International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

Analysis of Different American Streaming Services and Shows

Abhijay Paliwal[1], Lobhas Sangai[1], Vikas Khare[2]
[1] MBA TECH III YEAR STME, NMIMS, Indore, India
[2] Associate Professor, STME, NMIMS, Indore, India

***Abstract*: United states has one of the largest Over-the-top (OTT) subscribers in the world, with more than 340 Million subscribers, much more than population of 330 Million. Demand for content is at an all-time high. Netflix leads other OTT services with more than 170 Million paying subscribers. Generating over 76 Billion USD annually, it is one of the highest revenue-generating field in country. United states has currently over 25+ OTT platforms, facing strong competition with each other. With large number of options available, high subscription rates, budget user finds it difficult to choose correct OTT platform to subscribe. Many factors such as showing wrong recommendations/content, misunderstanding audiences need, high prices, content challenge and discovery etc. are some of the challenges OTT platforms are facing today.**

Keywords: Shows, movie, OTT, ratings, revenue

1. INTRODUCTION-

Mainly users choose platform which has high rated shows, popularity etc. this does not mean that their demands get fulfilled. Users have different show preferences and sometimes they did not find the shows they are looking for, eventually wasting both money and time. Proper analysis of different OTT platforms plays a major role in selecting correct platform as per user's requirements US entertainment market consist of over 200 companies, which comprises of satellite broadcast, IPTV and Cable TV. Company with largest number of satellite and cable tv broadcast is comcast, with 21.2 million customers. US market has seen drastic change in resent decade, with introduction of Netflix and prime video streaming services, it led to customer turnover of many broadcasting services. In 2017, Netflix surpassed 50 Million subscribers, surpassing total sum of both comcast and DirecTV subscribers. In recent years, customers of cable tv is declining as more and more customers are shifting to online streaming services.

In latest years, the appearance of diverse OTT systems has brought a singular issue: the problem in deciding on which OTT platform to subscribe to. Netflix, Amazon Prime, hulu and Disney+ are a number of the various OTT offerings which might be famous to the public [4]. These systems are arising with new approaches to face out amongst competition through imparting authentic content material, it's miles obtrusive that extra clients are being misplaced in finding out which platform could be appropriate for his or her use. Moreover, maximum of the to be had advice structures are targeted on suggesting the content material however now no longer the systems that maintain and offer the ones contents [6]. To ease the selection dilemma, our take a look at ambitions to give a tenet for deciding on the suitable OTT platform that suits one's non-public preferences. Lots of researchers already paintings in the direction of this area and in line with Reddy et al, on advice structures primarily based totally on content material-primarily based totally filtering suggests, film metadata, in particular style and rating, are key determinants in predicting what a consumer might also additionally need to look at withinside the future [7].

Vidooly et al [2] analysed that younger audiences between 24–34 years of age prefer the romance genre, whereas drama, crime thriller and reality shows are consistent among all age groups. At the age of 34, interest in the romance genre begins to decline, and the focus is on action and adventure films. Documentaries capture the attention of viewers aged 24 to 45. According to the survey, there is a difference in content consumption between the elderly and young people. Adult audiences prefer movies, youngsters prefer series. As per KPMG et al analysis [3], The business models opted by the services is also very much important, Nowadays many OTT providers provide initial free trial when a user signs upon their platform and demand incremental premium fee at a later stage after consumer behaviour is in favour. So, the business model that is prevalently employed by OTT providers around the globe is B2C. Shin et al suggested that OTT providers should work on making high-quality content that can compete with the material that is available in OTT.

Sant Singh et al[9] suggested that the Creators want to assume some thing out of container that is innovative and innovative. The maximum crucial process for creators is to create mindboggling content material which makes customers hold to binge watch on numerous OTT platforms. The opposition in this marketplace section will growth in destiny so entrepreneurs ought to awareness on imposing accurate techniques on the proper vicinity and time. Minzheong et al [8] analysed that Collaborations with other not so known platforms and media houses from foreign countries also increase their user base and popularity. For example, in 2019, CJ ENM, a Korean entertainment company, and its subsidiaries Studio Dragon and Netflix have signed a strategic partnership through a multi-year content production and distribution contract. As part of a three-year partnership that began in 2020, Studio Dragon and its creators have released an original series available to Netflix members worldwide.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

METHODOLOGY-

We have used python language libraries in this analysis, Scikit-learn (previously scikits.learn and additionally called sklearn) is a unfastened software program gadget getting to know library for the Python programming language. It capabilities diverse classification, regression and clustering algorithms together with support-vector machines, random forests, gradient boosting, k-approach and DBSCAN, and is designed to interoperate with the Python numerical and libraries NumPy and SciPy. Matplotlib is a visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform records visualization library constructed on NumPy arrays and designed with the wider SciPy stack. Word Cloud is a records visualization method used for representing textual content records wherein the scale of every phrase shows its frequency or importance. Significant textual records factors may be highlighted the usage of a phrase cloud. Word clouds are broadly used for analysing records from social community websites. pandas is a Python package deal imparting fast, flexible, and expressive records systems designed to make operating with "relational" or "labeled" records each smooth and intuitive. It pursuits to be the essential high-stage constructing block for doing practical, real-global records evaluation in Python. Coming to streaming sites, we have used top 4 most popular streaming sites of united states, namely Netflix, Amazon prime, Hulu, and Disney+ and sorted them by the shows which they offer. As some titles are available on one or more platforms, we have binarized the field with 1 as available and 0 as unavailable.

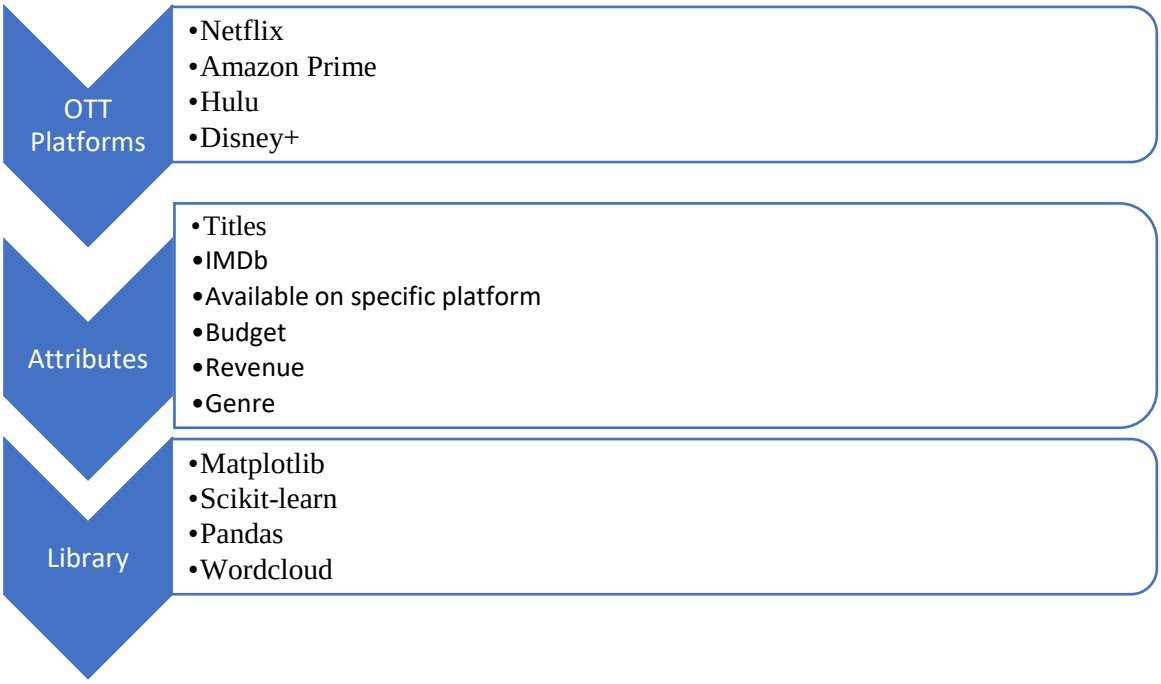

Fig 1: Different methodologies used in analysis

ID	Title	Year	Age	IMDb	Rotten Tomatoes	Netflix	Hulu	Prime video	Disney+	Type	Budget(Million Dollars)	Revenue (Million dollars)	Genre
1	Breaking Bad	2008	18	9.4/10	100/100	1	0	0	0	1	195	250	Drama, Dark comedy, Suspense, Crime film, Thriller genre
2	Stranger Things	2016	16	8.7/10	96/100	1	0	0	0	1	300	378	Supernatural, Horror, Drama, Science fiction, Horror fiction, Historical film
3	Attack on Titan	2013	18	9.0/10	95/100	1	1	0	0	1	375	201	Action fiction, Dark fantasy, Apocalyptic and post Apocalyptic fiction, Drama
4	Better Call Saul	2015	18	8.8/10	94/100	1	0	0	0	1	150	180	Comedy, Drama, Tragedy, Dark comedy, Crime film, Legal drama, Crime TV genre

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

5	Dark	2017	18	8.8/10	93/100	1	0	0	0	1	195	260	Thriller, Drama, Science fiction, Mystery, Supernatural
6	Avatar: The Last Airbender	2005	7	9.3/10	93/100	1	0	1	0	1	150	319	Animated series, Comedy, Adventure, Action fiction, Animation

Table 1: Data on which analysis has been performed

In Table 1, Title column represents title of the show, year represents the release year of the show, Age represents the certification age rating of the movie, IMDb represents the IMDb rating, Rotten tomatoes represent rotten tomatoes rating of the movie, Netflix, Hulu, prime video and Disney+ columns represents the platforms where show are present, Budget column represents the total budget of movie in million dollars, revenue column represents the revenue generated for the movie and genre column represents which genre the movie belongs to.

ANALYSIS-

Based on the dataset, following analysis is performed to get insights from data and to answer the questions which people have on streaming services and shows

1. **Number of shows available across platforms comparing to the age ratings-**

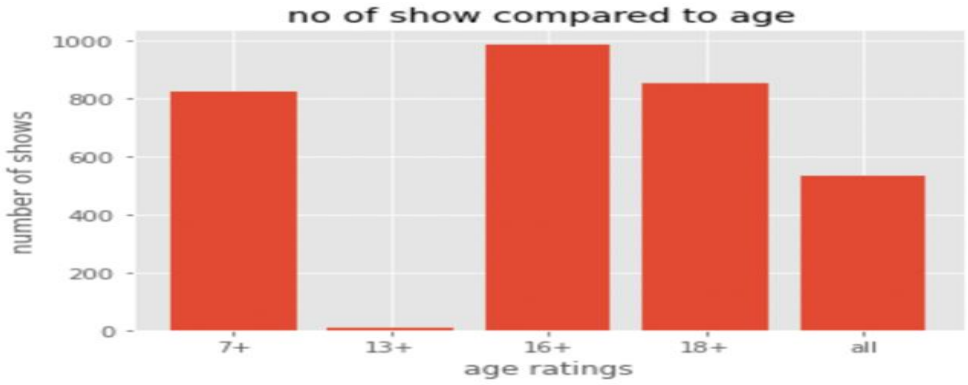

Fig 2: Number of shows compared to age ratings

- According to fig.2, about 25.6 % lies on shows which have 7+ rating
- 0.2% data lies on 13+ shows
- 30.7% data lies on 16+ shows
- 26.3% data lies on 18+ shows
- 17% data have all age ratings
- This shows that the mature shows dominate the American films industry

2. **High IMDb and Low IMDb rated shows present in the data-**

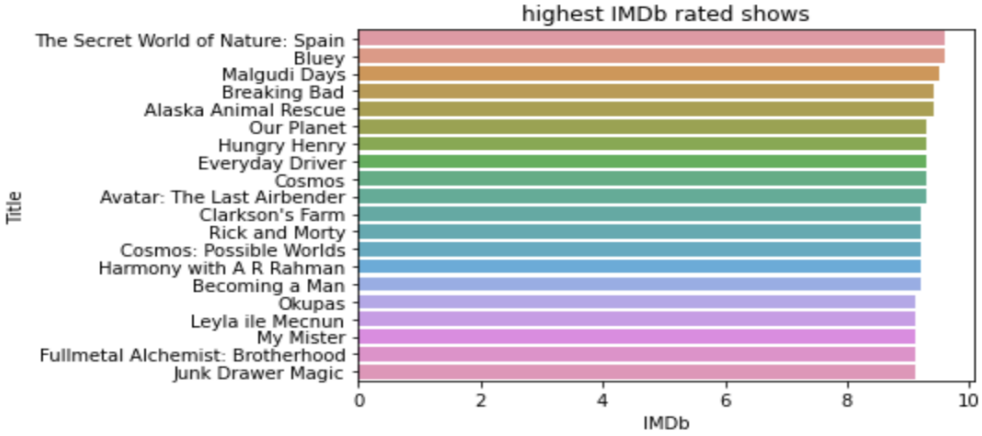

Fig 3: Titles of highest IMDb rated shows and their ratings

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

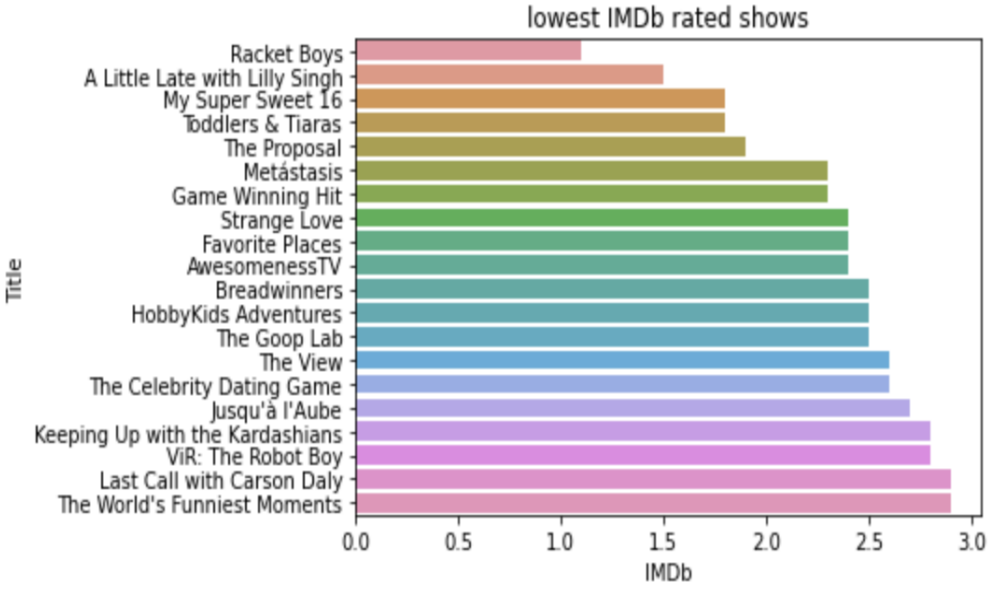

Fig 4: Titles of lowest IMDb rated shows and their ratings

2nd analysis is about highest imdb rating and lowest imdb rated shows from all streaming platforms. in the fig 3 and fig.4, the show ‘ racket boys’ has got the lowest imdb rating that is 1.6 followed by a little late with lilly singh, my super sweet 16 and many more. Also ‘the secret world of nature spain’ has got the highest imdb rating that is 9.9 followed by bluey, breaking bad etc.

3. Is IMDb ratings and rotten tomatoes are connected? -

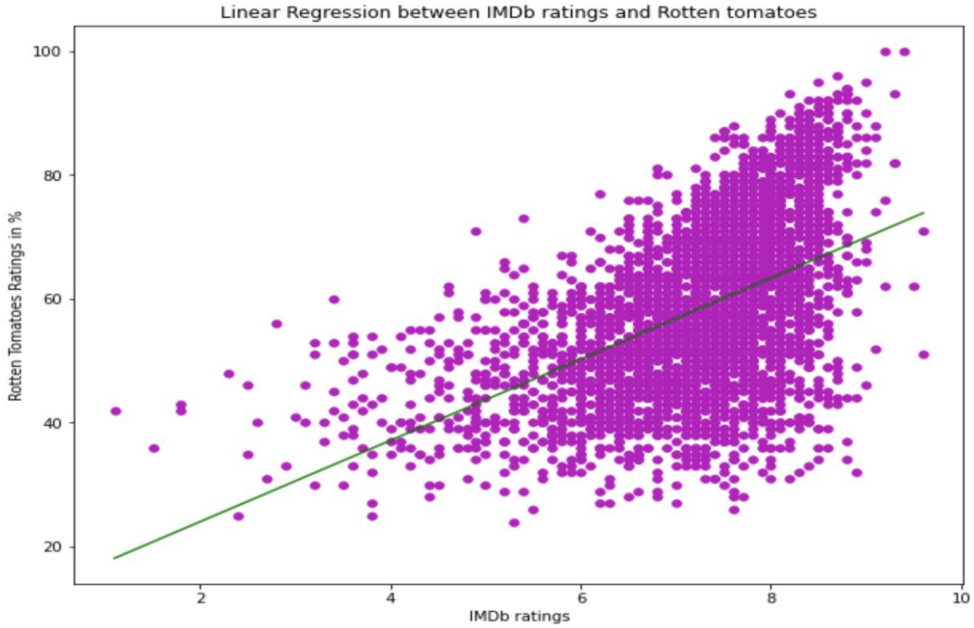

Fig 5: Linear regression plot between IMDb ratings and rotten tomatoes ratings

Most of the audience perceive that IMDb and rotten tomatoes ratings are similar, but according to our analysis, we can see that the regression line is not perfectly linear, which shows that if IMDb rating is changed by a unit, then rotten tomatoes rating does not change by same unit. Hence the rotten tomatoes rating and IMDb does not have a linear relationship.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

4. Which streaming site contains high IMDb rated shows?

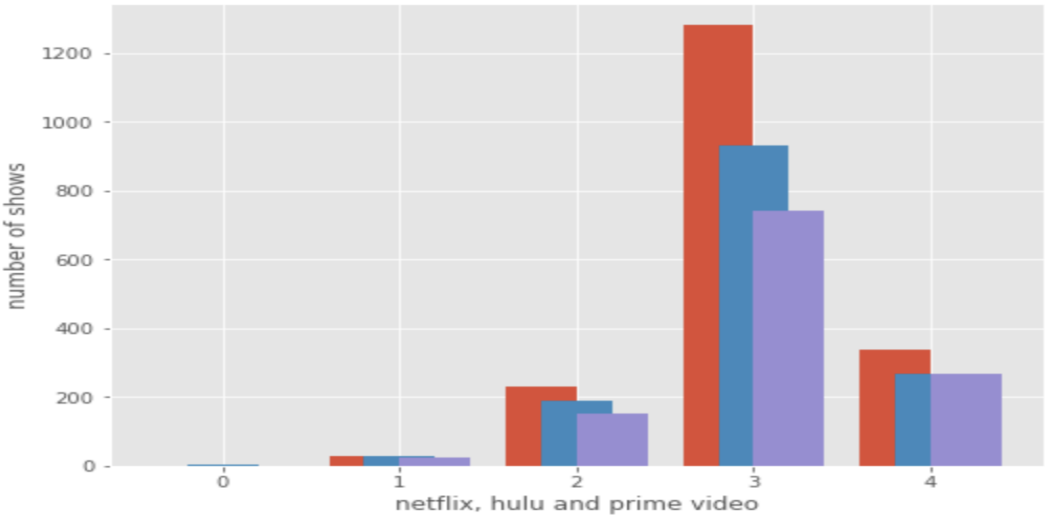

Fig 6: Total number of shows on OTT platforms and their IMDb ratings ranges

In fig.6, on the X axis, 0 defines the IMDb ratings between 0 and 2, 1 represent rating between 2 and 4, 2 between 4 and 6, 3 between 6 and 8 and 4 above 8.
The red graph represents Netflix, Blue represent Hulu and violet represent prime video. Most of the users prefer the streaming site which has high number of shows. Also, with high IMDb rated shows, we can clearly see Netflix lead them all with more than 1250+ shows having IMDb ratings between 6 and 8 and 300+ shows which have more than 8+ IMDb ratings, followed by Hulu, another American streaming platform with 850+ between 6 and 8 and 300+ above 8 IMDb ratings respectively. This shows that Netflix provides high amounts of quality shows to their subscribers with more reasonable rates.

5. Correlation between different stats

a. Is IMDb and Rotten Tomatoes correlated?

```
from scipy.stats import pearsonr
corr, _ = pearsonr(data["IMDb"], data["Rotten Tomatoes"])
print('Pearsons correlation(R): %.3f' % corr)
```

```
Pearsons correlation(R): 0.517
```

from above data, the correlation coefficient is very low, so they are not related.

b. Does high IMDb rated shows are present in Netflix?

```
from scipy.stats import pearsonr
corr, _ = pearsonr(data["IMDb"], data["Netflix"])
print('Pearsons correlation(R): %.3f' % corr)
```

```
Pearsons correlation(R): 0.016
```

from above data, the correlation coefficient is very low, so they are not related.

The next analysis is about the correlation, to explain this we have taken some examples like, is IMDb ratings and rotten tomatoes rating are correlated to each other i.e. if IMDb rating is higher then rotten tomatoes is also higher or does high IMDb rating shows are always present on Netflix, Hulu etc. so Pearson correlation coefficient is used in python to analyse correlation coefficient of the specific data and we observed that the correlation coefficient for question “Is IMDb and Rotten Tomatoes correlated?” found

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

to be very low about only 0.517, and also the correlation coefficient of question “Does high IMDb rated shows are present in Netflix” has also found to be very low i.e. about 0.016 only so in conclusion, the IMDb rating and rotten tomato rating, is not related and also shows present on Netflix are not always of high IMDb rating.

6. Wordcloud for the common titles present in data

Fig 7: Word cloud image of common titles of shows present in the data

Fig.7 is the wordcloud analysis of the most common titles present in the data, according to this, a Word Cloud provides the ability to analyse any text quickly and depicts valuable information just by visualizing. In Fig.7- the words “Love”, “world” and “show” occur most of the time in titles and are very common, so it occurs as a big words in the image, and followed by the words like marvel, adventure, girl, life, city etc. as lower no of occurrences.

7. Common words present in Titles

a. Common words in all of titles present-

```
print("most common words are ",filtered_list)
```

```
most common words are  [['Love', 40], ['World', 36], ['Show', 36], ['Life', 35], ['Little', 23], ['House', 23], ['A
merican', 23]]
```

b. Common words present In shows streaming on Netflix-

```
[['Love', 21], ['Show', 21], ['World', 16], ['Little', 15], ['American', 14], ['New', 14], ['Life', 14], ['Me', 13]
]
```

c. Common words present In shows streaming on Hulu-

```
[['Love', 46], ['Life', 19], ['Little', 17], ['Black', 16], ['Girls', 16], ['Last', 14], ['Show', 14], ['House', 13
], ['Next', 12], ['Time', 12], ['City', 12], ['Stories', 12], ['Story', 11]]
```

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

d. Common words present In shows streaming on Prime Video-

```
[['Star', 15], ["Marvel's", 12], ['Wild', 12], ['Disney', 10], ['Wars:', 9], ['World', 8], ['Marvel', 8], ['Adventu
res', 8], ['Secrets', 8], ['Animal', 8], ['Mickey', 7], ['Life', 6], ['Wars', 6]]
```

e. Common words Present in Shows streaming on Disney+

```
[['World', 30], ['Love', 26], ['Life', 23], ['City', 21], ['Show', 21], ['War', 20], ['Adventures', 17]]
```

filtering operation among all the titles of the shows present in the data is performed to know what the most used word in the titles of American shows is. Titles like “love”, “life” and “little” are most used words. In Netflix, “Love”, “world”, “show” is most used. In Hulu, “love”, “show” and “world” are common, In Prime Video,” world”, “love” and “life” are common and in Disney+, “star”, “marvel’s” and “wild” were most common words.

Title tells most about the movie, first thing a customer knows is the title of the movie, a catchy title makes a movie unique

8. Availability on different platforms-

a. Netflix-

```
#decision tree on nflx
feature_cols = ['Movie_type','certificate_rating_movie','IMDb_type','rotten_tomato_type']
X = data_new[feature_cols]
y= data_new['available_on_netflix']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1)

clf = DecisionTreeClassifier()

# Train Decision Tree Classifer
clf = clf.fit(X_train,y_train)

#Predict the response for test dataset
y_pred = clf.predict(X_test)

print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
```

```
Accuracy: 0.9169397889509622
```

b. Hulu-

```
#decision tree on hulu
feature_cols = ['Movie_type','certificate_rating_movie','IMDb_type','rotten_tomato_type']
X = data_new[feature_cols]
y= data_new['available_on_hulu']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1)

clf = DecisionTreeClassifier()

# Train Decision Tree Classifer
clf = clf.fit(X_train,y_train)

#Predict the response for test dataset
y_pred = clf.predict(X_test)

print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
```

```
Accuracy: 0.8487585350713842
```

decision tree model is made to predict if the columns “Movie-type”, “Certificate_rating_movie”, “IMDb type” and “Rotten tomato type” were given to model, could it predict whether the movie would belong to Netflix or not. Here the model got accuracy of 91.6% and same was done with hulu, where we got accuracy as 84.8%

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

This model is useful for the audiences who could not subscribe to all the streaming sites and wants to subscribe a site which fulfills their likes of a movie

9. Which movie genre generates high revenue?

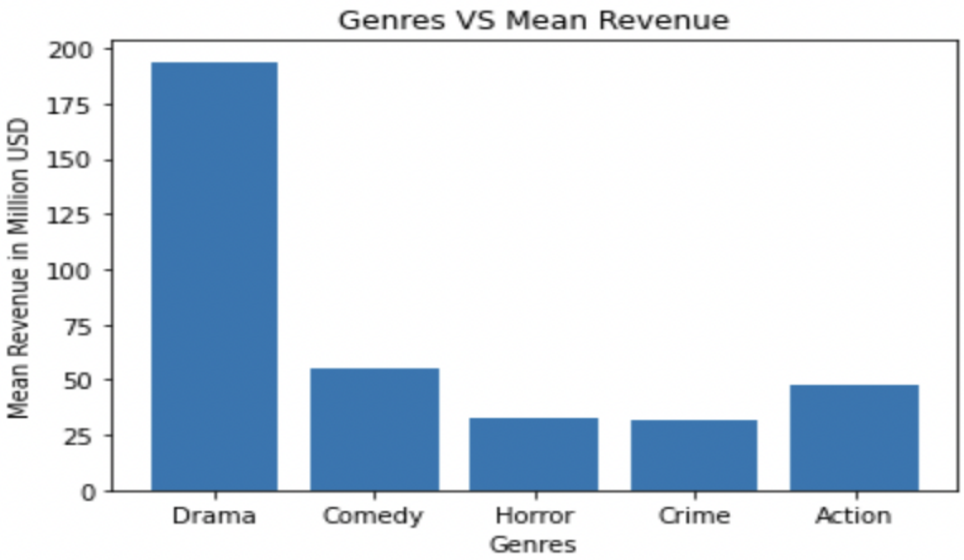

Fig. 8: Movie and shows genres compared to mean revenue

According to fig. 8, Drama movies generated very high revenue with mean of USD 193 Million, followed by Comedy with USD 55 Million and Action with USD 47 Million.

This analysis is about which type of movies generates high revenue, for analysis, genres of the movies are taken on x axis and their mean revenues on y axis. From fig. 8, it is clearly identifiable that in general the dramatical type of movies generates the most revenue like about 193 million us dollars. The action and comedy movies come after that generating about 50 million us dollars as revenue. And last came the crime and horror shows.

This graph tells that "Drama" movies are mostly likes by the audiences and for a producer, producing a drama movie would probably be more profitable for them.

10. Does high Budget means High revenue?

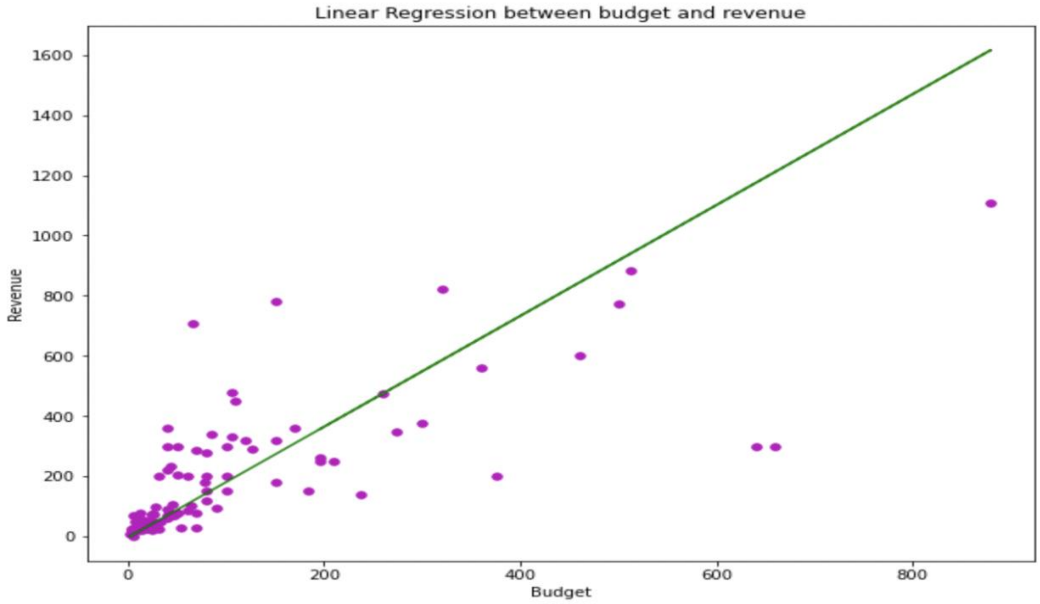

Fig. 9: Linear regression between Budget of shows and movies

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

For this, we have used linear regression and plotted a graph for it. From fig.9, It is seen that Budget and Revenue follow linear relationship, so it could be said that if Budget of movie is high, then chances are its revenue would be high.
This is the analysis to identify that the high budget means high revenue –, it is seen that Budget and Revenue follow linear relationship, also if number of episodes in a particular movie/series is high then the watch time for the audience becomes high, so revenue would be higher. could be said that if Budget of movie is high, then chances are its revenue would be high.
Also to plot the linear regression, we have removed the outliers present in the graph for more clear analysis

11. Is High Budget and High revenue means high Rating?

We have considered high IMDb rating as >7 and high revenue >100

a. does high budget means high rating?

```
#does high budget means high rating?
count = 0
for i in range(100):
    if data_new['IMDb'][i] >= 7.0 and data_updated['Budget (million dollar)'][i] >= 100:
        count +=1

print(count,"% movies have high budget and High Rating")
```

```
29 % movies have high budget and High Revenue
```

According to our analysis, 29% of movies have high budget and high revenue, so we cannot conclude that high budget movies does not mean high revenue generated for the movie

b. Does high revenue means high rating?

```
#does high revenue means high rating?
count = 0
for i in range(100):
    if data_new['IMDb'][i] >= 7.0 and data_updated['Revenue (million dollar)'][i] >= 100:
        count +=1

print(count,"% movies have high revenue and High Rating")
```

```
48
```

According to our analysis, 48% of movies have high revenue and high rating, so it is concluded that high revenue movies does not mean high ratings for the movie

CONCLUSION-

Based on all the analysis performed above, we have answered many common questions and misconceptions which people have about the shows, OTT platforms and cinema ratings. From our analysis, we conclude that Netflix is the best OTT platform among all its competitors till date, having high show choices and high rating movies across genres. Also, it is the oldest platform among all competitors. “Love”, “world” and ”show” are the most common words present in the show titles across all OTT platforms, which shows the likings of the user. For clearing common misconceptions, in our analysis, IMDb and rotten tomatoes rating are not related. Also it does not means that high budget, revenue means high rating for the movie.

REFERENCES:

[1] Sundaravel, E., and N. Elangovan. "Emergence and future of Over-the-top (OTT) video services in India: an analytical research."
[2] Vidooly (2019). Indian OTT Landscape: A Survey Report by Vidooly. Retrieved January 18, 2020 from https://vidooly.com/blog/wpcontent/uploads/2019/12/Indian-OTT-Landscape- Report-FREE-SAMPLE-1.pdf
[3] KPMG (2019). India’s Digital Future. Retrieved from https://assets.kpmg/content/dam/kpmg/in/pdf/2019/08/india-media-entertainment- report-2019.pdf.
[4] Gevers, A. (2019). Video Streaming in India: Hot hothot! Comscore, Inc. Retrieved February 3, 2020, from https://www.comscore.com/Insights/Blog/Video-Streaming-in-India .
[5] Hutchins, B. Li, B. and Rowe, D. (2019). Over-the-top sport: live streaming services, changing coverage rights markets and the growth of media sport portals. Media, Culture & Society, 41 (7),975-994. https://doi.org/10.1177/0163443719857623
[6] Matrix, S. (2014). The Netflix effect: Teens, binge-watching, and on-demand digital media trends. Jeunesse: Young People, Texts, Cultures, 6(1), 119-138. https://doi.org/10.1353/jeu.2014.0002

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

[7] Zakrzewski, C. (2015). Apple's Tim Cook: "We Believe the Future of TV Is Apps." from https://blogs.wsj.com/personal-technology/2015/09/09/apples-tim- cook-we-believe-the-future-of-tv-is-apps/
[8] Minzheong, Song. "A Case Study on Partnership Types between Network Operators & Netflix: Based on Corporate Investment Model." *International Journal of Internet, Broadcasting and Communication* 12.1 (2020): 14-26.
[9] Dasgupta, Dr, and Dr Grover. "Understanding adoption factors of over-the-top video services among millennial consumers." *International Journal of Computer Engineering and Technology* 10.1 (2019).