Streaming Service Data Exploration

Dhruv Nasit Urvi Acharya Ayushman Choudhury

2025-05-05

Disney+ Dataset

The Disney+ dataset, sourced from Kaggle, contains metadata about movies and TV shows available on the Disney+ streaming platform. It includes information such as title, release year, genre, content rating, runtime, and a brief description. This dataset is useful for analyzing trends in Disney+ content, such as the distribution of genres, the age of titles, and duration patterns. Link

Netflix Dataset

The Netflix dataset, also sourced from Kaggle, includes detailed metadata for movies and TV shows available on Netflix. It provides fields such as title, release year, genre, type (movie or TV show), content rating, and runtime. This dataset enables exploration of content trends on Netflix, such as genre dominance, recent content releases, and runtime comparisons. Link

Data Provenance

The data used in this project was collected and published by Shivam Bansal on Kaggle. It consists of two separate datasets: one for Disney+ and one for Netflix. These datasets were compiled to provide an overview of the content libraries on the respective streaming platforms. Each row in the datasets represents a unique title—either a movie or a TV show—and includes information such as release year, genre, content rating, and runtime. The purpose of the data is to enable analysis of content trends across platforms, supporting comparisons related to genre distribution, content length, and date of releases.

FAIR and CARE Principles

The datasets used in this project align well with the **FAIR** (Findable, Accessible, Interoperable, and Reusable) principles:

- **Findable:** The datasets are publicly available on Kaggle, a well-known data repository, with unique and persistent URLs.
- **Accessible:** The data can be downloaded freely by anyone with a Kaggle account, without pay-walls or restrictions.
- **Interoperable:** The data is provided in standard CSV format, making it easy to use across various programming languages and tools such as R and Python.
- **Reusable:** The datasets include clear documentation and metadata, allowing users to understand and re-purpose the data for similar exploratory or comparative analyses.

The data also aligns with the spirit of the **CARE** principles (Collective Benefit, Authority to Control, Responsibility, and Ethics), though not directly applicable in full since this data does not pertain to Indigenous communities. However:

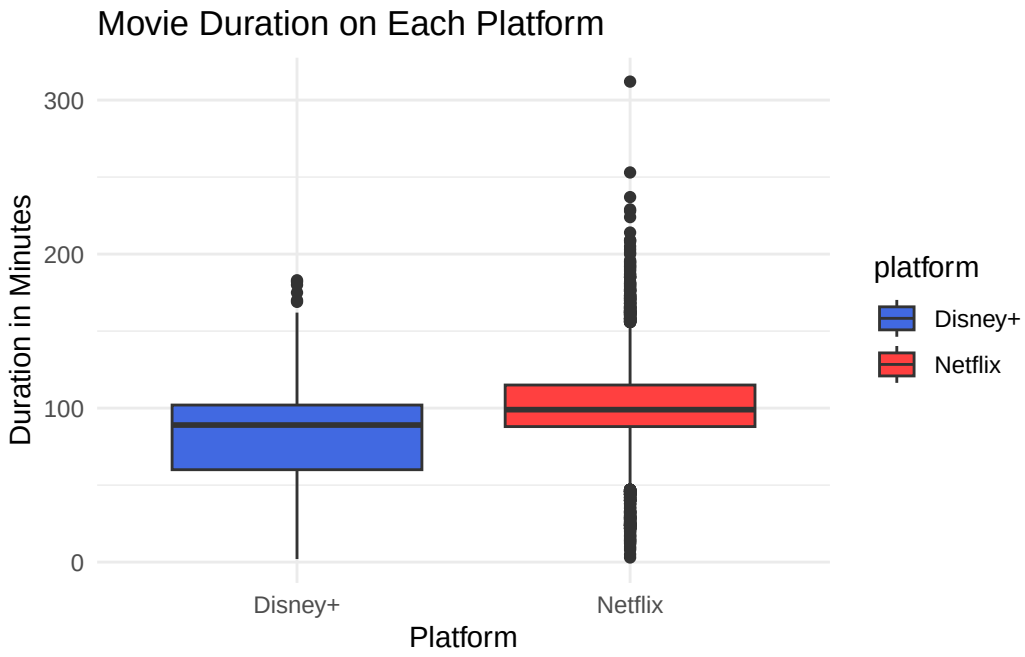
- **Collective Benefit:** The datasets contribute to public knowledge about streaming trends and allow educational use.
- **Responsibility and Ethics:** The use of the data is ethical, as it is anonymized and focuses on publicly available media titles rather than individuals.

Overall, the data supports transparent, responsible, and collaborative academic research.

Research Questions

1. Are Disney+ movies shorter or longer compared to Netflix movies?
2. Which genres dominate each platform? (E.g., Disney+ = Animation/Kids, Netflix = Drama/Thriller?)
3. Which same Movies and TV shows are available on both the platforms?

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

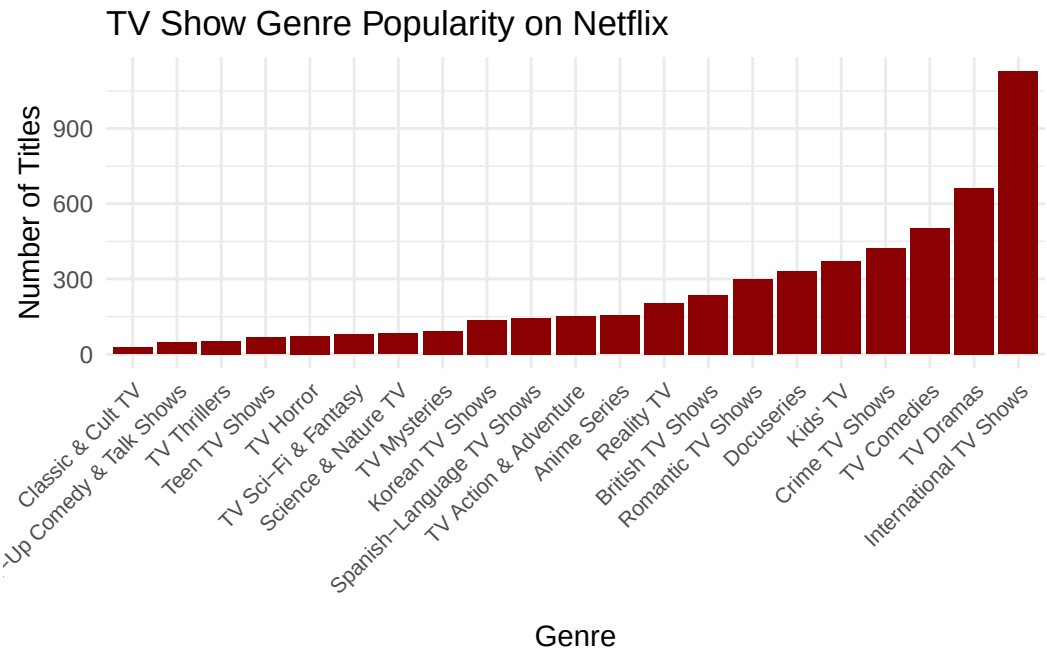


Insights

The box-plot shows that Netflix movies have a higher median duration compared to Disney+ movies. The IQR(inter quartile range) for Disney+ is slightly wider, showcasing that there is more variability in movie lengths. Additionally, Netflix exhibits more outliers with some movies going beyond 300 minutes. Disney+ movies are shorter and more consistent, with fewer outliers and a lower spread. This aligns with Disney's focus on family content, which favors more concise time. Overall, the box-plot suggests that Netflix offers a broader variety of movie lengths including longer and shorter experiences, but Disney+ maintains a uniform duration.

Popularity Plots

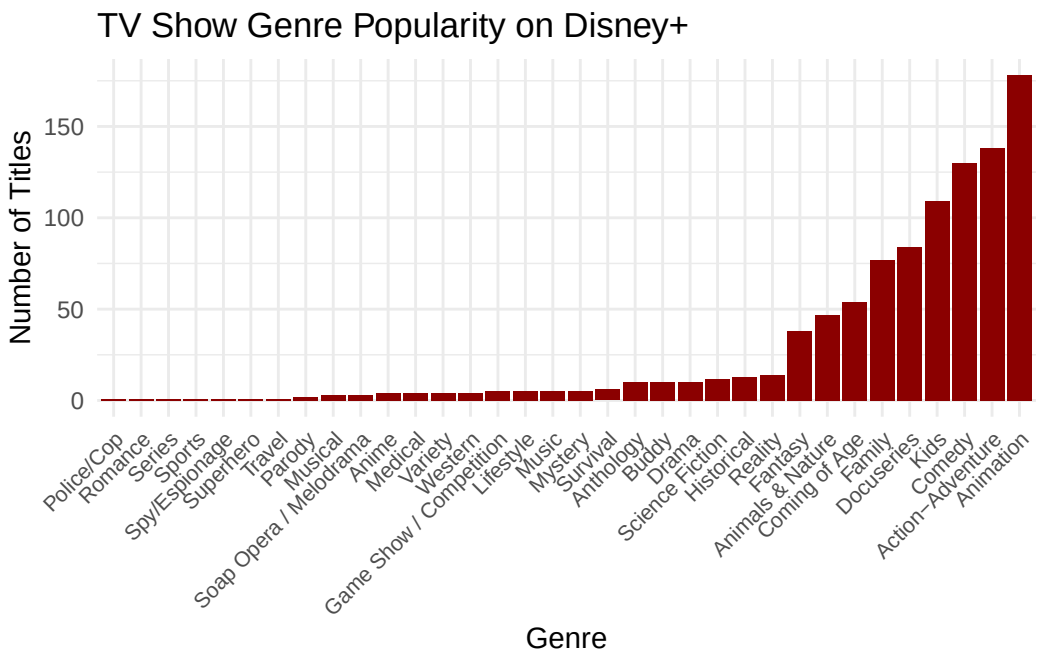
TV Show Genre Popularity on Netflix



Insights

The bar chart "TV Show Genre Popularity on Netflix" provides an overview as to what TV genres are available on the platform as well as counting the number of titles associated with each genre. International TV shows dominate the catalog, with over 1000 titles. This shows Netflix's global availability (Several countries have content made locally as well as internationally on the platform). Additionally, TV Dramas, TV comedies and Crime TV shows also have a lot of entries that are popular for the service. This shows that these cater to Netflix's strategy of providing access to diverse content. Other genres such as Kid's TV, Docuseries and Romantic TV have a significant presence, displaying Netflix's appeal to family friendly content. Genres like Stand UP Comedy and TV thrillers have fewer titles, suggesting these are for niche audiences, and don't take up much of Netflix's user base. This chart reveals Netflix's prioritization of appealing to global audiences while still trying to maintain a variety of other genres.

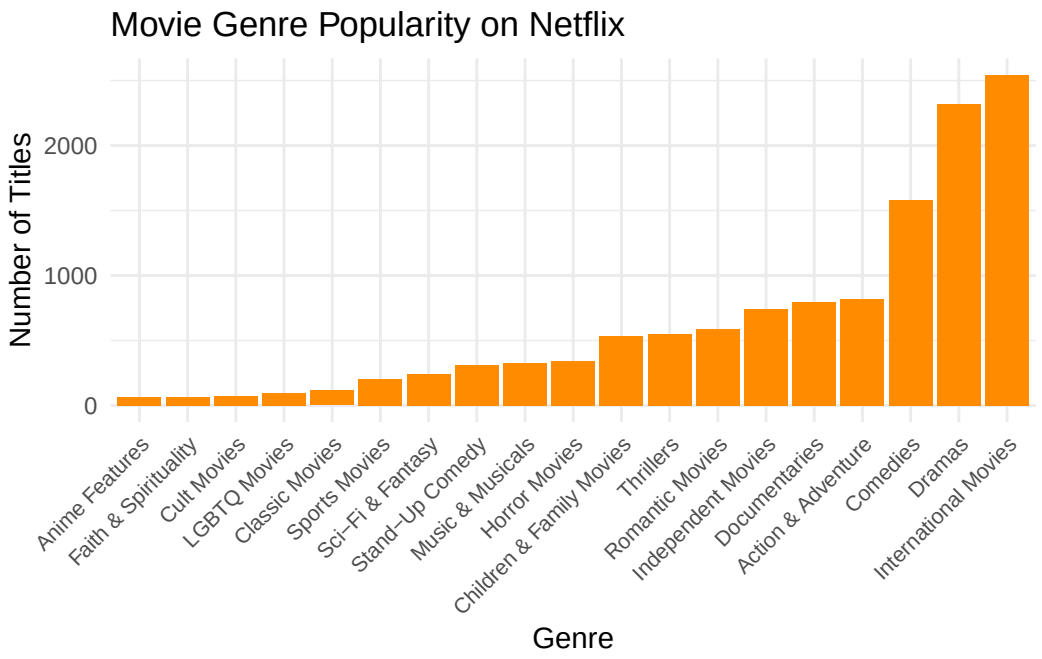
TV Show Genre Popularity on Disney+



Insights

This chart reveals that Animation is the most dominant genre in Disney+, followed closely by Action-Adventure, Comedy and Kids content. This aligns with Disney's brand which centers around family friendly content. Some other well represented genres include Documentaries, Family, and Fantasy which suggests that Disney also focuses on educational and invigorating content for a larger age demographic. Genres like Reality, Science Fiction and Drama have moderate representation, indicating slight to middling emphasis on mature storytelling. Genres such as Romance, Sports, Satire, and Podcast, are barely represented showing that Disney doesn't really focus on these categories. Overall, this chart shows Disney+'s core strategy of family friendly content with animation and adventure, while adding variation with other genres like Documentaries, Family, and Fantasy.

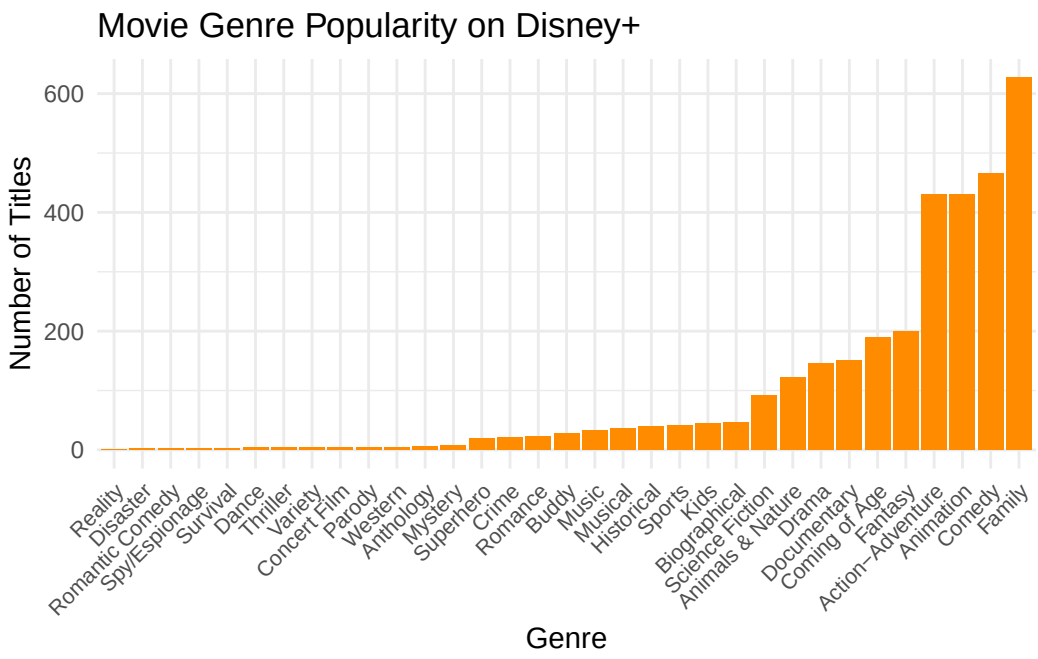
Movie Genre Popularity on Netflix



Insights

The chart shows that International movies, dramas and comedies dominate the Netflix movie catalog with International Movies and Dramas containing over 2000 titles. This once again demonstrates Netflix's strong emphasis on global availability. Dramas and Comedies are next most popular genres on the platform. Action & Adventure, Documentaries, Independent movies, Romantic Movies, Thrillers, and Family Movies, also have a lot of representation, presenting that Netflix puts in effort to provide a variety of unique content. Niche categories like Anime Features, Cult movies and Faith & Spirituality, have less titles. Pointing to less demand for these niches or not having the licensing for certain titles. This distribution reflects Netflix's strategy of prioritizing International Movies, Dramas, and Comedies while still supporting secondary and niche genres.

Movie Genre Popularity on Disney+



Insights

This bar chart reveals that family, Comedy, Animation, and Action-Adventure are by far the most prominent movie genres on Disney+. Each of them has around 400-600 titles. This aligns with Disney's family friendly brand. Fantasy and Coming of Age are also quite popular, suggesting that these are still important for a broader range of audiences. Documentaries and Dramas are middle tier showing middling focus for those genres. Reality, Disaster, and Romantic comedy are the least represented showing less demand for the genre. This showcases Disney+'s commitment to wholesome, family friendly content as well as adding diversity through niche genres.

Repeated Titles across streaming platforms (Movies and TV shows)

Table 1: Titles Found Both on Netflix and Disney Plus (Top 20)

title	count	title_type	countries
Mulan II	15	Movie	United States, South Korea, Singapore, Russia, Malaysia, Kazakhstan, Taiwan, Hong Kong, Japan, China, India, Syria, Iran, Egypt, Pakistan
Bonkers	11	TV Show	United States, Hong Kong, South Korea, France, Canada, China, United Kingdom, Australia, Japan, Taiwan, Philippines
Thumbelina	8	Movie	Ireland, United States, Canada, United Kingdom, Denmark, Spain, Poland, Hungary
Chip 'n Dale's Rescue Rangers	6	TV Show	United States, Taiwan, South Korea, China, Japan, United Kingdom
Strange Magic	6	Movie	United States, Canada, United Kingdom, Singapore, Australia, Thailand
Disneynature Oceans	5	Movie	France, Switzerland, Spain, United States, United Arab Emirates
Henry Hugglemonster	5	TV Show	Ireland, United Kingdom, United States, South Korea, Canada
Into the Okavango	5	Movie	United States, Angola, Botswana, Namibia, South Africa
Iron Man Armored Adventures	5	TV Show	Canada, United States, United Kingdom, France, Luxembourg
The Chronicles of Narnia: Prince Caspian	5	Movie	United States, Poland, Slovenia, Czech Republic, United Kingdom

Table 1: Titles Found Both on Netflix and Disney Plus (Top 20) *(continued)*

title	count	title_type	countries
Tomorrowland	5	Movie	United States, Spain, France, Canada, United Kingdom
A Goofy Movie	4	Movie	United States, Australia, France, Canada
Around the World in 80 Days	4	Movie	United States, Germany, Ireland, United Kingdom
Dumbo	4	Movie	United States, United Kingdom, Australia, Canada
Gnomeo & Juliet	4	Movie	United Kingdom, United States, Canada, Ireland
MUPPET BABIES	4	TV Show	South Korea, United States, China, Japan
Miraculous: Tales Of Ladybug & Cat Noir	4	TV Show	France, South Korea, Japan, United States
Miss Peregrine's Home for Peculiar Children	4	Movie	United States, United Kingdom, Belgium, Canada
Phineas and Ferb	4	TV Show	United States, South Korea, China, Taiwan
Saving Mr. Banks	4	Movie	United States, United Kingdom, Australia, United States, United Kingdom, Australia

Insights

This table highlights the most frequently overlapping content across Netflix and Disney+, with "Mulan II" topping the list, being available in 15 different country-platform combinations. The majority of the shared titles are movies, although there are a few notable TV shows such as Bonkers, Chip 'n Dale's Rescue Rangers, and Iron Man Armored Adventures. The content spans a wide geographic range, with the United States appearing consistently across nearly all entries, followed by other countries such as the United Kingdom, Canada, South Korea, and France. This reflects both platforms' shared strategy to cater to international audiences with popular family-oriented and animated content. Notably, many of these titles are classic Disney properties, reinforcing Disney's strong global brand recognition and the potential for licensing

overlap before exclusive rights are enforced. The table also shows that while some titles have extensive international distribution, others appear in only a few select countries, indicating variability in regional licensing agreements. Overall, this tables provides insight into content commonality and geographical reach between the two major streaming services.

Project Repo (Github)

project link

Citations

Bansal, S. (2021). *Netflix Movies and TV Shows* [Data set]. Kaggle.

https://www.kaggle.com/datasets/shivamb/netflix-shows

Bansal, S. (2021). *Disney+ Movies and TV Shows* [Data set]. Kaggle.

https://www.kaggle.com/datasets/shivamb/disney-movies-and-tv-shows

Code Appendix

```
# Load necessary packages
library(tidyverse)
library(dplyr)
library(knitr)
library(kableExtra)
library(ggplot2)

# import data
netflixRaw <-read.csv(file="/Users/dhruvnasit/Downloads/netflix_dataset.csv")
disneyRaw <-read.csv(file="/Users/dhruvnasit/Downloads/disney_plus_dataset.csv")

## Tidying data
# remove and rename columns as needed
netflixTidy <- netflixRaw %>%
  rename(
    genre = listed_in
  ) %>%
  select(
    -show_id, -director,
    -cast, -description
```

```
  )

# Tidying Disney+
# remove null values
netflixTidy <- netflixTidy %>%
  filter(country != "")

disneyTidy <- disneyRaw %>%
  rename(
    genre = `listed_in`,
    type = `type`,
    country = `country`
  ) %>%
  select(-show_id, -director, -cast, -description) %>%
  filter(!is.na(country), country != "") %>%
  separate_rows(country, sep = ", ")

# filter the combined data set to only include the rows that are movies
moviesOnly <- read.csv(
  file="/Users/dhruvnasit/Downloads/combined_streaming_data.csv"
  ) %>%
  filter(type == "Movie")

# movie_duration_mins is a new column that is created by getting the duration,
# from the duration columns
# "min" is re moved from the text and the duration is turned into a numeric

moviesOnly <- moviesOnly %>%
  mutate(movie_duration_mins = as.numeric(str_replace(duration, " min", "")))

moviesOnlyPlot <- ggplot(
  moviesOnly,
  aes(
    x = platform,
    y = movie_duration_mins,
    fill = platform
  )
) +
  geom_boxplot() +
  scale_fill_manual(values = c("Disney+" = "royalblue", "Netflix" = "brown1")) +
  labs(
    title = "Movie Duration on Each Platform",
```

```
    x = "Platform",
    y = "Duration in Minutes"
  ) +
  theme_minimal()

print(moviesOnlyPlot)
# function counts the occurrences of each genre
genreCount <- function(genre_occurence, type) {
  genre_occurence %>%
    separate_rows(genre, sep = ", ") %>%
    filter(genre != "TV Shows" & genre != "Movies") %>%
    count(genre, name = "count") %>%
    arrange(desc(count)) %>% # arranged in descending order
    mutate(type = type)
}

# create subset dataset of tv shows for Netflix
netflixTvTable <- netflixTidy %>%
  filter(type == "TV Show")


# create subset dataset of movies
netflixMovieTable <- netflixTidy %>%
  filter(type == "Movie")
# count tv show genres Netflix
tvGenre <- genreCount(netflixTvTable, "TV Show")

# creating plot for tv genre
tvGenrePlot <- ggplot(
  tvGenre,
  aes(
    x = reorder(genre, count),
    y = count
  )
) +
  geom_bar(stat = "identity", fill = "darkred") +
  # darkred makes all the bars the same color
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
    legend.position = "none"
    # legend removed since all the bars are the same color
```

```
  ) +
  labs(
    title = "TV Show Genre Popularity on Netflix",
    x = "Genre",
    y = "Number of Titles"
  )
print(tvGenrePlot)
# Plot for Disney+
genreCount <- function(df, type_label) {
  df %>%
    separate_rows(genre, sep = ", ") %>%
    filter(genre != "TV Shows" & genre != "Movies") %>%
    count(genre, name = "count") %>%
    arrange(desc(count)) %>%
    mutate(type = type_label)
}
disneyTV <- disneyTidy %>% filter(type == "TV Show")
disneyMovies <- disneyTidy %>% filter(type == "Movie")
# Apply genre count
tvGenres <- genreCount(disneyTV, "TV Show")
movieGenres <- genreCount(disneyMovies, "Movie")
tvPlot <- ggplot(tvGenres, aes(x = reorder(genre, count), y = count)) +
  # darkred makes all the bars the same color
  geom_bar(stat = "identity", fill = "darkred") +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
    legend.position = "none"
  ) +
  labs(
    title = "TV Show Genre Popularity on Disney+",
    x = "Genre",
    y = "Number of Titles"
  )
print(tvPlot)
# count movie genres
movieGenre <- genreCount(netflixMovieTable, "Movie")

# creating plot for movies
movieGenrePlot <- ggplot(
  movieGenre,
  aes(
```

```
    x = reorder(genre, count),
    y = count
  )
) +
  # darkorange makes all the bars the same color
  geom_bar(stat = "identity", fill = "darkorange") +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
    # legend removed since all the bars are the same color
    legend.position = "none"
  ) +
  labs(
    title = "Movie Genre Popularity on Netflix",
    x = "Genre",
    y = "Number of Titles"
  )

print(movieGenrePlot)
# plot for disney+
moviePlot <- ggplot(movieGenres, aes(x = reorder(genre, count), y = count)) +
  # darkorange makes all the bars the same color
  geom_bar(stat = "identity", fill = "darkorange") +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
    legend.position = "none"
  ) +
  labs(
    title = "Movie Genre Popularity on Disney+",
    x = "Genre",
    y = "Number of Titles"
  )
print(moviePlot)
# counting instances of repeated titles
repeatedTitles <- read_csv(
  file="/Users/dhruvnasit/Downloads/combined_streaming_data.csv"
)%>%
  group_by(title) %>%  # grouping by movie and tv show title
  summarise(
    count = n(), # count the occurrences of each title in disneyNetflixdata
    #title type and title turned into one string
```

```
    title_type = paste(unique(type), collapse = ", "),
    # country and title turned into one string
    countries = paste(unique(country), collapse = ", ")
  ) %>%
  arrange(desc(count)) # arranged in descending order

# visualizing in table form
# visualizing in table form
repeatedTitles %>%
  head(20) %>%
  knitr::kable(
    caption = "Titles Found Both on Netflix and Disney Plus (Top 20)",
    format = "latex",
    longtable = TRUE,
    booktabs = TRUE
  ) %>%
  kable_styling(
    latex_options = c("repeat_header"),  # no scale_down!
    full_width = FALSE,
    position = "center"
  ) %>%
  column_spec(4, width = "12em",
              extra_css = "word-wrap: break-word; white-space: normal;")
```