Amitabh S Virk

AGENDA

01. CONTEXT & OBJECTIVE

02. ABOUT THE TEAM

03. ANALYSIS

04. RECOMMENDATIONS & VISUALIZATIONS

WHY DATA?

1. **In 2021, 65.1% of the Population has access to internet compared to 0.4% in the 1995**

2. **This rise has led increase in the Data Generation and by 2022 it is expected to rise to 400EB(exabyte) of data produced per month.**
   (1EB = 1000000000GB)

3. **How Hollywood is utilizing Data Analysis**
   a. **Determining Production Costs**
   b. **Targeted Marketing Campaigns**
   c. **Recommendation engines**

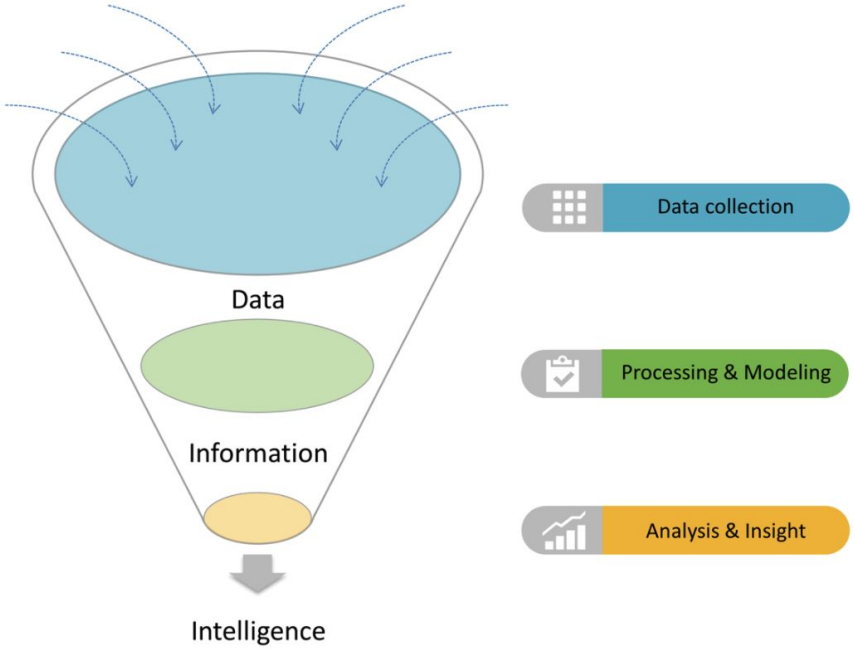
Big Time Hollywood Production Company, is interested in producing a bunch of movies. Due to the rise seen in recent years in the popularity of Hollywood movies across the globe, they are planning on expanding their horizons outside the US. They have been quite successful in the past and want to keep the winning streak alive along with expansion of the Business.

And this is where UOP Analytics Inc, comes into play.

QUESTIONS WE ARE TRYING TO ANSWER:

1. **How does this data relate to the business and how does it justify our objective?**
2. **What factors affect the revenues made by previous films?**
3. **Does the budget of the movie affect the success of the movie?**
4. **What factors lead to the not so good ratings and good ratings?**
5. **What effects does the geography of release have on the revenue numbers or what areas are the most profitable for the movies?**
6. **What genres are most popular ?**
7. **What affects the Revenue most, critics ratings or IMDB ratings?**
8. **Who is best at predicting the Revenue and ratings of the movie?**

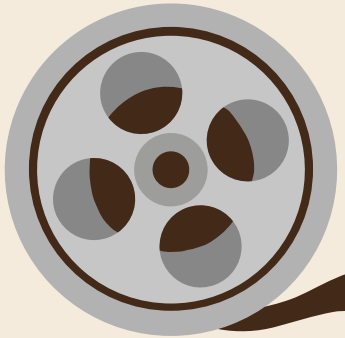
PROCESS PHILOSOPHY



1. **Raw Data Collection**
2. **Processing**
   a. Data Cleaning
   b. Adaptation
3. **Modeling**
   a. Correlations
   b. Creating Linear Regression models
4. **Visualizations**
5. **Recommendations and Insights**

DATA COLLECTION

The first and the foremost part of a successful analysis is data collection to ensure that we have enough insights to support our decisions. Data is scattered all over the web and we used various techniques to collect and clean the data to meet our needs



IMDB

This data contains information like movies, release years, budget, Net Gross Income, votings

Exchange Rates

This is used to convert the different currency information to US dollars to make comparisons

Ticket Inflation

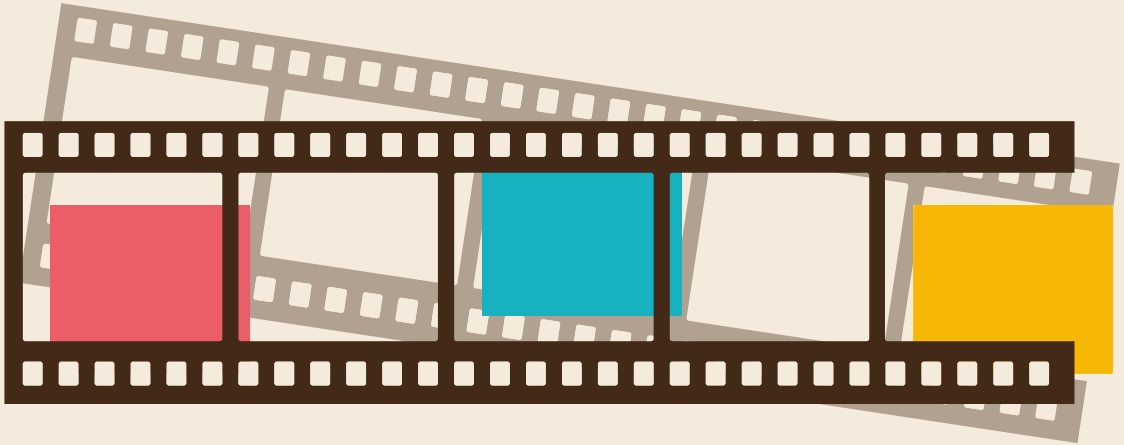
Used to tackle the Ticket prices and inflation inflation over the years

Survey

We wanted some more data to back our decisions so we did additional survey

“HE WHO WOULD SEARCH THE PEARLS MUST DIVE BELOW “

- JOHN DRYDEN



ANALYSIS AND MODELS

What factors affect revenue and ratings?

1. Budget
2. Duration
3. Demographic group ratings

QUICK HIGHLIGHTS

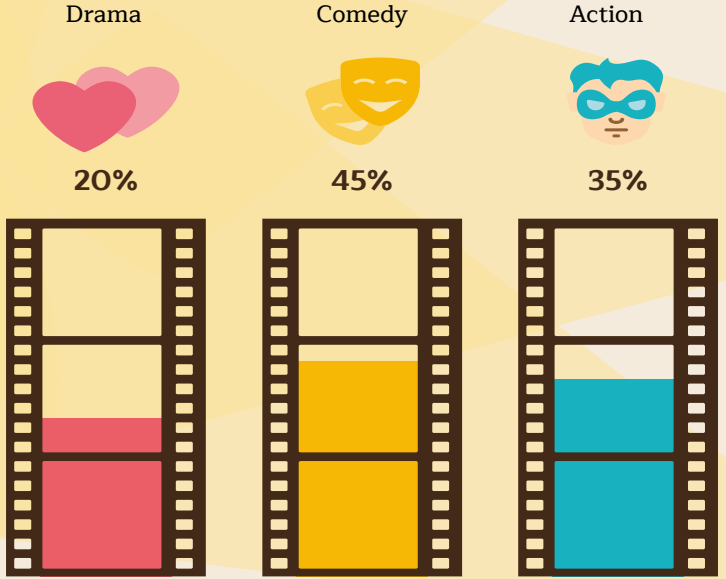
Country

US is the country with the most number of movies followed closely by UK, France and India

Production Houses

MGM, Warner Bros and Universal studios are best

Genres

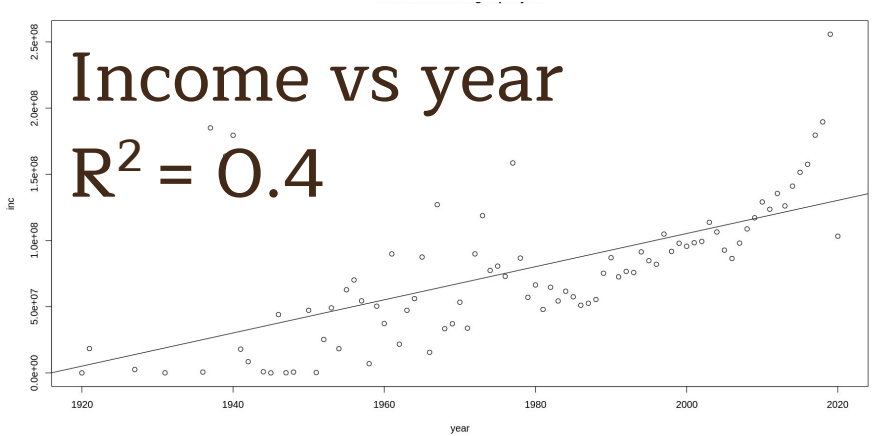


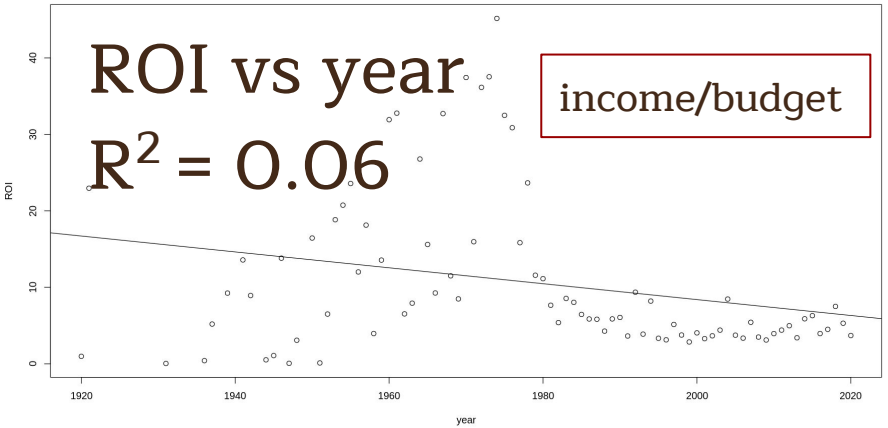
CALCULATING ROI

Year	Currency	In USD
1998	Rupees	1/41
1999	Rupees	1/43

Film	Year	Currency	Budget	In USD
ABCD	1999	Rupees	100	100/43

INCOME: IS IT UNBIASED?

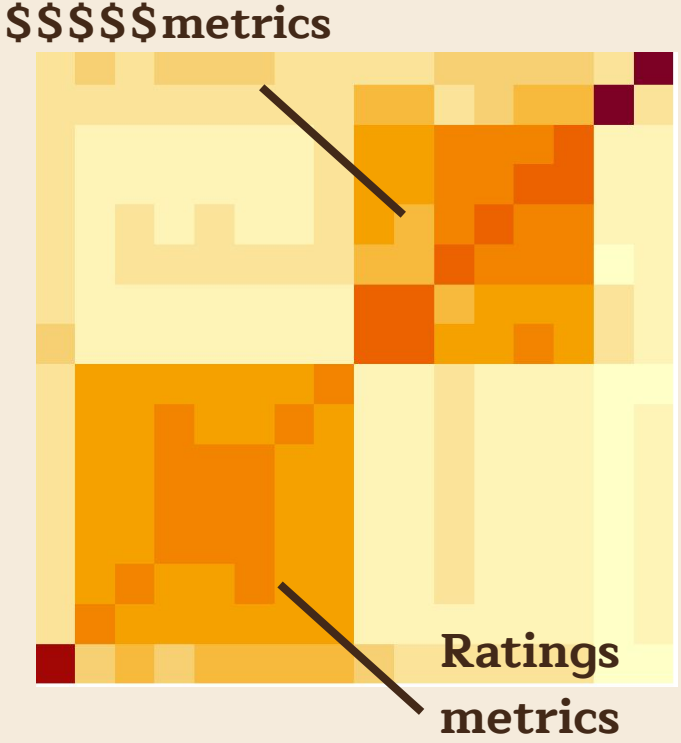




Income is influenced by year = inflation

ROI not influenced by year = unbiased

ROI: IS IT PREDICTABLE?

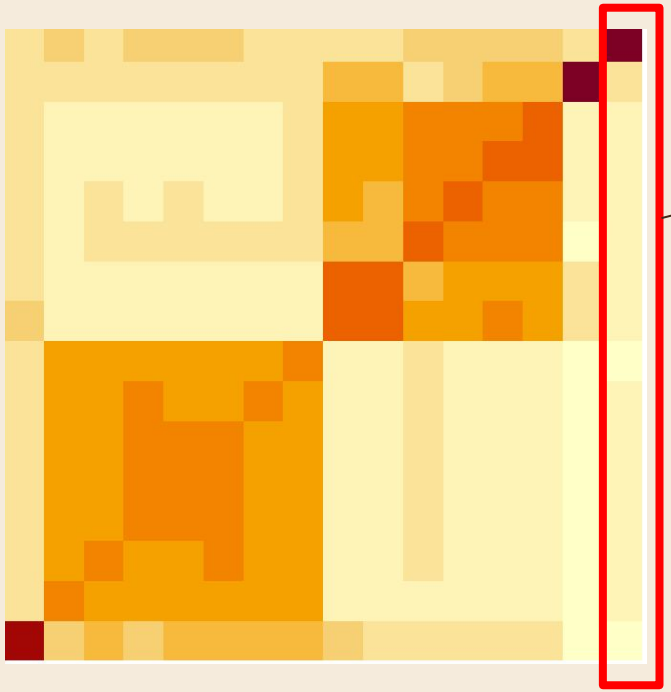


Heatmap of correlation strength:

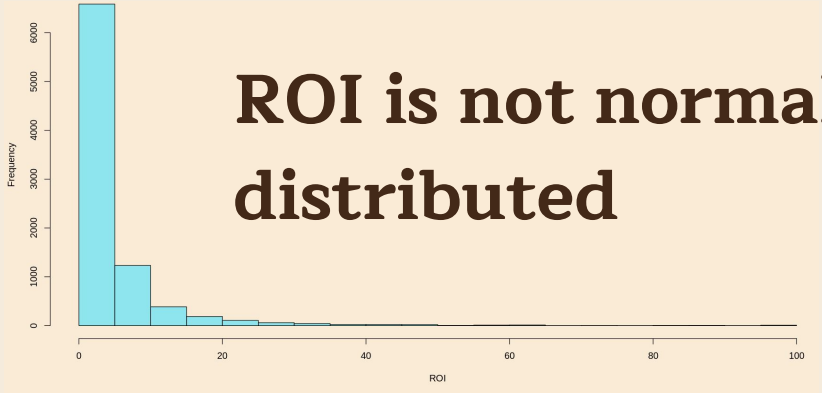
*Budget and income metrics are correlated

*All ratings measures are correlated

ROI: IS IT PREDICTABLE?



ROI does not correlate with other measures in our dataset



ROI is not normally distributed

ROI VS VIEWER RATING: MLR

ROI $R^2 = 0.12$

Duration	**
Avg_vote	***
Raw budget	***
Median_vote	***
18-30 vote	
30-45 vote	***
45+ vote	
Female vote	***
Male vote	
Budget	***

ROI VS VIEWER RATING: MLR

ROI

R^2 = 0.12

Duration	**
Avg_vote	***
Raw budget	***
Median_vote	***
18-30 vote	
30-45 vote	***
45+ vote	
Female vote	***
Male vote	
Budget	***

Avg vote

R^2 = 0.98

Duration	*
Raw budget	
18-30 vote	***
30-45 vote	***
45+ vote	***
Female vote	***
Male vote	***
Year	**

ROI VS VIEWER RATING: MLR

Avg vote

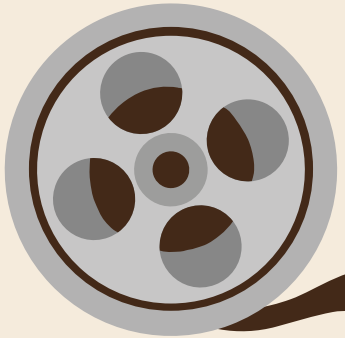
$R^2 = 0.98$

Duration	*
Raw budget	
18-30 vote	***
30-45 vote	***
45+ vote	***
Female vote	***
Male vote	***
Year	**

Can predict avg vote from any demographic, but not from budget

ANALYSIS SUMMARY

Correlation analyses, single and multiple linear regression



Inflation

Our data was not corrected for inflation, we added in two inflation-correcting measures

ROI

This was not predictable using budget, rating, demographic group, and duration data

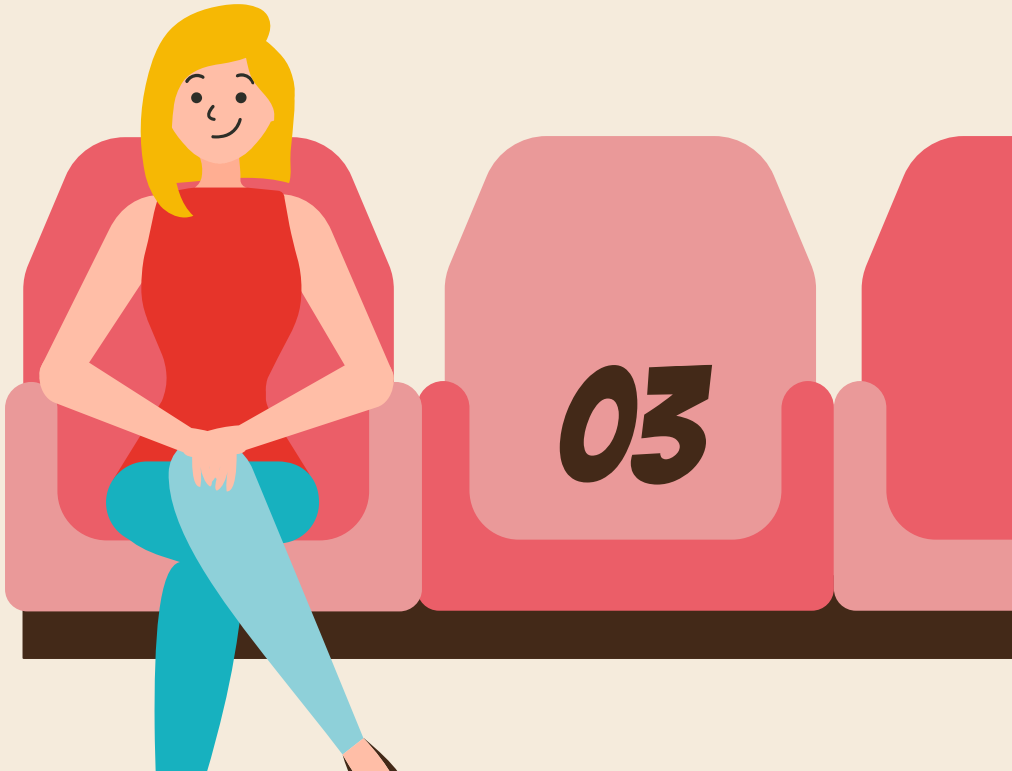
Viewer Rating

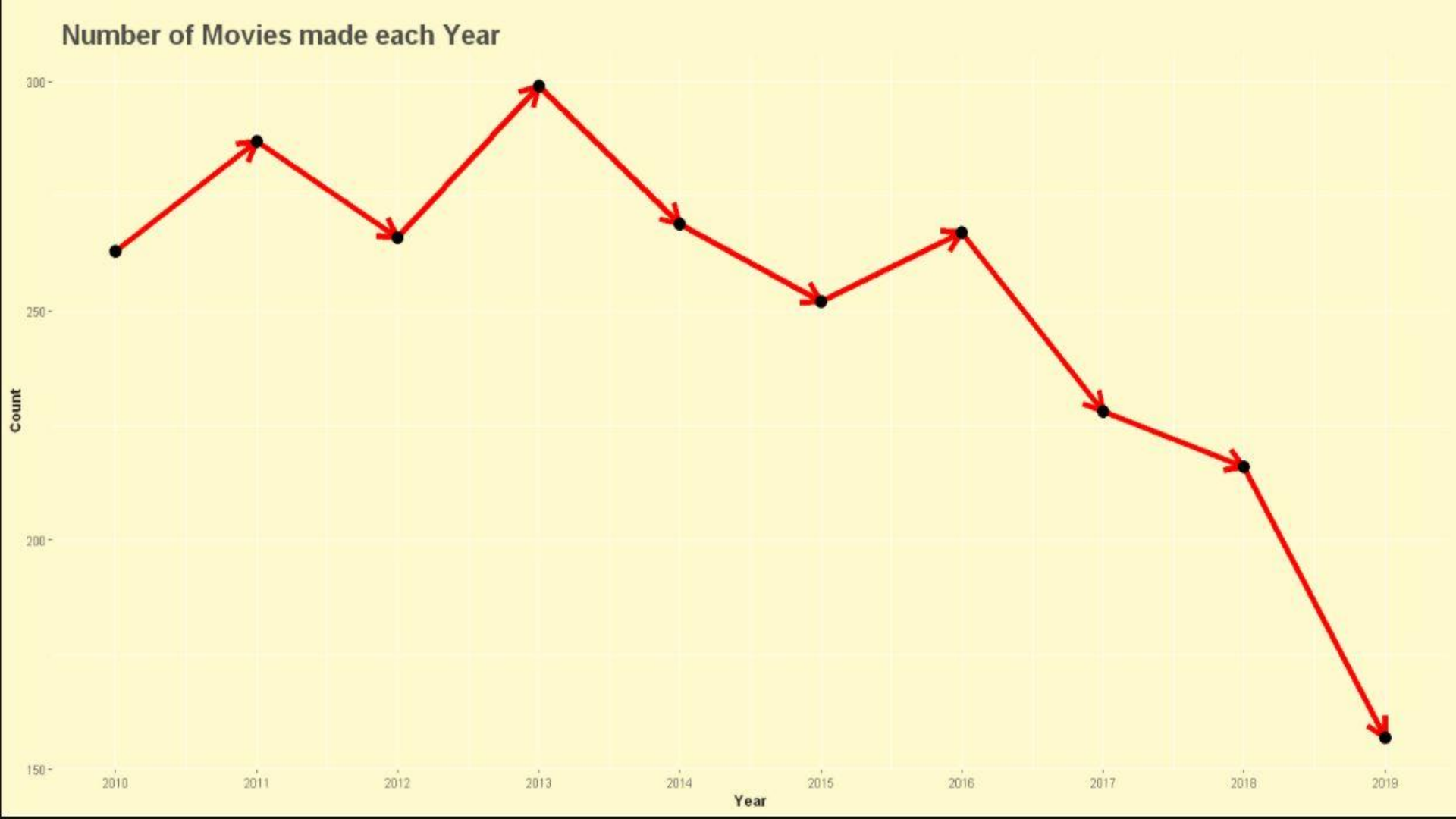
Each demographic group could predict the overall rating

What's next?

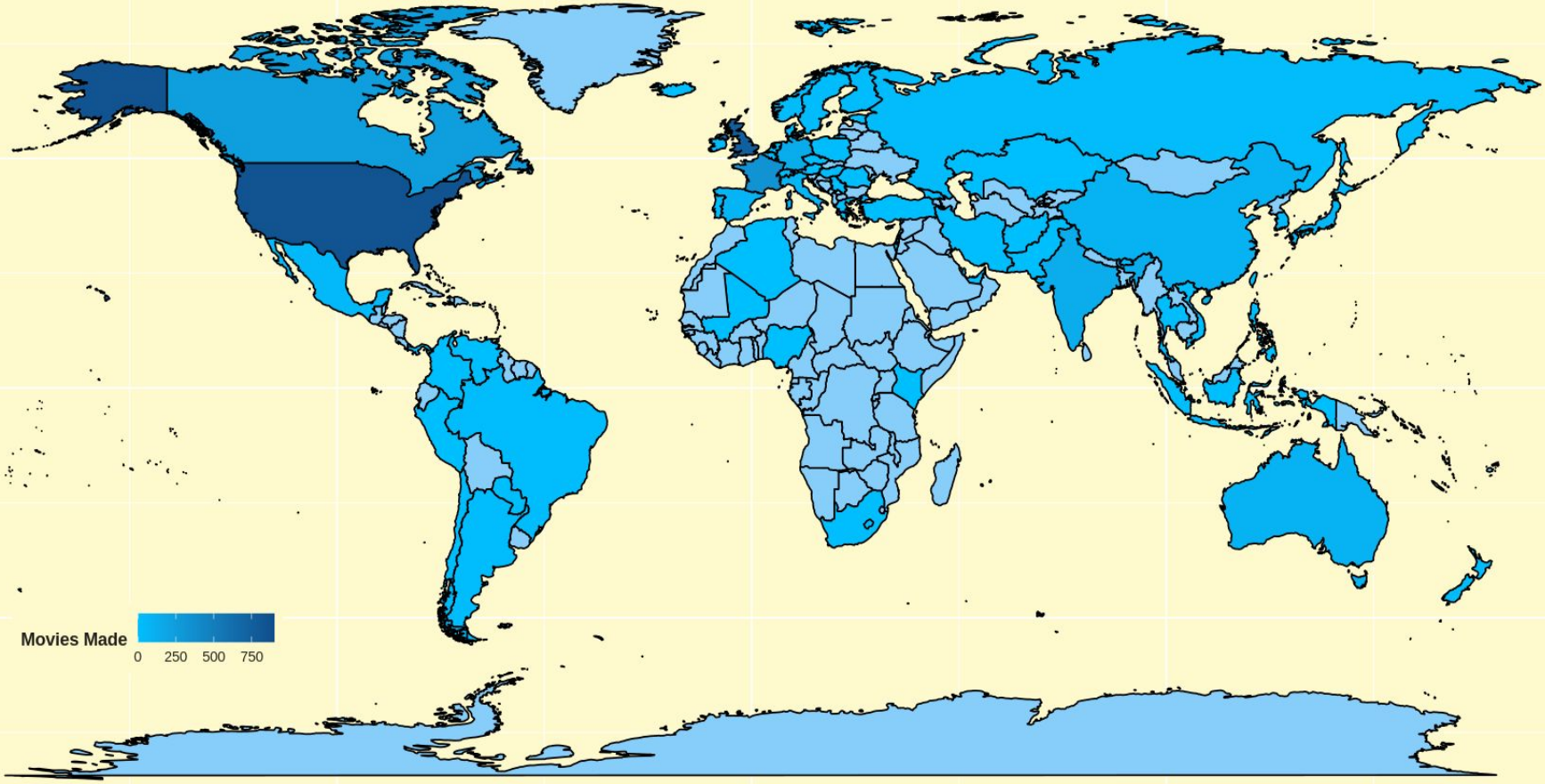
*Look at grouped data (production house and genre)
*Focus on 21st century revenue sources

RECOMMENDATION, PLOTS & SUMMARY

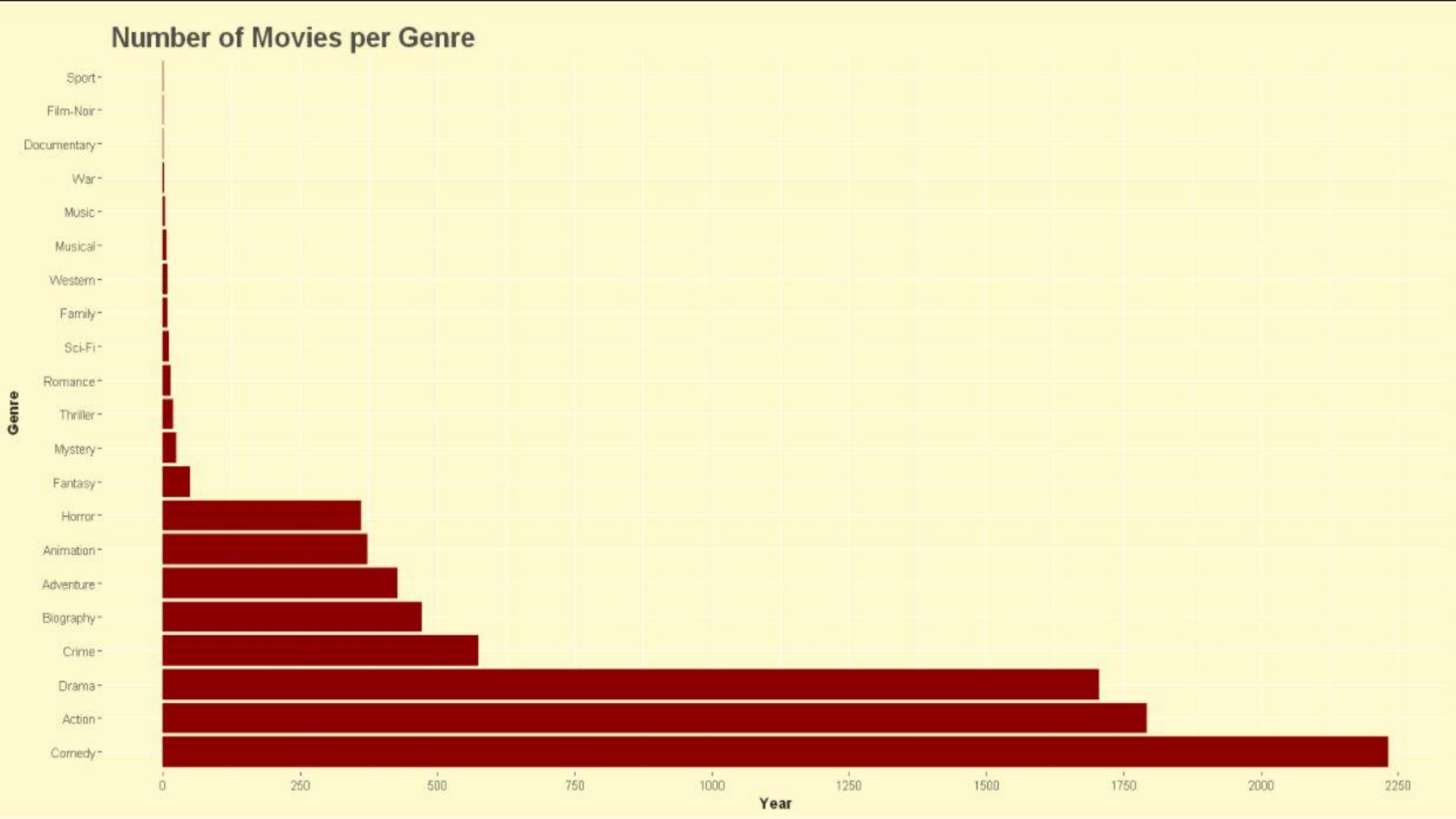


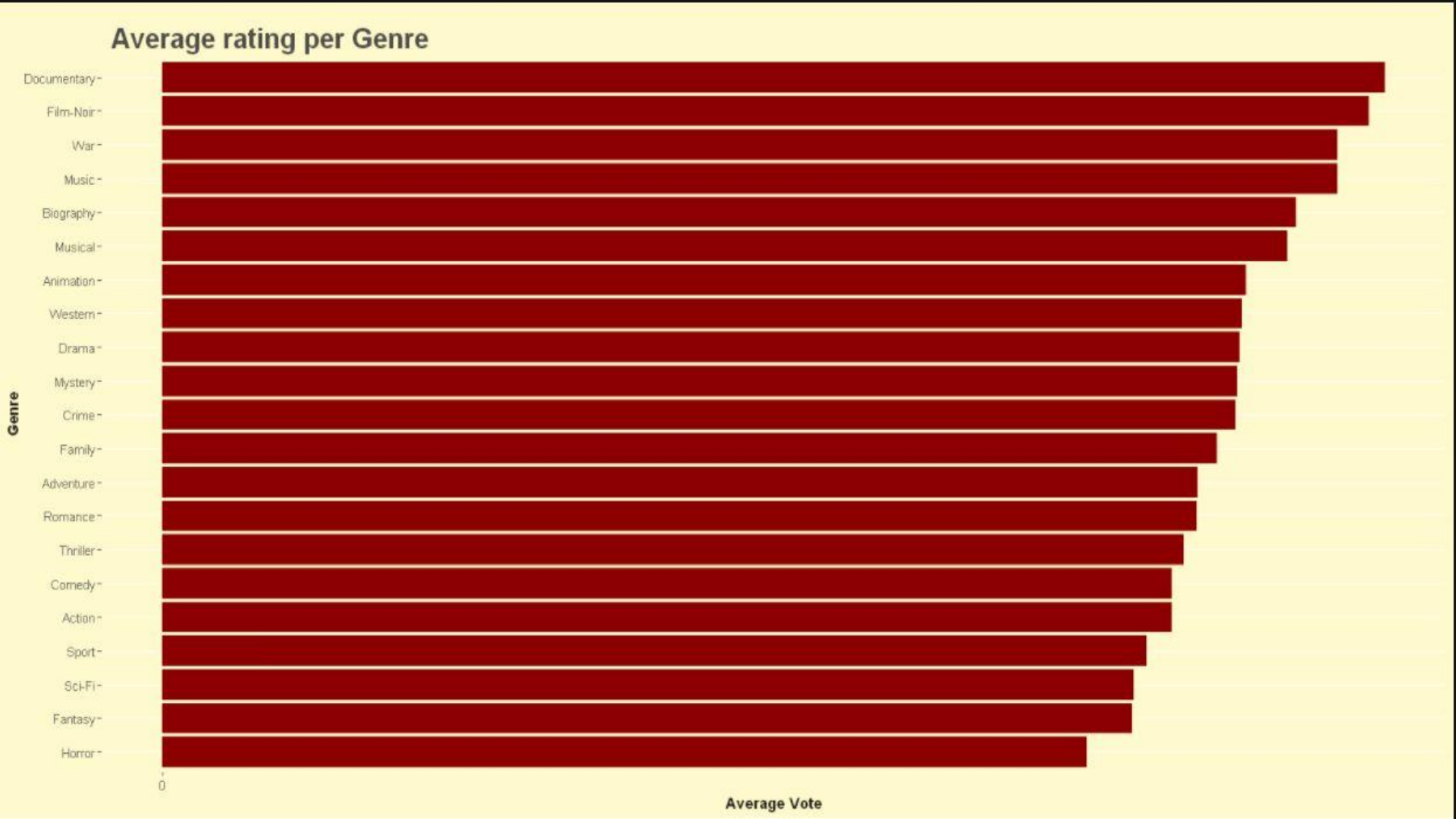

Number of Movies made each Year
300
250
200
150
Count
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Year

Number of Movies by Country

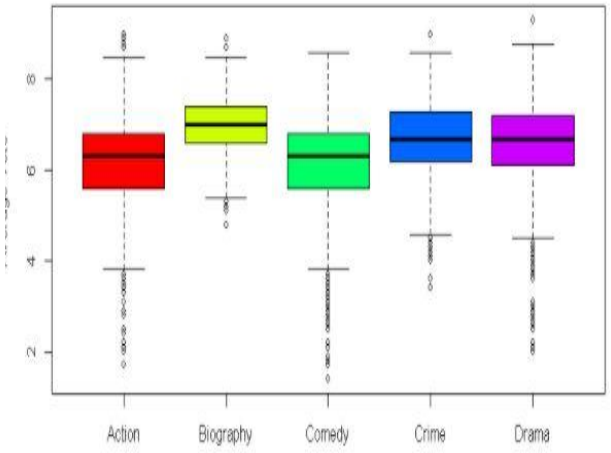


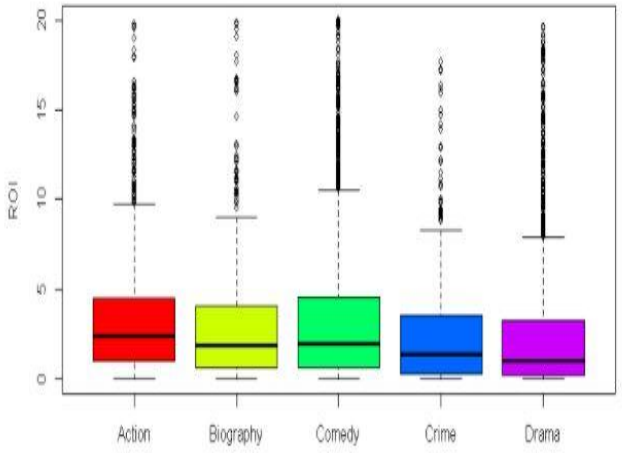
Data Source: IMDB movies dataset


Number of Movies per Genre
Sport
Film-Noir
Documentary
War
Music
Musical
Western
Family
Sci-Fi
Romance
Thriller
Mystery
Fantasy
Horror
Animation
Adventure
Biography
Crime
Drama
Action
Comedy
Genre
0
250
500
750
1000
1250
1500
1750
2000
2250
Year


Average rating per Genre
Documentary
Film-Noir
War
Music
Biography
Musical
Animation
Western
Drama
Mystery
Crime
Family
Adventure
Romance
Thriller
Comedy
Action
Sport
Sci-Fi
Fantasy
Horror
Genre
0
Average Vote

Genre Comparisons

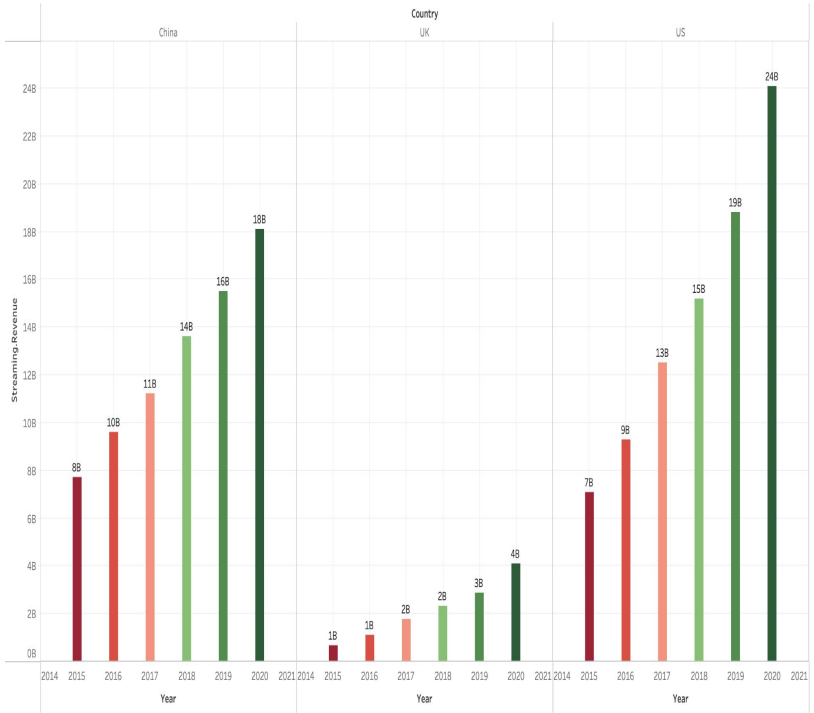


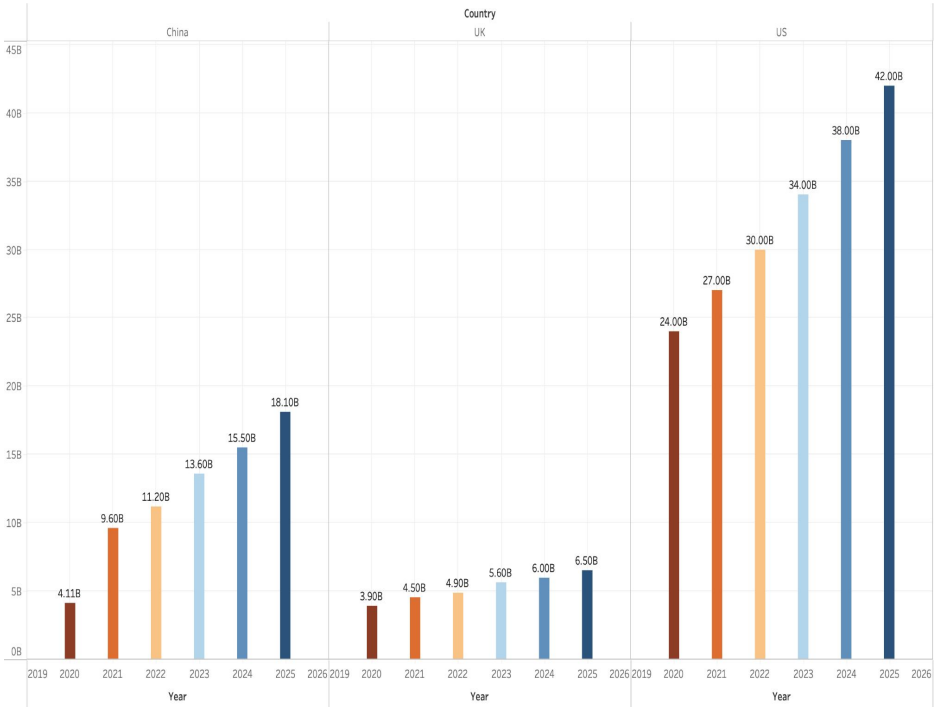


CUSTOMER SURVEY

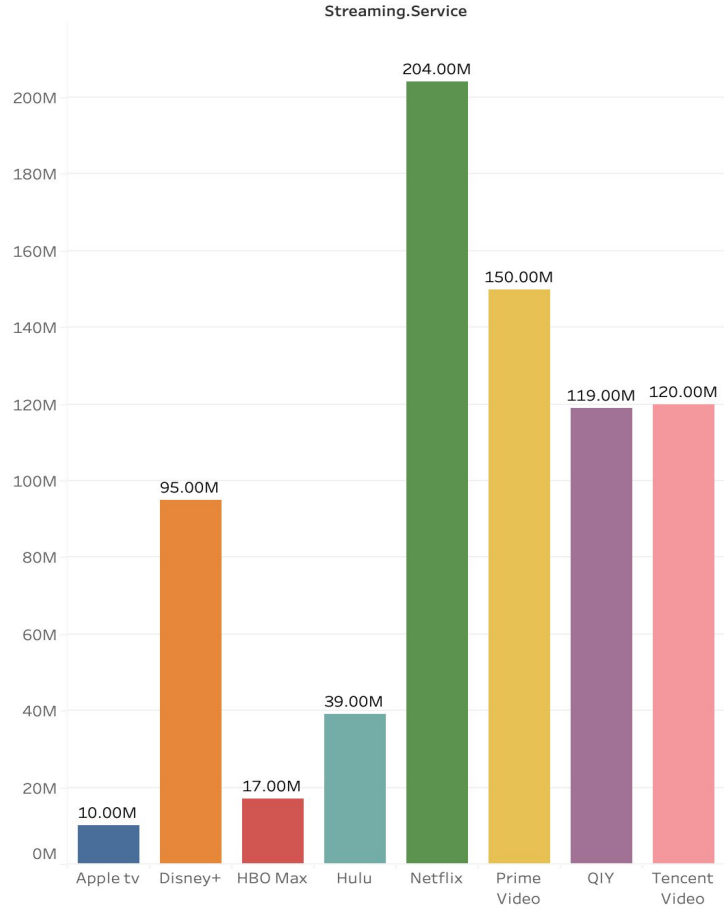
	YES	NO	OTHER
Do you watch the movie and shows in foreign language using subtitles or dubbing?	88.2 %	11.8 %	0%
Are you comfortable going to the theatres after covid ?	7.1%	88%	4.9%
What streaming service you prefer?	Netflix (51.8%)	Prime video (29.4%)	18.8%

Revenue trends of Streaming services and predicted growth

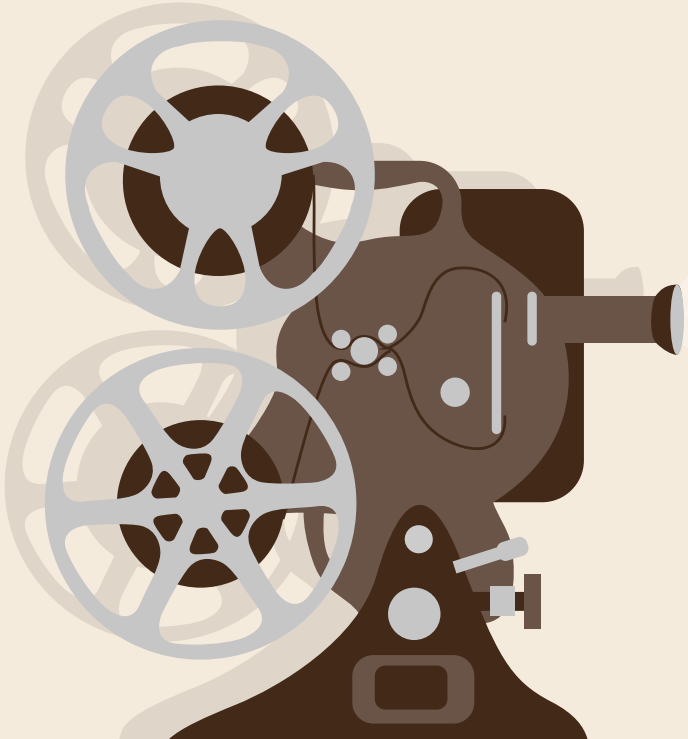




Streaming Services Summary



- The streaming service has reached 50.11 Billion Globally in 2020
- This Revenue is expected to reach 70 Billions by 2025
- Top Three streaming services have 449 Million subscribers i.e. movies on millions of devices in one click
- Streaming services take movies based on ratings



RECOMMENDATIONS

- Country -> USA
- Genre -> Comedy / Biography
- Production House -> Universal Studios
- Screening enriched for 30-45 age group people
- Use a streaming service for the launch. Preferably Netflix & Amazon Prime