supported by highest title counts.
- Expected outcome : Higher audience engagement.

Recommendation 2 :

Focus production and acquisition in top-performing countries

- Supported by country map distribution.
- Expected outcome: Stronger regional market presence.

Recommendation 3 :

Re-evaluate post-2019 content strategy

- Supported by declining yearly additions.
- Expected outcome: Optimized release planning and growth recovery.

Conclusion

- Successfully delivered an end-to-end Netflix content analytics project.
- Cleaned and transformed raw data into actionable insights.
- Built an executive-ready Tableau dashboard.
- Demonstrated strong SQL, Python, and BI skills for real-world decision support.

THANK YOU!

FOR YOUR ATTENTION

By - Subrata Dey
Data Analyst Intern
Unified Mentor Private Limited
Date :