

# NETFLIX DATA: CLEANING, ANALYSIS AND VISUALIZATION

By Himanshu Gautam



NETFLIX



[Linkedin](#)



[Github](#)

**Netflix is a streaming service that offers a wide selection of movies, TV shows, documentaries, anime, and more. It's available in many languages and in over 190 countries, with 283 million paid memberships as of December 2024**

**This project involves loading, cleaning, analyzing, and visualizing data from a Netflix dataset. We'll use Structured Query Language (SQL) to write codes in MySQL Workbench and use Tableau to create charts and dashboards for data visualization.**

## Raw Data:

- Click [here](#) to access the database
- Netflix is a popular streaming service that offers a vast catalog of movies, TV shows, and original content.
- This dataset is a cleaned version of the original version which can be found [here](#)
- This dataset will be cleaned with MySQL and visualized with Tableau.

## Data Collection Period:

- The data consists of content added to Netflix from 2008 to 2021
- The oldest content is as old as 1925 and the newest as 2021

## Importing Data in MySQL and Data Manipulation:

- Importing data from CSV file to MySQL workbench
- Cleaning data to remove duplicates, missing values, and outliers for accurate analysis.

## Tasks:

- Data Cleaning
- Analysis
- Visualization

## Technologies used:

- MySQL
- Tableau

## Creating Database and importing data in MySQL:

```
create database Netflix;  
use Netflix;  
CREATE TABLE mytable (  
show_id VARCHAR(1024),  
type VARCHAR(1024),  
title VARCHAR(1024),  
director VARCHAR(1024),  
country VARCHAR(1024),  
release_year BIGINT,  
rating VARCHAR(1024),  
duration VARCHAR(1024),  
listed_in VARCHAR(1024)  
);
```

```
INSERT INTO mytable  
(show_id,type, title, director, country , release_year ,rating, duration,  
listed_in) VALUES(.....)
```

## CONTENT DISTRIBUTION ON NETFLIX

**SQL Code:** use Netflix ;  
select count(type) as total  
from mytable ;

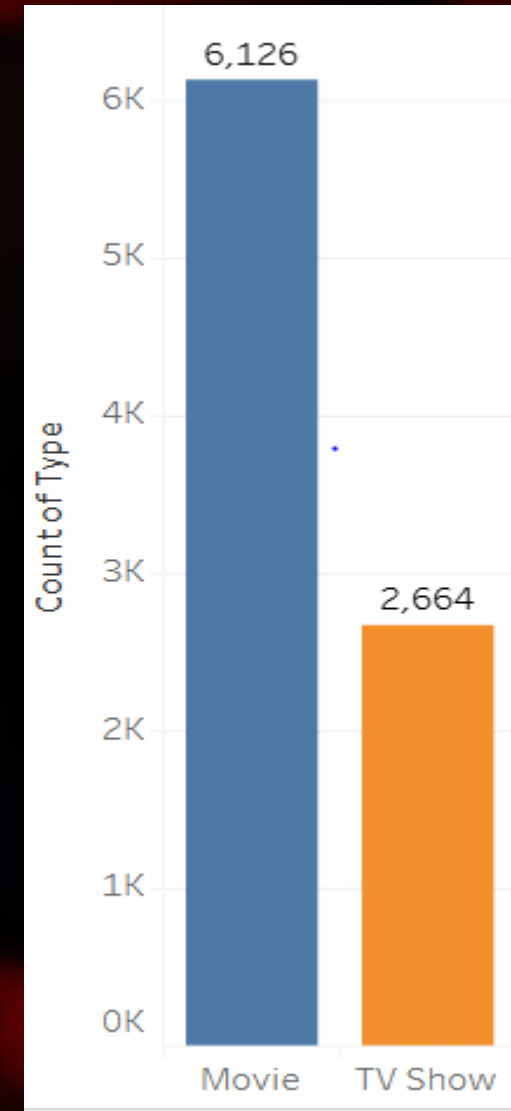
**Total:** 8790

**SQL Code:** use Netflix ;  
select count(type) as total from mytable  
where type = "Movie";

**Total:** 6126

**SQL Code:** use Netflix ;  
select count(type) as total from mytable  
where type = " TV Show";

**Total:** 2664



## RATING ON NETFLIX:

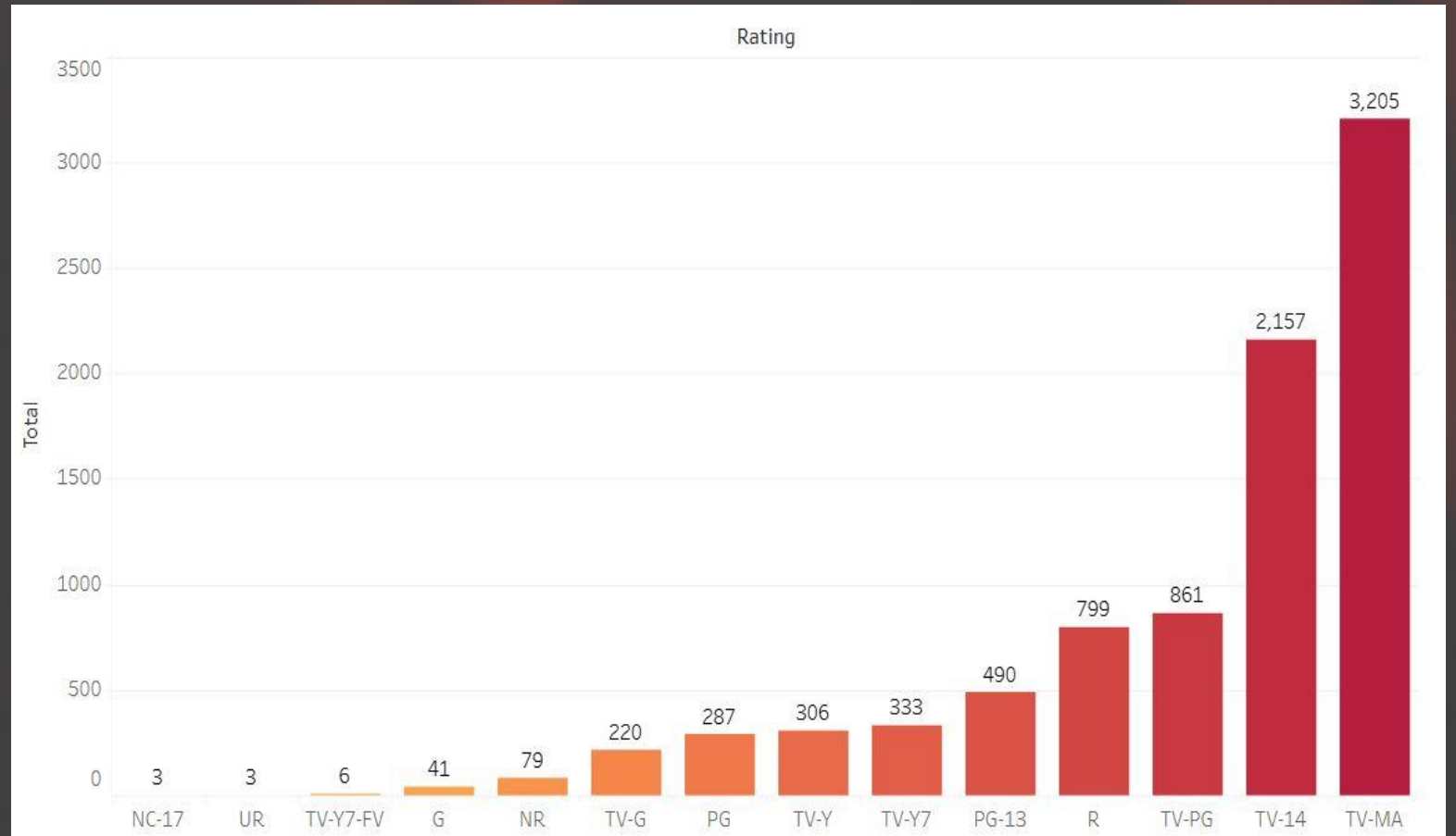
**SQL Code:** use Netflix;

select rating , count(type) as Total from mytable

group by rating

order by count(type) desc;

rating	Total
TV-MA	3205
TV-14	2157
TV-PG	861
R	799
PG-13	490
TV-Y7	333
TV-Y	306
PG	287
TV-G	220
NR	79
G	41
TV-Y7...	6
NC-17	3
UR	3





## TOP 10 COUNTRIES WITH THE MOST CONTENT ON NETFLIX :

**SQL Code:** use Netflix;

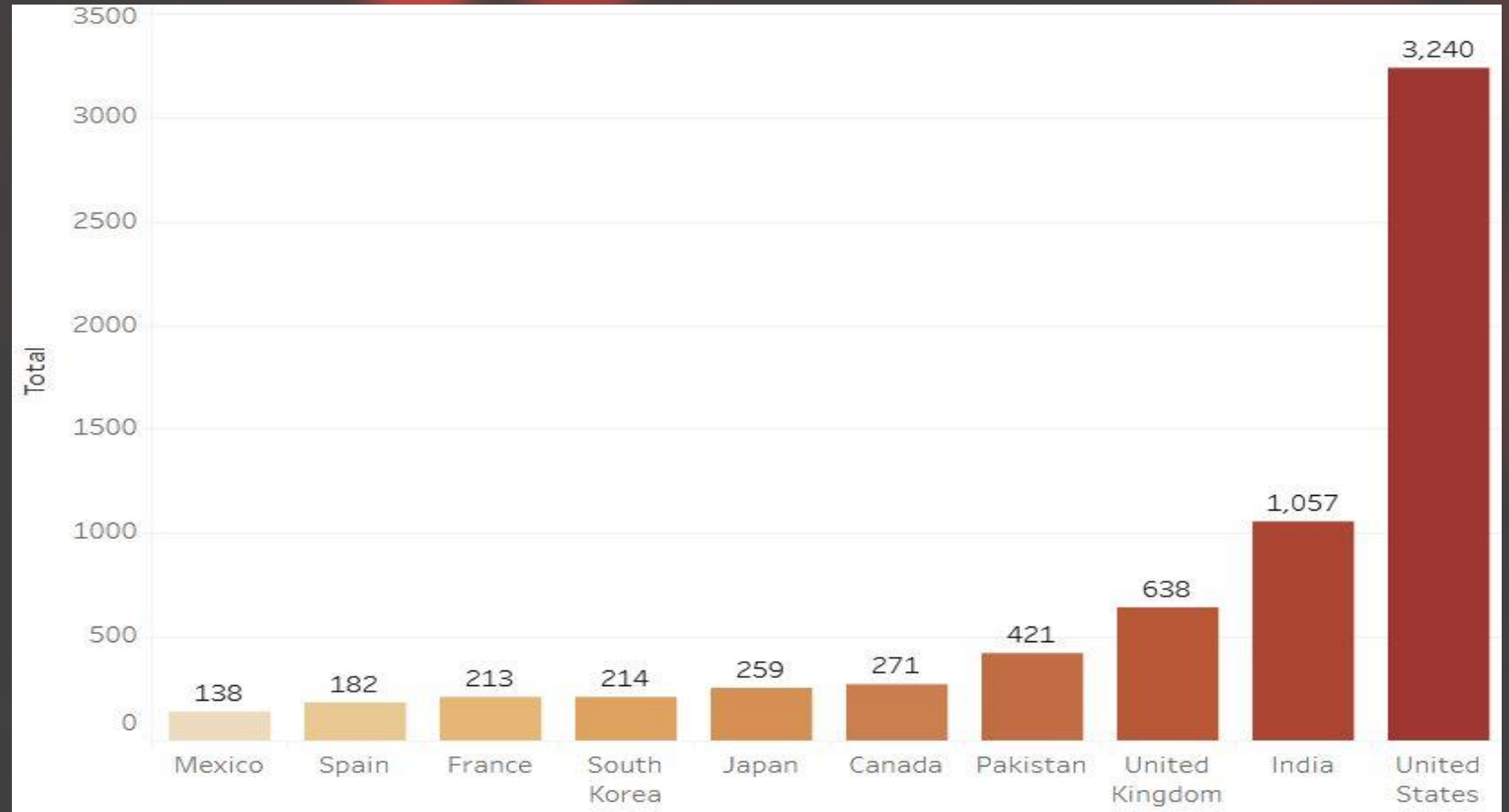
select country , count(type) as Total from mytable

group by country

order by count(type) desc

limit 10;

country	Total
United States	3240
India	1057
United Kingdom	638
Pakistan	421
Canada	271
Japan	259
South Korea	214
France	213
Spain	182
Mexico	138





# YEARLY RELEASES OF MOVIES AND TV SHOWS ON NETFLIX:

**SQL Code:** use Netflix;

select release\_year, type , count(type) as Total\_no from mytable

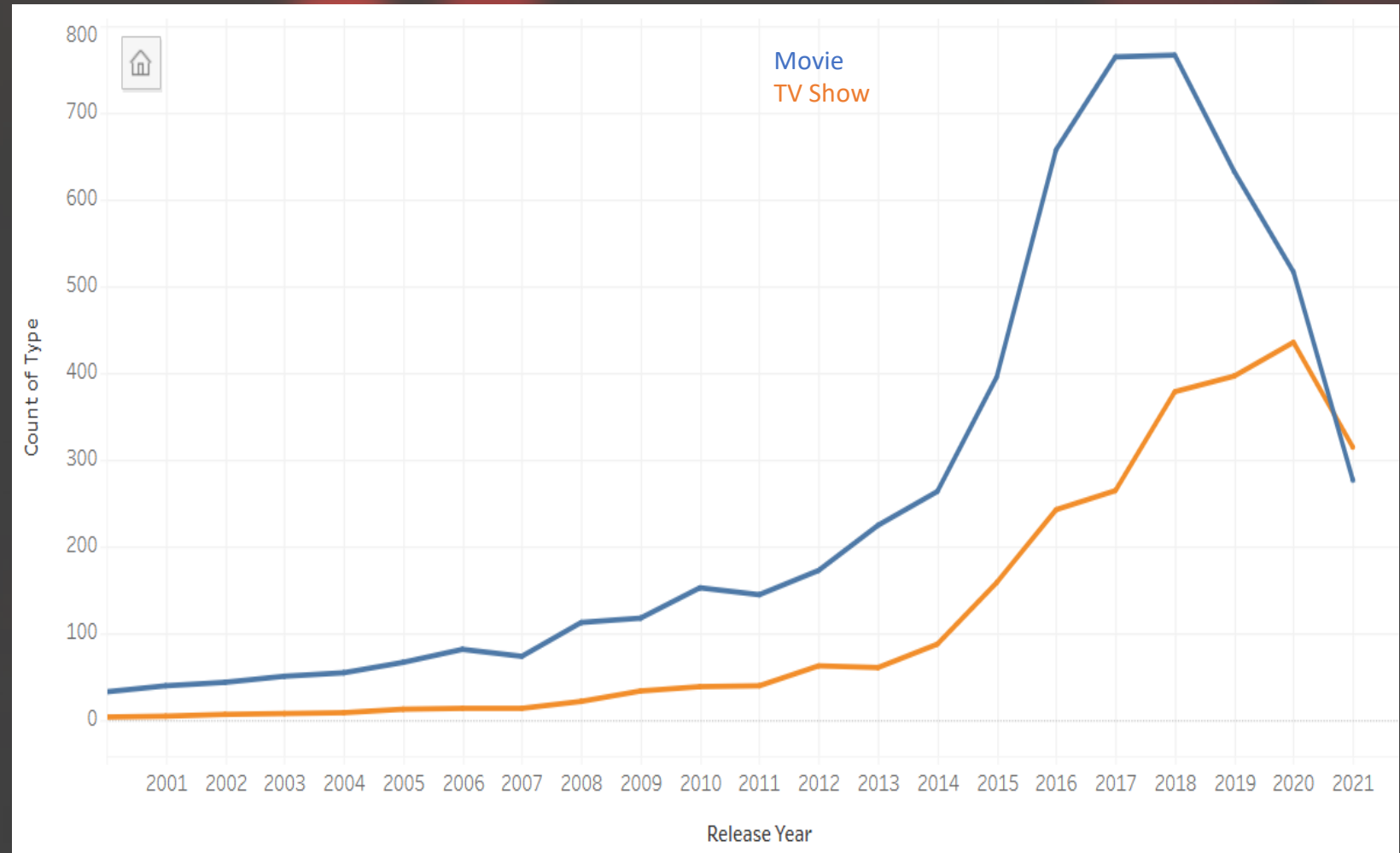
where release\_year >2000

group by type, release\_year

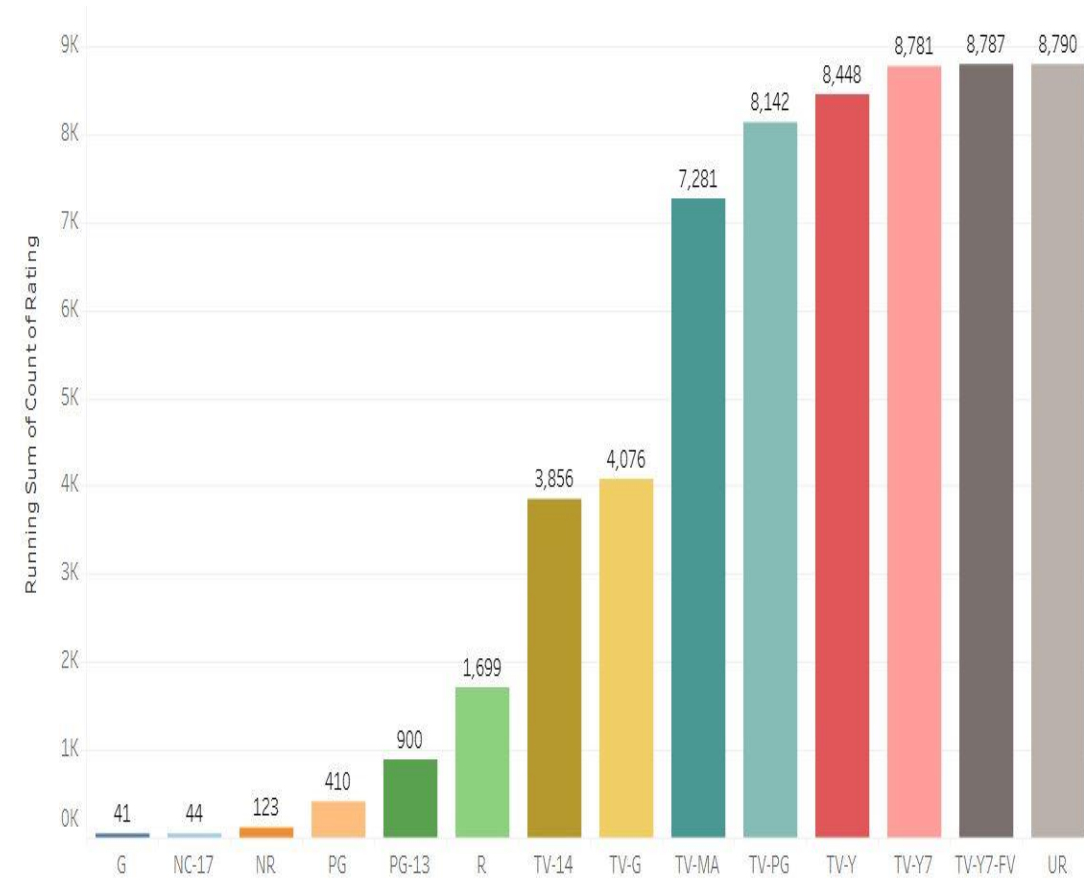
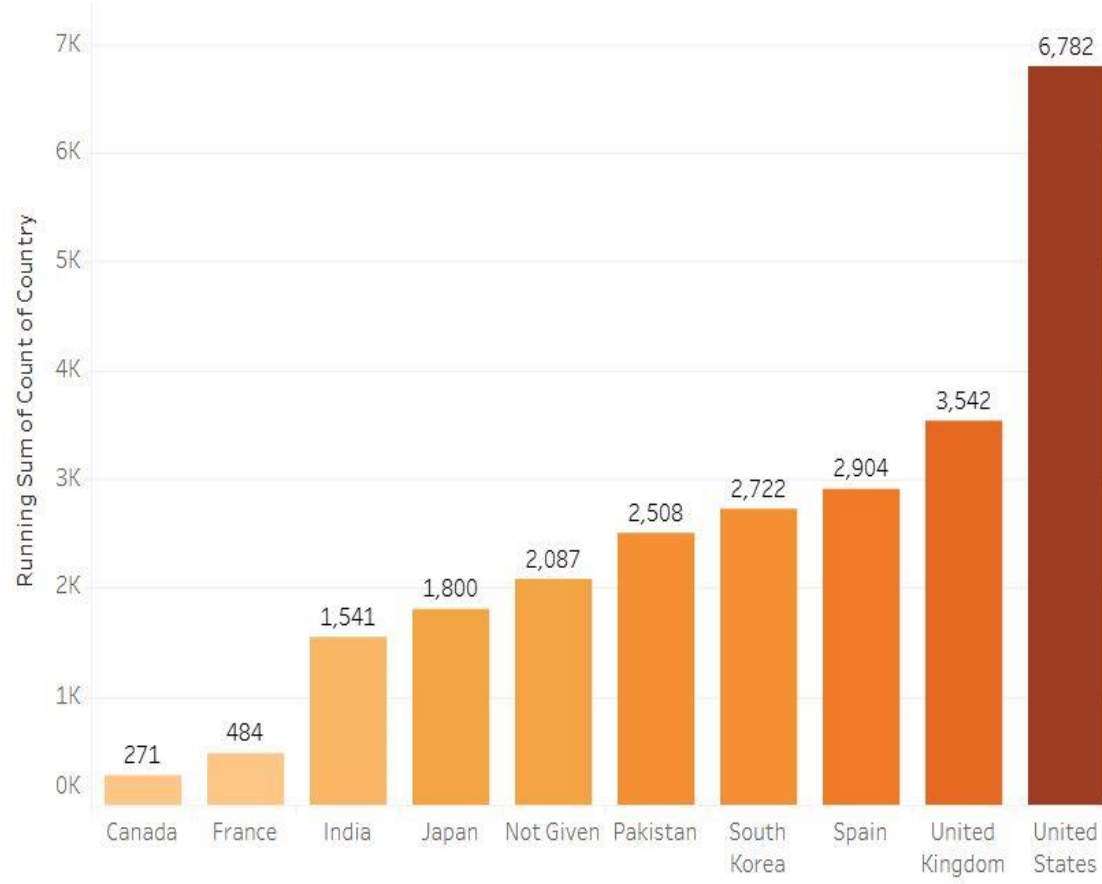
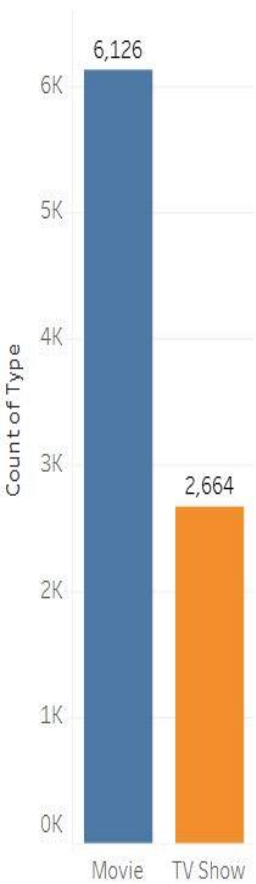
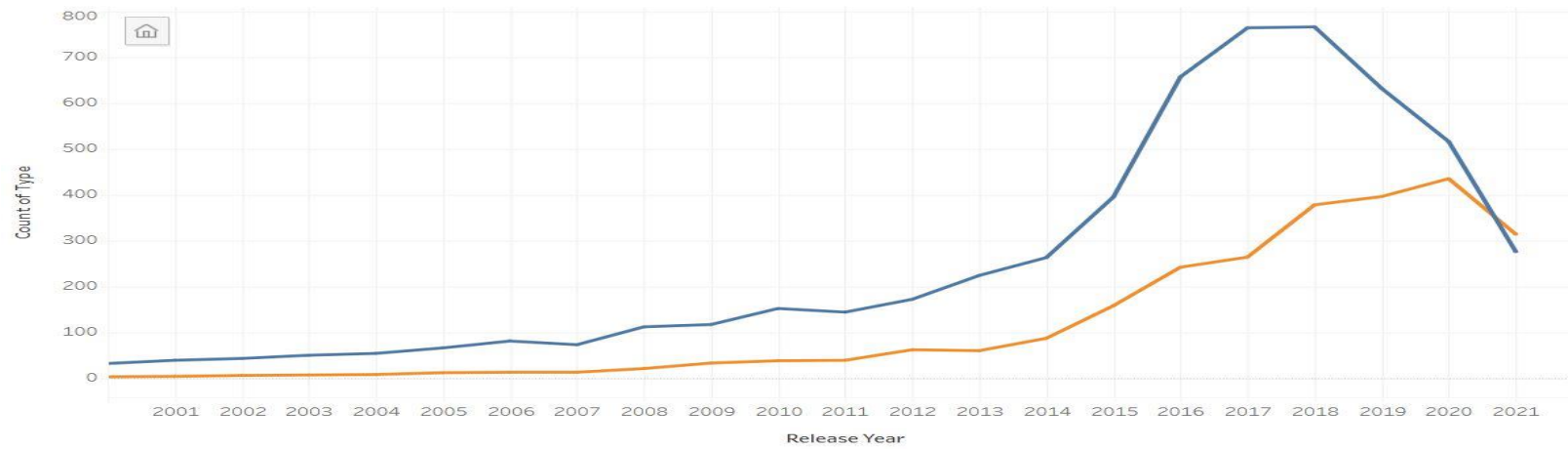
order by release\_year desc;

release_year	type	Total_no
2021	TV Show	315
2021	Movie	277
2020	Movie	517
2020	TV Show	436
2019	Movie	633
2019	TV Show	397
2018	TV Show	379
2018	Movie	767
2017	TV Show	265
2017	Movie	765
2016	Movie	658
2016	TV Show	243
2015	TV Show	159
2015	Movie	396
2014	Movie	264
2014	TV Show	88
2013	Movie	225
2013	TV Show	61
2012	Movie	173
2012	TV Show	63
2011	Movie	145
2011	TV Show	40
2010	Movie	153
2010	TV Show	39
2009	Movie	118
2009	TV Show	34

2008	TV Show	22
2008	Movie	113
2007	Movie	74
2007	TV Show	14
2006	Movie	82
2006	TV Show	14
2005	Movie	67
2005	TV Show	13
2004	Movie	55
2004	TV Show	9
2003	Movie	51
2003	TV Show	8
2002	Movie	44
2002	TV Show	7
2001	Movie	40
2001	TV Show	5



# NETFLIX TABLEAU DASHBOARD



## SAMPLE INSIGHTS:

- *The majority of Netflix content is Movie, with a significant increase in Movie releases over the year*
- *TV Shows are the second most common type of content, with a steady release pattern over the year*
- *The top 5 ratings are TV-MA, TV-14, TV-PG, R, And PG-13*
- *The Top 3 countries with the maximum viewership are the United States, India, and the United Kingdom*
- *Maximum number of viewership likes to watch Movies over TV Shows*
- *There is a clear trend of increasing content released over the years, with a significant spike in 2013*