Disney+ and Netflix TV Shows and Movies Analysis

Jacqueline Maloney & Aruushi Naik

Trends and Patterns in Disney+ and Netflix TV Shows and Movies: A Comparative Analysis

Figure 1: Netflix and Disney+

Introduction

Disney+ and Netflix are two of the most well-known streaming services in the world, providing millions of users with access to a vast variety of TV series and films. We conduct an exploratory data analysis (EDA) of the platforms and their films and television series that are available on Netflix and Disney+ in this analysis. Given the dynamic nature of digital entertainment, knowing how content trends and patterns have changed over time offers important insights regarding platform expansion, production strategy, and consumer preferences.

Key patterns in Disney+ and Netflix content will be examined in this analysis, such as variations in the number of movies and TV shows each year, average running times, regional production distribution, content ratings, etc. Disney+ and Netflix's future content choices and marketing plans will be influenced by the insights this analysis provides into consumer preferences, strategic planning for content, and platform evolution.

Broader research question for our analysis is :

What trends and patterns can be observed in Disney Plus and Netflix TV shows and movies?

We are interested in comparing these two platforms to determine if one is better than the other or has more to offer to their consumers. We can determine these trends and patterns by analyzing the number of TV shows and movies, duration lengths, directors, regional distribution, content rating and comparing release and added years of content. These six topics and attributes will be our areas of focus in helping us determine any trends and patterns from our dataset.

Dataset Overview

To address the research question, we will conduct an in-depth analysis of the Disney+ and Netflix data sets, examining various aspects of the content offered on both platforms. These two datasets were taken from Kaggle. Shivam Bansal had provided both of our sets containing 12 columns of data with many entries across Netflix and Disney+. Each case in our data set is a movie or TV show from either Netflix or Disney+. This data set supports the FAIR principles as it has public access in a CSV format. The variables and data are interoperable and reusable for further analysis. It also aligns with CARE by providing collective benefit and data collection was transparent. The dataset contains factual data strictly relating to the Netflix and Disney+ streaming platform that can be used and analyzed by anyone.

Disney+ Data Preview

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
s1	Movie	Duck the Hal...	Alonso Ramir...	Chris Diaman...		November 26,...	2016	TV-G	23 min	Animation, F...	Join Mickey ...
s2	Movie	Ernest Saves...	John Cherry	Jim Varney, ...		November 26,...	1988	PG	91 min	Comedy	Santa Claus ...
s3	Movie	Ice Age: A M...	Karen Disher	Raymond Albe...	United States	November 26,...	2011	TV-G	23 min	Animation, C...	Sid the Slot...
s4	Movie	The Queen Fa...	Hamish Hamilton	Darren Criss...		November 26,...	2021	TV-PG	41 min	Musical	This is real...
s5	TV Show	The Beatles:...		John Lennon,...		November 25,...	2021		1 Season	Docuseries, ...	A three-part...
s6	Movie	Becoming Cou...	Liz Garbus	Jacques Yves...	United States	November 24,...	2021	PG-13	94 min	Biographical	An inside lo...

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
s7	TV Show	Hawkeye		Jeremy Renne...		November 24,...	2021	TV-14	1 Season	Action-Adven...	Clint Barton...
s8	TV Show	Port Protect...		Gary Muehlbe...	United States	November 24,...	2015	TV-14	2 Seasons	Docuseries, ...	Residents of...
s9	TV Show	Secrets of t...		Dr. Ray Ball...	United States	November 24,...	2019	TV-PG	2 Seasons	Animals & Na...	A day in the...
s10	Movie	A Muppets Ch...	Kirk R. That...	Steve Whitmi...	United States	November 19,...	2008	G	45 min	Comedy, Fami...	Celebrate th...

Netflix Data Preview

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
s1	Movie	Dick John-son...	Kirsten Johnson		United States	September 25...	2020	PG-13	90 min	Documentaries	As her fathe...
s2	TV Show	Blood & Water		Ama Qamata, ...	South Africa	September 24...	2021	TV-MA	2 Seasons	Internationa...	After crossi...
s3	TV Show	Ganglands	Julien Leclercq	Sami Boua-jil...		September 24...	2021	TV-MA	1 Season	Crime TV Sho...	To protect h...
s4	TV Show	Jailbirds Ne...				September 24...	2021	TV-MA	1 Season	Docuseries, ...	Feuds, flirt...
s5	TV Show	Kota Factory		Mayur More, ...	India	September 24...	2021	TV-MA	2 Seasons	Internationa...	In a city of...
s6	TV Show	Midnight Mass	Mike Flana-gan	Kate Siegel,...		September 24...	2021	TV-MA	1 Season	TV Dramas, T...	The arrival ...
s7	Movie	My Little Po...	Robert Culle...	Vanessa Hudg...		September 24...	2021	PG	91 min	Children & F...	Equestria's ...
s8	Movie	Sankofa	Haile Gerima	Kofi Ghan-aba...	United State...	September 24...	1993	TV-MA	125 min	Dramas, Inde...	On a photo s...
s9	TV Show	The Great Br...	Andy Devon-shire	Mel Giedroyc...	United King-dom	September 24...	2021	TV-14	9 Seasons	British TV S...	A talented b...
s10	Movie	The Starling	Theodore Melfi	Melissa McCa...	United States	September 24...	2021	PG-13	104 min	Comedies, Dr...	A woman adju...

The tables above were built using the head() method and display the first few values from the data sets. It provides a quick preview of the data set's structure, column names, and value types, which is important for understanding the data's overall format and content.

Disney+ Data Summary

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
Length:1450	Length:1450	Length:1450	Length:1450	Length:1450	Length:1450	Length:1450	Min. :1928		Length:1450	Length:1450	Length:1450
Class :charac-ter	Class :charac-ter	Class :charac-ter	Class :charac-ter	Class :charac-ter	Class :charac-ter	Class :charac-ter	1st Qu.:1999		Class :charac-ter	Class :charac-ter	Class :charac-ter	Class :character
Mode :charac-ter	Mode :charac-ter	Mode :charac-ter	Mode :charac-ter	Mode :charac-ter	Mode :charac-ter	Mode :charac-ter	Median :2011		Mode :charac-ter	Mode :charac-ter	Mode :charac-ter	Mode :character
NA	NA	NA	NA	NA	NA	NA	Mean :2003		NA	NA	NA	NA

NA	NA	NA	NA	NA	NA	NA	3rd Qu.:2018	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	Max. :2021	NA	NA	NA	NA

Netflix Data Summary

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
Length:8807	Length:8807	Length:8807	Length:8807	Length:8807	Length:8807	Length:8807	Min. :1925	Length:8807	Length:8807	Length:8807	Length:8807
Class :charac-ter	Class :charac-ter	Class :charac-ter	Class :charac-ter	Class :charac-ter	Class :charac-ter	Class :charac-ter	1st Qu.:2013	Class :charac-ter	Class :charac-ter	Class :charac-ter	Class :character
Mode :charac-ter	Mode :charac-ter	Mode :charac-ter	Mode :charac-ter	Mode :charac-ter	Mode :charac-ter	Mode :charac-ter	Median :2017	Mode :charac-ter	Mode :charac-ter	Mode :charac-ter	Mode :character
NA	NA	NA	NA	NA	NA	NA	Mean :2014	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	3rd Qu.:2019	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	Max. :2021	NA	NA	NA	NA

Additionally, we used the summary() function to create a summary table. Each column in the data data frame is summarized in the table that is shown above.

Data Cleaning and Wrangling

To streamline the analysis, the datasets for Disney+ and Netflix were merged into a single dataset. This joining process ensures a consistent structure for comparative analysis and enables efficient exploration of trends across both the platforms.

```
# Add a platform column to the Disney+ dataset
disney_plus <- disney_data %>%
  mutate(platform = "Disney Plus")
# Add a platform column to the Netflix dataset
netflix <- netflix_data %>%
  mutate(platform = "Netflix")
# Join the two datasets
moviesandtv <- bind_rows(disney_plus, netflix)
```

Upon looking at the *moviesandtv* dataset, we came across empty cells and decided to replace it with N/A values to account for missing data. This would maintain uniformity throughout the dataset and allow for more accurate analysis by clearly identifying missing information across the Disney+ and Netflix datasets.

To facilitate targeted analysis for each type of content, we separated the *moviesandtv* dataset into two distinct tables: one for movies and one for TV shows. For movies, we changed the

format for duration by removing the "min" label, while for TV shows, we converted the number of seasons into integers to ensure clear and specific analysis of content.

TV Shows Data

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	platform
s5	TV Show	The Beat-les:...	NA	John Lennon,...	NA	November 25,...	2021	NA	1	Docuseries, ...	A three-part...	Disney Plus
s7	TV Show	Hawkeye	NA	Jeremy Renne...	NA	November 24,...	2021	TV-14	1	Action-Adven...	Clint Barton...	Disney Plus
s8	TV Show	Port Pro-tect...	NA	Gary Muehlbe...	United States	November 24,...	2015	TV-14	2	Docuseries, ...	Residents of...	Disney Plus
s9	TV Show	Secrets of t...	NA	Dr. Ray Ball...	United States	November 24,...	2019	TV-PG	2	Animals & Na...	A day in the...	Disney Plus
s14	TV Show	Dr. Oakley, ...	NA	Dr. Michelle...	United States	November 17,...	2013	TV-PG	10	Action-Adven...	Meet Dr. Mic...	Disney Plus
s18	TV Show	Disney Fancy...	NA	Mia Jen-ness,...	United State...	November 12,...	2018	TV-PG	3	Animation, Kids	Nancy makes ...	Disney Plus
s19	TV Show	Disney Inter...	NA	Carolina Dom...	NA	November 12,...	2021	TV-PG	1	Comedy, Comi...	Allegra is r...	Disney Plus
s29	TV Show	Olaf Presents	NA	Josh Gad	NA	November 12,...	2021	TV-PG	1	Animation, F...	Olaf goes fr...	Disney Plus
s52	TV Show	Disney Am-phibia	NA	Justin Felbi...	United State...	November 3, ...	2018	TV-Y7	3	Animation, C...	Anne Boonchu...	Disney Plus
s53	TV Show	Photo Ark	NA	Joel Sartore	United States	November 3, ...	2017	TV-PG	1	Animals & Na...	National Geo...	Disney Plus

The above table showcases the TV Shows dataset which contains data from both the platforms.

Movies Data

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	platform
s1	Movie	Duck the Hal...	Alonso Ramir...	Chris Dia-man...	NA	November 26,...	2016	TV-G	23	Animation, F...	Join Mickey ...	Disney Plus
s2	Movie	Ernest Saves...	John Cherry	Jim Varney, ...	NA	November 26,...	1988	PG	91	Comedy	Santa Claus ...	Disney Plus
s3	Movie	Ice Age: A M...	Karen Disher	Raymond Albe...	United States	November 26,...	2011	TV-G	23	Animation, C...	Sid the Slot...	Disney Plus
s4	Movie	The Queen Fa...	Hamish Hamil-ton	Darren Criss...	NA	November 26,...	2021	TV-PG	41	Musical	This is real...	Disney Plus
s6	Movie	Becoming Cou...	Liz Garbus	Jacques Yves...	United States	November 24,...	2021	PG-13	94	Biographical	An inside lo...	Disney Plus
s10	Movie	A Muppets Ch...	Kirk R. That...	Steve Whitmi...	United States	November 19,...	2008	G	45	Comedy, Fami...	Celebrate th...	Disney Plus
s11	Movie	Adventure Th...	John Gleim	Don Hahn, Ka...	NA	November 19,...	2020	TV-PG	59	Documentary	Explore the ...	Disney Plus
s12	Movie	Puppy for Ha...	NA	NA	NA	November 19,...	2020	TV-G	4	Comedy, Fami...	Check out Da...	Disney Plus
s13	Movie	The Pixar Story	Leslie Iwerks	Stacy Keach,...	United States	November 19,...	2007	G	91	Documentary	A.. ground-brea...	Disney Plus
s15	Movie	America the ...	NA	Michael B. J...	NA	November 12,...	2021	TV-PG	2	Animals & Na...	Epic, grand ...	Disney Plus

The above table showcases the Movies dataset which contains data from both the platforms.

Exploratory Data Analysis(EDA)

How many TV shows and movies from a specific year are available on Netflix and Disney+ ?

TV Shows

Number of TV Shows on Netflix vs. Disney+ From a Particular Year
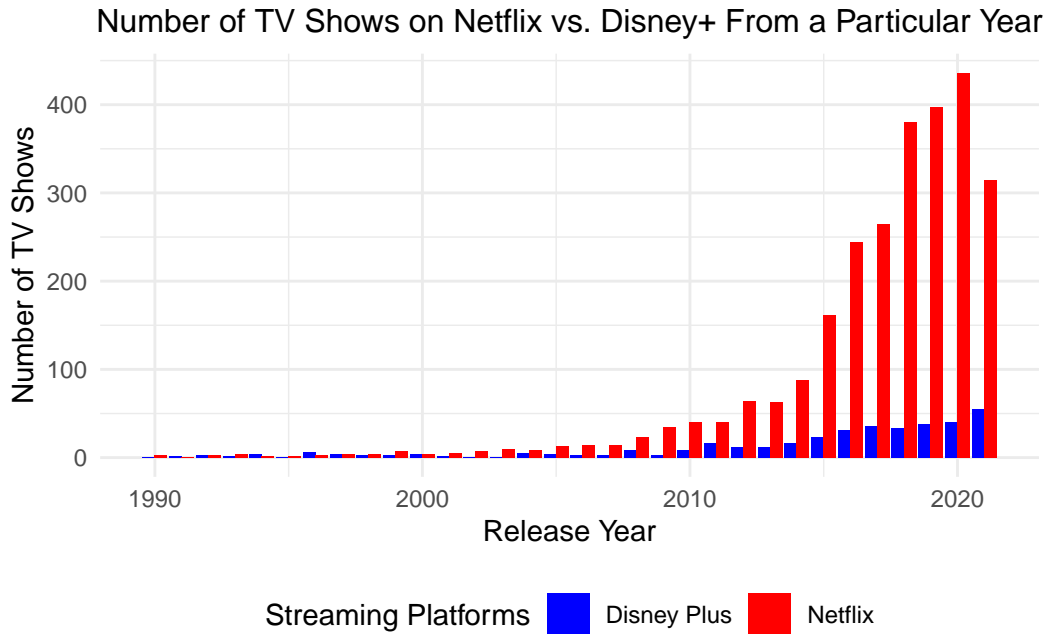
The bar chart illustrates a number of TV series that are accessible on Netflix and Disney+ from each particular year. When it comes to TV shows selection, Netflix constantly outperforms Disney+, especially after the 2000s, which is an indication of its larger selection of TV shows from the year 2000 to 2021. Netflix's content strength is demonstrated its peak at 436 TV shows from the year 2020 alone. Disney+, on the other hand, peaks at 55 TV shows from the year 2021 but still lags behind Netflix. It is interesting to note that Netflix still surpasses Disney+ in 2021 despite a minor decline in TV series from that year. This confirms Netflix's domination in the streaming industry by showing that it offers a wider range of TV shows than Disney+.

Movies

Number of Movies on Netflix vs. Disney+ From a Particular Year
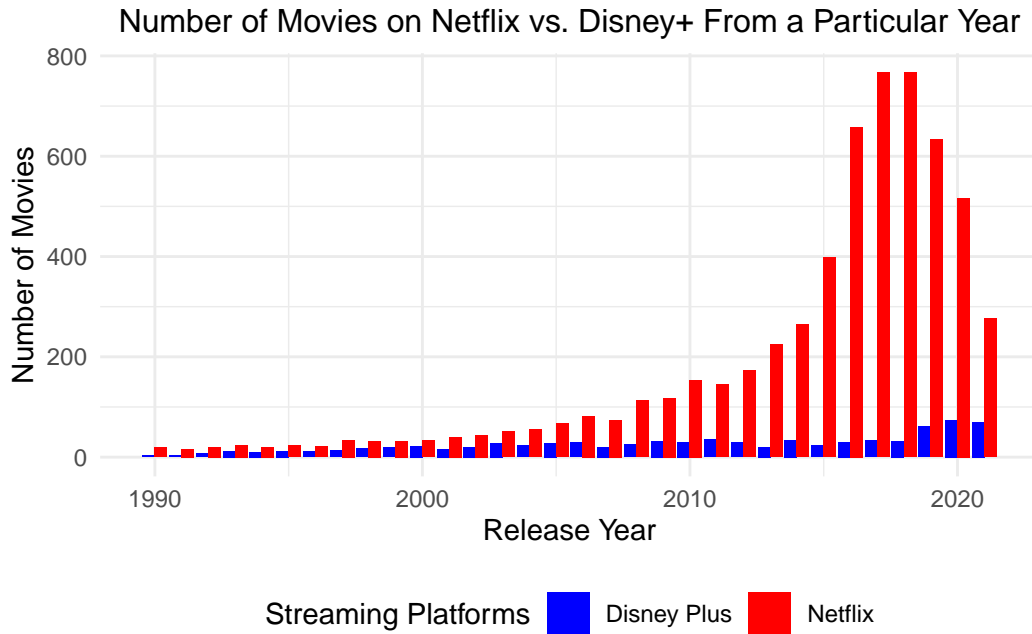
The bar chart shows the quantity of films that have been made available on Netflix and Disney+ throughout time. In terms of movie collection, Netflix significantly surpasses Disney+ since 1990. Netflix demonstrates its extensive content library by reaching its pinnacle in 2018 and 2019 movies, with 767 films annually. But after these years, Netflix sees a drop in movie content, much like their TV show patterns. Disney+, in contrast, has a relatively smaller collection of movies, but it has been steadily growing with its content from the recent years, reaching a peak of 74 films from 2020. Netflix's dominance in the movie streaming industry is highlighted by this trend, although Disney+ is progressively growing its collection, especially with more recent movies.

What is the average duration of movies and TV shows?

TV Shows

Average Seasons of TV Shows on Netflix vs. Disney+
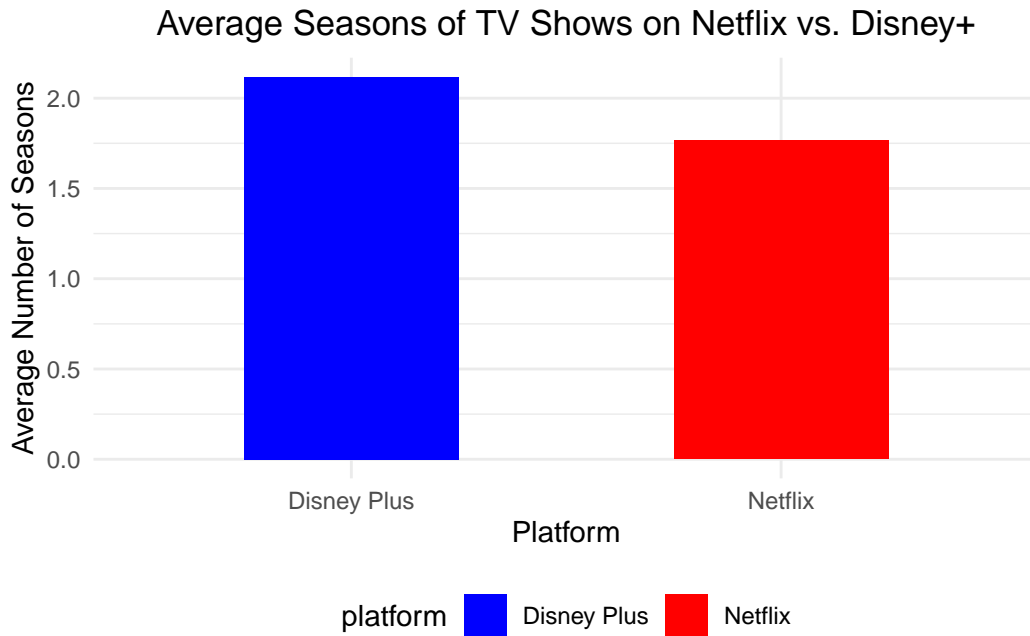
The average number of seasons for Netflix and Disney+ TV series is compared in the plot. It illustrates that, in contrast to Netflix, which normally offers shows with fewer than two seasons, Disney+ usually offers shows with a greater average number of seasons - 2 or more. Disney+'s emphasis on family-friendly franchises like Marvel, Star Wars, Pixar, and Disney Channel series, which frequently have several seasons to appeal to younger viewers who prefer longer-running series, is in line with this trend. Netflix, on the other hand, appeals to a wider audience, including elderly viewers who might like short stories, by offering a varied mix of original content and shorter series.

Movies

Average Time of Movies on Netflix vs. Disney+
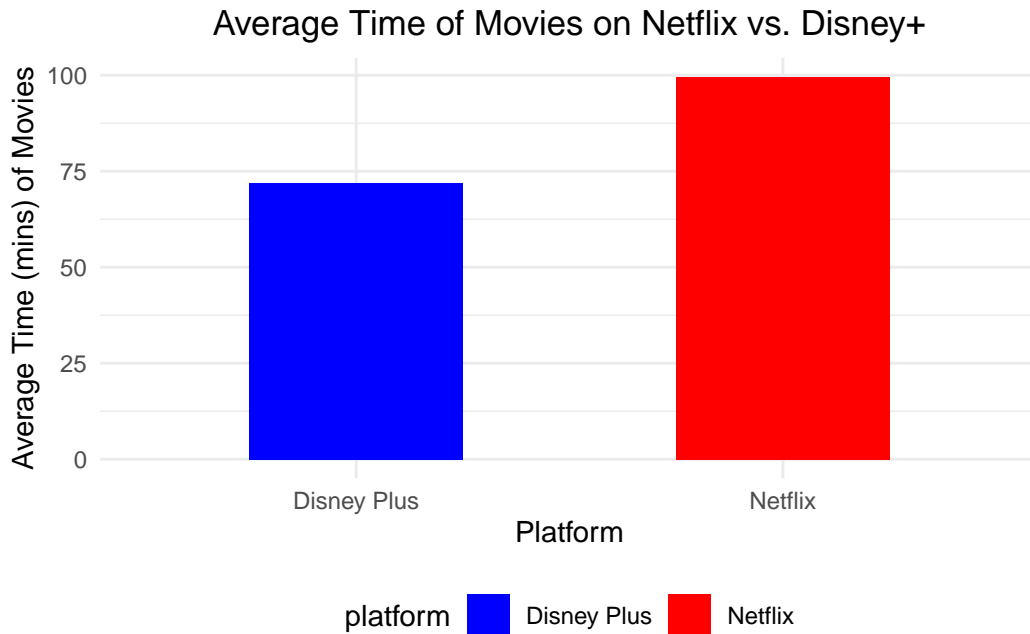
The plot compares with the average running times of Netflix and Disney+ films. It demonstrates that the average runtime of Netflix films is higher i.e. 99 minutes than that of Disney+ films i.e. 71 minutes. This variation can be explained by Netflix's emphasis on a wide variety of movies, including highly regarded, award-winning productions which typically entail lengthier running periods. Furthermore, Netflix offers a greater selection of adult-oriented live-action movies, which are usually lengthier in length. Disney+, on the other hand, concentrates more on animation and franchise films aimed at younger audiences, which are typically shorter in order to keep viewers' attention spans intact. The two platforms' different target audiences and content strategies are reflected in this trend.

Which directors are most frequently featured on Disney+?

TV Shows

Top TV Shows Directors on Disney+ and Netflix
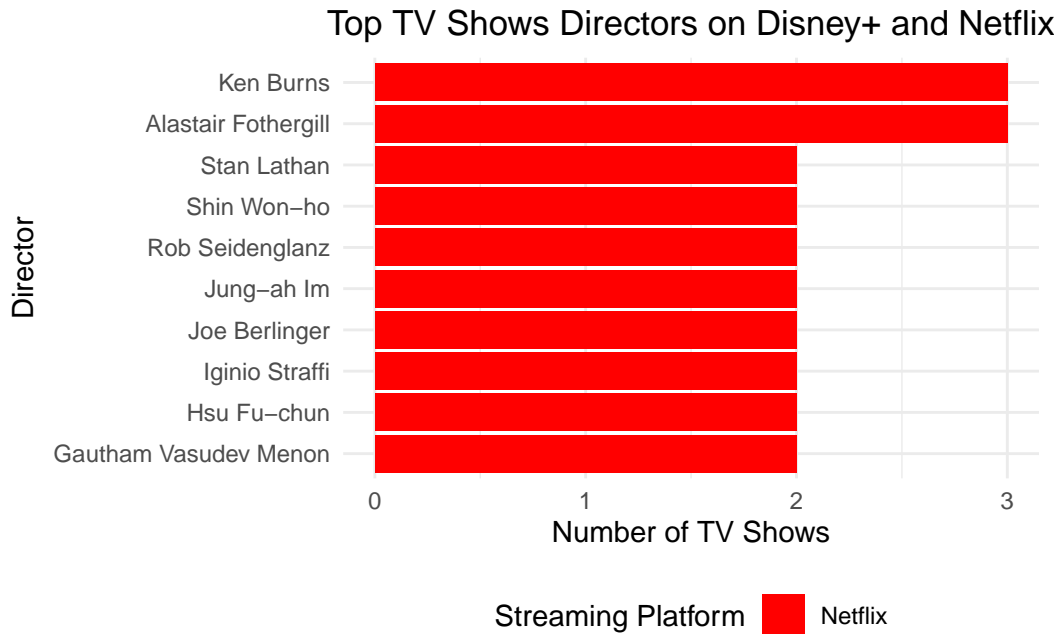
The top ten TV show directors on Netflix and Disney+ can be seen in the graph. Given that Disney+ has no directors included in the data set (only "NA" values are present), it is evident from the plot that all of the top directors in the graph are connected to Netflix. With three TV shows each available on Netflix, directors Ken Burns and Alastair Fothergill are clearly the most significant contributors. This implies that Netflix features a wider variety of content from a wider spectrum of directors. A data set restriction in the Disney+ data set is the reason for the lack of directors for Disney+.

Movies

Top Movie Directors on Disney+ vs. Netflix
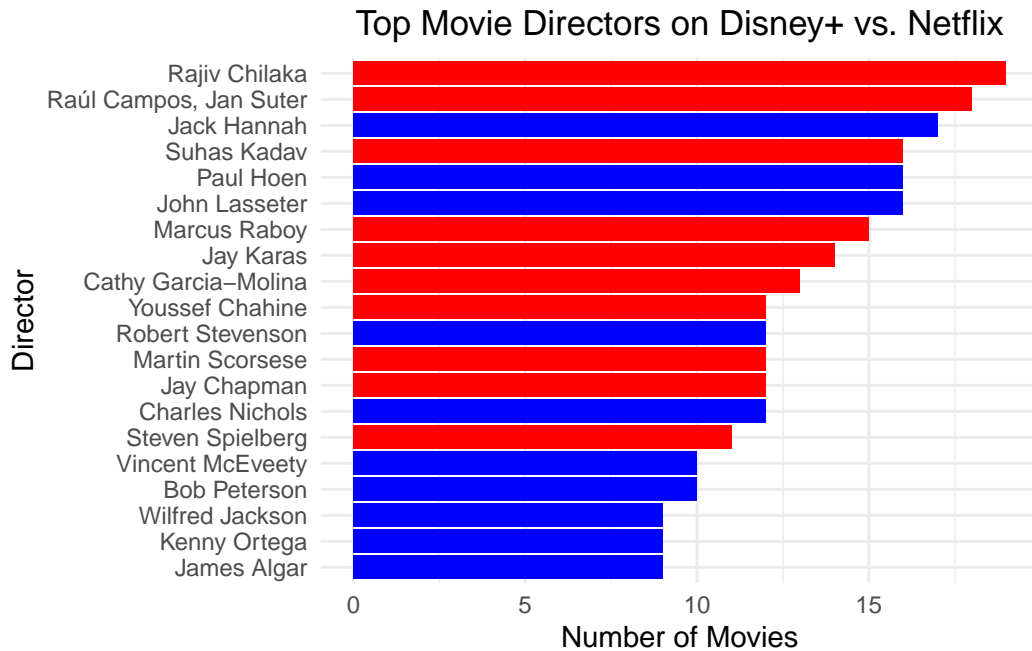
The top ten film directors on Netflix and Disney+ are shown in the graph. It shows that Rajiv Chilaka, whose 19 films are available on Netflix, is Netflix's main dominator. On the other side, Jack Hanna is in charge of Disney+, which offers 17 films that he has directed. This demonstrates a considerable difference between the movie selections on the two platforms, with Netflix displaying a greater number of Chilaka-directed productions. Hanna makes a substantial contribution to Disney+. The plot emphasizes how different the kinds of films that are available on each platform are, with Netflix providing a wider selection of filmmakers and genres.

What is the regional distribution of Disney+ content (based on country/region of production)?

Regional Distribution of Content on Disney+

Regional Distribution of Content on Netflix
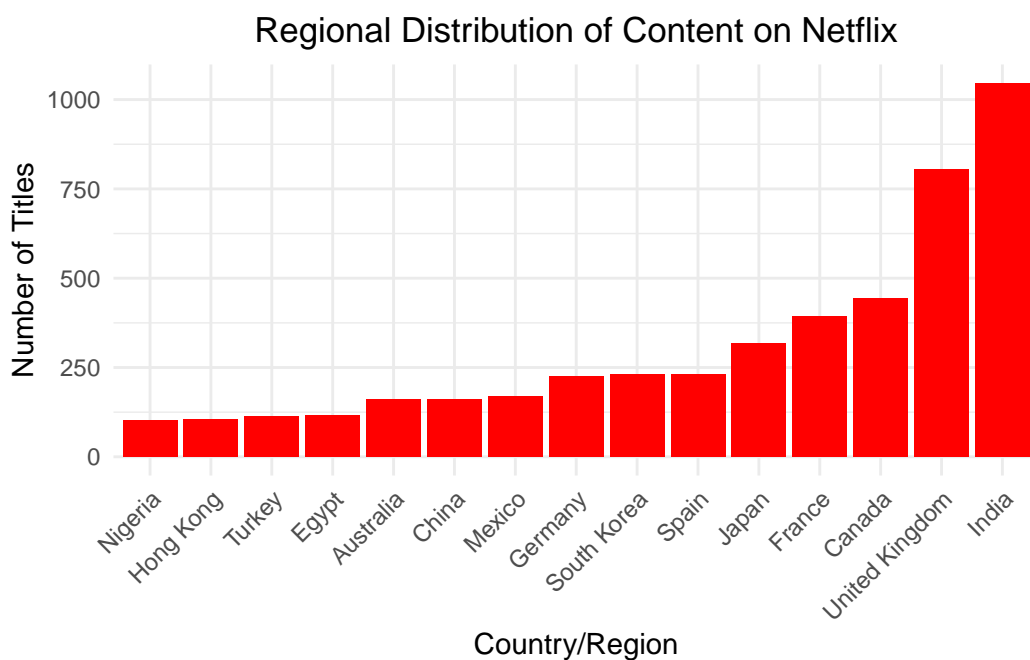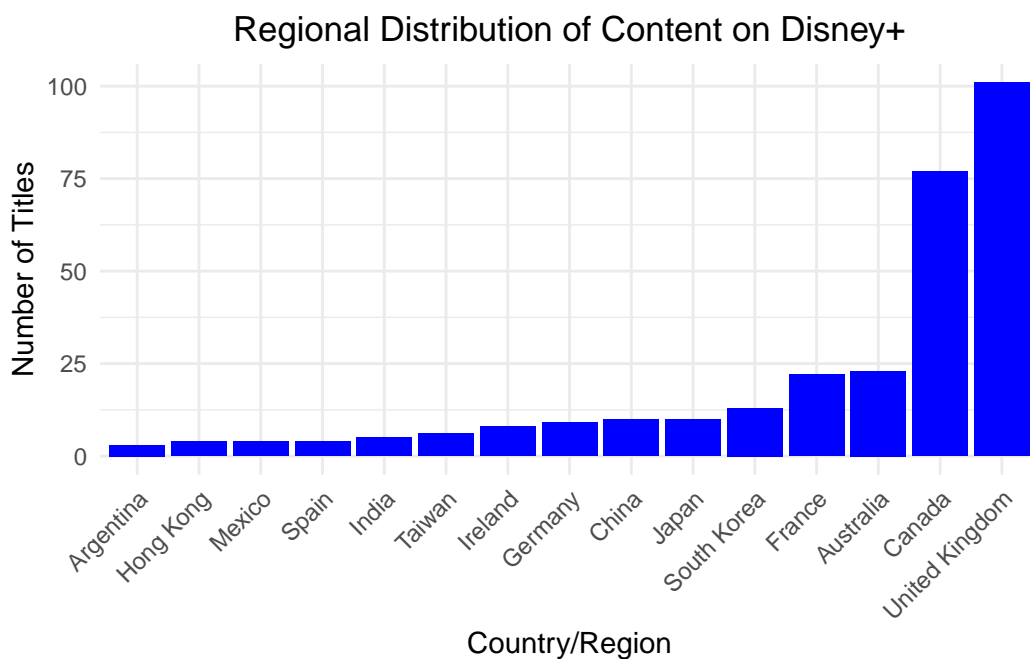
These two plots display the regional distribution for Disney+ and Netflix. We are able to determine which locations are most population for production of TV shows and movies for each streaming service. As the United States was determined to be the most popular amongst all platforms, we did not include it in our plots in order to have a better visual of the other

regions. We are able to determine that Disney+ does most of its work in the UK and Canada with Australia trailing behind by over 50 titles. Netflix primary works with India and the UK. Canada is a close third but way far behind. The largest difference to note is how much wider of a network Netflix has compared to Disney+. With being a much larger streaming platform, India has been a location for over 1000 titles whereas Disney+'s top location only is around 100. The distribution is relatively similar with similar regions being more popular for production than others.

Is there a trend in content ratings?

Content Ratings on Disney Plus and Netflix
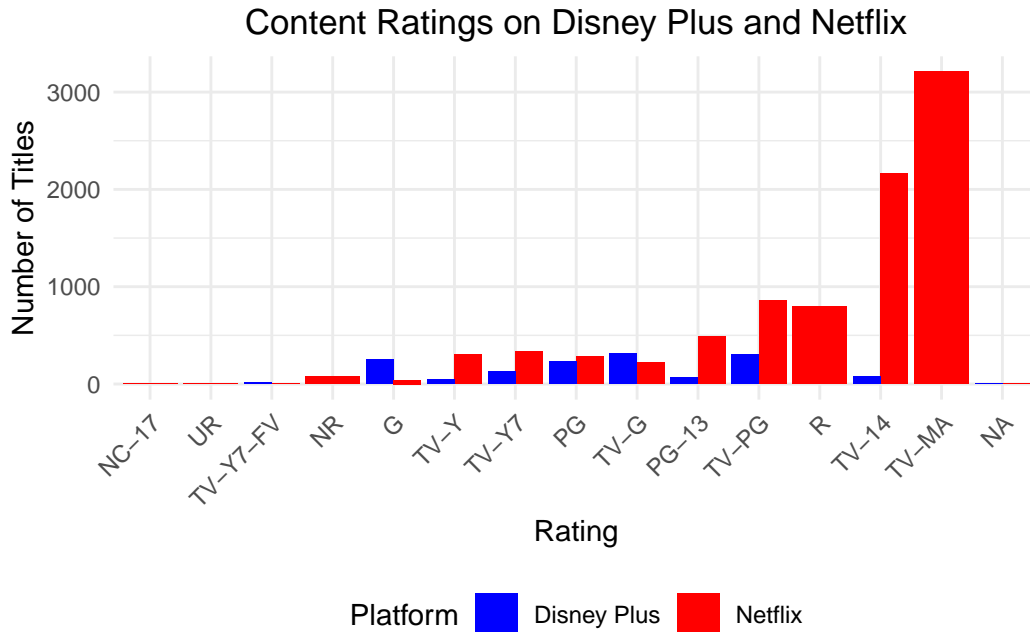
Disney Plus tailors to a younger audience whereas Netflix tailors to teens and young adults. As seen in our plot, Netflix has significantly more movies and tv shows that are rated for young adults and adults. Both TV-MA and R Netflix contains many titles while Disney+ seems to not even contain any. Looking closely at the child and family friendly ratings, Disney+ has more if its popularity within PG and TV-G. Netflix also has titles with these ratings but as it is a larger streaming service less than 500 is very minimal compared to over 3000 TV-MA titles. Disney+ is drawing a younger less mature audience so it is catering to that by providing more age appropriate shows and movies. Netflix has the young adult and adult audience so it tailors its titles more towards this older audience.

How does the time between a movie's release and its addition to Disney+ and Netflix vary?

Release Year vs. Year Added for Movies and TV Shows

Platform ● Disney Plus ● Netflix
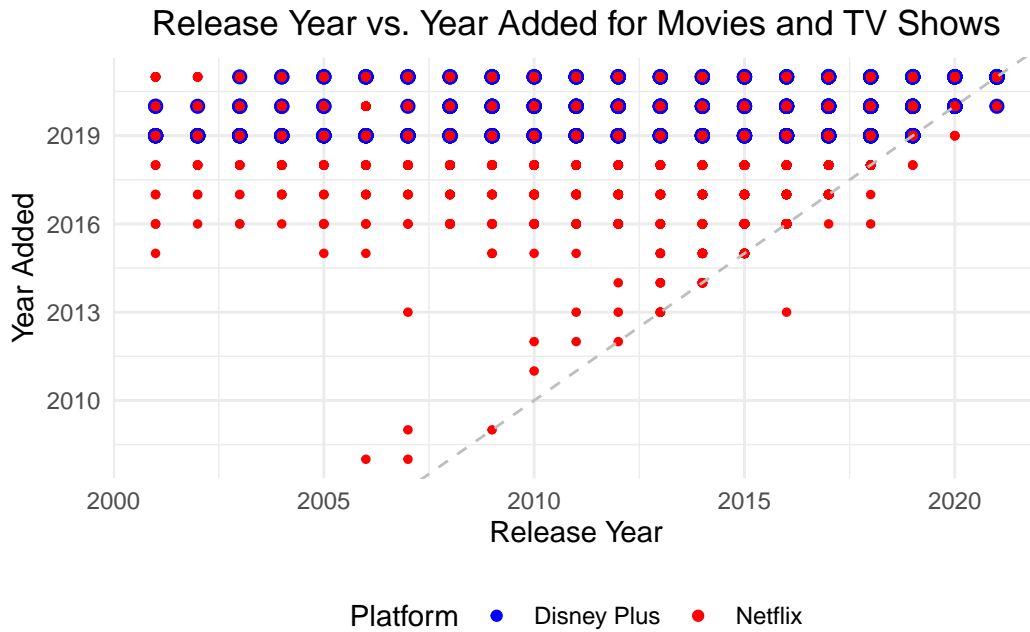
In comparing movies and tv shows release year to the year they are added to streaming platforms it seems that newer films are being added much faster to streaming platforms. The grey dashed line is a reference line showing where release year = year added. With the exception of some outliers below the line from incorrect data input we can see many films added after their year of release. The graph starts in 2000 as Netflix was not founded until 1997 and Disney+ in 2019. Both platforms have provided a variety of content from many years dating back to the 1900s as well (not provided in the graph). Also, it is interesting to see how with time more titles seem to be added much quicker. Looking at titles from the early 2000s, most were not added to Netflix until 2016 or later. Looking at more recent titles from 2015 to 2021 are being added right after production or within a year or two. Now that both these streaming platforms have been around for a little while we are able to see that there are plenty of movies and shows being added to the platforms within the same year of release as shown on the reference line.

Summary/ Conclusion

This report examines the content data sets from Netflix and Disney+ to identify key trends and differences in their TV shows and movie offerings. The data sets include seasons of TV series, average runtimes of movies, the number of TV series and films available from various years, and the contributions of the best directors on each platform. A brief overview of the platforms' growth trends and content strategy is as follows:

1. Content Availability: Netflix constantly outperforms Disney+ in terms of TV shows and movies, with the highest performance in 2020 and 2018–2019. Disney+ continues to expand steadily, reaching a peak in 2021, but its volume still lags behind that of Netflix. Netflix still dominates the market even though it saw a slight decline after reaching its peak. However, Disney+ shows a determined effort to increase the quantity of its library.

2. Content Characteristics: Disney+ emphasizes family-friendly, multi-season TV franchises and shorter animated films, while Netflix offers diverse, shorter TV series and longer, live-action films. This is an indication of the platforms' different approaches, with Netflix aiming for a wider, more diverse audience and Disney+ targeting younger audiences and families.

3. Directorial Influence: Netflix showcases a broader variety of directors across TV shows and movies, led by Rajiv Chilaka, while Disney+ highlights Jack Hanna's contributions. Netflix's diverse range of directors underscores its extensive and varied content offerings.

4. Regional Distribution: Excluding the United States, the UK and Canada are two of the top locations for both Disney+ and Netflix. The UK is Disney+'s top while India is the top location for Netflix. Generally, many of the same locations are utilized across the board for both platforms when producing content.

5. Ratings: While Netflix has R and TV-MA rated content Disney+ is not seen to contain any titles in that rating. It can be concluded that Netflix is catering to an older more mature audience. Disney+ is more focused on the younger and family friendly audience producing mostly G, PG, and TV-G content.

6. Release Year vs. Year Added: It is more common today to see movies and tv shows being added to streaming platforms within the first year or two of release. Many of the titles from the year 2000s were not added to Netflix until the later in. Netflix came out in 1977 and Disney+ in 2019. Both platforms added a variety of content dating back to years ago including in the 1900s. Additionally Netflix and Disney+ more recently started added content within the same year or within a year of release.

These observations draw attention to the unique strategies used by Netflix and Disney+ to reach their target markets and expand their libraries. As Netflix and Disney+ provide service and content for different audiences it becomes difficult to determine which platform is better. It is clear that Disney+ is designed for younger viewers and Netflix is more for older but also provides younger content. After making comparison of of the platforms and content we can see that Netflix seems to be the larger more board platform. It contains more titles, a variety of ratings, locations, and directors across all their content. Both platforms are great streaming services and more so depend on what the audience is looking for in a platform.

References

Image Reference:

"Netflix vs Disney Plus." *Gamerant.* Accessed 14 Dec. 2024. static0.gamerantimages.com/wordpress/wp-content/uploads/2022/08/Netflix-vs-Disney-Plus.jpg

Website References:

Bansal, Shivam. "Netflix Movies and TV Shows." *Kaggle*, 27 Sept. 2021, www.kaggle.com/datasets/shivamb/netf shows. Accessed 14 Dec. 2024.

Bansal, Shivam. "Netflix Movies and TV Shows." *Kaggle*, 27 Sept. 2021, www.kaggle.com/datasets/shivamb/netf shows. Accessed 14 Dec. 2024.

Code Appendix

```r
# Load the necessary packages
library(dplyr)
library(tidyr)
library(kableExtra)
library(stringr)
library(ggplot2)
library(scales)
# Load the Disney+ & Netflix data from the provided URL
disney_url <- "https://docs.google.com/spreadsheets/d/e/2PACX-1vT2CVS1o_R5Dq-ATuNlVRInKOWHG9(

disney_data <- read.csv(disney_url)

netflix_url <- "https://docs.google.com/spreadsheets/d/e/2PACX-1vRhYpV_lEupTdc9VgHeH2814Qtwa

netflix_data <- read.csv(netflix_url)
# Preview Disney+ (first 10 entries) with truncated text and adjusted font size for the table
head(disney_data, 10) %>%
  # Truncate character columns
  mutate(across(where(is.character), ~str_trunc(., width = 15, ellipsis = "..."))) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = FALSE) %>%
  column_spec(1:12, width = "1cm") %>%  #Adjust column width
  row_spec(0, bold = TRUE, font_size = 4.5)%>% # Adjust header font size
  kable_styling(font_size = 6) #Modify cell font size
# Preview Netflix the data
# Preview Disney+ (first 10 entries) with truncated text and adjusted font size for the table
head(netflix_data, 10) %>%
  # Truncate character columns
```

```r
  mutate(across(where(is.character), ~str_trunc(., width = 15, ellipsis = "...")))) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = FALSE) %>%
  column_spec(1:12, width = "1cm") %>%  #Adjust column width
  row_spec(0, bold = TRUE, font_size = 4.5)%>% # Adjust header font size
  kable_styling(font_size = 6) #Modify cell font size

# Summarize Disney+ the data
summary(disney_data) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = FALSE) %>%
  column_spec(1:12, width = "1cm") %>%
  row_spec(0, bold = TRUE, font_size = 4.5) %>%
  kable_styling(font_size = 6)%>%
  column_spec(1, extra_css = "text-align: left;")
# Summarize Netflix the data
summary(netflix_data) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = FALSE) %>%
  column_spec(1:12, width = "1cm") %>%
  row_spec(0, bold = TRUE, font_size = 4.5) %>%
  kable_styling(font_size = 5)%>%
  column_spec(1, extra_css = "text-align: left;")
# Add a platform column to the Disney+ dataset
disney_plus <- disney_data %>%
  mutate(platform = "Disney Plus")
# Add a platform column to the Netflix dataset
netflix <- netflix_data %>%
  mutate(platform = "Netflix")
# Join the two datasets
moviesandtv <- bind_rows(disney_plus, netflix)
# Add N/A values to replace missing data
moviesandtv <- moviesandtv %>%
  mutate(across(where(is.character), ~na_if(., "")))
#Separate the moviesandtv table into two distinct tables

movies <- moviesandtv %>%
  filter(type == 'Movie') %>% # Filter the type column
  mutate(duration = as.integer(str_remove(duration, " min"))) # Remove the string 'min' and

tvshows <- moviesandtv %>%
  filter(type == 'TV Show') %>% # Filter the type column
```

```r
  mutate(duration = str_replace(duration, "Seasons", "Season")) %>% # Replace the Seasons to
  mutate(duration = as.integer(str_remove(duration, " Season"))) # Remove the string 'Season
head(tvshows, 10) %>%
  # Truncate character columns
  mutate(across(where(is.character), ~str_trunc(., width = 15, ellipsis = "..."))) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = FALSE) %>%
  column_spec(1:12, width = "0.95cm") %>%  #Adjust column width
  row_spec(0, bold = TRUE, font_size = 4.5)%>% # Adjust header font size
  kable_styling(font_size = 5) #Modify cell font size
head(movies, 10) %>%
  # Truncate character columns
  mutate(across(where(is.character), ~str_trunc(., width = 15, ellipsis = "..."))) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = FALSE) %>%
  column_spec(1:12, width = "0.95cm") %>%  #Adjust column width
  row_spec(0, bold = TRUE, font_size = 4.5)%>% # Adjust header font size
  kable_styling(font_size = 5) #Modify cell font size


### How many TV shows and movies from a specific year are available on Disney+ and Netflix ?

# For TV Shows
# Step 1: Wrangle the TV shows dataset
# Filter the dataset to include only TV shows released between 1950 and 2025
tv_trend <- tvshows %>%
  filter(release_year >= 1990 & release_year <= 2025) %>%
  group_by(platform, release_year) %>% # Group by platform and release year
  summarise(count = n()) # Count the number of TV shows

# Step 2: Create the trend line plot
ggplot(tv_trend, aes(x = release_year, y = count, fill = platform)) +
  geom_bar(stat = "identity", position = "dodge") + # Create bar plot with grouped bars
  labs( title = "Number of TV Shows on Netflix vs. Disney+ From a Particular Year",
    x = "Release Year",
    y = "Number of TV Shows",
    fill = "Streaming Platforms" ) + # Add titles, labels, and key for the plot
  scale_fill_manual(values = c("Disney Plus" = "blue", "Netflix" = "red")) + # Set custom co.
  theme_minimal() + # Add a clean theme
  theme(
    plot.title = element_text(hjust = 0.5, size = 12), #center the title and change the fron
    legend.position = "bottom") # Positions the legend at the bottom of the plot
```

```r
# For Movies
# Step 1: Wrangle the Movies dataset
# Filter the dataset to include only Movies released between 1950 and 2025
movie_trend <- movies %>%
  filter(release_year >= 1990 & release_year <= 2025) %>%
  group_by(platform, release_year) %>% # Group by platform and release year
  summarise(count = n()) # Count the number of movies

# Step 2: Create the bar graph plot
ggplot(movie_trend, aes(x = release_year, y = count, fill = platform)) +
  geom_bar(stat = "identity", position = "dodge") + # Create bar plot with grouped bars
  labs( title = "Number of Movies on Netflix vs. Disney+ From a Particular Year",
    x = "Release Year",
    y = "Number of Movies",
    fill = "Streaming Platforms") + # Add titles, labels, and key for the plot
  scale_fill_manual( values = c("Disney Plus" = "blue", "Netflix" = "red")) + # Set custom co
  theme_minimal()+ # Add a clean theme
  theme(
    plot.title = element_text(hjust = 0.5, size = 12), #center the title and change the front
    legend.position = "bottom")  # Positions the legend at the bottom of the plot
### What is the average duration of movies and TV shows on Disney+ and Netflix?
# For TV Shows
# Step 1: Wrangle the TV Shows dataset
average_duration_tv <- tvshows %>%
  group_by(platform) %>% # Group by platform
  summarise(avg_seasons = mean(duration, na.rm = TRUE)) # Calculate mean number of seasons

# Step 2: Plot the bar graph plot
ggplot(average_duration_tv, aes(x = platform, y = avg_seasons, fill = platform)) +
  geom_bar(stat = "identity", width = 0.5) +  # Create a bar plot
  labs(title = "Average Seasons of TV Shows on Netflix vs. Disney+",
    x = "Platform",
    y = "Average Number of Seasons" ) + # Add titles, labels, and key for the plot
  scale_fill_manual(values = c("Disney Plus" = "blue", "Netflix" = "red")) + # Set custom col
  theme_minimal() + # Add a clean theme
  theme(
    plot.title = element_text(hjust = 0.5), #center the title
    legend.position = "bottom")  # Positions the legend at the bottom of the plot
# For Movies
# Step 1: Wrangle the Movies dataset
average_duration_movies <- movies %>%
  group_by(platform) %>% # Group by platform
```

```r
  summarise(avg_time = mean(duration, na.rm = TRUE)) #Calculate mean number of time

# Step 2: Plot the bar plot
ggplot(average_duration_movies, aes(x = platform, y = avg_time, fill = platform)) +
  geom_bar(stat = "identity", width = 0.5) +  # Create a bar plot
  labs( title = "Average Time of Movies on Netflix vs. Disney+",
    x = "Platform",
    y = "Average Time (mins) of Movies"
  ) + # Add titles, labels, and key for the plot
  scale_fill_manual(values = c("Disney Plus" = "blue", "Netflix" = "red")) + # Set custom col
  theme_minimal() + # Add a clean theme
  theme(
    plot.title = element_text(hjust = 0.5), #center the title
    legend.position = "bottom")  # Positions the legend at the bottom of the plot
### Which directors are most frequently featured on Disney+ and Netflix?
# For TV shows
# Step 1 : Wrangle the TV shows dataset
director_data_tv <- tvshows %>%
  filter(!is.na(director)) %>% # Remove rows with NA in the director column
  separate_rows(director, sep = ",") %>% # Separate rows with a separator
  group_by(platform, director) %>% # Group by platform and director
  summarise(count = n()) # Count the number of rows

# Step 2: Get the top directors for each platform
top_directors_tv <- director_data_tv %>%
  group_by(platform) %>% # Group by platform
  top_n(10, count) %>% # Select top 10 directors by count
  ungroup() # Ungroup the data

# Step 3: Plot the bar data
ggplot(top_directors_tv, aes(x = reorder(director, count), y = count, fill = platform)) +
  geom_bar(stat = "identity") + # Create a bar plot
  coord_flip() + # Flipping coordinates for better readability
  labs(title = "Top TV Shows Directors on Disney+ and Netflix",
    x = "Director",
    y = "Number of TV Shows",
    fill = "Streaming Platform" ) + # Add titles, labels, and key for the plot
  scale_fill_manual(values = c("Disney Plus" = "blue", "Netflix" = "red")) + # Set custom col
  theme_minimal()+ # Add a clean theme
  theme(
    plot.title = element_text(hjust = 0.5), #center the title
    legend.position = "bottom")  # Positions the legend at the bottom of the plot
```

```r
# For movies
# Step 1 : Wrangle the Movies dataset
director_data_movies <- movies %>%
  filter(!is.na(director)) %>% # Remove rows with NA in the director column
  group_by(platform, director) %>% # Group by platform and director
  summarise(count = n()) # Count the number of rows

# Step 2: Get the top directors for each platform
top_directors <- director_data_movies %>%
  group_by(platform) %>% # Group by platform
  top_n(10, count) %>% # Select top 10 directors by count
  ungroup() # Ungroup the data

# Step 3: Plot the data
ggplot(top_directors, aes(x = reorder(director, count), y = count, fill = platform)) +
  geom_bar(stat = "identity", show.legend = FALSE) + # Bar plot without legend
  coord_flip() + # Flip coordinates for better readability
  labs(title = "Top Movie Directors on Disney+ vs. Netflix",
    x = "Director",
    y = "Number of Movies",
    fill = " Streaming Platform" ) + # Add titles, labels, and key for the plot
  scale_fill_manual(values = c("Disney Plus" = "blue", "Netflix" = "red") ) + # Set custom c
  theme_minimal() + # Add a clean theme
  theme(
    plot.title = element_text(hjust = 0.5), #center the title
    legend.position = "bottom") # Positions the legend at the bottom of the plot

# What is the regional distribution of Disney+ content (based on country/region of production

#Step 1: Wrangle the data to select Disney+ and sort by country of production
disney_distribution <- moviesandtv %>%
  filter(platform == "Disney Plus") %>% #filter only Disney Plus entries
  separate_rows(country, sep = ", ") %>% #Separate the list of country entries by the comma a
  group_by(country) %>%
  summarise(count = n(), .groups = "drop") %>%
  arrange(desc(count)) %>%
  filter(!is.na(country) & country != "United States") %>% #remove NA values and not include
  slice_head(n = 15) #Select only the top 15 production countries

#Step 2: Wrangle the data to select Netflix and sort by country of production
netflix_distribution <- moviesandtv %>%
  filter(platform == "Netflix") %>% #filter only Netflix entries
```

```r
  separate_rows(country, sep = ", ") %>% #Separate the list of country entries by the comma a
  group_by(country) %>%
  summarise(count = n(), .groups = "drop") %>%
  arrange(desc(count)) %>%
  filter(!is.na(country) & country != "United States") %>% #remove NA values and not include
  slice_head(n = 15) #Select only the top 15 production countries

#Step 3: Create a bar graph displaying the regional distribution for Disney+
ggplot(disney_distribution, aes(x = reorder(country, count), y = count, fill = "Disney Plus")
  geom_bar(stat = "identity", position = "dodge", show.legend = FALSE) + #creates a bar grap
  labs( #provides title and axis labels
    title = "Regional Distribution of Content on Disney+",
    x = "Country/Region",
    y = "Number of Titles",
    fill = "Platform"
  ) +
  scale_fill_manual(values = c("blue")) + #display the data in blue for Disney+
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5), #center the title
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1) #adjust the y axis content
  )

#Step 3: Create a bar graph displaying the regional distribution for Netflix
ggplot(netflix_distribution, aes(x = reorder(country, count), y = count, fill = "Netflix")) -
  geom_bar(stat = "identity", position = "dodge", show.legend = FALSE) + #creates a bar grap
  labs( #provides title and axis labels
    title = "Regional Distribution of Content on Netflix",
    x = "Country/Region",
    y = "Number of Titles",
    fill = "Platform"
  ) +
  scale_fill_manual(values = c("red")) + #display the data in red for Netflix
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5), #center the title
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1) #adjust the y axis content
  )

# Is there a trend in content ratings? For instance, does Disney Plus predominantly feature

#Step 1: Wrangle the moviesandtv dataset to analyze the ratings with their count
```

```r
ratings <- moviesandtv %>%
  filter(!grepl("min", rating))%>% #removed random values in ratings containing "min" entries
  group_by(platform, rating) %>% #group the data to display both Netflix and Disney+ and the
  summarise(count = n(), .groups = "drop") %>%
  arrange(desc(count))

#Step 2: Create a bar plot displaying ratings and their count for both Netflix and Disney+
ggplot(ratings, aes(x = reorder(rating, count), y = count, fill = platform)) +
  geom_bar(stat = "identity", position = "dodge") + #creates a bar graph
  scale_fill_manual(values = c("Disney Plus" = "blue", "Netflix" = "red")) + #fills the bars
  labs( #provides title, axis labels, and legend
    title = "Content Ratings on Disney Plus and Netflix",
    x = "Rating",
    y = "Number of Titles",
    fill = "Platform"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5), #center the title
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1), #adjust the y axis content
    legend.position = "bottom" #position the legend on the bottom of the plot
  )
# How does the time between a movie's release and its addition to Disney+ and Netflix vary?

#Step 1: Wrangle the dataset so that the date_added is a integer of the year and not a string
moviesandtv_datechange <- moviesandtv %>%
  filter(!is.na(date_added), !is.na(release_year), release_year > 2000) %>% #filters out any
  mutate(added_year = as.integer(str_trim(str_sub(date_added, -4)))) #mutates the table to di

#Step 2: Create a plot displaying the correlation between the year releasted and year added t
ggplot(moviesandtv_datechange, aes(x = release_year, y = added_year, color = platform, size =
  geom_point() + #Creates a scatter plot with both Disney+ and Netflix data
  geom_abline(slope = 1, intercept = 0, linetype = "dashed", color = "grey") + #This is refer
  scale_color_manual(values = c("Netflix" = "red", "Disney Plus" = "blue")) + #setting Netfli
  scale_size_manual(values = c("Netflix" = 1, "Disney Plus" = 2)) +  # Custom size for each p
  labs( #provides title, axis labels, and legend
    title = "Release Year vs. Year Added for Movies and TV Shows",
    x = "Release Year",
    y = "Year Added",
    color = "Platform"
  ) +
  theme_minimal() +
```

```r
theme(
  plot.title = element_text(hjust = 0.5),  # Center the title
  legend.position = "bottom" #position the legend on the bottom of the plot
)+
guides(size = "none") #removes the second legend displaying size as a key
```