

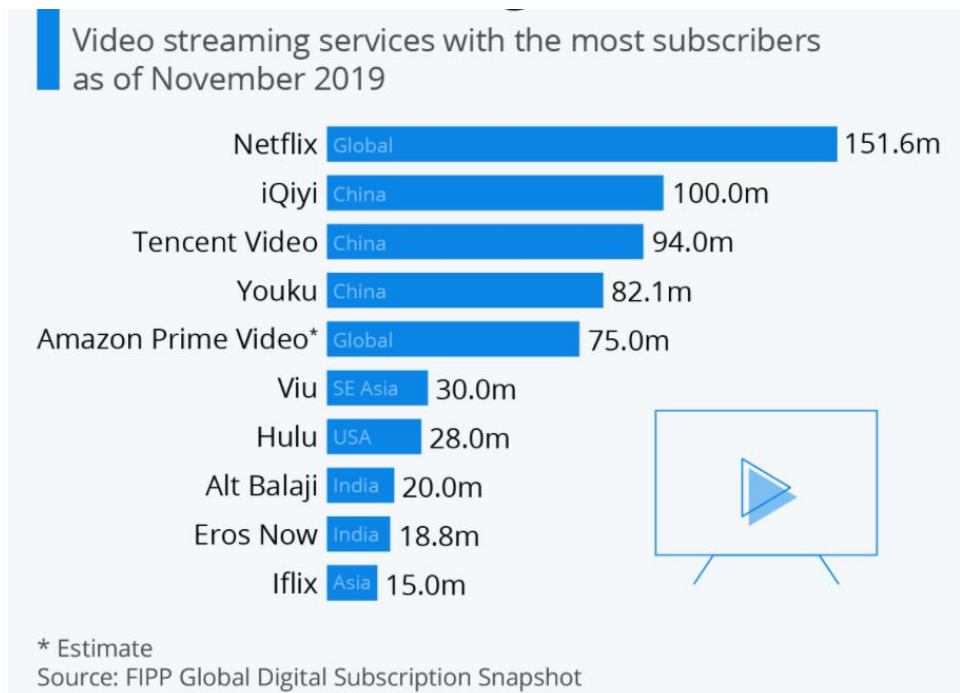
SMU Data Science Bootcamp
Project 1 – Group 4
Mark Burton, Juveriya Baig, Jenette Smith, Angie Tran, Jammy Lo

Streaming Services Data Analysis for Netflix, Hulu, Prime Video and Disney+

Inspiration:

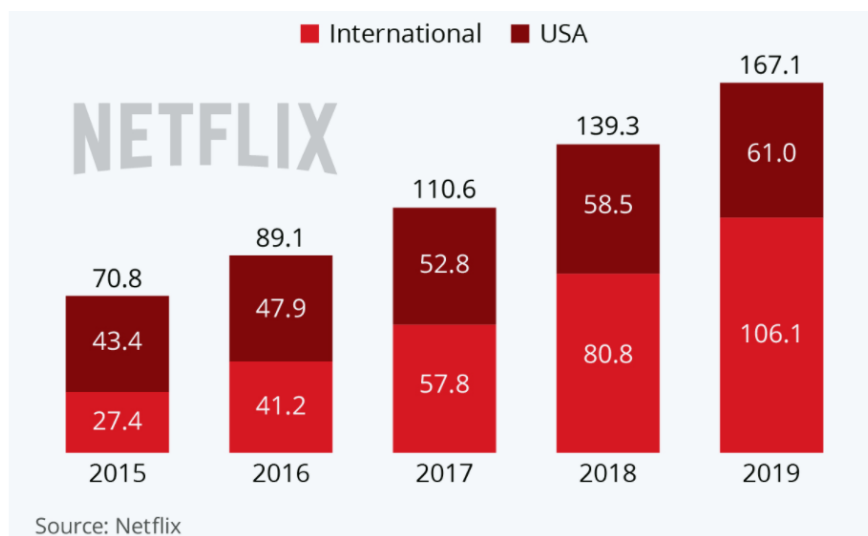
We decided as a team to focus on Netflix because with the COVID-19 pandemic more people and families are having to social distance and stay home. Therefore, many people are turning to a form of streaming service (i.e. Netflix, Hulu, Prime Video, Disney+) for entertainment and as a distraction from the everyday reality of living through a health crisis with unemployment rates and morale as a whole affected.

Netflix was created in 1997 and since this has morphed from a DVD rental service to a monthly subscription service. The affordability of DVD players in 2005, the implementation of streaming service capabilities in 2007 and the introduction of original content in 2012 helped to propel Netflix over the 15 years. Worldwide, Netflix has seen a growth in subscribers with COVID-19 from 167.1 million (Q4 2019) to 183 million (April 2020) to 193 million (July 2020). Amazon's Prime Video service is the closest competitor to Netflix with 75 million subscribers, comparing Netflix to Hulu, Prime Video and Disney+.



Number of Users for Netflix (in the U.S.) By Year:

2000	300000
2002	600000
2005	4.2 million
2011	20.15 million
2012	25.47 million
2013	31.71 million
2014	37.7 million
2015	43.4 million
2016	47.91 million
2017	52.81 million
2018	58.49 million
2019	61.04 million
2020 (Q1)	69.96 million
2020 (Q2)	72.90 million
2020 (Q3)	73.08 million



Our color scheme for the inspiration for our visuals is based off of the HEX colors for Netflix, Hulu, Prime Video and Disney+.

netflix #E50914 ■

hulu #3DBB3D ■

prime #3B7A57 ■

disney #006E99 ■

We also referred to Kaggle.com as an inspiration because there have been other projects and visualizations created for Netflix, Hulu, Prime Video and Disney+. With COVID-19 being a worldwide pandemic, people from a variety of backgrounds, cultures, ethnicity, household income, etc. have access to streaming services to provide educational value and entertainment. Netflix is available in over 190 countries but is not available in China, Crimea, North Korea and Syria.

Netflix is available for **streaming** in over 190 countries. Our library of TV shows and movies varies based on the country and will change from time to time.



NOTE: Netflix is not yet available in China, Crimea, North Korea, or Syria.

Background:

For this project, we examined data from Netflix movies and tv shows to observe how this data compares to Netflix's main competitors (i.e. Hulu, Disney+, Prime Video). We used two sources of CSV files from Kaggle.com for our data sets:

<https://www.kaggle.com/ruchi798/movies-on-netflix-prime-video-hulu-and-disney>

<https://www.kaggle.com/ruchi798/tv-shows-on-netflix-prime-video-hulu-and-disney>

Research Questions:

- 1) How does Netflix's content offering compare to its competitors? (Amazon, Hulu, Disney+, etc.)
- 2) Is Netflix's content rating better or worse than other streaming services using ratings of Rotten Tomatoes or IMDb ratings?
- 3) Is the suggested age group of content different across streaming services?

- 4) How has the content that is offered changed over time?
- 5) Is there a correlation between the rating and the number of streaming services a movie/tv show is on?

Data Cleaning:

After this data was cleaned, we removed the data for movie and tv genre, run-time, Rotten Tomatoes rating, country of origin and the language for each respective movie or tv show. We also removed the director category because this category applied to only movies and not tv shows. Also, due to time constraints for this project, we did not have time to investigate directors into further details. However, one future recommendation would be to examine the top 1-15 directors to see how the directors compare across Netflix, Hulu, Prime Video and Disney+. Therefore, we focused on data on title, year, age, IMDb rating, Netflix, Hulu, Prime Video and Disney+. The type category for both data sets is distinguished with a 1 for movies and a 0 for tv shows.

The data for genre was too broad since there are so many different types of movie categories and this would be harder to quantify with the rest of the data. Secondly, only the data for movies showed a genre category. We decided to eliminate the data for Rotten Tomatoes since our team is already using the data for the IMDb rating to rank and rate Netflix to these three competitors. Additionally, we omitted the data for run-time, country of origin and the language spoken during the tv show or movie to centrally focus on data of information as this data would also be very broad with so many different languages spoken and countries of origin. It would also be difficult to distinguish run-time length for a movie or a tv because run-time is not pre-set by a certain criteria or metrics for streaming service. While a feature film at a movie theatre must run at least 70-75 minutes and be no less than 40 minutes but no more than 3 hours, this same rule doesn't apply to streaming service programming.

Our team initially hypothesized that there would be higher ratings and more tv shows on Netflix compared to its competitors since Netflix has been the "preferred" streaming service for most people.

Project Analysis:

In our analysis, we analyzed 16,744 movies using 12 columns of data. We added a 'Service' column to identify if a movie appears on Netflix, Hulu, Prime Video and Disney+. Any movies appearing under more than one 'Service' are identified as 'Multiple.' Upon further research, there are 11,758 movies on Prime Video; 3,188 movies on Netflix; 639 movies on Hulu; 639 movies listed on multiple platforms and 539 movies on Disney+. We discussed dropping the data under the 'Multiple' category since this doesn't consist of much data. Prime Video has more services for older movies and there are also more movies that you can buy or rent directly

through Prime Video. Disney+ was launched on Nov. 12. 2019 in the United States, Canada and the Netherlands. Hulu was launched Oct. 29, 2007, while Prime Video launched on Sept. 7, 1996.

ID	Title	Year	Age	IMDb	Netflix	Hulu	Prime Video	Disney+	Service	Directors	Type
0	1	Inception	2010	13.0	8.8	1	0	0	0	Christopher Nolan	0
1	2	The Matrix	1999	18.0	8.7	1	0	0	0	Lana Wachowski, Lilly Wachowski	0
2	3	Avengers: Infinity War	2018	13.0	8.5	1	0	0	0	Anthony Russo, Joe Russo	0
3	4	Back to the Future	1985	7.0	8.5	1	0	0	0	Robert Zemeckis	0
4	5	The Good, the Bad and the Ugly	1966	18.0	8.8	1	0	1	0	Sergio Leone	0
...
16739	16740	The Ghosts of Buxley Hall	1980	NaN	6.2	0	0	0	1	Bruce Bilson	0
16740	16741	The Poof Point	2001	7.0	4.7	0	0	0	1	Neal Israel	0
16741	16742	Sharks of Lost Island	2013	NaN	5.7	0	0	0	1	Neil Gelinas	0
16742	16743	Man Among Cheetahs	2017	NaN	6.6	0	0	0	1	Richard Slater-Jones	0
16743	16744	In Beaver Valley	1950	NaN	NaN	0	0	0	1	James Algar	0

16744 rows x 12 columns

ID	Title	Year	Age	IMDb	Netflix	Hulu	Prime Video	Disney+	Service	Directors	Type	
0	1	Inception	2010	13.0	8.8	1	0	0	0	Netflix	Christopher Nolan	0
1	2	The Matrix	1999	18.0	8.7	1	0	0	0	Netflix	Lana Wachowski,Lilly Wachowski	0
2	3	Avengers: Infinity War	2018	13.0	8.5	1	0	0	0	Netflix	Anthony Russo,Joe Russo	0
3	4	Back to the Future	1985	7.0	8.5	1	0	0	0	Netflix	Robert Zemeckis	0
4	5	The Good, the Bad and the Ugly	1966	18.0	8.8	1	0	1	0	Multiple	Sergio Leone	0

▶ ML

df.Service.value_counts()

Prime Video11758
Netflix3188
Hulu639
Multiple627
Disney+532
Name: Service, dtype: int64

We created age bins to further help distinguish the groups for movies and TV shows:

G: 0-6

PG: 8-12

PG-13: 14-16

R: 17+

Movies:

▶ MI

```
# Create the bins on Age column in which Data will be held
bins = [0,7, 13, 17, 99]
```

```
# Create the names for the five bins
group_names = ["G", "PG", "PG-13", "R"]
```

▶ MI

```
df["Age Bin"] = pd.cut(df["Age"], bins, labels=group_names, include_lowest=True)
df
```

	ID	Title	Year	Age	IMDb	Netflix	Hulu	Prime Video	Disney+	Service	Directors	Type	Age Bin
0	1	Inception	2010	13.0	8.8	1	0	0	0	Netflix	Christopher Nolan	0	PG
1	2	The Matrix	1999	18.0	8.7	1	0	0	0	Netflix	Lana Wachowski,Lilly Wachowski	0	R
2	3	Avengers: Infinity War	2018	13.0	8.5	1	0	0	0	Netflix	Anthony Russo,Joe Russo	0	PG
3	4	Back to the Future	1985	7.0	8.5	1	0	0	0	Netflix	Robert Zemeckis	0	G
4	5	The Good, the Bad and the Ugly	1966	18.0	8.8	1	0	1	0	Multiple	Sergio Leone	0	R
...
16739	16740	The Ghosts of Buxley Hall	1980	NaN	6.2	0	0	0	1	Disney+	Bruce Bilson	0	NaN
16740	16741	The Poof Point	2001	7.0	4.7	0	0	0	1	Disney+	Neal Israel	0	G
16741	16742	Sharks of Lost Island	2013	NaN	5.7	0	0	0	1	Disney+	Neil Gelinas	0	NaN
16742	16743	Man Among Cheetahs	2017	NaN	6.6	0	0	0	1	Disney+	Richard Slater-Jones	0	NaN
16743	16744	In Beaver Valley	1950	NaN	NaN	0	0	0	1	Disney+	James Algar	0	NaN

16744 rows x 13 columns

TV Shows:

		Title	Year	Age	IMDb	Netflix	Hulu	Prime Video	Disney+	type	Service	Age Bin
0		Breaking Bad	2008	18.0	9.5	1	0	0	0	1	Netflix	R
1		Stranger Things	2016	16.0	8.8	1	0	0	0	1	Netflix	PG-13
2		Money Heist	2017	18.0	8.4	1	0	0	0	1	Netflix	R
3		Sherlock	2010	16.0	9.1	1	0	0	0	1	Netflix	PG-13
4		Better Call Saul	2015	18.0	8.7	1	0	0	0	1	Netflix	R
...	
5606		Tut's Treasures: Hidden Secrets	2018	NaN	NaN	0	0	0	1	1	Disney+	NaN
5607		Paradise Islands	2017	NaN	NaN	0	0	0	1	1	Disney+	NaN
5608		Wild Russia	2018	NaN	NaN	0	0	0	1	1	Disney+	NaN
5609		Love & Vets	2017	NaN	NaN	0	0	0	1	1	Disney+	NaN
5610		United States of Animals	2016	NaN	NaN	0	0	0	1	1	Disney+	NaN

5611 rows x 11 columns

Similarly, we also added a column to calculate TV shows by 'Service' to determine if a TV show appeared on more than one streaming service or not. Our dataset for TV shows consisted of 5,611 shows with 10 columns of data. Prime Video consists of 1,889 TV shows; Netflix has 1,748 TV shows; Hulu has 1,452 TV shows and Disney+ has 156 TV shows. There is also a number of TV shows that are flagged under 'Multiple' since more than one TV show could also appear under multiple streaming services. We also discussed omitting the 'Multiple' category from our data since this number represents a small percentage.

```

▶ MI
# add one more column for the type of streaming service
df.insert(9,"Service","",True)
df

```

	Title	Year	Age	IMDb	Netflix	Hulu	Prime Video	Disney+	type	Service
0	Breaking Bad	2008	18.0	9.5	1	0	0	0	1	
1	Stranger Things	2016	16.0	8.8	1	0	0	0	1	
2	Money Heist	2017	18.0	8.4	1	0	0	0	1	
3	Sherlock	2010	16.0	9.1	1	0	0	0	1	
4	Better Call Saul	2015	18.0	8.7	1	0	0	0	1	
...
5606	Tut's Treasures: Hidden Secrets	2018	NaN	NaN	0	0	0	1	1	
5607	Paradise Islands	2017	NaN	NaN	0	0	0	1	1	
5608	Wild Russia	2018	NaN	NaN	0	0	0	1	1	
5609	Love & Vets	2017	NaN	NaN	0	0	0	1	1	
5610	United States of Animals	2016	NaN	NaN	0	0	0	1	1	

5611 rows x 10 columns

```

▶ MI
df.Service.value_counts()

```

```

Prime Video    1889
Netflix        1748
Hulu           1452
Multiple        366
Disney+        156
Name: Service, dtype: int64

```

```

▶ MI
aveService = df.groupby('Service').IMDb.mean()
aveService

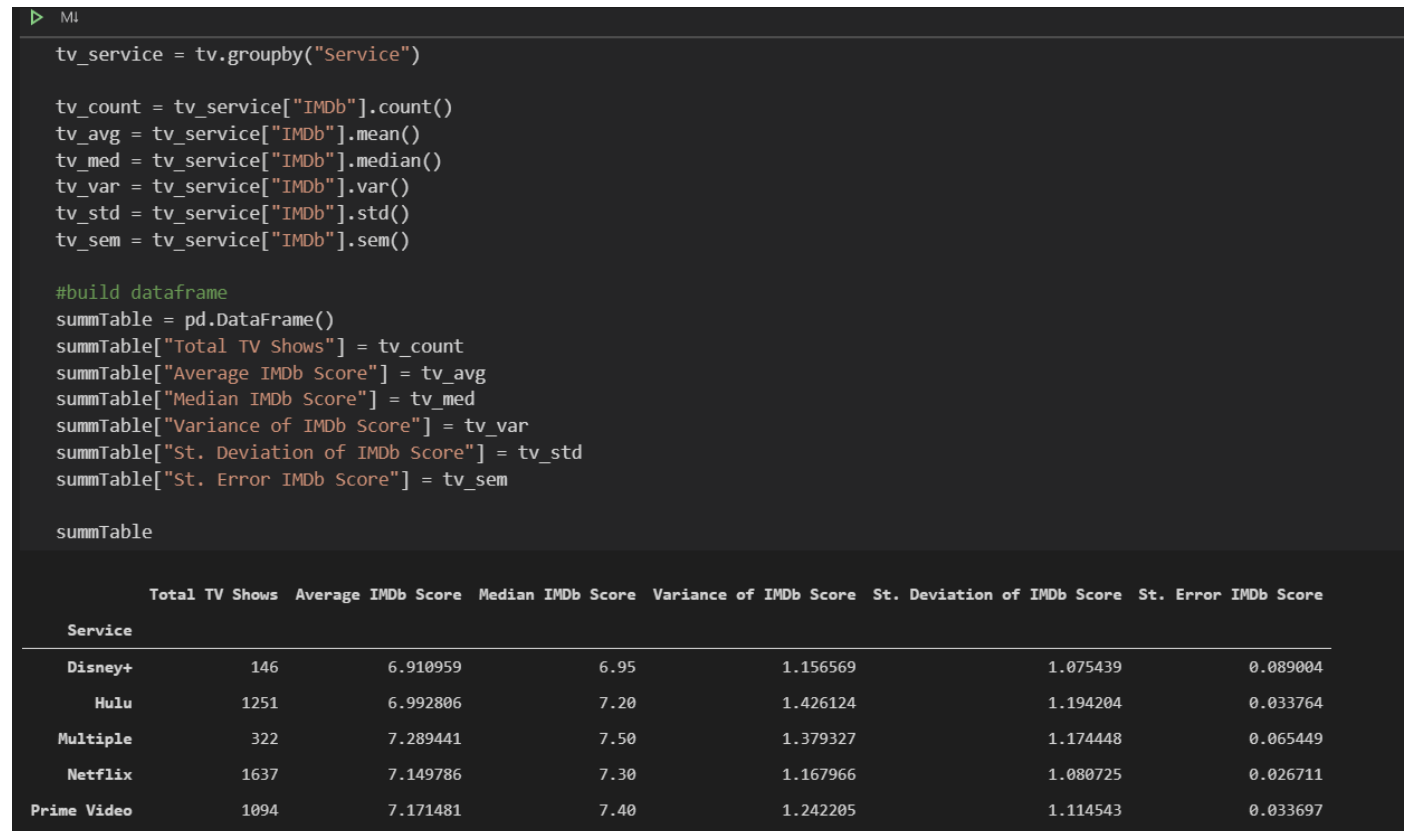
```

```

Service
Disney+    6.910959
Hulu       6.992806
Multiple   7.289441
Netflix    7.149786
Prime Video 7.171481
Name: IMDb, dtype: float64

```

We created a data frame for TV shows to further analyze total TV shows; average IMDb score; median IMDb score; variance of IMDb score; standard deviation of IMDb score and the standard error of IMDb score across Disney+, Hulu, Multiple Streaming Services, Netflix and Prime Video.



We also created a box plot to visually indicate how TV shows differ from one streaming service to another. There are left outliers for our box plot and Prime Video skews the most out of all of the streaming services we analyzed. This same information can further be illustrated with our violin plot. However, with our violin plot Hulu and Prime Video both have a longer tail but Disney+ has more content. As a side note, IMDb is a well-known database for Amazon.

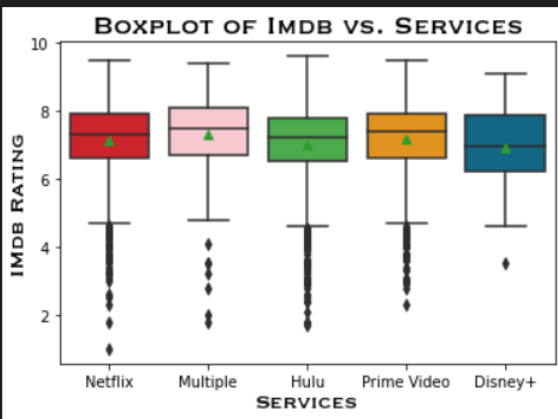
MI

```
# Make Boxplot
my_pal = {"Netflix": "#E50914", "Hulu": "#3DBB3D", "Prime Video": "#FF9900", "Disney+": "#006E99", "Multiple": "pink"}

sns.boxplot(x="Service", y="IMDb", data=tv, palette=my_pal, showmeans=True)

# Give our chart some labels and a title
plt.title("Boxplot of Imdb vs. Services", fontweight="bold", fontsize=20, name="Copperplate")
plt.xlabel('Services', fontsize=16, name="Copperplate")
plt.ylabel('IMDb Rating', fontsize=16, name="Copperplate")

plt.show()
```



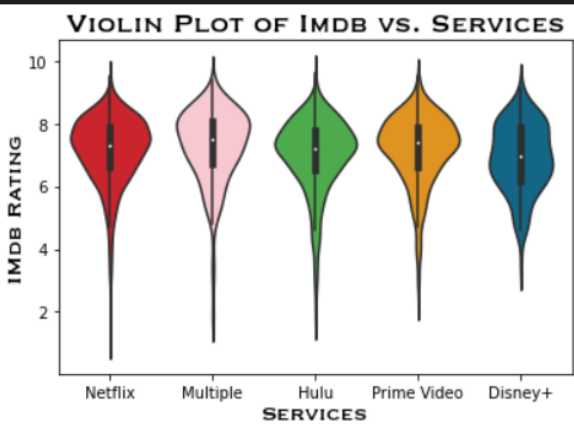
MI

```
# Make violin plot
my_pal = {"Netflix": "#E50914", "Hulu": "#3DBB3D", "Prime Video": "#FF9900", "Disney+": "#006E99", "Multiple": "pink"}

sns.violinplot(x="Service", y="IMDb", data=tv, palette=my_pal, showmeans=True)

# Give our chart some labels and a title
plt.title("Violin Plot of Imdb vs. Services", fontweight="bold", fontsize=20, name="Copperplate")
plt.xlabel('Services', fontsize=16, name="Copperplate")
plt.ylabel('IMDb Rating', fontsize=16, name="Copperplate")

plt.show()
```



For this group project, we didn't focus on doing heavy regression but we did focus on conducting a T-test and ANOVA. We compared Netflix first to the other 3 streaming services in an ANOVA and then compared Hulu to the other 3 streaming services in a T-test. Lastly, we compared Prime Video in a T-test to Disney+. Since Group 5 of this dataset identified any 'Multiple' services, we did not use this group of information in our analysis.

Comparing Group 1 for Netflix to Group 3 for Prime Video:

Prime Video was the most significant to Netflix with a P-value of 0.613 with 0.05 being the industry standard for comparison. Since Netflix compared to both Disney+ and Hulu have a smaller P-value further away from 0.05 then there is stronger evidence in favor of rejecting the null hypothesis and accepting the alternate hypothesis.

Comparing Group 2 for Hulu to Group 4 for Disney+:

Disney+ was the most significant to Hulu with a P-value of 0.039 with 0.05 being the industry standard for comparison.

Comparing Group 3 for Prime Video to Group 4 for Disney+:

This comparison is not significant because the P-value is 0.006.

We found that IMDb ratings between Netflix and Prime Video are significant but there is no difference in ratings between Hulu and Disney+. We are more confident that a T-test is significant for all streaming services showing that Netflix has better rated movies than Hulu. There are four distinct movie ratings: Disney+, Netflix, Hulu and Prime Video. However, for TV shows, Netflix and Prime Video are better than Hulu and Disney+ but we can't confidently say that Hulu is better than Disney+. We are not able to reject the null hypothesis that one is better than the other for TV shows comparing Netflix and Prime Video to Hulu and Disney+.

```
▶ M1

# Extract individual groups
group1 = tv[tv["Service"] == "Netflix"]["IMDb"]
group2 = tv[tv["Service"] == "Hulu"]["IMDb"]
group3 = tv[tv["Service"] == "Prime Video"]["IMDb"]
group4 = tv[tv["Service"] == "Disney+"]["IMDb"]
group5 = tv[tv["Service"] == "Multiple"]["IMDb"]
```

```

▶ MI

# ANOVA test
stats.f_oneway(group1, group2, group3, group4)

F_onewayResult(statistic=7.701546310794659, pvalue=3.955907843985844e-05)

▶ MI

# T-tests between Netflix versus others
print(stats.ttest_ind(group1, group2, equal_var=False))
print(stats.ttest_ind(group1, group3, equal_var=False))
print(stats.ttest_ind(group1, group4, equal_var=False))

Ttest_indResult(statistic=3.6463060547800894, pvalue=0.00027137441595873345)
Ttest_indResult(statistic=-0.5045325817570749, pvalue=0.6139356461863421)
Ttest_indResult(statistic=2.570089063461629, pvalue=0.011013958199669294)

▶ MI

# T-tests between Hulu versus others
print(stats.ttest_ind(group2, group3, equal_var=False))
print(stats.ttest_ind(group2, group4, equal_var=False))

Ttest_indResult(statistic=-3.745669680387388, pvalue=0.00018427206685950345)
Ttest_indResult(statistic=0.8597998243815425, pvalue=0.390987056685557)

▶ MI

# T-tests between Prime Video versus Disney+
print(stats.ttest_ind(group3, group4, equal_var=False))

Ttest_indResult(statistic=2.7374605550047564, pvalue=0.006782744991005879)

```

We analyzed Regression in our project analysis to determine if there is a linear regression present between Netflix, Disney+, Hulu and Prime Video. Our research indicated that there is not a linear regression present with a negative correlation. Regressions are not significant from the datasets that we analyzed for TV shows or movies.

We were also able to determine that by using a scatter plot for prediction vs. regression models that this information is bad. If our group was able to make a perfect prediction then the scatter plot would have appeared more in a linear direction, instead of having data appear everywhere once plotted.

► M4

```
#Regression
tv_year = tv.iloc[:,1]
tv_imdb = tv.iloc[:,2]
tv_correlation = stats.pearsonr(tv_year,tv_imdb)

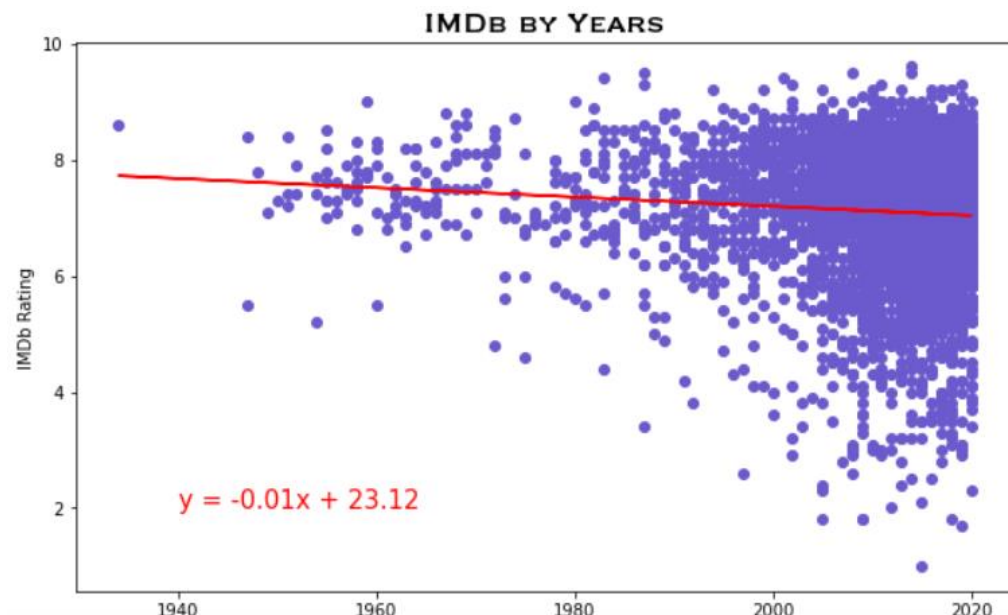
(slope, intercept, rvalue, pvalue, stderr) = linregress(tv_year,tv_imdb)

regress_values = tv_year * slope + intercept
line_eq = "y = " + str(round(slope,2)) + "x + " + str(round(intercept,2))
print(line_eq)
plt.figure(figsize = (10,6))
plt.scatter(tv_year,tv_imdb,color="slateblue")
plt.plot(tv_year,regress_values,"r-")
plt.annotate(line_eq,(1940,2),fontsize=15,color="red")

# Give our chart some labels and a title
plt.title("IMDb by Years",fontweight ="bold",fontsize=20, name="Copperplate")
plt.xlabel('Year')
plt.ylabel('IMDb Rating')
print(f"The correlation between both factors is {round(tv_correlation[0],2)}")

plt.show()
```

y = -0.01x + 23.12
The correlation between both factors is -0.08

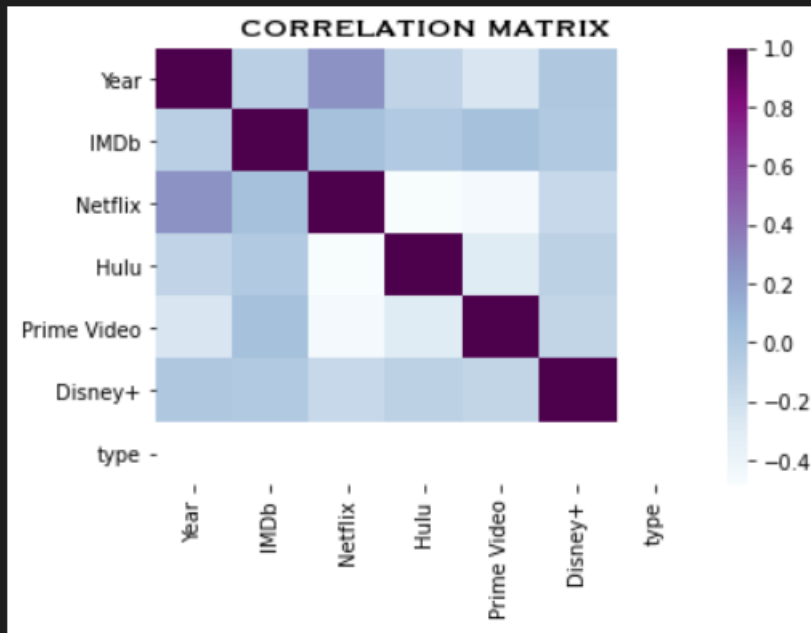


Our group also analyzed all four streaming services using a Seaborn Heat Map. This Seaborn Heat Map indicated that Netflix and year are more positive related and there are more TV shows as the years increase over time. IMDb has more correlation with all services, while Prime Video and year are negative. Additionally, Prime Video and Netflix do not share much in common but Hulu and Disney+ have more in common.

▶ MI

```
corrs = tv.corr()
sns.heatmap(corrs,cmap="BuPu")
title = 'CORRELATION MATRIX'
plt.title(title, loc='center', fontsize=16, fontweight='bold', name="Copperplate")
```

Text(0.5, 1.0, 'CORRELATION MATRIX')



▶ MI

```
feature = tv[['Year', 'Netflix', 'Hulu', 'Prime Video', 'Disney+']]
target = tv["IMDb"]
X = sm.add_constant(feature)
Y = target
model = sm.OLS(Y,X)
results = model.fit()
results.summary()
```

OLS Regression Results

Dep. Variable:	IMDb	R-squared:	0.012
Model:	OLS	Adj. R-squared:	0.011
Method:	Least Squares	F-statistic:	11.16
Date:	Sun, 15 Nov 2020	Prob (F-statistic):	1.05e-10
Time:	21:42:40	Log-Likelihood:	-6838.0
No. Observations:	4450	AIC:	1.369e+04
Df Residuals:	4444	BIC:	1.373e+04
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	24.5903	3.281	7.495	0.000	18.158	31.023
Year	-0.0088	0.002	-5.389	0.000	-0.012	-0.006
Netflix	0.2363	0.064	3.690	0.000	0.111	0.362
Hulu	0.0613	0.061	0.999	0.318	-0.059	0.182
Prime Video	0.1720	0.064	2.696	0.007	0.047	0.297
Disney+	-0.0494	0.104	-0.476	0.634	-0.253	0.154

Omnibus:	900.314	Durbin-Watson:	1.504
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1887.198
Skew:	-1.182	Prob(JB):	0.00

Kurtosis: 5.142 Cond. No. 3.91e+05

Warnings:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 3.91e+05. This might indicate that there are strong multicollinearity or other numerical problems.

▶ MI

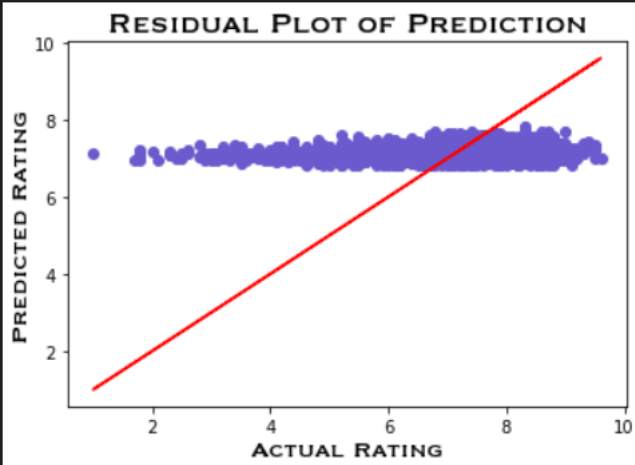
```
predictions = results.predict()
residuals = Y - predictions
```

ML

```
plt.scatter(Y, predictions, color="slateblue")
plt.plot(Y,Y,color="r")

# Give our chart some labels and a title
plt.title("Residual Plot of Prediction",fontweight ="bold",fontsize=20, name="Copperplate" )
plt.xlabel('Actual Rating',fontsize=16, name="Copperplate")
plt.ylabel('Predicted Rating',fontsize=16, name="Copperplate")

plt.show()
```

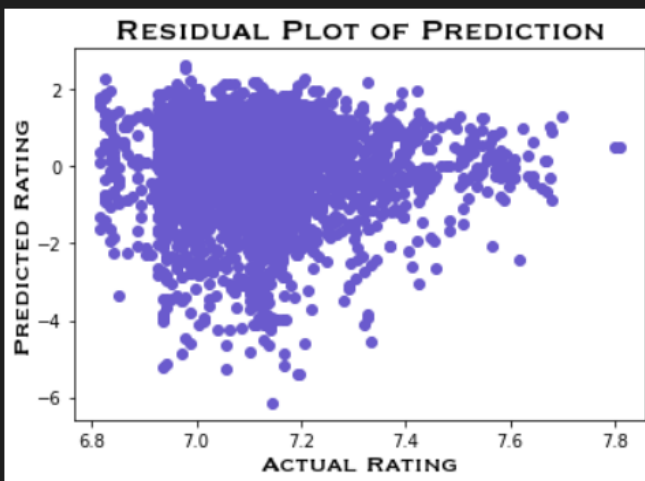


ML

```
plt.scatter(predictions,residuals,color="slateblue")

# Give our chart some labels and a title
plt.title("Residual Plot of Prediction",fontweight ="bold",fontsize=20, name="Copperplate" )
plt.xlabel('Actual Rating',fontsize=16, name="Copperplate")
plt.ylabel('Predicted Rating',fontsize=16, name="Copperplate")

plt.show()
```



Limitations:

As previously mentioned, we decided to omit data for directors because of limitation on time to code as much as we would have preferred (i.e. comparing the number of directors for the top 1-15 movies to determine any trends in data).

Since we omitted country of origin, languages spoken during the movie or tv show and run-time, this data was not analyzed during this project. This project also didn't examine how multiple countries use streaming services or internationally how streaming services compare/differ globally (i.e. domestic services within the United States compared to international countries who are permitted to use streaming services).

There was also a limitation in using age because there are a highly significant number of NULL values for movies and tv shows so we were unable to truly portray age as a variable with our data analysis. 7,300+ are NULL values when we cleaned this data.

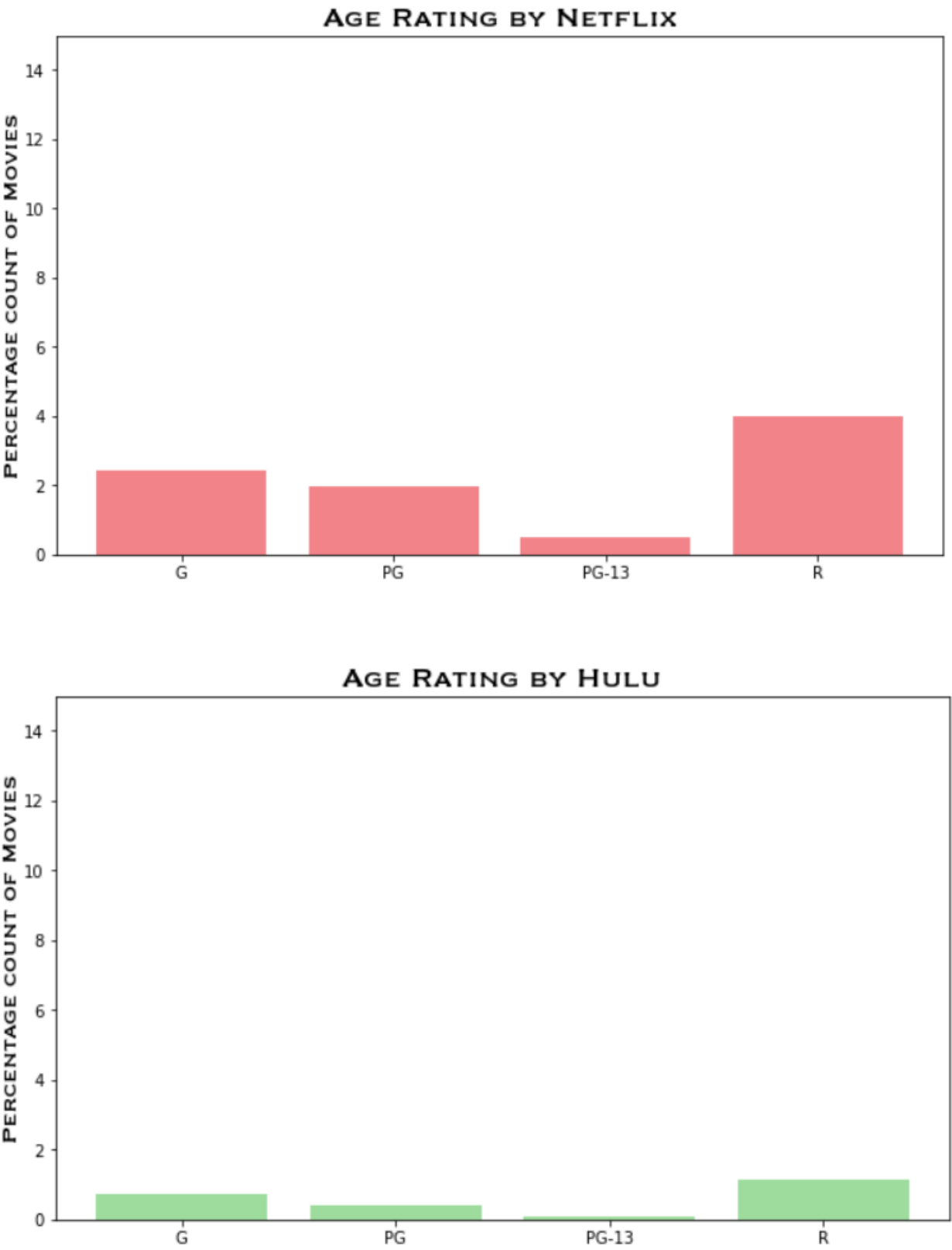
Since Disney+ is a relatively new streaming service (starting in Nov. 2019), there isn't as much data available to analyze compared to Netflix, Hulu or Prime Video. Within the scope of this project, we were not able to find the number of Netflix subscribers given our datasets analyzed.

Future Recommendations and Analysis for Research:

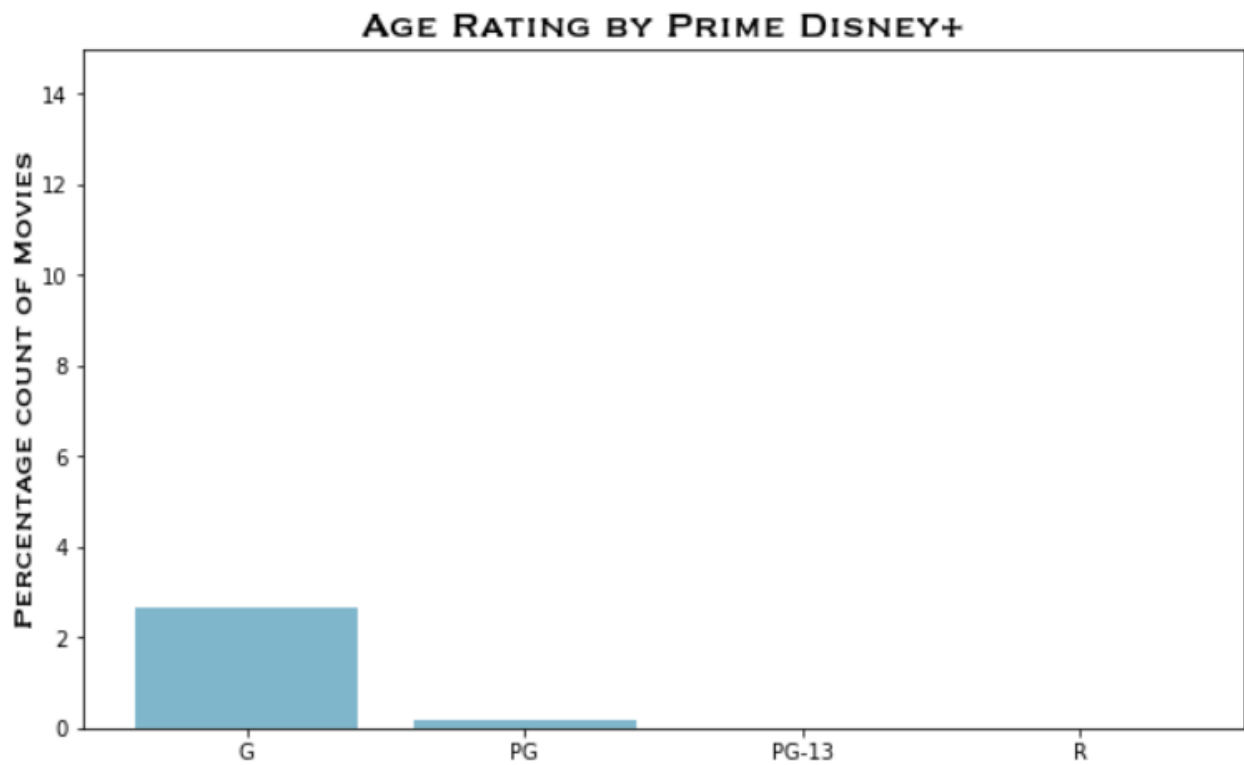
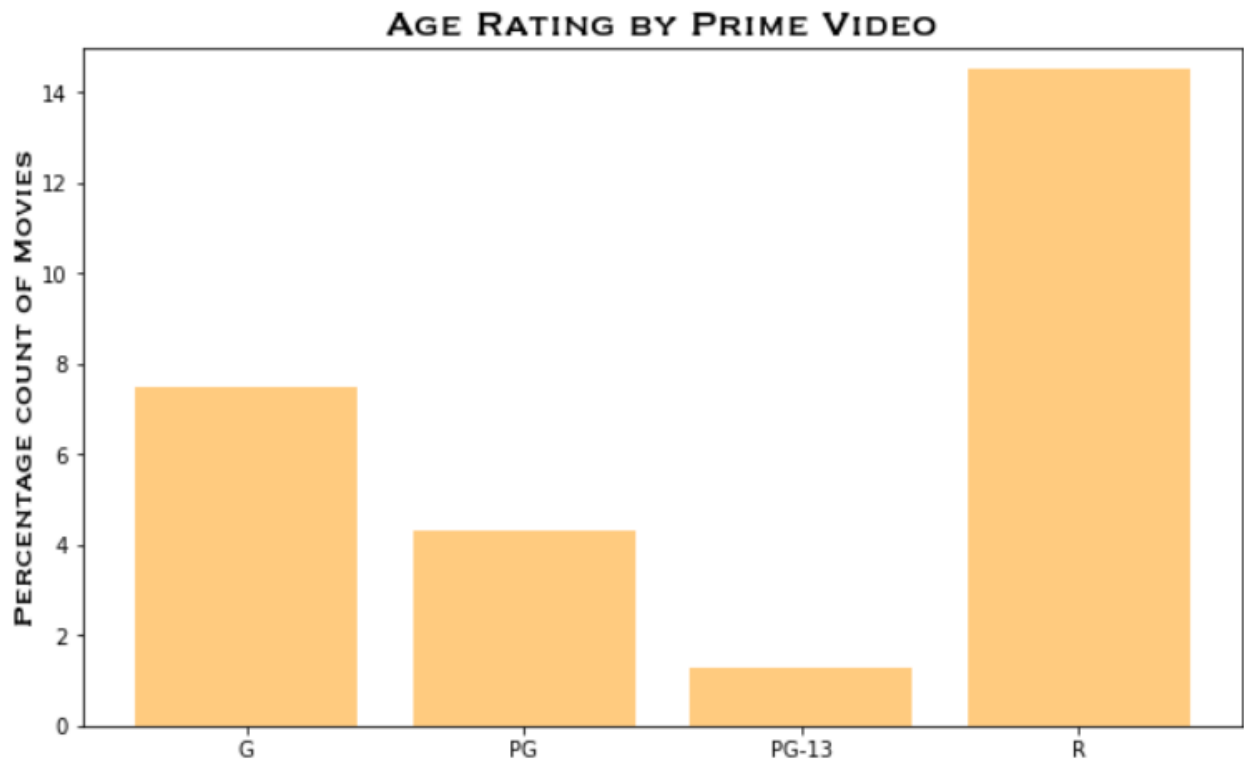
- Netflix, Hulu, Disney+ and Prime Video subscribers by
 1. age of subscribed user subscribing to streaming service
 2. education level
 3. household income
- directors and how data changes year to year for the top 1-15 movies across streaming services
- analyzing data for country of origin, languages spoken and run-time for tv shows or movies to further develop a more detailed analysis
- investigate how United States compares to Canada, Latin America, Asian Pacific, etc. for streaming services
- conducting a similar study with Disney+ dataset in 5 years after Disney+ has been on the market for several years as Disney+ launched Nov. 19, 2019

In conclusion, Netflix and Prime Video are recommended above Hulu and Disney+ based upon on data analysis from these two data sets for movies and TV shows. Netflix and Prime Video were found to have higher ratings and more significant data looking at T-test and ANOVA.

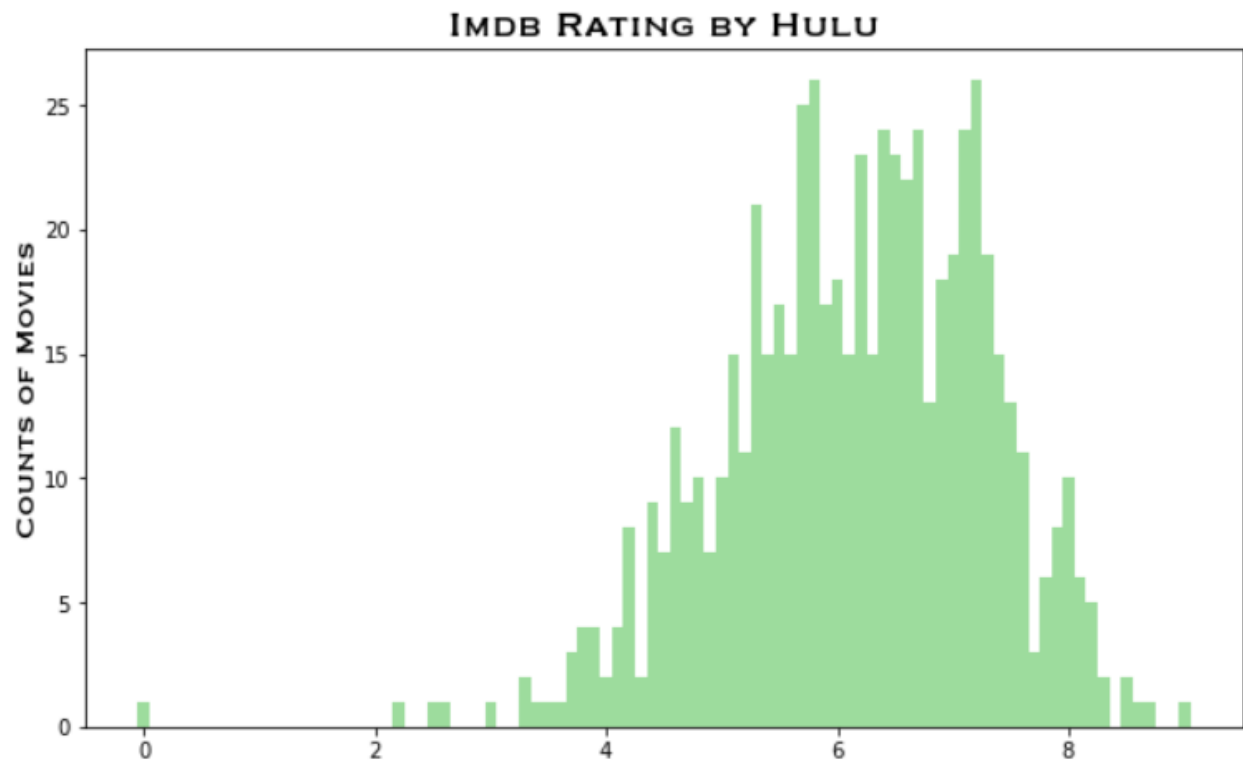
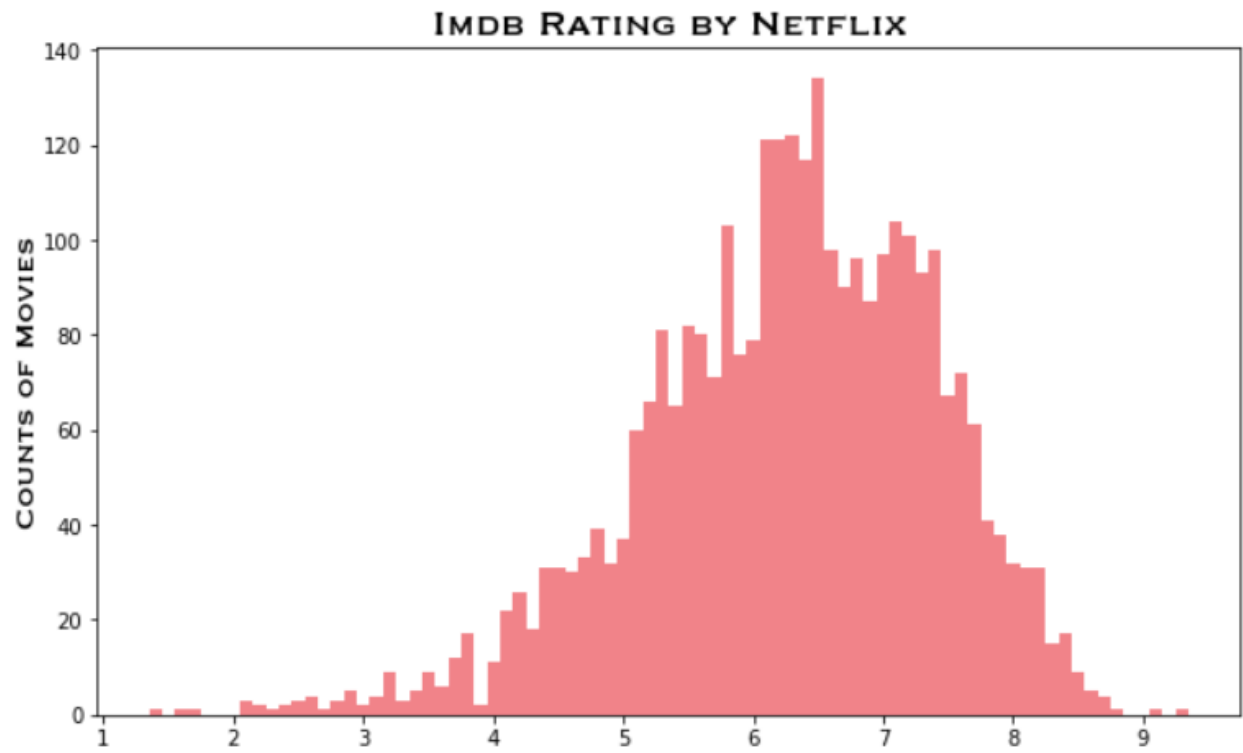
Analysis:
Bar Chart Between Age & Services - Figure 1-2:



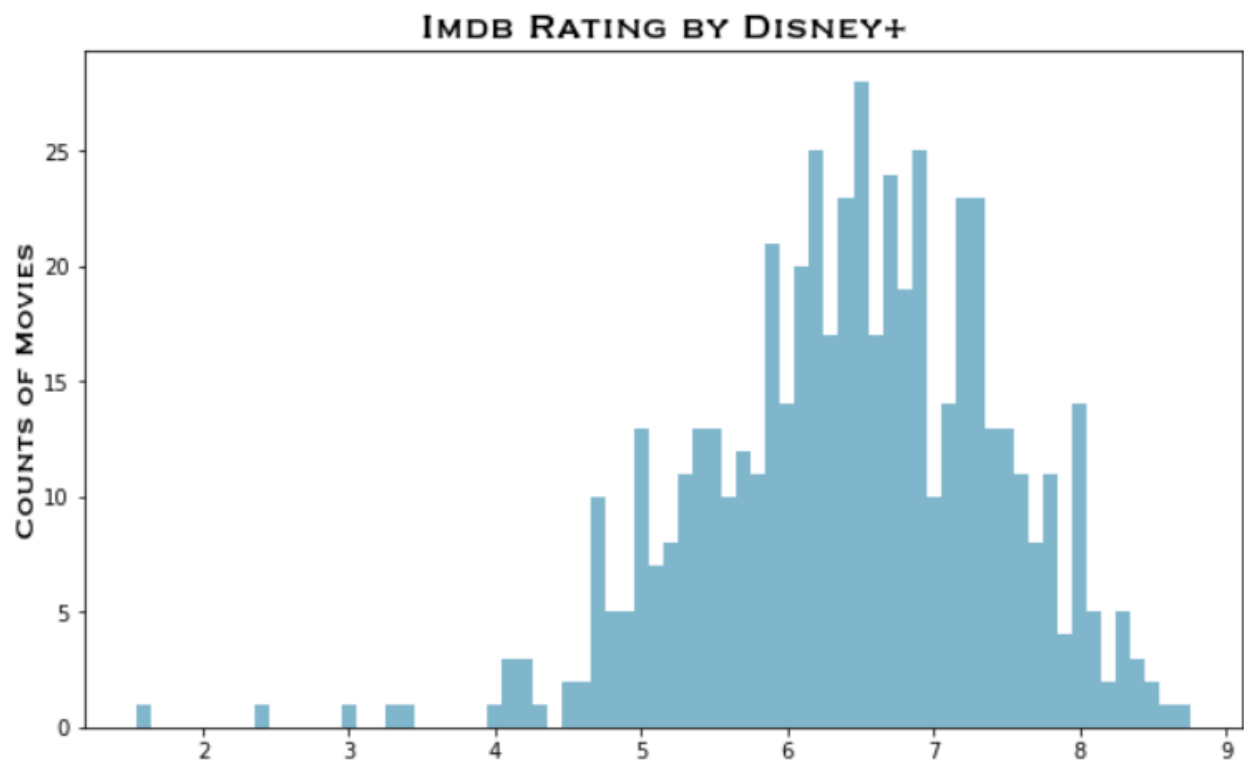
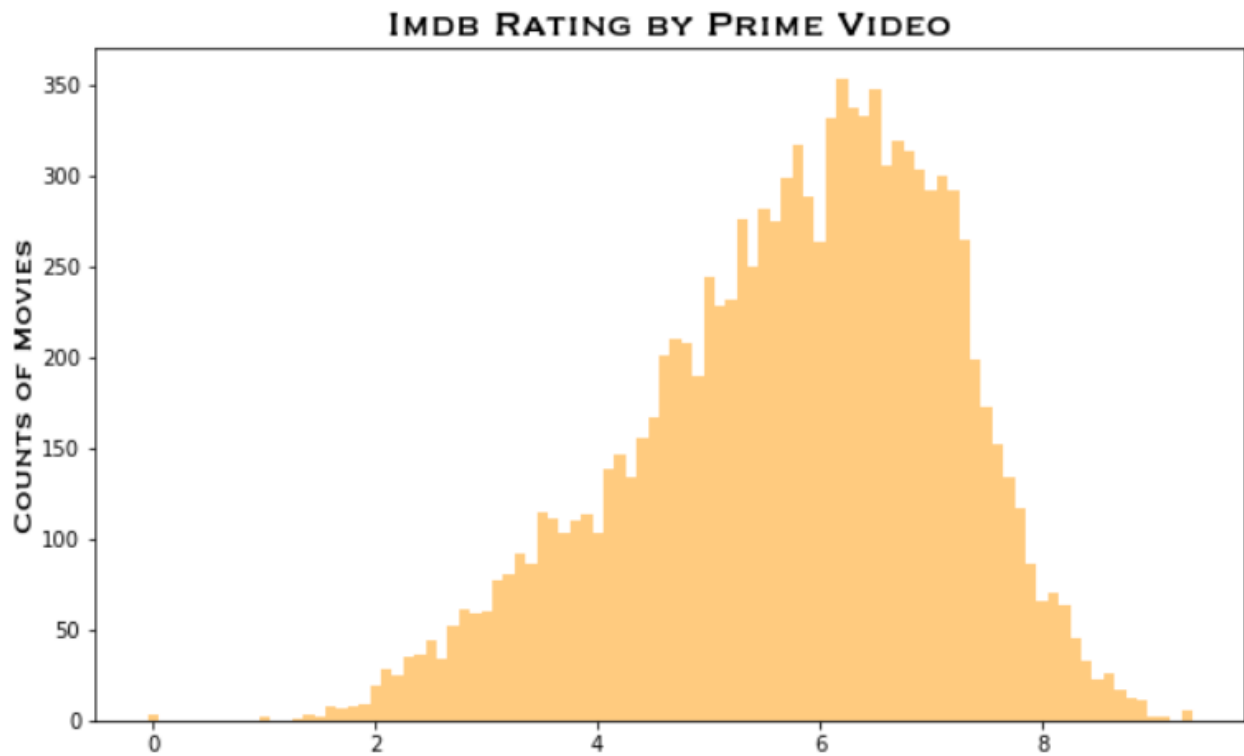
Bar Chart Between Age & Services - Figure 3-4:



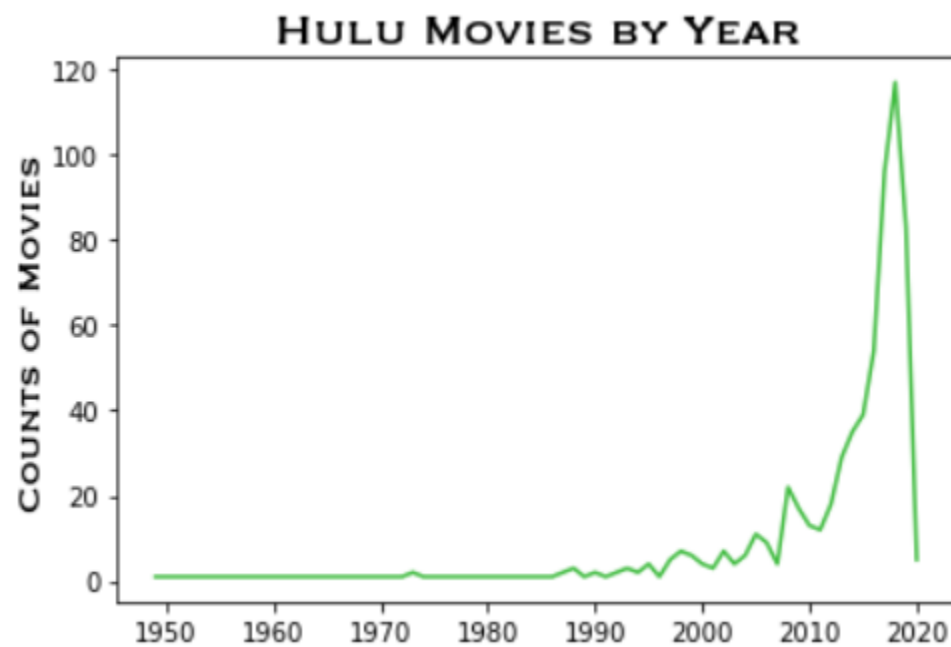
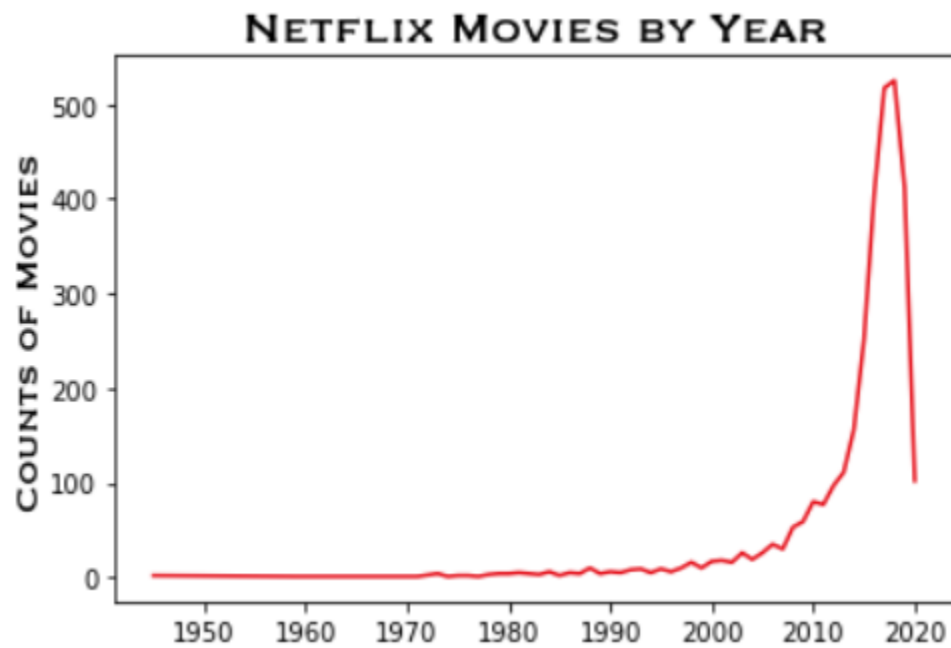
Bar Chart Between IMDb Rating & Services - Figure 1-2:



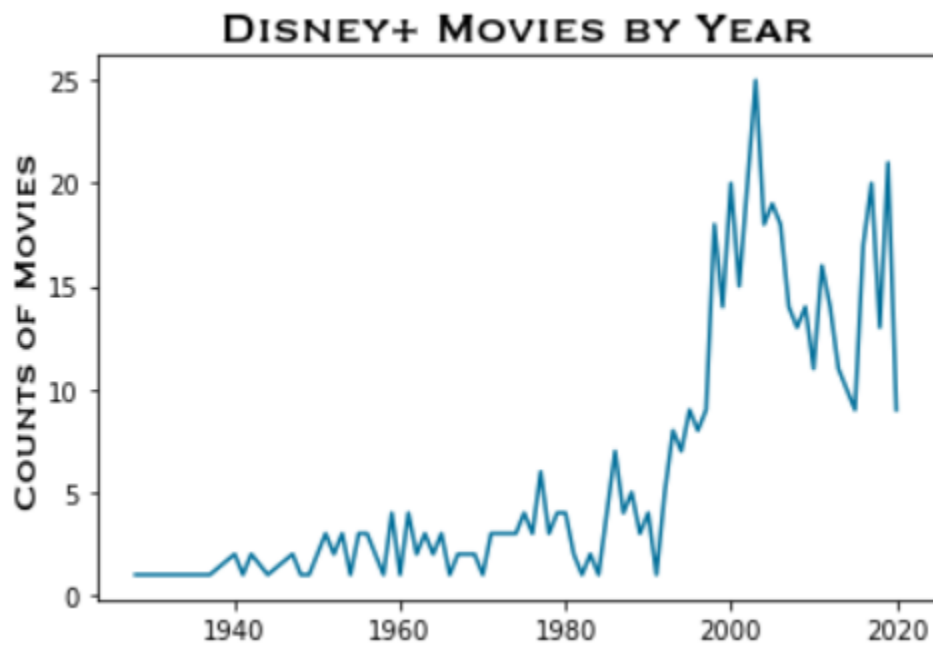
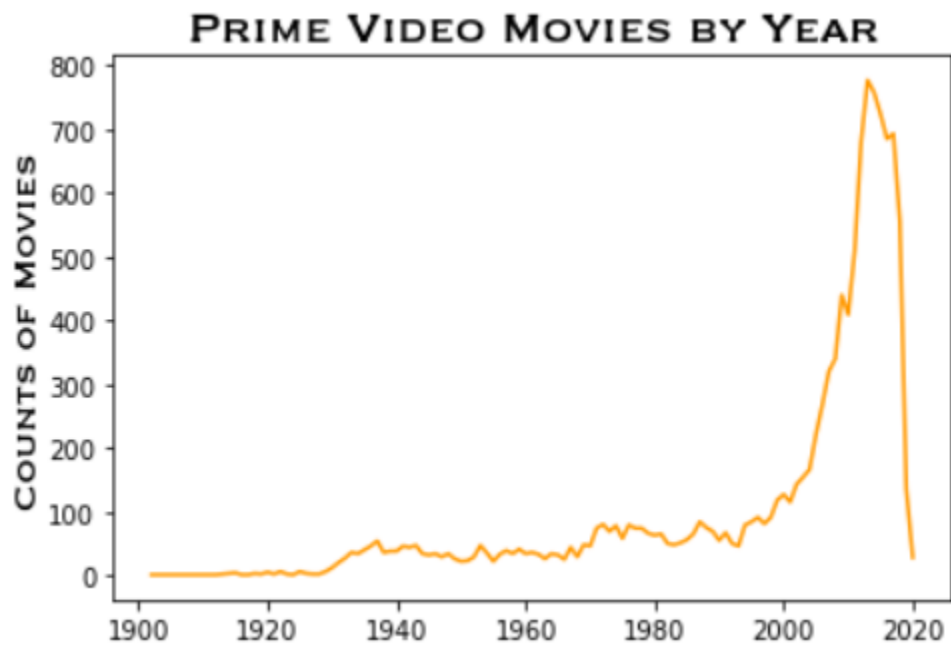
Bar Chart Between IMDb Rating & Services - Figure 3-4:



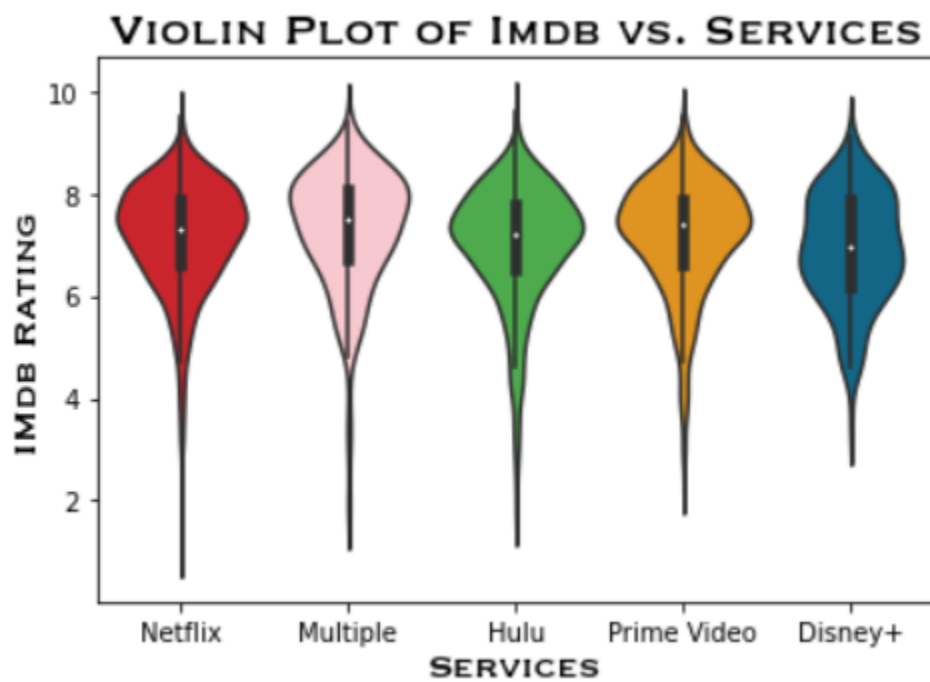
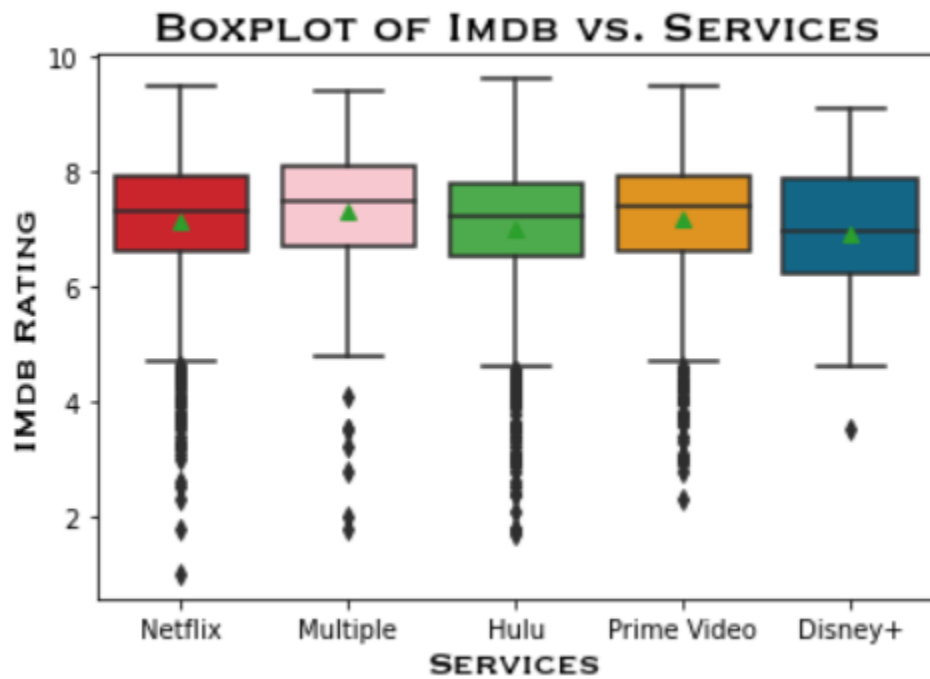
Line Graphs Year vs. Services - Figures 1-2:



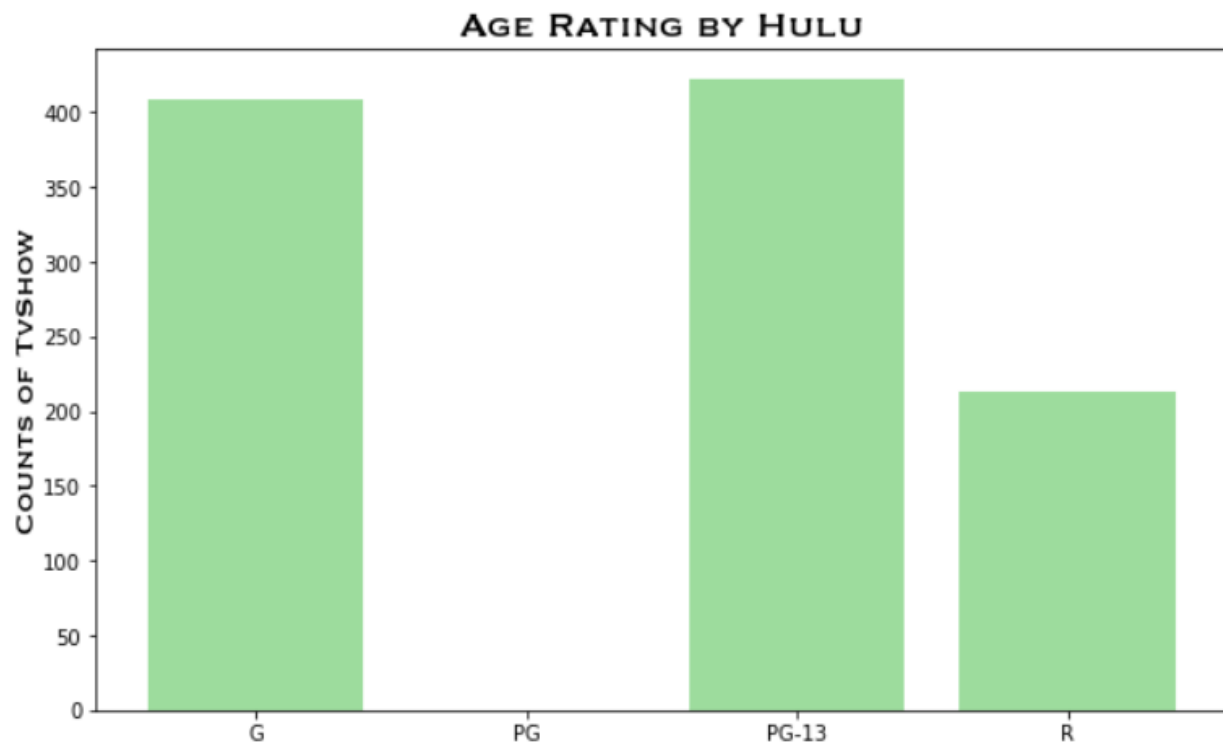
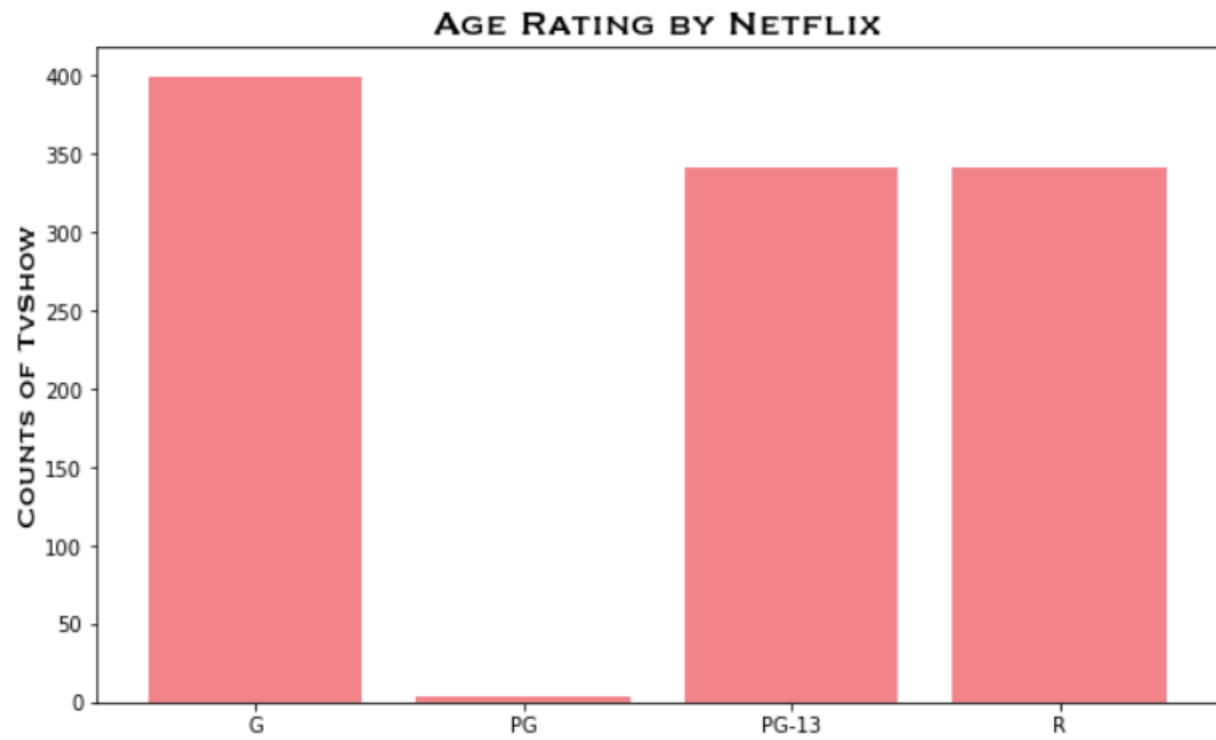
Line Graphs Year vs. Services - Figures 3-4:



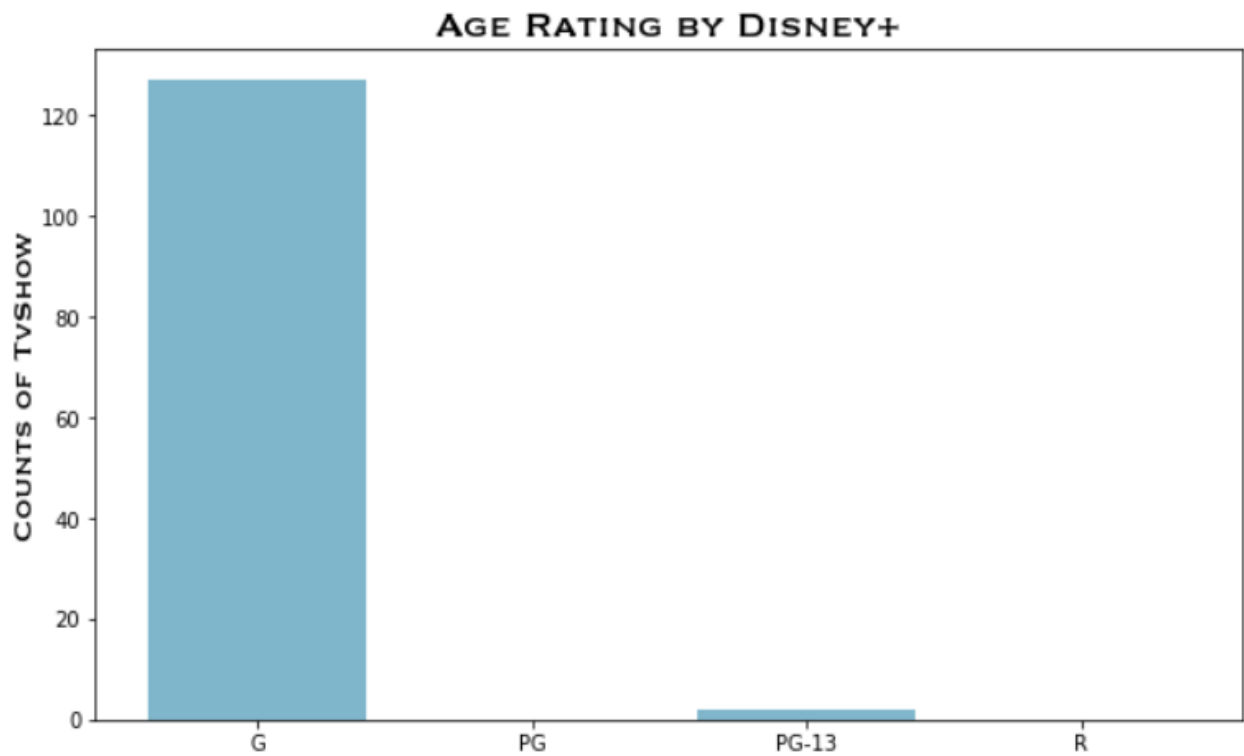
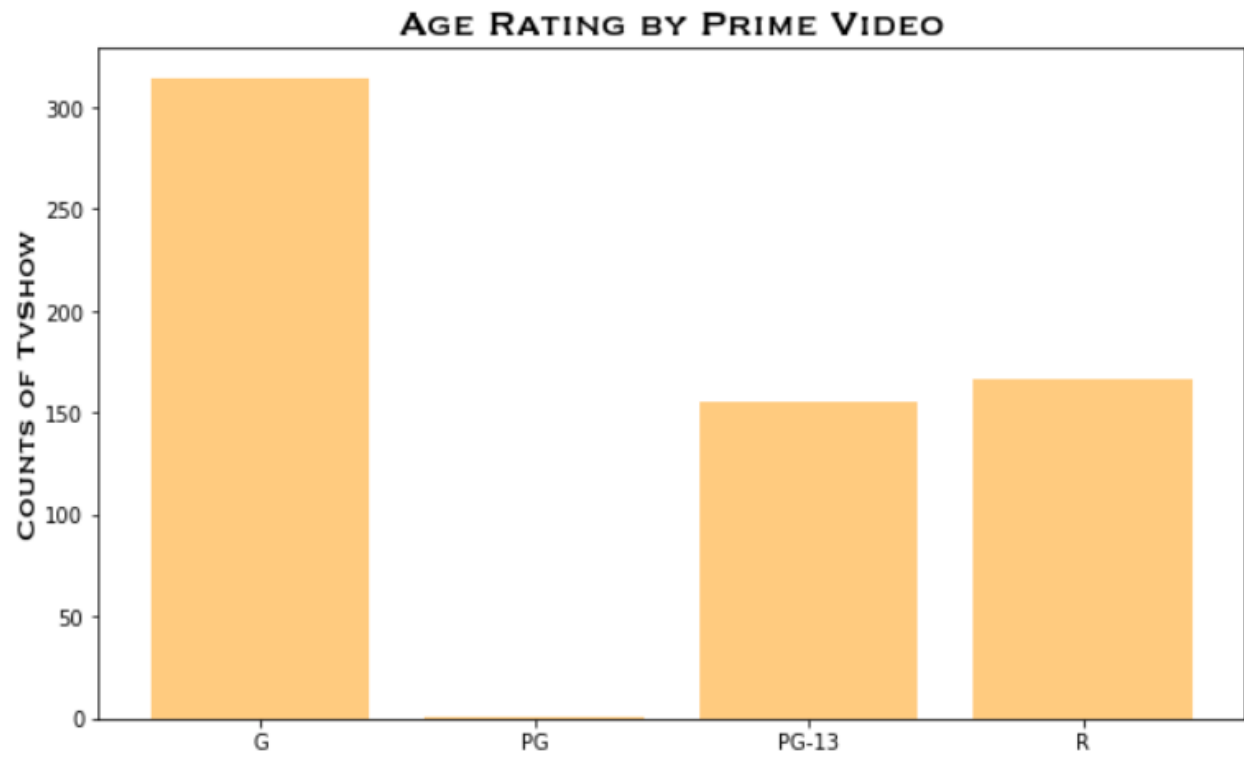
Box Plot Vs. Violin Plot – IMDb Rating Across Streaming Services for TV Shows:



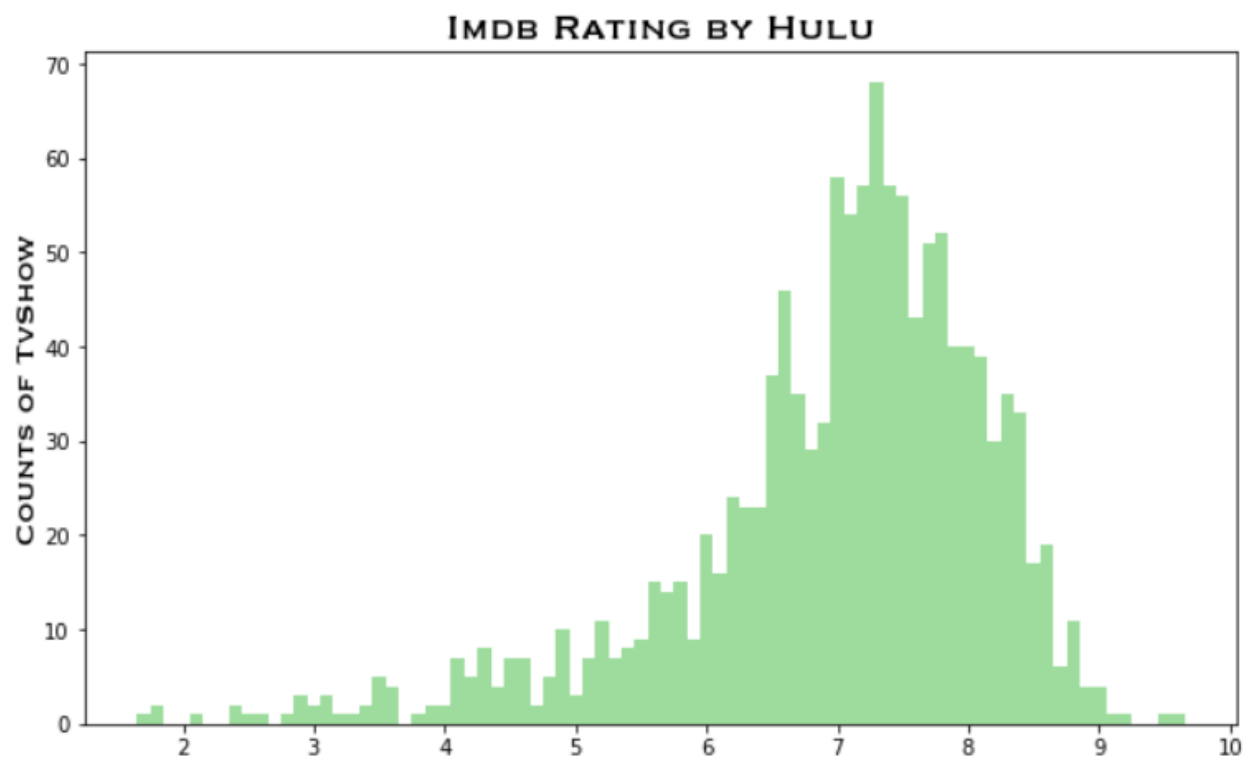
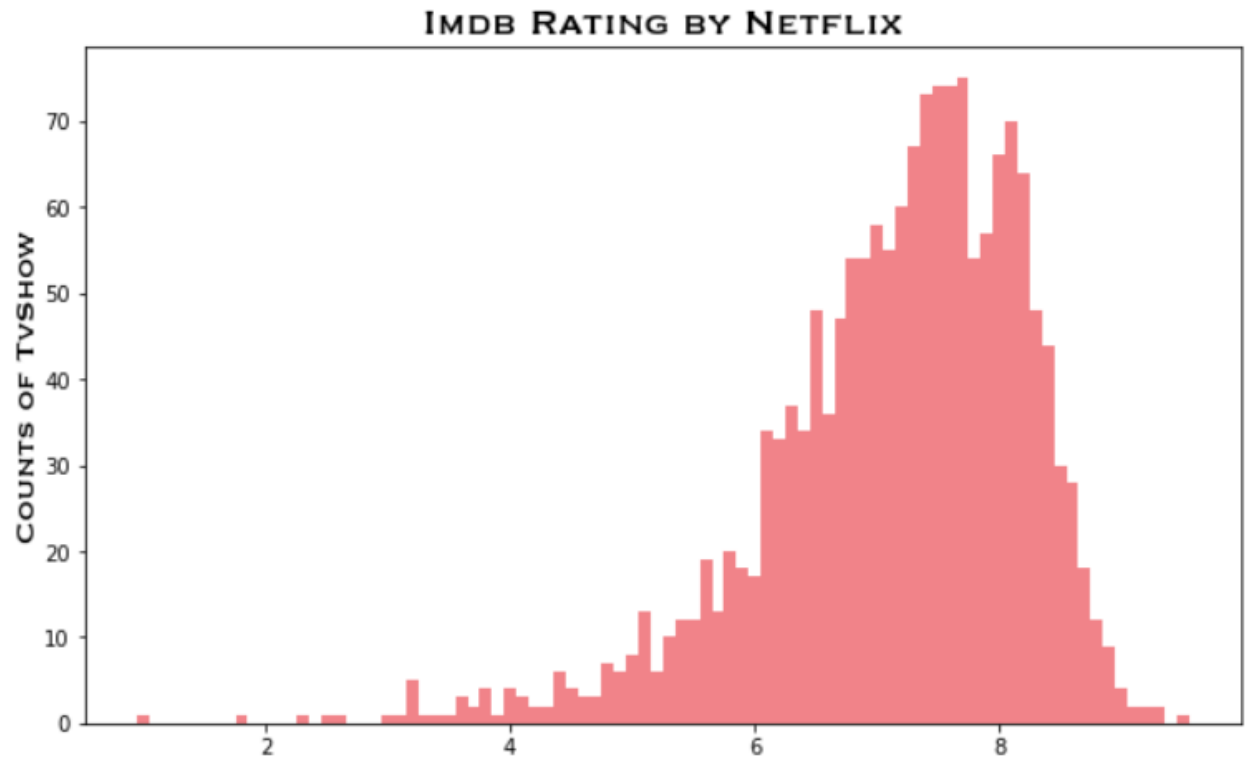
Bar Chart Between Age & Services - Figure 1-2:



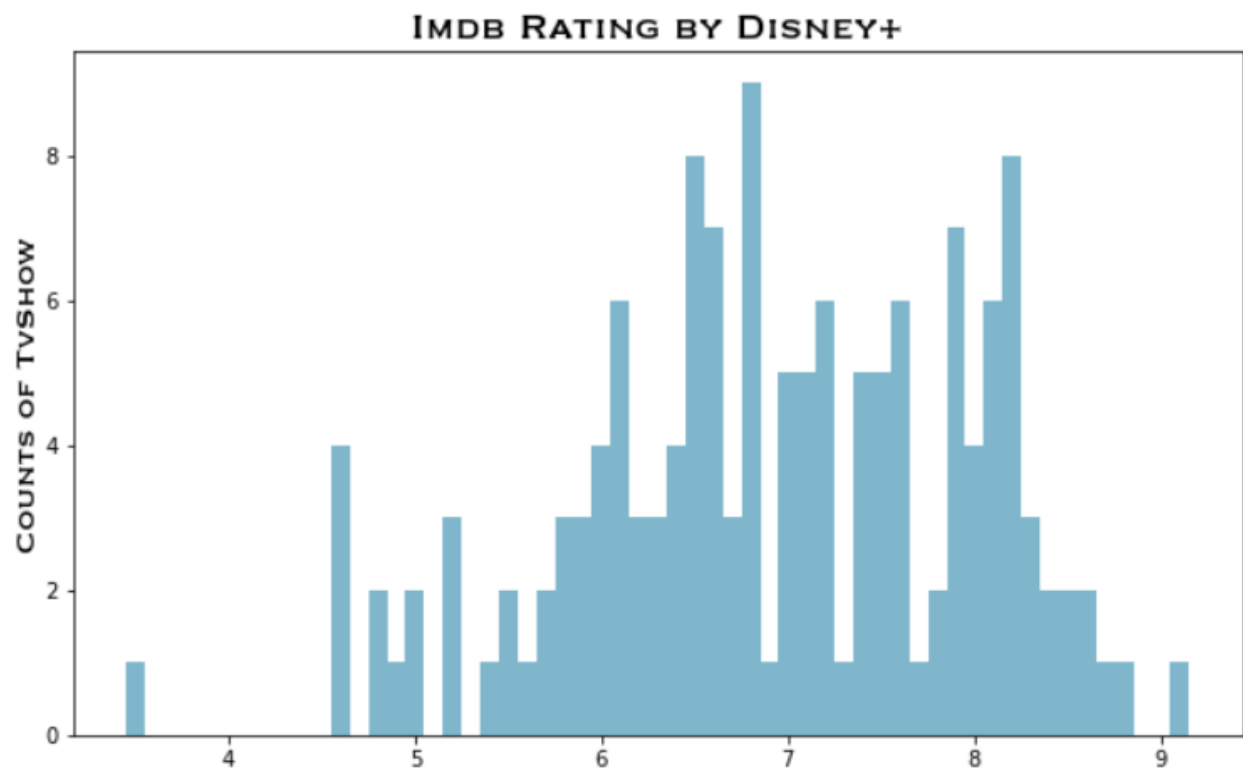
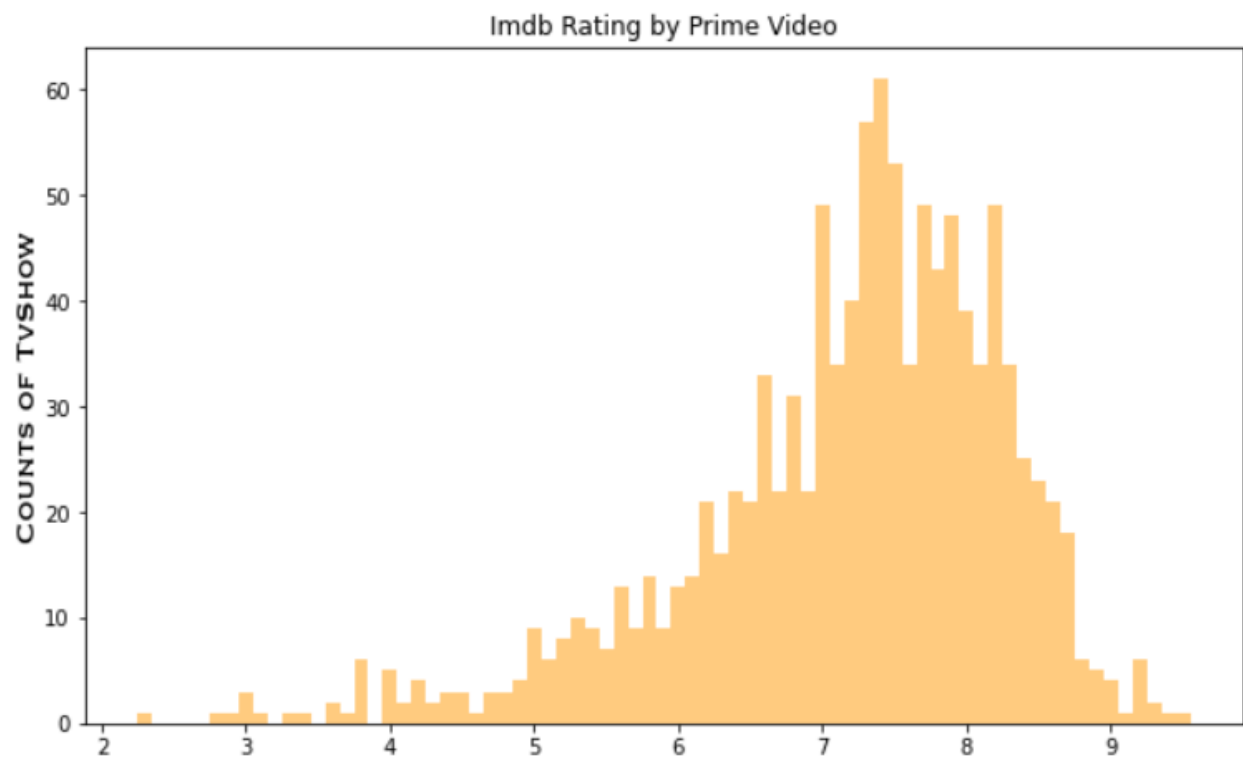
Bar Chart Between Age & Services - Figure 3-4:



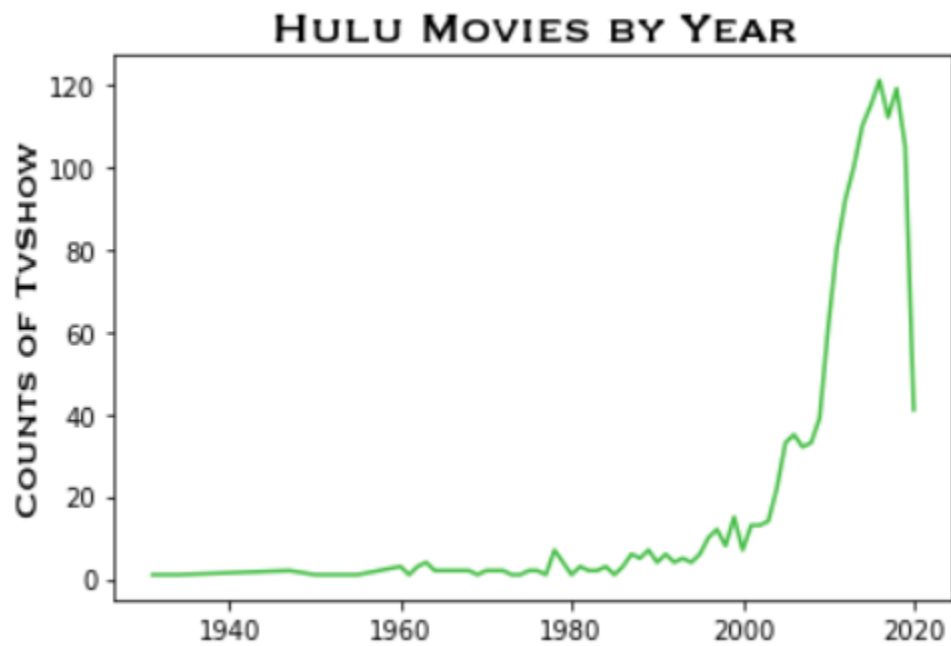
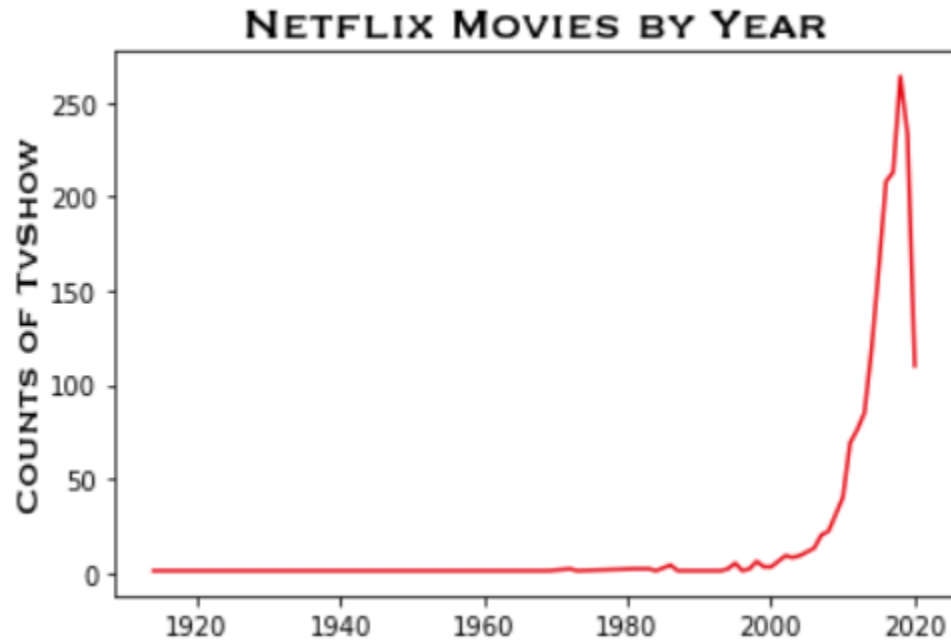
Bar Chart Between IMDb Rating & Services - Figure 1-2:



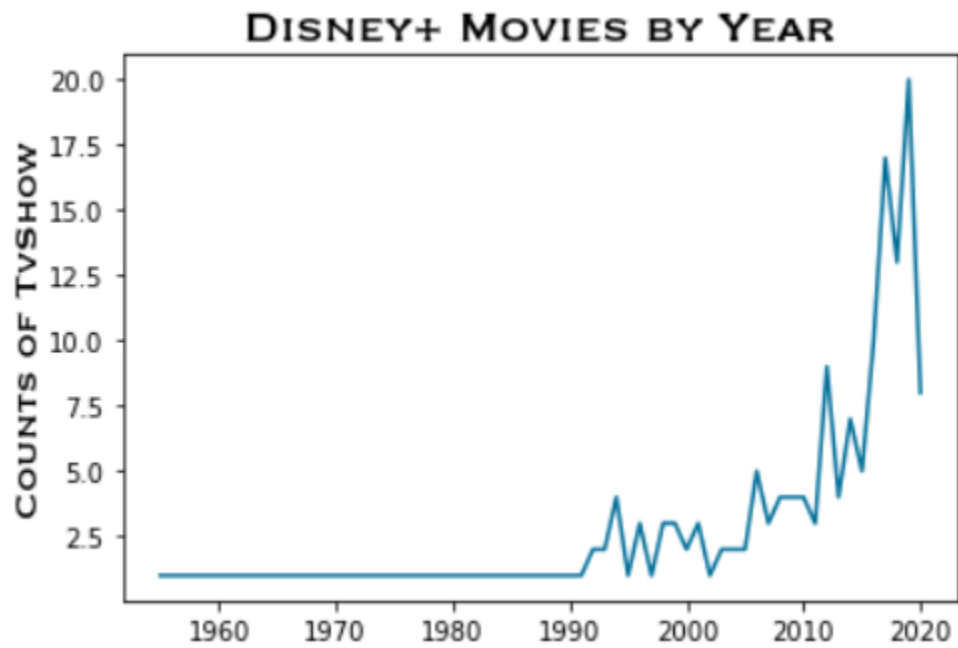
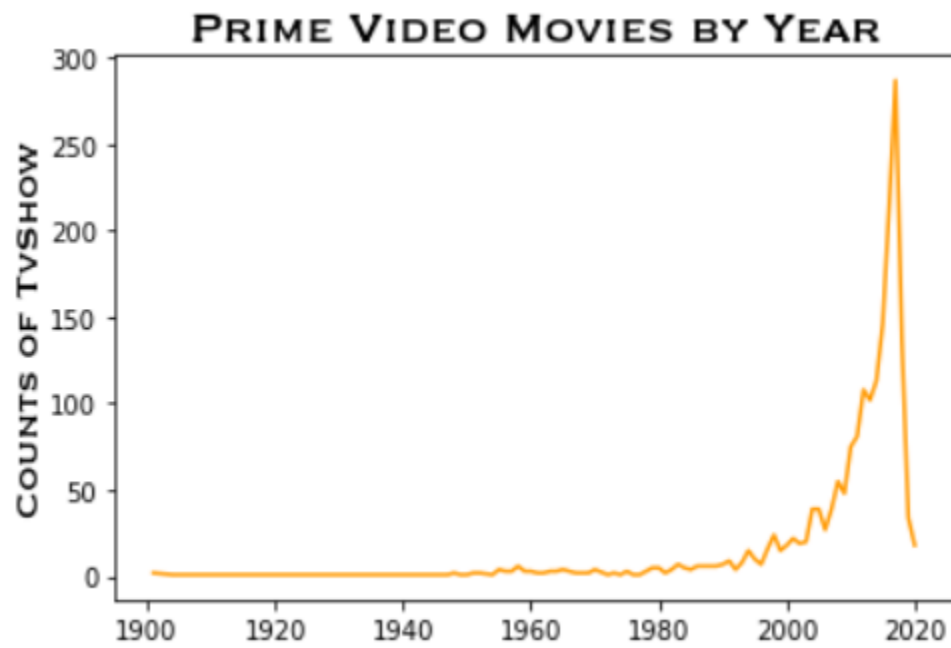
Bar Chart Between IMDb Rating & Services - Figure 3-4:



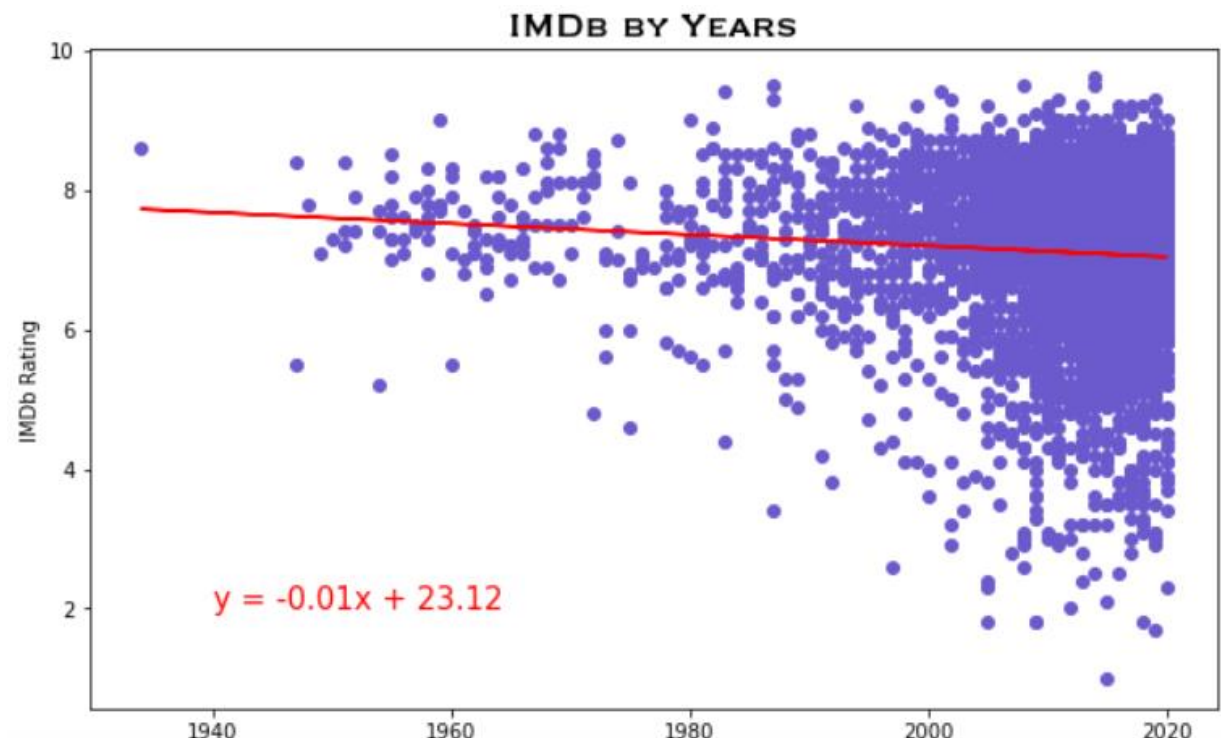
Line Graphs Year vs. Services - Figures 1-2:



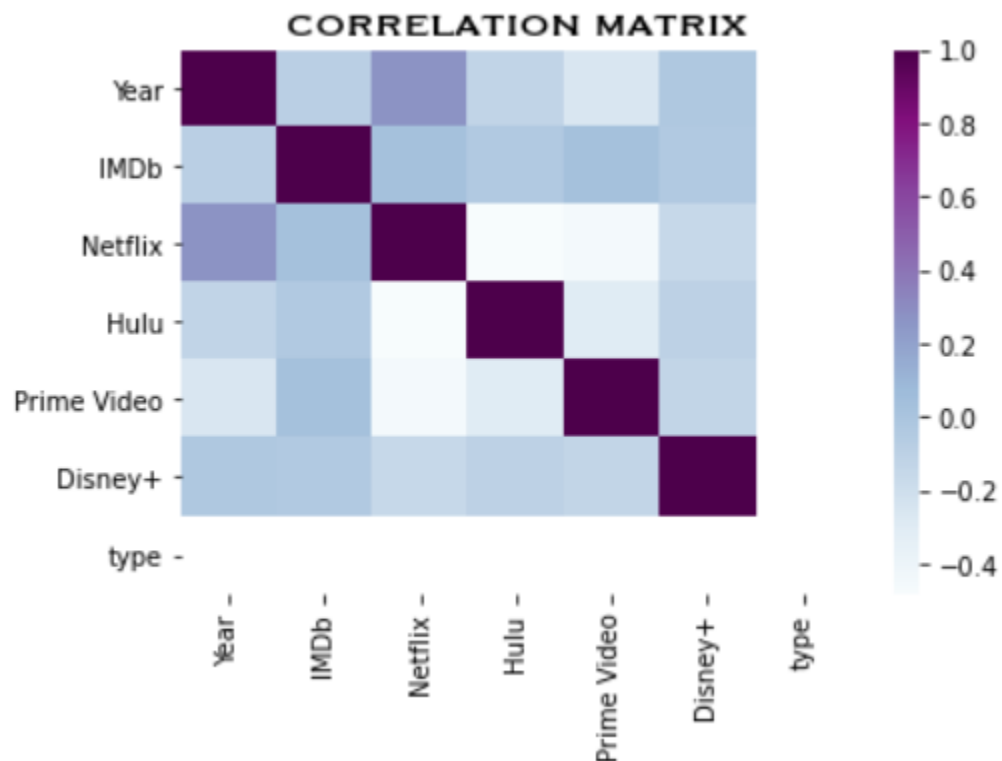
Line Graphs Year vs. Services - Figures 3-4:



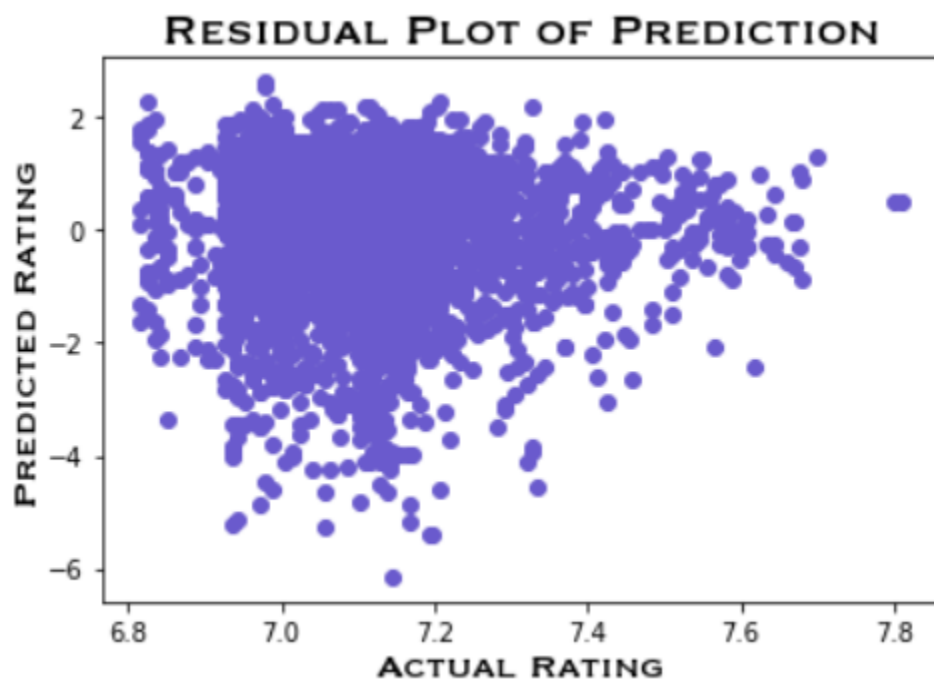
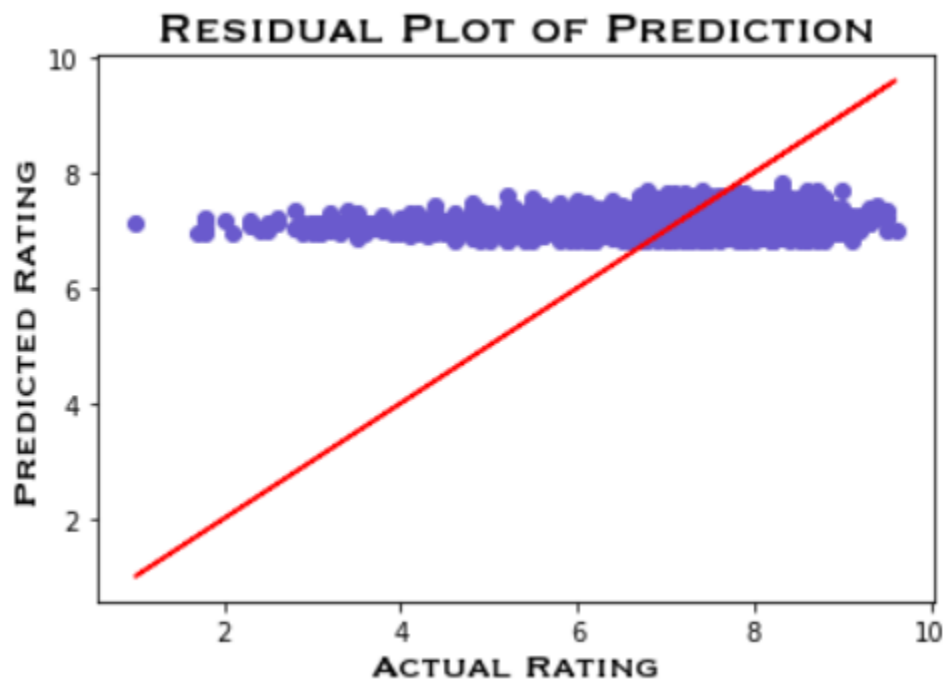
Regression for IMDb By Years:



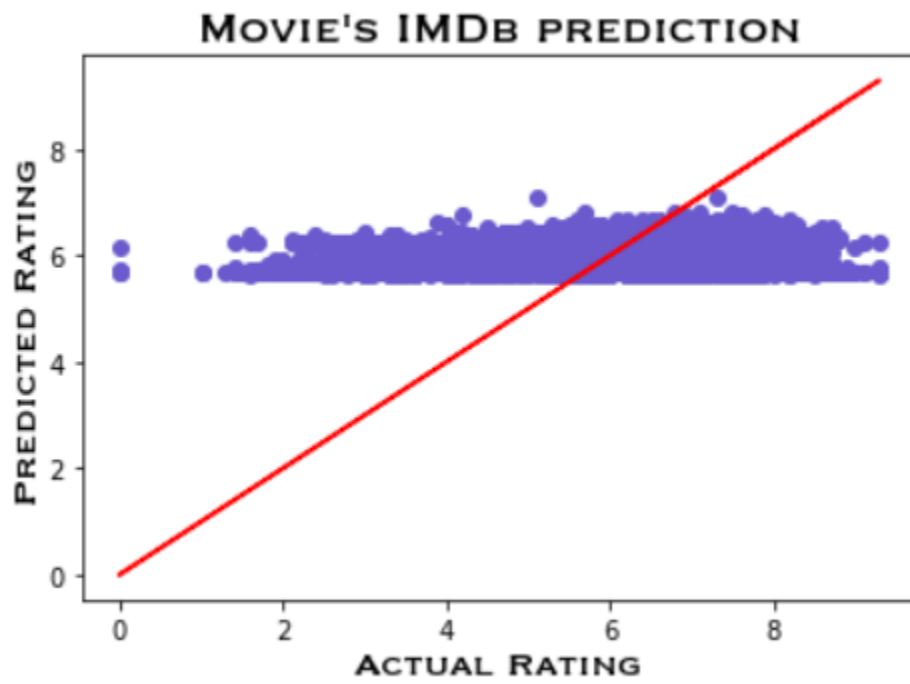
Correlation Matrix – Seaborn Heat Map Across All Four Streaming Services:



Residual Plot of Prediction:



Movie's IMDb Prediction:



References:

1. <https://www.quora.com/How-are-movie-lengths-determined#:~:text=on%20this%2C%20however.-,Most%20theaters%20will%20not%20show%20a%20film%20shorter%20than%20about,as%20a%20E2%80%9Cfeature%20film%20E2%80%9D.&text=The%20Academy%20of%20Motion%20Picture,long%20a%20movie%20can%20be.>
2. <https://www.businessofapps.com/data/netflix-statistics/>
3. <https://www.statista.com/statistics/250937/quarterly-number-of-netflix-streaming-subscribers-in-the-us/>
4. <https://help.netflix.com/en/node/14164#:~:text=Netflix%20is%20not%20yet%20available,%2C%20North%20Korea%2C%20or%20Syria.>
5. https://www.google.com/search?q=date+Disney+%2B+launched&rlz=1C1CHBF_enUS813US815&oq=date+Disney+%2B+launched&aqs=chrome..69i57j33i160.15355j1j7&sourceid=chrome&ie=UTF-8
6. https://www.google.com/search?rlz=1C1CHBF_enUS813US815&sxsrf=ALeKk01uoCw64Ev8P1oH3LtymwCesPTZsg%3A1606012744107&ei=SM-5X8-TBsGC5wK59YTQDQ&q=date+Hulu+launched&oq=date+Hulu+launched&gs_lcp=CgZwc3ktYWlQAzIGCAAQCBAeOhEIABCwAxCKAxC3AxDIhCLA1DjlQdYgggHYIavB2gBcAB4AIABR4gBgQGSAQEymAEAoAEBqgEHZ3dzLXdpesgBCrgBAsABAQ&sclient=psy-ab&ved=0ahUKEwjP1_fYj5XtAhVBwVvKHbk6AdoQ4dUDCA0&uact=5
7. https://www.google.com/search?rlz=1C1CHBF_enUS813US815&sxsrf=ALeKk00PspGfIEbTsTuLjgGw0brvuDFkwwg%3A1606012869111&ei=xc-5X7WmBoH45gKBspxQ&q=date+Prime+video+launched&oq=date+Prime+video+launched&gs_lcp=CgZwc3ktYWlQAzoRCAAQsAMQigMQtwMQ5QIQiwM6BggAEAcQHjoICAAQBxAKEB46BggAEAoQHjoICAAQBxAFEb46BggAEAUQHjoKCAAQCBAHEAoQHjoICAAQCBAHEB46BAghEApQodMDWK7vCWDv-AloAnAAeAOAAbsgiAHccZIBezIuMC4xLjIuMS4xLjAuMS4wLjOYAQCgAQGqAQdnd3Mtd2l6yAEKuAECwAEB&sclient=psy-ab&ved=0ahUKEwj1nMWUkJXtAhUBvFkKHQEZBwoQ4dUDCA0&uact=5