**Project Report**

# **GitHub URL**

www.github.com/lisacsryan/UCDPA\_LisaRyan

# **Abstract**

A look at Disney Plus streaming service contents, to see if there are patterns between the content that is being put up on the platform, age ratings and year of release of content.

# **Introduction**

For the purpose of this project, I will be examining the contents of the Disney Plus streaming service, and analysing this data to see if there are patterns between the content that is being put up on the platform, age ratings and duration of content . Disney Plus is a streaming service which launched in late 2019 in three launch countries, where it replaced the pre-existing DisneyLife in the UK which had been available for the previous four years. (Forbes, 2019). Disney Plus curates movies and television shows both created by and owned by Walt Disney Studios. As of January 1st, 2023, the streaming service has 161.8 million subscribers globally (The Walt Disney Company, 2023). This was something which I found easily accessible data available for which I found an interesting topic, the factors which make up the decisions of the streaming providers as to which content they will continue to distribute.

# **Dataset**

The data that I am using for this analysis is available through Kaggle in a CSV file. This data set runs to the end of 2021.

This dataset is available here: <https://www.kaggle.com/datasets/shivamb/disney-movies-and-tv-shows?select=disney_plus_titles.csv>

# **Implementation Process**

Importing Data

I imported the relevant packages, Pandas, MatPlotLib, Numpy, and Seaborn at the start of my process which would allow me to analyse and visualize the data going forward.

I imported a CSV file called “Disney\_Plus\_Titles.csv” from Kaggle, which I downloaded to my hard drive and imported using this code:

**disney = "disney\_plus\_titles.csv"**

I then created a dataframe for this project, using the Disney file as imported above, telling it to read the columns first.

**df = pd.read\_csv(disney, low\_memory=False)**

Data Cleaning

Initially, I focused on checking whether there were any null values, which were quite plentiful in the director, cast and country categories of data.

**df.isnull().sum()**

Once that was confirmed, I then took action to change any null values as follows:

**df['director'].replace(np.nan, 'No Director',inplace=True)**

**df['cast'].replace(np.nan, 'No Cast',inplace=True)**

**df['country']=df['country'].fillna(df['country'].mode()[0])**

This ensured that for every piece of content listed in the file, we would not come up with any blanks when it came to analysis and everything would be included.

Data Exploration

In looking at this data, I was really interested to see not only what types of content were contained in the streaming service, but also if there were any trends to be found in the duration of streaming content, age ratings and type of content.

I looked at splitting the content of Disney Plus into the two types which are already set out in the data, movies and television shows.

Firstly, I looked to get the Value Counts of each type, using this code:

**df['type'].value\_counts()**

I then plotted this graph as follows:

**sns.countplot(x='type',data=df,palette="ch:s=.25,rot=-.25")**

**plt.show()**

I also broke the two types of data down into separate data frames for further analysis using this code:

**df1= df[df['type']=='Movie']**

**df2= df[df['type']=='TV Show']**

I also did some of this analysis using the Pandas GroupBy, using the two separate data frames to look in more detail at the data, for example, looking to see how many TV shows of each rating were released in each year.

**df2.groupby(["rating","release\_year"]).count().**

I also used this to show that I could define a custom function to create reusable code. Instead of each time having to type in the code above, I created this:

**def ReleaseYearRatings ():**

**print(df2.groupby(["rating","release\_year"]).count())**

**ReleaseYearRatings()**

Secondly, I looked at the years listed for content release to see if there were any years which particularly stood out for content creation/release.

**plt.figure(figsize = (12,6))**

**sns.countplot(x='release\_year',order = df['release\_year'].value\_counts().index[0:10],data = df,palette='mako')**

**plt.title('Content Release in Years on DisneyPlus Vs Count')**

I further broke this down using the two extra data sets created above, using this code:

**plt.figure(figsize = (12,6))**

**sns.countplot(x='release\_year',order = df1['release\_year'].value\_counts().index[0:10],data = df,palette='mako')**

**plt.title('Movie Release in Years on DisneyPlus Vs Count')**

**plt.figure(figsize = (12,6))**

**sns.countplot(x='release\_year',order = df2['release\_year'].value\_counts().index[0:10],data = df,palette='mako')**

**plt.title('Movie Release in Years on DisneyPlus Vs Count')**

I was interested in looking at the TV series on Disney Plus and seeing if there was a pattern to how many series were on the streaming service. This count runs until the end of 2021 so does not take into account any further series which are awaiting a new series to be added, therefore this data does not confirm “one and done” status or discontinuation but is interesting to look at. I used the following code to see the numbers

**df2['duration'].value\_counts()**

and then plotted it using this code:

**plt.figure(figsize = (15,6))**

**sns.countplot(x='duration',order = df2['duration'].value\_counts().index[0:20],data = df,palette='mako')**

**plt.title('How many Seasons Does Each Series Have?')**

And lastly, I looked at how much of the Disney Plus content fits in each individual age rating category.

**plt.figure(figsize = (12,6))**

**sns.countplot(x='rating',order = df['rating'].value\_counts().index[0:10],data = df,palette='mako')**

**plt.title('How Much of Disney Plus Content is in which Age Rating Category?')**

Again, I was interested to see if there were any differences in the types of content being made for different age categories, to see if there were any where more tv shows were being developed than movies, or vice versa. I used the following code to plot this out:

**plt.figure(figsize = (12,6))**

**sns.countplot(x='rating',order = df1['rating'].value\_counts().index[0:10],data = df,palette='mako')**

**plt.title('How Much of Disney Plus Content is in which Age Rating Category?')**

**plt.figure(figsize = (12,6))**

**sns.countplot(x='rating',order = df2['rating'].value\_counts().index[0:10],data = df,palette='mako')**

**plt.title('How Much of Disney Plus Content is in which Age Rating Category?')**

# **Results**

**What is the Breakdown of Disney Plus content by type?**

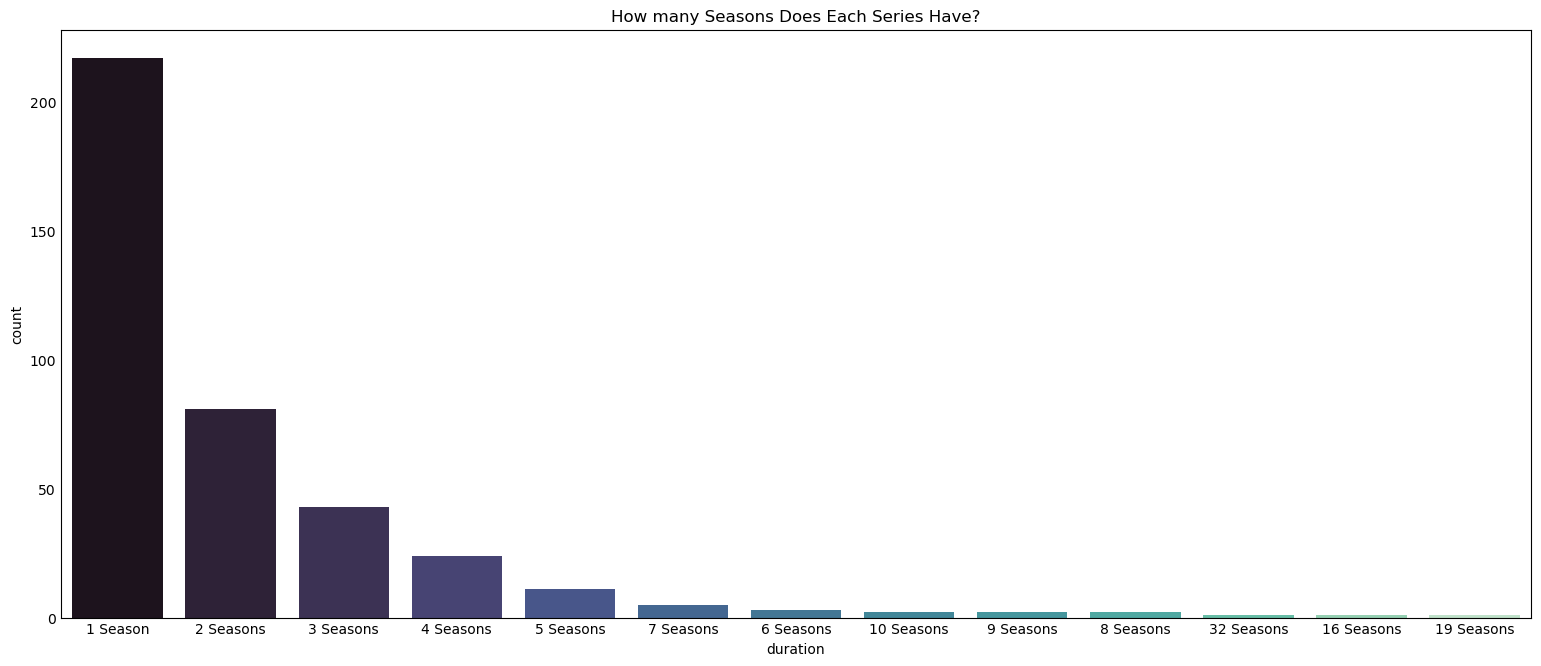
When broken down by type, Disney Plus contains 1051 movies and 393 television shows.

Chart, bar chart

Description automatically generated

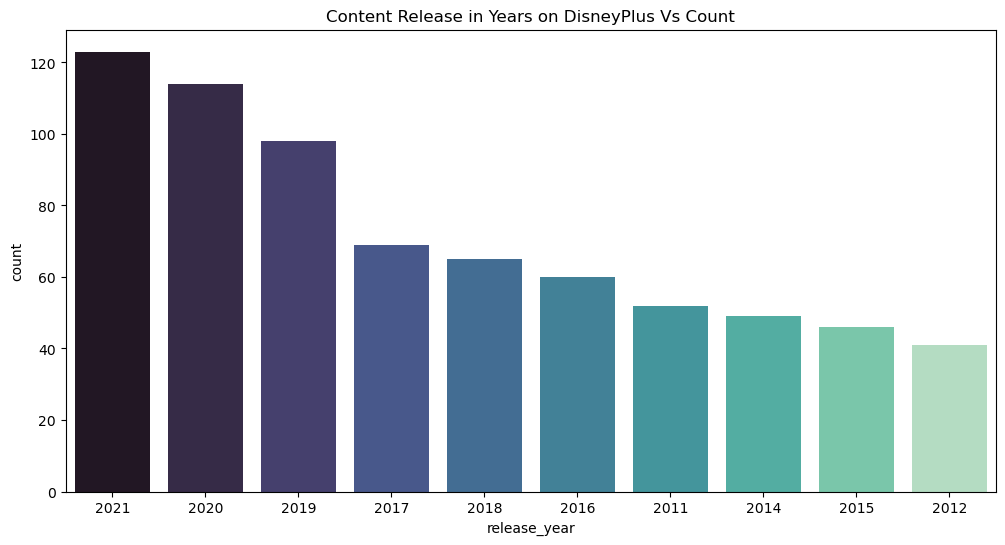
**How many TV Series are once-off, compared to ongoing?**

217 of the 393 television series are only available for one series on Disney Plus (55%), with 20% having two series and just under 11% having 3 series. The longest running series on Disney Plus has 32 series currently.

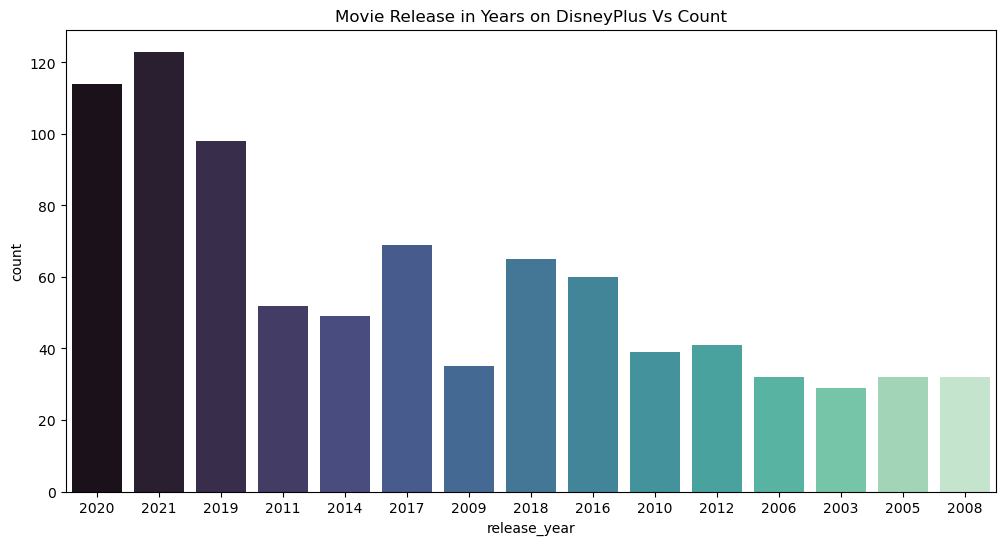


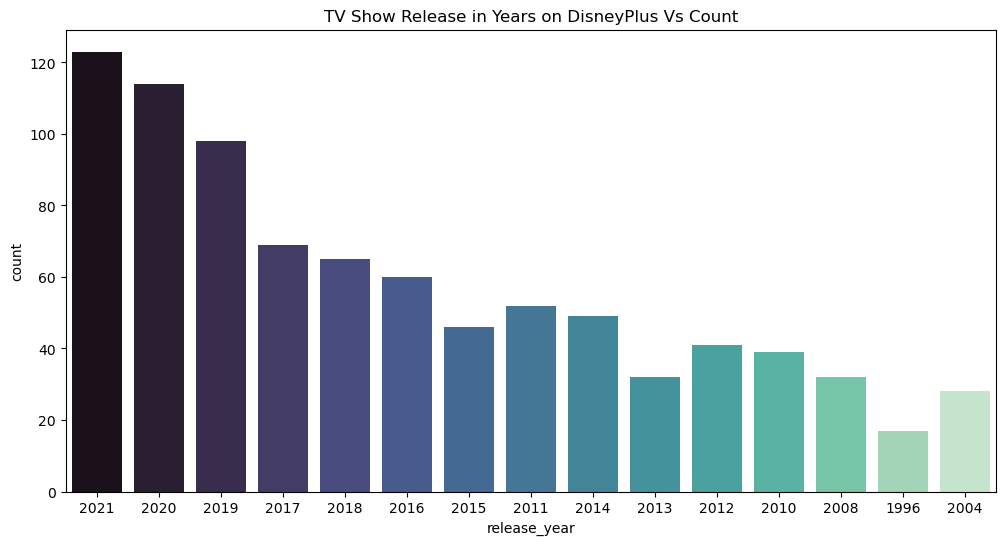
**Is there much of a difference by year to the amount of Disney Plus content being released?**

2021 (the most recent year used in this dataset) is the year with the most content released, with 123 releases, closely followed by 2020, where 114 releases of content took place. These sharp increases from lower-count years (2018 was 65 pieces of content) are likely due to not only the increase in content being specifically made for Disney Plus customers, but also the incorporation of other media companies and their content by the Walt Disney Company.



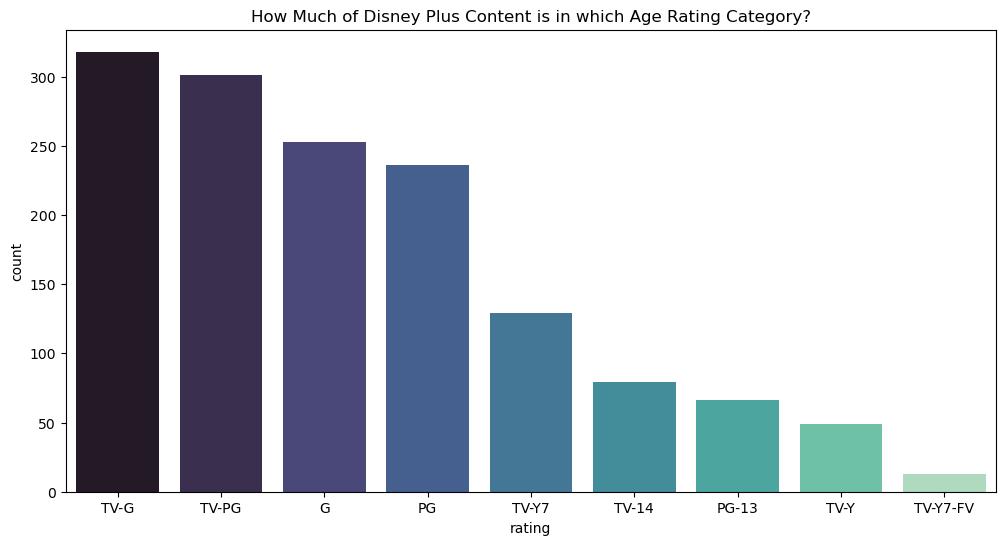
When broken down by content type, 2021 was still the busiest year yet for new releases in both film and tv show content. I decided to plot the top 15 years of content for each to see if there were any bigger differences. While in each case, generally the number of television shows per year were less than the movies, for example, although 2021 was the high point for both content types, it meant there were 53 new TV shows and 70 new movies. Movie content release dates start much further back (1944) although the vast majority of content has been since the year 2000. Meanwhile, the oldest TV content dates back to the 1970s, with an increase in content being seen from the mid 1990s. Unlike the Movie listing, where the top 15 did not go before the year 2000, 1996 was one of the bigger years for television series which are shown on Disney Plus, with six series from that year being available on the streaming service.



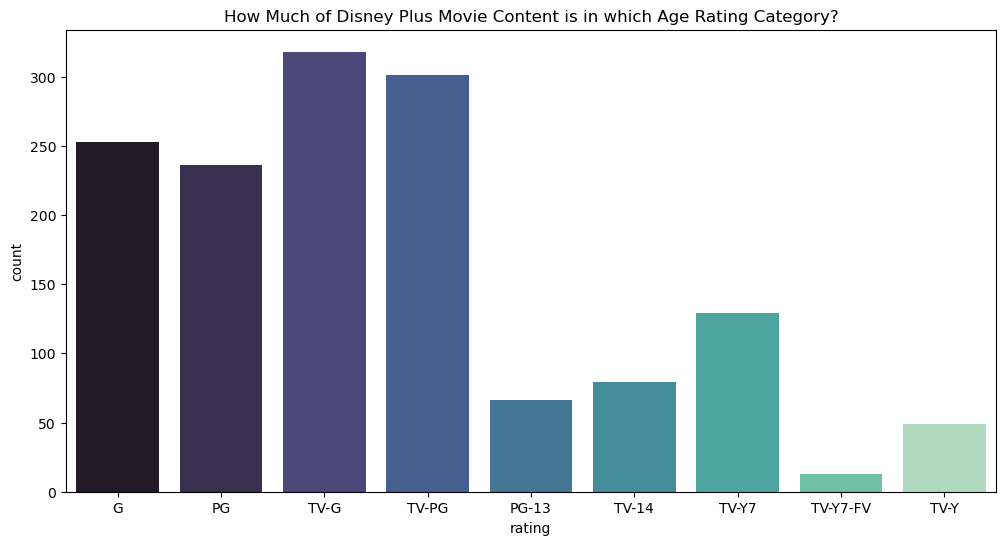
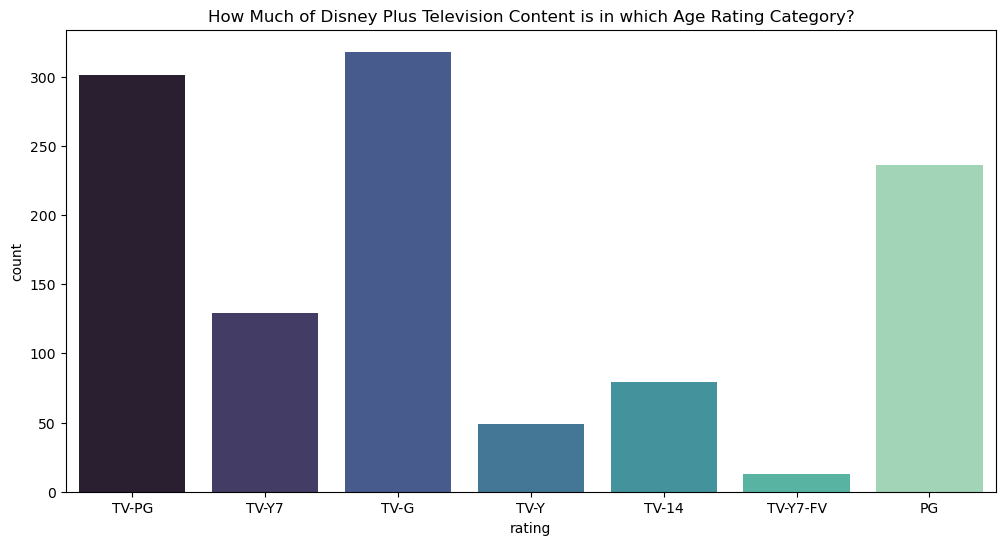


**How much Content is suitable for which Age Rating?**

While Disney is notorious for creation of content for children, in both movie and television programming, the content on Disney Plus goes from a TV-G rating all the way up to TV-14 (Parents Strongly Cautioned), and for films from G to PG13. The vast majority of content on Disney Plus is suitable for all audiences, with 318 items rated TV-G, and another 253 rated G, while parental guidance is recommended for 301 television programmes and 236 films. Comparatively, just 79 items of television content were given the rating “Parents Strongly Cautioned”, while 66 film pieces were rated the same. While Disney Plus is definitely moving more towards the adult market as well as for children and young people, they still are balanced firmly towards family friendly programming.



When solely looking at Disney Plus television content, the largest proportion of programming is rated PG (121), followed by Y7 (93). A third of the amount rated PG was deigned only suitable for 14 and up (42).



Similarly, when looking at the movie and age rating break downs, the majority of content is suitable for all audiences, with 253 G rated items, and 235 PG rated items. Just 66 items were considered only suitable for those aged over 13.

# Insights

* 72% of the content currently available on the Disney Plus Streaming Service is in Movie Format, while 28% is made up of television series.
* 55% of TV series on Disney Plus have just one season. While another 20% has 2 series, and 11% have 3 series, the numbers with anything more than 5 series are miniscule. The longest running series on Disney Plus is The Simpsons, which is currently at 32 series.
* With each passing year, the amount of content created for the streaming service/commissioned to be put up for subscribers increases, regardless of content type. This suggests that Disney is not just purchasing the rights to older television shows or movies which they do not currently own, and are more likely to spend on development of new content themselves where they own the intellectual property.
* Disney Plus has done a lot of work on changing the brand of Disney from something solely aimed at children to a brand for the entire family. However, their previous catalogue of work being so aimed at a younger audience does skew the amount of content available for adult users and this is an area where they should be focusing to increase adult subscribers without children.

# References

1. Forbes, 2019. *Everything You Need To Know About Disney - Pricing, Release Date, Content and More.* [Online]   
   Available at: https://www.forbes.com/sites/erikkain/2019/11/11/everything-you-need-to-know-about-disney--pricing-release-date-content-and-more/  
   [Accessed 26 March 2023].

2. The Walt Disney Company, 2023. " THE WALT DISNEY COMPANY REPORTS FIRST QUARTER EARNINGS FOR FISCAL 2023" [Online] Available at: https://thewaltdisneycompany.com/app/uploads/2023/02/q1-fy23-earnings.pdf

[Accessed 26 March 2023]