

Analysis of content produced by Netflix over the past years

Sanjay chauhan

2/21/2022

Introduction

This report has been prepared for descriptive analysis regarding content added by the online streaming platform giant Netflix over the past years. We focused on some of the key factors such as program type for instance whether they are focusing more on movies or T.V. shows and at which rating type their majority content lies. We also extracted the list of countries who are actively contributing to the streaming giant platform. And also analyzed the popularity of Netflix over the past years. While diving deep we also had a look on week days with most additions of content.

Exploratory Data Analysis

Now, we look at the [Data](#)

```
df <- read.csv("netflix.csv")
View(df)
```

Data wrangling process

Verification of data types assigned by default and rectifying those who are at fault

```
df$date_added <- dmy(df$date_added)
## Warning: 88 failed to parse.
```

Counting the data elements which are not available

```
df %>% map(~sum(is.na(.)))
## $show_id
## [1] 0
##
## $type
## [1] 0
##
## $title
## [1] 0
##
## $director
```

```
## [1] 0
##
## $cast
## [1] 0
##
## $country
## [1] 0
##
## $date_added
## [1] 98
##
## $release_year
## [1] 0
##
## $rating
## [1] 0
##
## $duration
## [1] 0
##
## $listed_in
## [1] 0
##
## $description
## [1] 0
```

Dropping the data rows which contain NAs if no other option or alternative available

```
df1 <- df %>% drop_na()
```

Filtering out the empty data cells and reframing data structure for accurate analysis

```
df2 <- df1 %>% select(-c(cast,director,description,listed_in)) %>%
  filter(country!="") %>% filter(duration!="") %>% filter(rating!="")
```

Now, I have added a new column called “weekday” for analyzing performance on each day of a week

```
df2$week_day <- format(as.Date(df2$date_added), "%A")
```

Here I reordered week days in the pervasive order

```
df2$week_day <-
ordered(df2$week_day, level=c("Sunday", "Monday", "Tuesday", "Wednesday",
"Thursday", "Friday", "Saturday"))
```

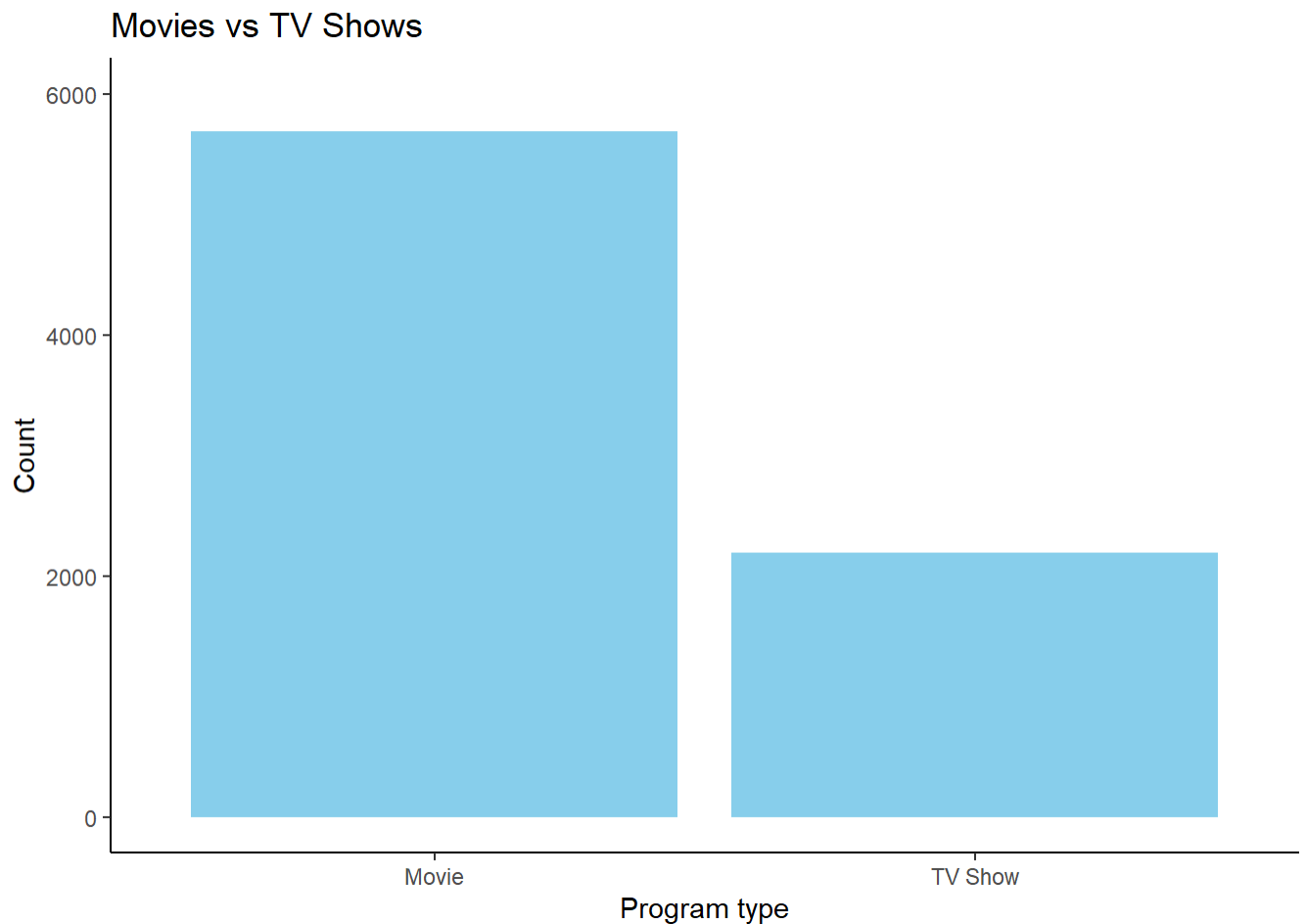
Now, if we look at the range of the data, we find that this data giving us a time span of **2008-2021** during which Netflix continuously focused on the type of content it has to provide to its audience. We also looked at the range of release years of those movies and we find that this time span is around **98 years** which indicates Netflix has added movies which released back in **1925**.

```
range(df2$date_added)
## [1] "2008-01-01" "2021-09-25"
range(df2$release_year)
## [1] 1942 2021
```

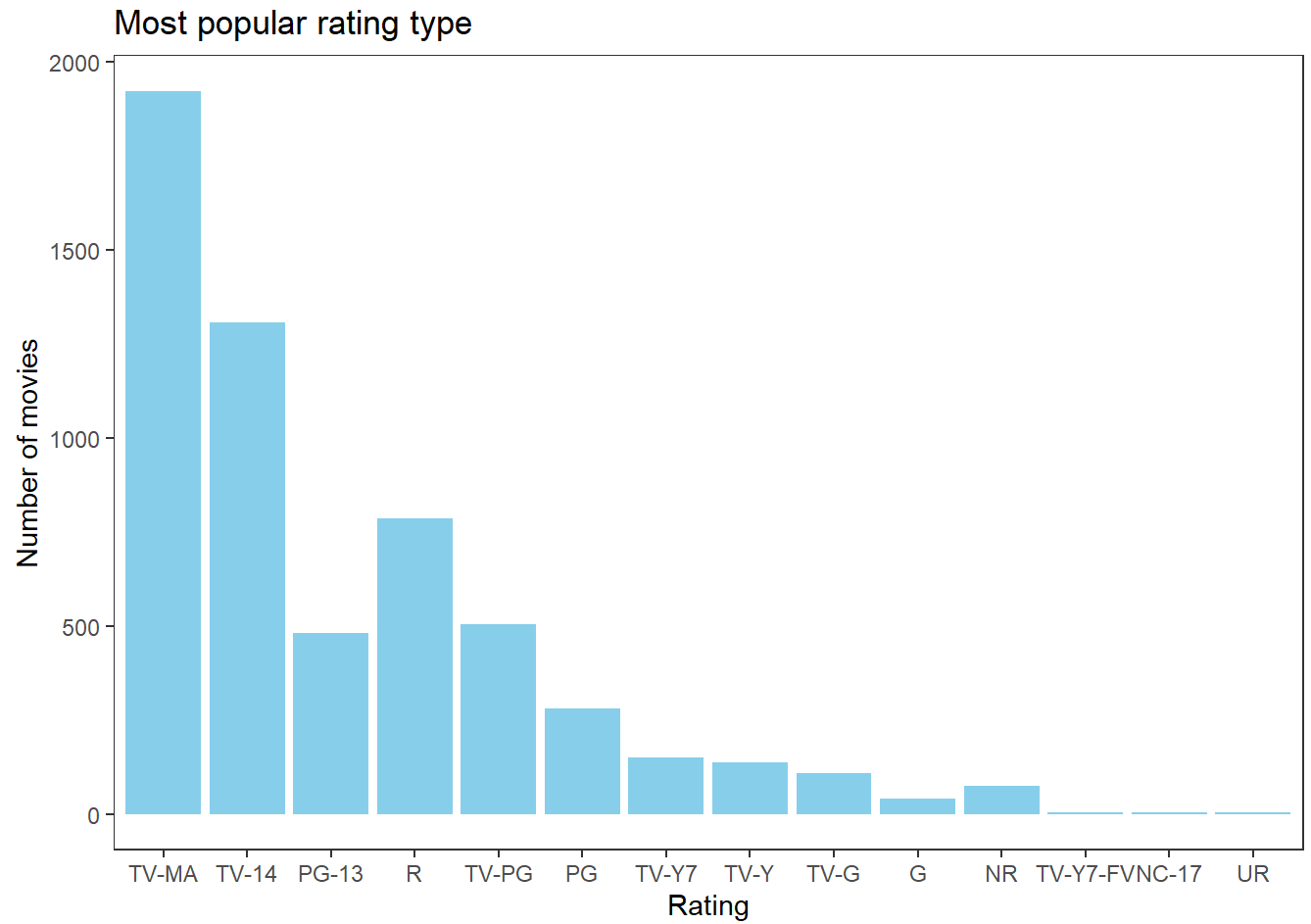
Descriptive Summaries and Visualization

Movies vs T.V. shows comparison-

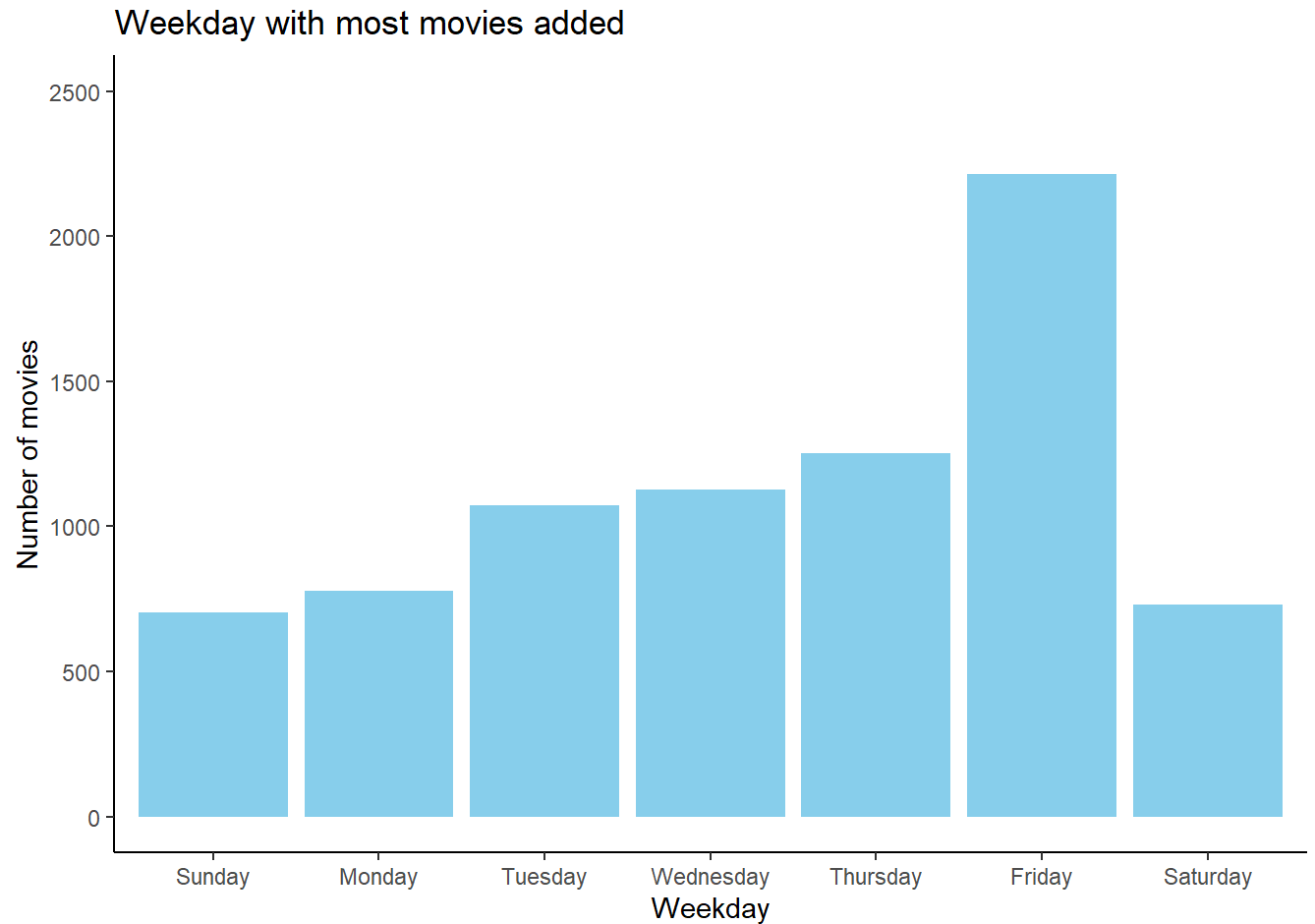
```
program_type <- df2 %>% group_by(type) %>%
  summarise(Number_of_movies=n())
head(program_type)
```



```
program_rating <- df2 %>% group_by(rating,type) %>% # summary 2
  summarise(count=n())
## `summarise()` has grouped output by 'rating'. You can override using the
## `.groups` argument.
program_rating[order(-program_rating$count),]
```



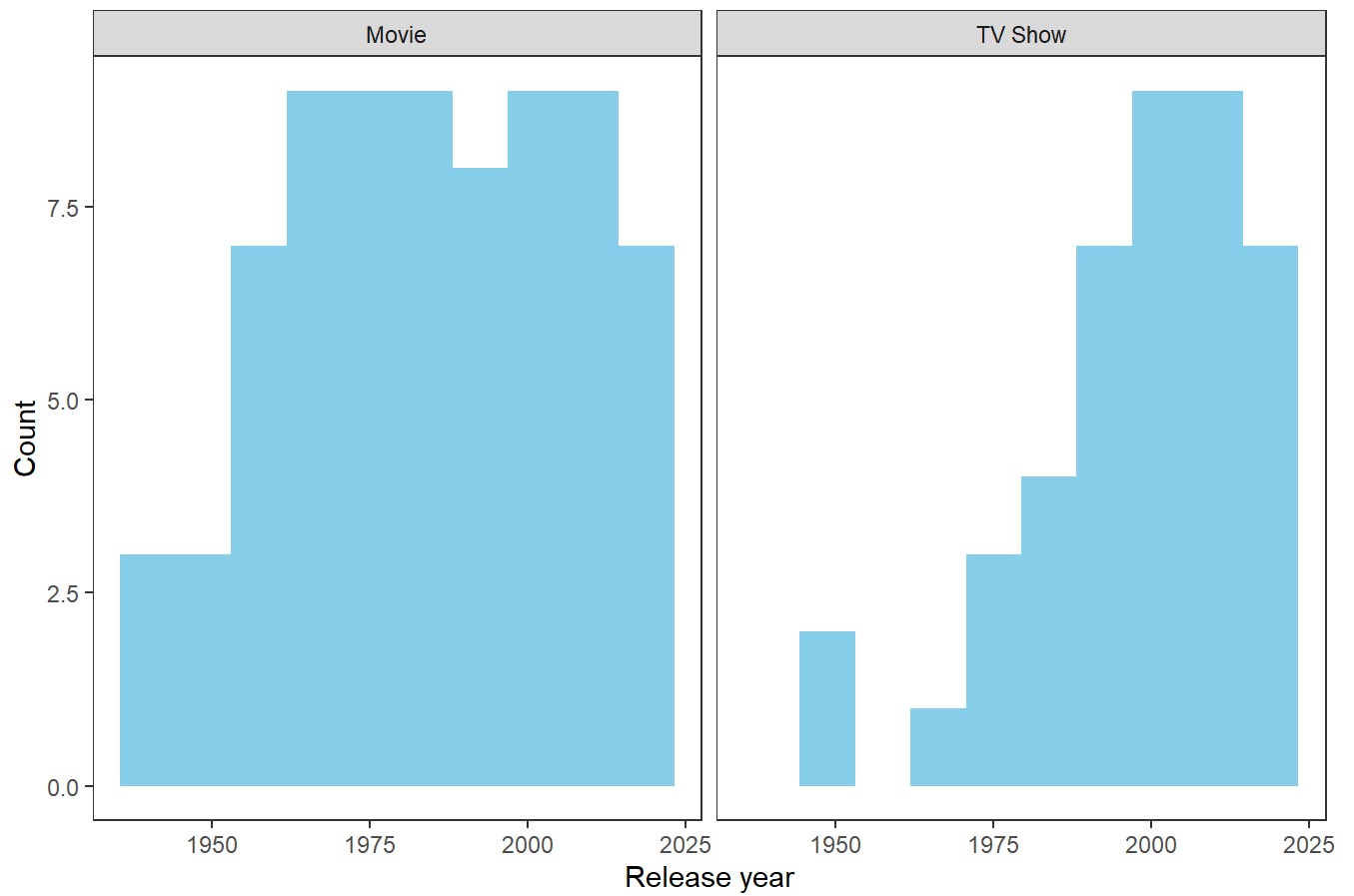
```
busiest_weekday <- df2 %>% group_by(week_day) %>%  
  summarise(count_of_movies_added=n())  
head(busiest_weekday,7)  
## Warning: Ignoring unknown parameters: postition
```



```
release_year <- df2 %>% group_by(release_year,type) %>%  
  summarise(count_of_movies_released=n())  
## `summarise()` has grouped output by 'release_year'. You can override using  
the `.groups` argument.
```

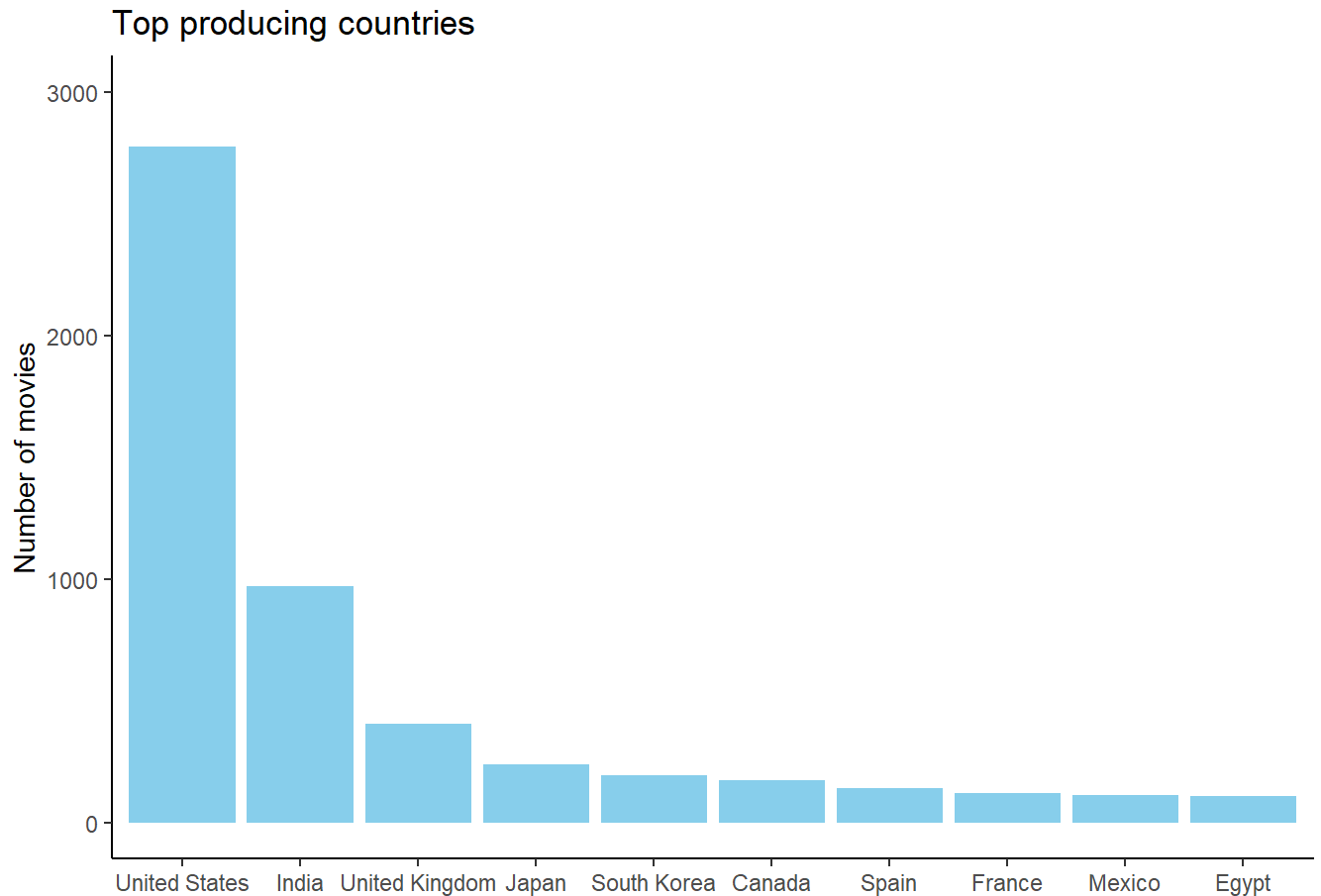
A distribution of release years of programs Netflix offers and they are categorized on the basis of show type.

Movies vs TV shows popularity: Timely comparison



Finally, the list of countries with most number of movies and T.V. shows available on Netflix

```
country <- df2 %>% group_by(country) %>%  
  summarise(top_producers=n())  
  
country <- country[order(-country$top_producers),]  
  
country_1 <- head(country,10)  
head(country_1,10)
```



Conclusion

- In the time span of around 13 years, Netflix has widely focused on movies if we compare it to the T.V. shows but with the time T.V. shows also becoming common on the platform. We have witnessed an increase in the addition of T.V. shows as well.
- Friday is the day with most content streaming giant added in their platform and the reason is quite obvious as it is the beginning of weekend.
- If we look at the distribution chart of release years of movies and T.V. shows currently available on Netflix, we would find that the T.V. programs are getting common with the time and most movies or shows Netflix has on its platform are released back in **2018**.
- Now, if we look at the top producers of content on Netflix we would find that the United states leads it with almost 53% of content while India ranks second with around 18.5% content and the common factor in top 2s are that both countries have a giant Film industry and an enormous local as well as international audience.
- Looking at the most common rating among all, TV-MA arrives frequently which indicates Netflix mostly focuses on mature/18+ audience.