

# Comparing Stream Services

2021

Thriller

TV-G

Can data provide the Team Cobra with the much needed insight before time runs out?

Directors:

Brennan, Ed  
Leopold, Madison

Cast:

Denq, Chris  
Nguyen, Ben  
Ye, Emily

▶ PLAY



# Menu Overview



# Goals

## Project Track: EDA in Digital Entertainment

Project Hypothesis:  
Are there unique differences between each service's brand, as told by their offerings?



# Process

## Data Sourcing:

- Kaggle

## Data Adjacency:

- Dataset “type”
- Same authors

kaggle



data.world

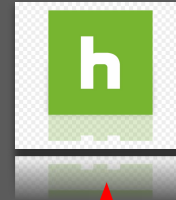
# Process

## Which Streaming Stream?:

- Top 3 focused services

## Team Distribution:

- One service / one member
  - cleaning
  - analysis
  - visualization



# Technical

## Cleaning:

`.merge`  
`.loc`, indexing  
`.dtype` / `.astype`  
`.dropna` / `.fillna` / `.isna()`  
`.split` / `.replace` / `.rename`

for loops

if-elif statements

user-defined functions

(9 databases: 30k -> ~5k)

```
#extract function to pull out categorical data
def extract(df, input_column):
    return df.loc[:, input_column].dropna(axis = 0).str.split(',', expand = False).tolist()
```

```
def nan_remover(input_list):
    ...
    loop through a provided list to see if the list contains nans
    if it does, nan_remover will remove the nans and return a new list
    ...
    # check for nulls or NaNs
    nan_check = [x for x in input_list if np.isnan(x) == True]
    nan_removed = []

    if len(nan_check) == 0:
        return "No nulls found"
    if len(nan_check) > 0:
        for value in input_list:
            if np.isnan(value) == False:
                nan_removed.append(value)
    return nan_removed
```



# Technical

## Visualization:

```
plt.plot  
plt.subplot  
bar, pie, line, hist, box
```

## Analysis:

```
.mean  
.median  
.mode
```

```
#determine how many bins  
print(earliest_year, latest_year, (latest_year - earliest_year)//5 + 1)  
cleaned_movie_netflix_df.hist('Year', bins = (latest_year - earliest_year)//5 + 1, grid = False)  
plt.title('Counts of Release Year 5')
```

```
# mean, median, mode for imdb and rt scores  
imdb_mean_tv = mean(imdbtv_nn)  
imdb_median_tv = median(imdbtv_nn)  
imdb_mode_tv = mode(imdbtv_nn)  
  
rt_mean_tv = np.mean(rttv_nn)  
rt_median_tv = median(rttv_nn)  
rt_mode_tv = mode(rttv_nn)
```

# Analysis by Platform

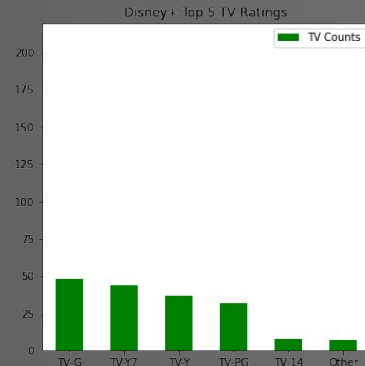
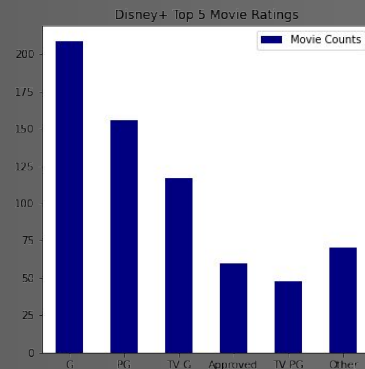
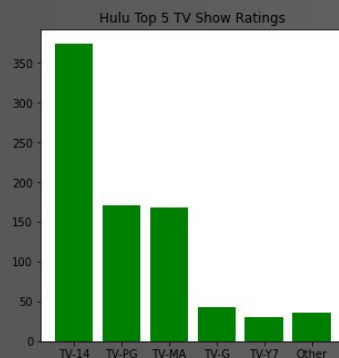
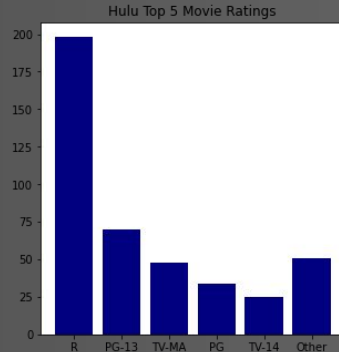
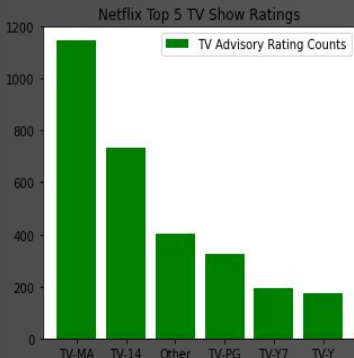
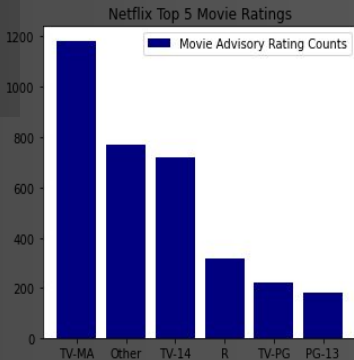
## Top 5 Advisory Ratings

### Movie:

- Netflix/Hulu - mature audiences
- Disney+ - children

### TV:

- Netflix - mature audiences
- Hulu - adolescent
- Disney+ - children





# Analysis by Platform

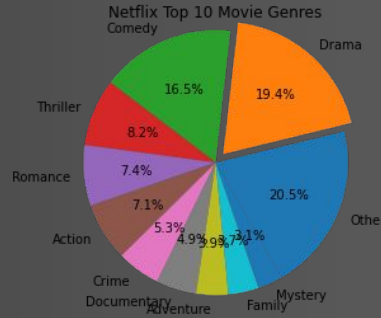
## Top 10 Genres

### Movie:

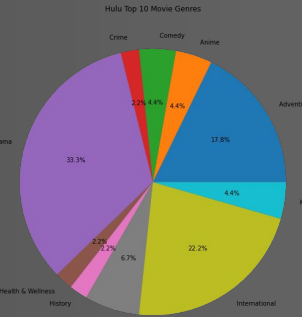
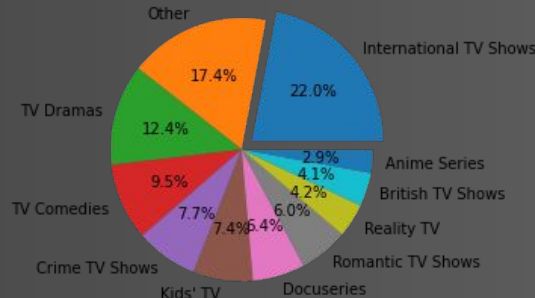
- Netflix/Hulu - Drama
- Disney+ - Family

### TV Shows:

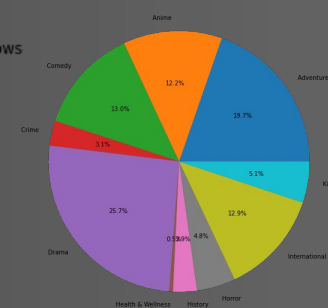
- Netflix - International
  - More accessibility
- Hulu - Drama
- Disney+ - Family



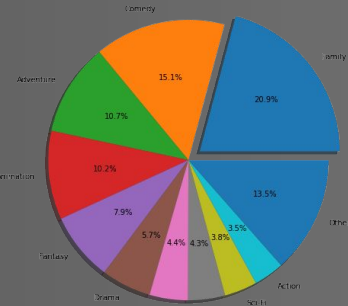
Netflix TV Top 10 Genres



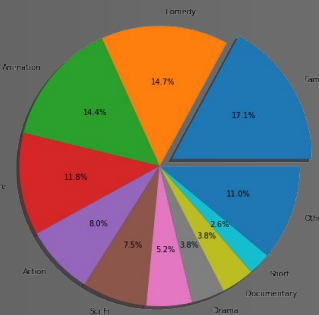
Hulu Top 10 TV Genres



Disney+ Top 10 Movie Genres



Disney+ Top 10 TV Genres

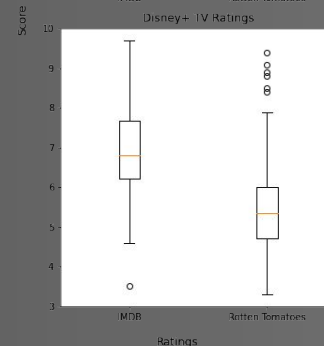
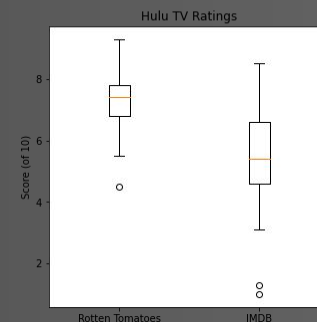
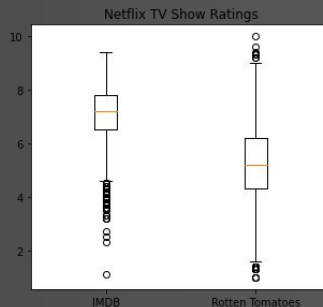
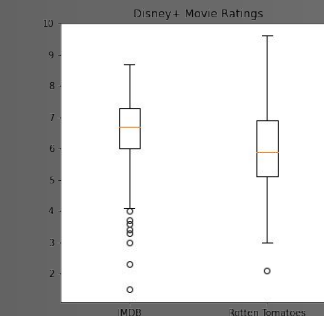
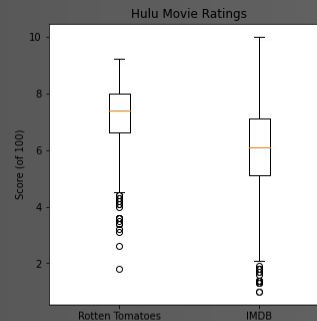
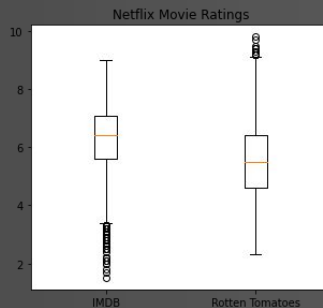


# Analysis by Platform

## Critical Reception

IMDB - general public reception  
Rotten Tomatoes - Critical reception

- For the most part, Rotten Tomatoes has a lower rating average
- Hulu Rotten Tomatoes Rating the least spread

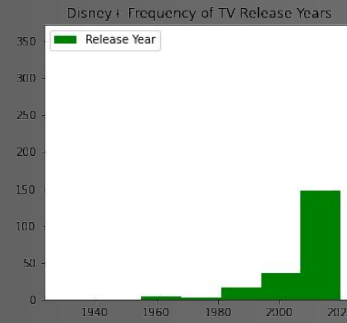
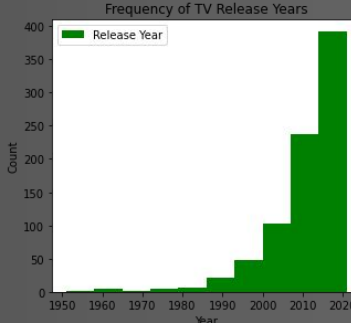
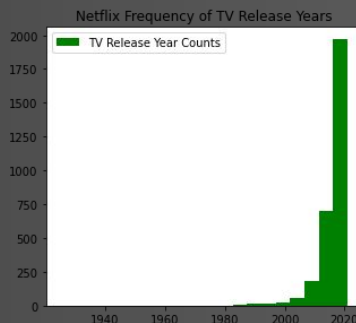
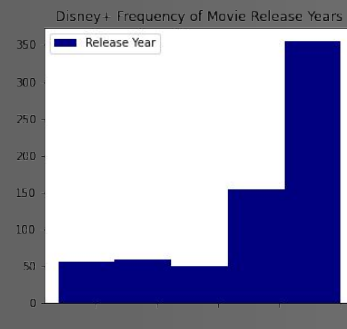
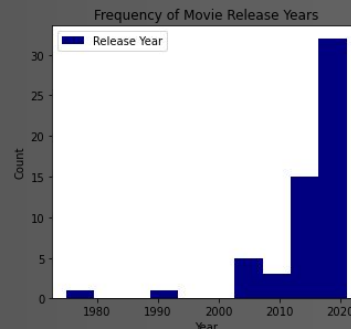
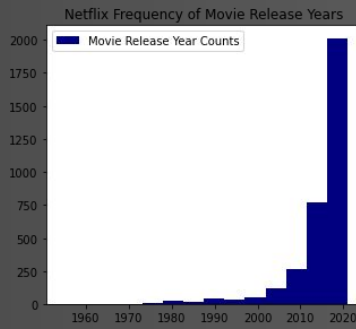


# Analysis by Platform



## Release Frequency by Year

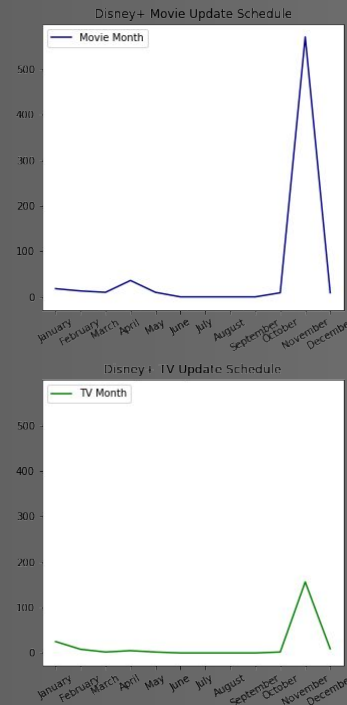
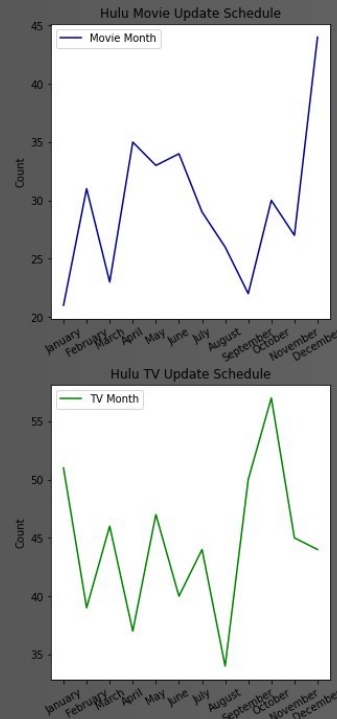
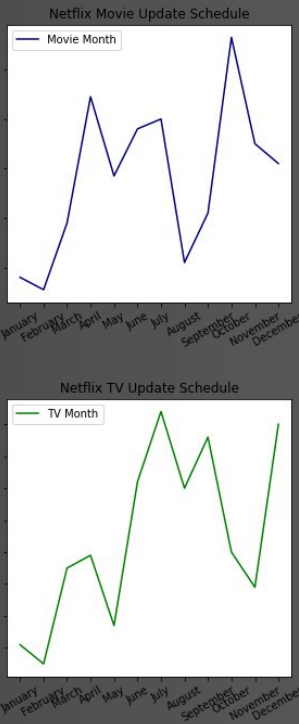
- Most of the offerings are up to date, within the past 10-20 years
- 21st century, film industry growth
- Earlier release dates, not as common, perhaps older audiences aren't as attracted



# Analysis by Platform

## Update Schedule

- Disney+ spike closer to the holidays
- General upward trend for Hulu and Netflix as the year progresses




# Reflections


## On our Data

- Limited data, limited scope
  - Mostly categorical data
- Difficult to verify data and analysis
- Limited insight into original source and context for data


### Movies on Netflix, Prime Video, Hulu and Disney+


A collection of movies found on these streaming platforms




 Ruchi Bhatia • updated 3 months ago (Version 3)

[Data](#) [Tasks \(4\)](#) [Code \(69\)](#) [Discussion \(10\)](#) [Activity](#) [Metadata](#) [Download \(1 MB\)](#) [New Notebook](#)

 Usability 10.0

 License CC0: Public Domain

 Tags arts and entertainment, online communities, movies and tv shows

Description

#### Content

The dataset is an amalgamation of:

- data that was scraped, which comprised a comprehensive list of movies available on various streaming platforms
- IMDb dataset

# Reflections

## On Topics for Further Exploration

- Content
  - Implications for e.g., consumer behavior, platform “success” (e.g. profit, subscription #s/growth, etc.)
  - Leverage for more expansive topics - e.g. entertainment, comparisons to more traditional landscapes for movies/tv shows, etc.
  - Longer term trends - e.g. Disney+’s maturation over time
  - Pair w/ qualitative analysis
- Technical
  - Code refactoring



# Conclusion

2021 [Thriller](#) [TV-G](#)

Thank you for tuning  
in! Open for questions!

♥ LIKE