Streaming Services Analysis

Team Members:

Salita Santiago

Shayna Chernak

Hima Gharat

Marcellis Valentin

Table of Contents:

Introduction:

Streaming platforms are on the rise and they have taken over the way the world views TV and Movies. Nowadays, more people are paying for streaming services versus cable TV- so it is safe to say that the future of TV is leaning towards streaming services. The various streaming platforms are available to consumers through apps, smart TVs, game consoles, streaming media players, and Amazon's Fire TV. Consumers have shown that they love things that are easily accessible and not expensive. Streaming platforms have made it so that the consumers get exactly that! Netflix is the most popular online streaming websites around with over 183 million paid subscribers across the globe.

Since Netflix's creation in 1997, it is safe to say that the platform has changed greatly- what started off as a DVD rental service has transformed into a streaming platform used by millions. This growth led us to wonder about how exactly Netflix has evolved since its creation and how other streaming services compare. Hulu, Disney +, and Amazon Prime have all become Netflix's main competition. Amazon Prime has about 150 million subscribers and is the fastest growing platform due to the amount of content they offer. Hulu has focused more on the shows currently on air and had only hit 30.4 million subscribers by the end of 2019. Lastly, with Disney + only launching one year ago- they already have 60 million subscribers and counting. With these diverse streaming platforms available to us at our fingertips, we want to know what people want more of: TV shows or Movies.

Extraction:

We used 2 datasets (both CSV files) from the public platform Kaggle.com. Each dataset was based on IMDB and Rotten Tomatoes ratings for TV shows and Movies on the top 4 streaming platforms: Hulu, Netflix, Amazon Prime, and Disney+. We used Pandas functions in Jupyter Notebook to load both CSV files.

Figure 1:

```
In [1]: import pandas as pd
        from sqlalchemy import create_engine

In [2]: movie_file = (r"C:\UPenn\ETL_Project\Netflix_Data_Analysis\movies.csv")
        movie_df = pd.read_csv(movie_file)
        movie_df.head()
```

Out[2]:

	ID	Title	Year	Age	IMDb	Rotten Tomatoes	Netflix	Hulu	Prime Video	Disney+	Type	Directors	Genres	Country	Language
0	1	Inception	2010	13+	8.8	87%	1	0	0	0	0	Christopher Nolan	Action,Adventure,Sci-Fi,Thriller	United States,United Kingdom	English,Japanese,Fre
1	2	The Matrix	1999	18+	8.7	87%	1	0	0	0	0	Lana Wachowski,Lilly Wachowski	Action,Sci-Fi	United States	English
2	3	Avengers: Infinity War	2018	13+	8.5	84%	1	0	0	0	0	Anthony Russo,Joe Russo	Action,Adventure,Sci-Fi	United States	English
3	4	Back to the Future	1985	7+	8.5	96%	1	0	0	0	0	Robert Zemeckis	Adventure,Comedy,Sci-Fi	United States	English
4	5	The Good, the Bad and the Ugly	1966	18+	8.8	97%	1	0	1	0	0	Sergio Leone	Western	Italy,Spain,West Germany	Italian

Figure 2:

```
In [3]: tv_file = (r"C:\UPenn\ETL_Project\Netflix_Data_Analysis\tv_shows.csv")
        tv_df= pd.read_csv(tv_file)
        tv_df.head()
```

Out[3]:

	ID_Number	Title	Year	Age	IMDb	Rotten Tomatoes	Netflix	Hulu	Prime Video	Disney+
0	0	Breaking Bad	2008	18+	9.5	96%	1	0	0	0
1	1	Stranger Things	2016	16+	8.8	93%	1	0	0	0
2	2	Money Heist	2017	18+	8.4	91%	1	0	0	0
3	3	Sherlock	2010	16+	9.1	78%	1	0	0	0
4	4	Better Call Saul	2015	18+	8.7	97%	1	0	0	0

Transformation:

Our first steps in cleaning up the datasets involved figuring out which variables were not relevant. To get the average rating on each streaming platform, we dropped the columns Rotten Tomatoes, Type, Directors, Genres, Country, Language, Runtime, Age, ID and Year on both of the datasets. After cleaning the datasets and making it workable, we started merging.

Figure 3:

```
In [4]: #Transform Movie DataFrame
        movie_cols= ["Title", "IMDb", "Netflix", "Hulu", "Prime Video", "Disney+" ]
        movie_transformed= movie_df[movie_cols].copy()

        #Rename column headers
        movie_transformed = movie_transformed.rename(columns={"Title": "movietitle_id",
                                                              "IMDb" : "movieimdb_id",
                                                              "Netflix": "movienetflix_id",
                                                              "Hulu": "moviehulu_id",
                                                              "Prime Video": "movieprime_id",
                                                              "Disney+": "moviedisney_id"})
        movie_transformed.head()
```

Out[4]:

	movietitle_id	movieimdb_id	movienetflix_id	moviehulu_id	movieprime_id	moviedisney_id
0	Inception	8.8	1	0	0	0
1	The Matrix	8.7	1	0	0	0
2	Avengers: Infinity War	8.5	1	0	0	0
3	Back to the Future	8.5	1	0	0	0
4	The Good, the Bad and the Ugly	8.8	1	0	1	0

Figure 4:

```
In [5]: #Transform TV DataFrame
tv_cols= ["Title", "IMDb", "Netflix", "Hulu", "Prime Video", "Disney+" ]
tv_transformed= tv_df[tv_cols].copy()

#Rename column headers
tv_transformed = tv_transformed.rename(columns={"Title": "tvtitle_id",
                                                "IMDb" : "tvimdb_id",
                                                "Netflix": "tvnetflix_id",
                                                "Hulu": "tvhulu_id",
                                                "Prime Video": "tvprime_id",
                                                "Disney+": "tvdisney_id"})
tv_transformed.head()
```

Out[5]:

	tvtitle_id	tvimdb_id	tvnetflix_id	tvhulu_id	tvprime_id	tvdisney_id
0	Breaking Bad	9.5	1	0	0	0
1	Stranger Things	8.8	1	0	0	0
2	Money Heist	8.4	1	0	0	0
3	Sherlock	9.1	1	0	0	0
4	Better Call Saul	8.7	1	0	0	0

Loading of data to database:

Once data was transformed, we created a database connection to PostgreSQL. In Postgres we did our analysis by averaging out all the ratings that each television show and movie received from IMBd. We then were able to see based off of the average ratings which performs better on which streaming service and overall which form of media is preferred.

Figure 5:

21 lines (19 sloc) | 818 Bytes

```
select * from movie
select movieimdb_id, movienetflix_id, from movie

select Avg(movieimdb_id) from movie where movienetflix_id = 1;
--Netflix Movie IMDb Rating: 6.25--
select Avg(movieimdb_id) from movie where moviehulu_id = 1;
--Hulu Movie IMDb Rating: 6.14--
select Avg(movieimdb_id) from movie where movieprime_id = 1;
--Prime Movie IMDb Rating: 5.77--
select Avg(movieimdb_id) from movie where moviedisney_id = 1;
--Disney+ Movie IMDb Rating: 6.44--
select * from tv;

select Avg(tvimdb_id) from tv where tvnetflix_id = 1;
--Netflix Show IMDb Rating: 7.16--
select Avg(tvimdb_id) from tv where tvhulu_id = 1;
--Hulu Show IMDb Rating: 7.06--
select Avg(tvimdb_id) from tv where tvprime_id = 1;
--Prime Show IMDb Rating: 7.18--
select Avg(tvimdb_id) from tv where tvdisney_id = 1;
--Disney+ Show IMDb Rating: 6.92--
```

Conclusion:

According to Time Magazine, streaming platforms do not release complete viewership data to the public. For example, Netflix only releases statistics on their top ten movies and TV shows. Our way around that was to examine the ratings that viewers gave to the movies and shows available on these various platforms. We chose to compare the ratings that are listed on IMBD.com to see what was most liked on these streaming platforms: TV shows or movies.

We used these datasets so we could identify the average rating IMBD users gave movies and TV shows. The final output helped us to recognize which streaming site had higher ratings for both movies and TV shows on each platform. Based on the output, we learned that on average, TV shows had higher ratings compared to movies. This leads us to believe that the consumer has a preference for TV shows, or that viewers tend to ‘like’ (give higher ratings) to TV shows moreso than movies. With these findings, we also predict that streaming platforms are in fact focusing more on providing TV show content versus movies.

Data Sources:

https://www.kaggle.com/ruchi798/tv-shows-on-netflix-prime-video-hulu-and-disney?select=tv_shows.csv

https://www.kaggle.com/ruchi798/movies-on-netflix-prime-video-hulu-and-disney?select=MoviesOnStreamingPlatforms_updated.csv

References:

https://time.com/5697802/most-popular-shows-movies-netflix/

https://www.investopedia.com/articles/markets/051215/who-are-netflixs-main-competitors-nflx.asp