A PROJECT REPORT

On

“ The Shift in Movie Consumption: A Comparative Analysis of Theatrical and OTT Platforms ”

Submitted to

University of Liverpool

DATA MINING AND MACHINE LEARNING (EBUS537)

BY **CHRISTINA ANN JACOB (201805307)**

Table of content

Table of figures

EXECUTIVE SUMMARY

Understanding whether viewers prefer watching movies in theaters or on OTT platforms is important in today's rapidly changing entertainment industry.With the emergence of streaming services and changes in audience habits, producers and distributors must determine what influences these preferences.The Apriori algorithm will help identify the combination of factors like high budget, famous actors, and action genres that make people go to the theaters.This report explains the data preparation, pattern discovery, and rule generation processes. The findings offer practical tips for producers and marketers to make better decisions, improve movie promotions, and align strategies with audience expectations.

INTRODUCTION

The entertainment industry is changing with new technology and viewer habits. Streaming platforms like Netflix are growing, and the OTT market could reach $332.5 billion by 2027. Theaters are coming back with a, earning $25 billion in 2022, 27% more than in 2021. Indeed, estimates suggest that revenues of the Indian box office might reach ₹15,300 crores ($1.9 billion USD) by 2023, accounting for almost 40% of the earnings in action and adventure movies. OTT subscriptions also grew 30% year-over-year (Ormax Media, 2024), driven by a rise in regional content.

Using the Apriori algorithm, a machine learning method, it uncovers hidden patterns in movie features that affect viewers' preference for theaters.This report explores audience behavior by applying data mining techniques. The Apriori algorithm is ideal for finding patterns in large datasets. It should provide producers, marketers, and distributors with some real, data-based views that would help them to make wiser decisions as to where to invest, and how to get connected with their audience.

IMPLEMENTATION

1. The Dataset

A virtual dataset was created to comprehensively analyze viewer preferences for theatrical and OTT movie platforms.

The columns in the dataset are as follows:

Title: Name of the movie.

Budget: Numeric, indicating the production cost of a movie.

HOA (History of Actors): Numeric, representing the number of successful movies an actor has been part of.

HOD (History of Directors): Numeric, indicating the number of successful movies directed.

THEATRE OR OTT: Categorical, identifying the platform type.

PREFERENCE: Categorical, indicating whether viewers prefer Theatrical (T) or OTT (O).

TIME: Numeric, indicating the season in which the movie was released

RATING: Numeric, representing the IMDb rating of the movie.

Year: Year of release.

Certificates: Categorical, indicating the certification (e.g., PG-13, R).

Genre: Categorical, covering genres such as Action, Drama, and Comedy.

Director: Name of the director.

Star Cast: List of key actors in the movie.

MetaScore: Numeric, representing critical scores from review aggregators.

Dataset Creation Process

For budget, genre, ratings, etc., information was sought from various sources like IMDb and The Numbers. Values for viewer preference and actor history were imputed in cases of missing values based on trends identified through industry reports published by EY FICCI and PwC.
The Apriori algorithm helps uncover meaningful patterns and relationships in the dataset to better understand viewer preferences.Frequent itemsets are progressively combined, and their confidence levels are evaluated to generate association rules. For example, the algorithm might reveal that high-budget action movies with renowned actors often lead to theatrical preferences. This ensures that only the most impactful patterns are identified.

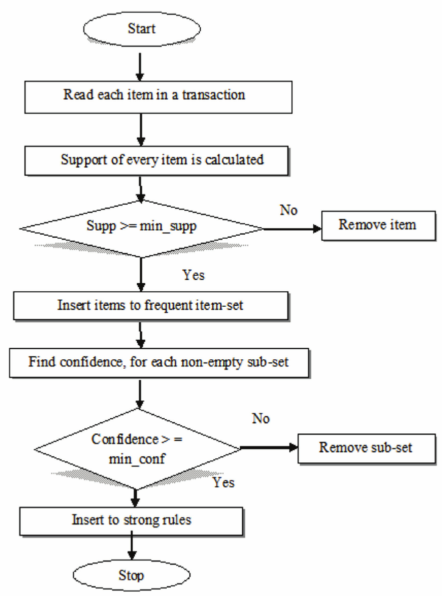


Fig 1: Flowchart of the Apriori Algorithm Process

Source : (Mittal, Pareek and Agarwal, 2014)

The flowchart illustrates this process and makes it easier to find important patterns, like if high-budget action blockbusters with popular actors tend to succeed in theaters, if R-rated movies often find a larger audience on OTT platforms, do action movies perform better in summer as theatrical releases or if successful actors and directors (high HOA and HOD values) strongly influence theater attendance.

The Apriori algorithm's balance of computational efficiency and interpretability helps in identifying actionable insights making it a powerful tool for producers, marketers, and distributors.

IMPLEMENTATION

To uncover patterns in viewer preferences, the dataset was prepared with essential attributes like **Budget**, **HOA**, and **PREFERENCE**. Preprocessing steps, including categorization and one-hot encoding, ensured the data was ready for analysis.

Preprocessing the Dataset

The first step was to clean the data by removing outliers, as they can skew results. Using the IQR method, outliers in key columns like Budget and HOD were removed, reducing the dataset from 99 to 90 rows. This made the data more accurate and easier to analyze. Figure 1 shows the process.

```python
# Define a function to remove outliers using the IQR method
def remove_outliers_iqr(df, column):
    Q1 = df[column].quantile(0.25)  # First quartile (25th percentile)
    Q3 = df[column].quantile(0.75)  # Third quartile (75th percentile)
    IQR = Q3 - Q1  # Interquartile range
    lower_bound = Q1 - 1.5 * IQR  # Lower bound
    upper_bound = Q3 + 1.5 * IQR  # Upper bound
    return df[(df[column] >= lower_bound) & (df[column] <= upper_bound)]

# Apply the function to 'Budget' and 'HOD' columns
filtered_df = remove_outliers_iqr(df, 'Budget')
filtered_df = remove_outliers_iqr(filtered_df, 'HOD')

# Display the size of the dataset before and after outlier removal
original_size = df.shape[0]
filtered_size = filtered_df.shape[0]

original_size, filtered_size
```

```
(99, 90)
```

After cleaning, the dataset was reviewed to see how the data looked without the outliers. For example, **Budget** was re-checked, and as seen in **Fig 2**, most movies now fall into the lower-budget range, especially between 0–25 million. This shows the dataset has more low-budget movies but still includes a few high-budget ones for balance.

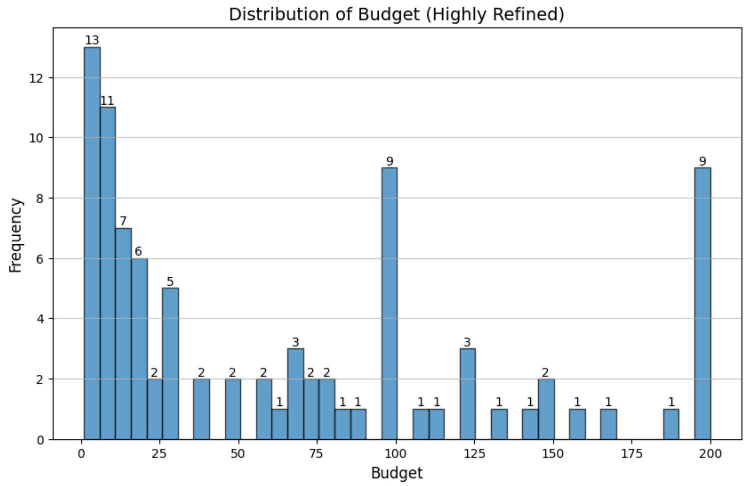


Fig 2: Refined 'Budget' distribution after outlier removal.

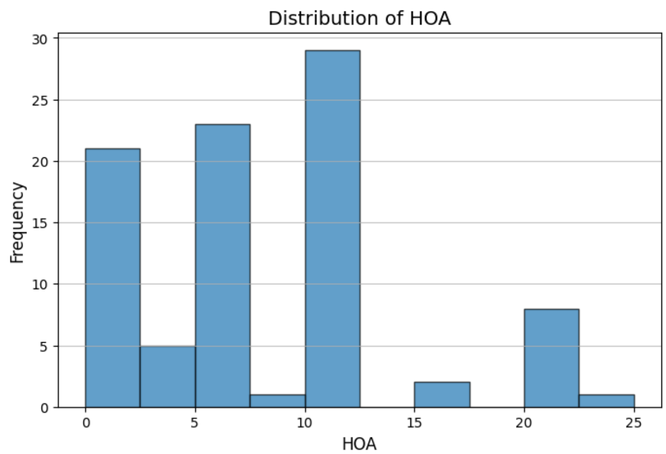


Fig 3: HOA distribution, showing most actors have medium success history with few having extensive track record.

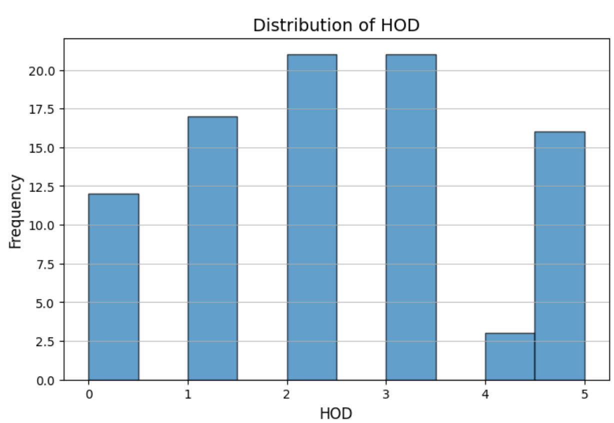


Fig 4: HOD distribution, showing a balanced spread of director success rates

The dataset also includes movie ratings and critic reviews. Most movies have ratings between 6 and 8, showing that the data focuses on moderately to highly rated films (Fig 5). Critic scores mostly range from 65 to 75, indicating that many of the movies are well-reviewed (Fig 6).

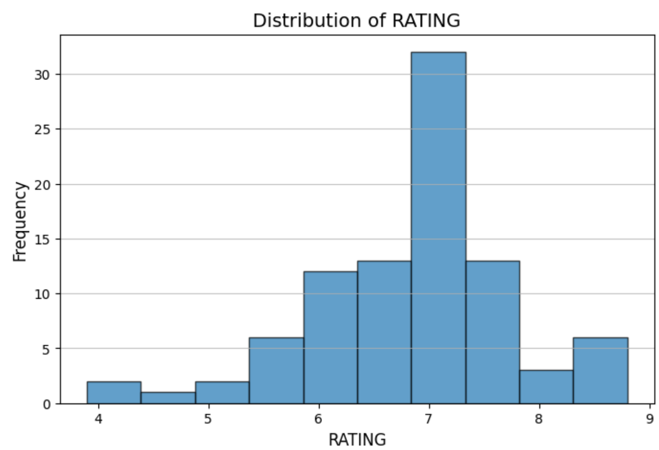


Fig 5: Movie ratings clustered around moderate-to-high ratings.

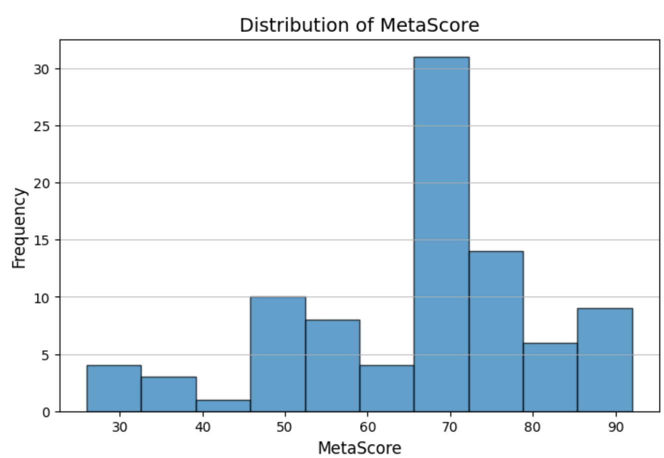


Fig 6: MetaScore distribution peaking in the critically acclaimed range.

By eliminating outliers and visualizing distributions, a clear understanding of the dataset's structure and patterns emerges, setting the stage for generating actionable insights through the Apriori algorithm.

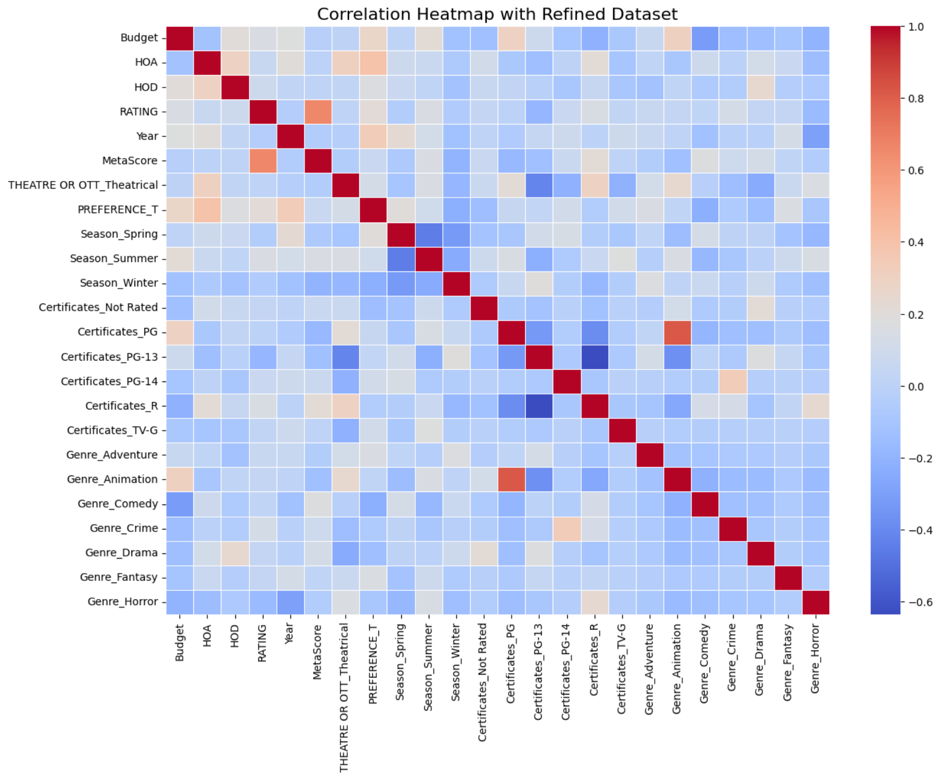


Fig. 7: Correlation Heatmap with Refined Dataset."

The **"Correlation Heatmap with Refined Dataset"** (**Fig. 7**) shows some interesting relationships in the data. For example,high-budget movies often have well-known actors, as seen in the positive link between **Budget** and **HOA**. This combination usually pushes viewers toward theatrical releases.**MetaScore** and **Ratings** are also positively linked, showing how critical acclaim connects to audience appreciation.On the other hand, things like **Genre** and **Certificates** don't show strong correlations, which means they might need more complex analysis to understand their impact.

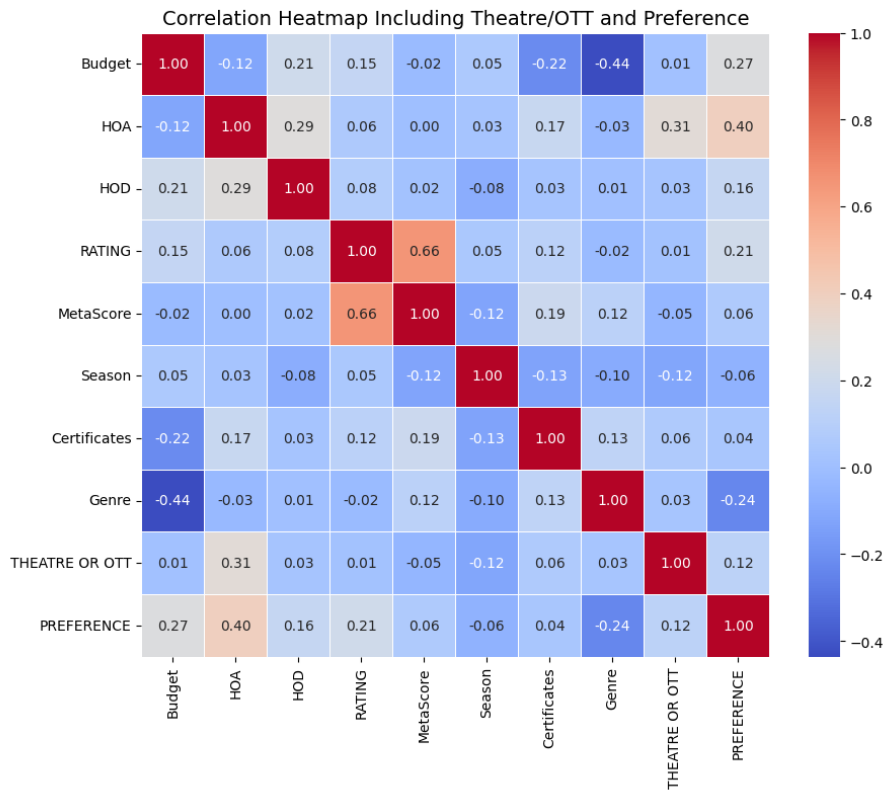


Fig. 8: Correlation Heatmap Including Theatre/OTT and Preference

The "Correlation Heatmap Including Theatre/OTT and Preference" (Fig. 8) shows that platform preferences are strongly linked to successful actors (HOA). This means popular actors bring more people to theaters. Successful directors with multiple hits (HOD) also strongly influence theater preference.

There's a negative link between genre and PREFERENCE. Drama and Comedy are more popular on OTT platforms, while Action and Adventure are usually preferred for theaters. Genre still matters. This means that certain seasons are linked to either theaters or OTT platforms. These heatmaps show how different factors influence platform preferences.

<u>Frequent Itemsets and association rules output</u>

```
Frequent Itemsets:
     support                   itemsets
0   0.488889              (Budget_High)
4   0.477778             (PREFERENCE_1)
1   0.466667                 (HOA_High)
2   0.444444                 (HOD_High)
5   0.366667                 (Season_1)
12  0.333333     (PREFERENCE_1, HOA_High)
9        0.3  (Budget_High, PREFERENCE_1)
6   0.255556                 (Season_2)
8   0.255556      (HOD_High, Budget_High)
14  0.244444     (HOD_High, PREFERENCE_1)
7   0.233333      (Budget_High, HOA_High)
11  0.233333         (HOD_High, HOA_High)
15  0.222222     (Season_1, PREFERENCE_1)
13  0.211111         (Season_1, HOA_High)
3        0.2                  (Genre_2)
10       0.2      (Season_1, Budget_High)

Association Rules:
      antecedents     consequents  antecedent support  consequent support  \
3      (HOA_High)  (PREFERENCE_1)            0.466667            0.477778
2  (PREFERENCE_1)      (HOA_High)            0.477778            0.466667
1  (PREFERENCE_1)   (Budget_High)            0.477778            0.488889
0   (Budget_High)  (PREFERENCE_1)            0.488889            0.477778
4      (Season_1)  (PREFERENCE_1)            0.366667            0.477778

    support  confidence      lift  representativity  leverage  conviction  \
3  0.333333    0.714286  1.495017               1.0  0.110370    1.827778
2  0.333333    0.697674  1.495017               1.0  0.110370    1.764103
1  0.300000    0.627907  1.284355               1.0  0.066420    1.373611
0  0.300000    0.613636  1.284355               1.0  0.066420    1.351634
4  0.222222    0.606061  1.268499               1.0  0.047037    1.325641

   zhangs_metric   jaccard  certainty  kulczynski
3       0.620833  0.545455   0.452888    0.705980
2       0.634043  0.545455   0.433140    0.705980
1       0.423956  0.450000   0.271992    0.620772
0       0.433172  0.450000   0.260155    0.620772
4       0.334211  0.357143   0.245648    0.535588
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async`
  and should_run_async(code)
/usr/local/lib/python3.10/dist-packages/mlxtend/frequent_patterns/fpcommon.py:161: DeprecationWarning: Da
  warnings.warn(
```

To exit full screen, pres

The Apriori algorithm uncovers critical insights into viewer preferences. High-budget movies with popular actors (Budget_High, HOA_High) strongly attract theatrical audiences, highlighting the combined appeal of star power and production value. A strong link exists between successful actors and theatrical preference (PREFERENCE_1, HOA_High), while summer releases (Season_1) further boost theater preference, emphasizing the role of timing. Key rules show that popular actors significantly drive theatrical viewing (confidence: 71.4%, lift: 1.50), high-budget productions align with theater preference (confidence: 62.7%), and summer releases favor cinemas (confidence: 60.6%). These insights emphasize the importance of strategic planning around cast, budget, and release timing to optimize audience engagement.

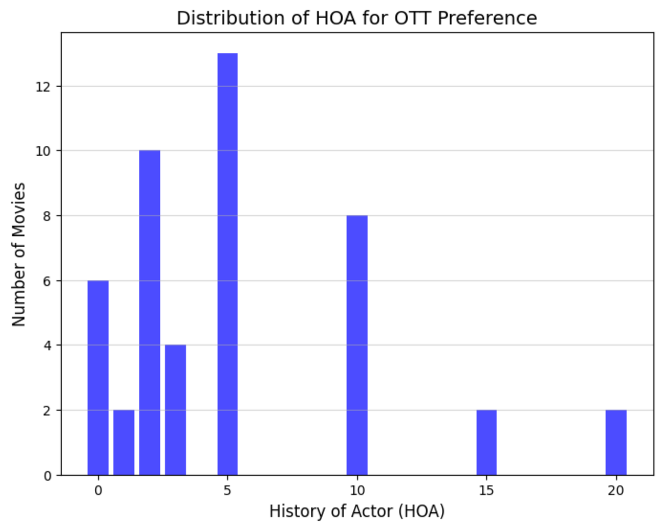


Fig. 9: Distribution of HOA for OTT Preference

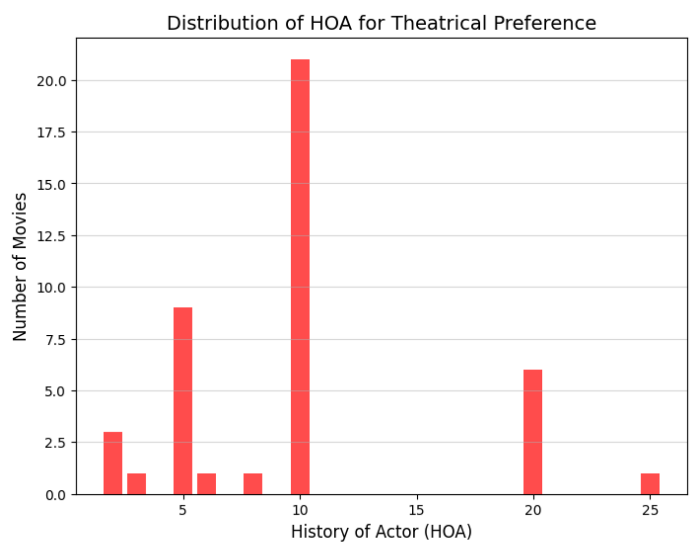


Fig. 10: Distribution of HOA for Theatrical Preference

OTT platforms mostly feature actors with moderate HOA values (5-10), focusing on diverse content without relying heavily on big stars. Theaters, however, favor higher HOA values (10+), showing a clear dependence on star power to draw audiences.

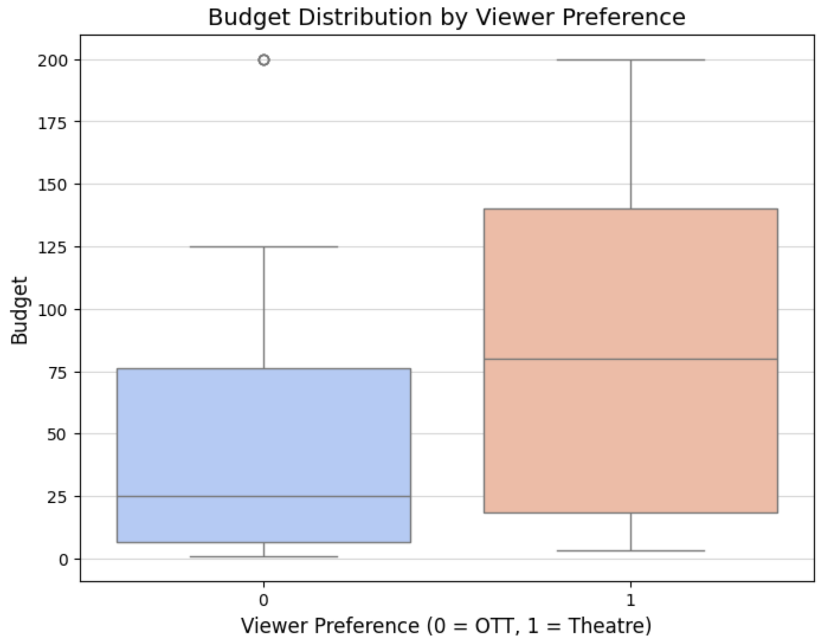


Fig.11: Budget Distribution Supporting Theatrical Preference for High-Budget Movies

Viewers lean towards high-budget movies for theater releases, likely because bigger budgets are seen creating a better cinematic experience. On the other hand, medium-budget films work well for OTT platforms, matching the convenience of watching at home.

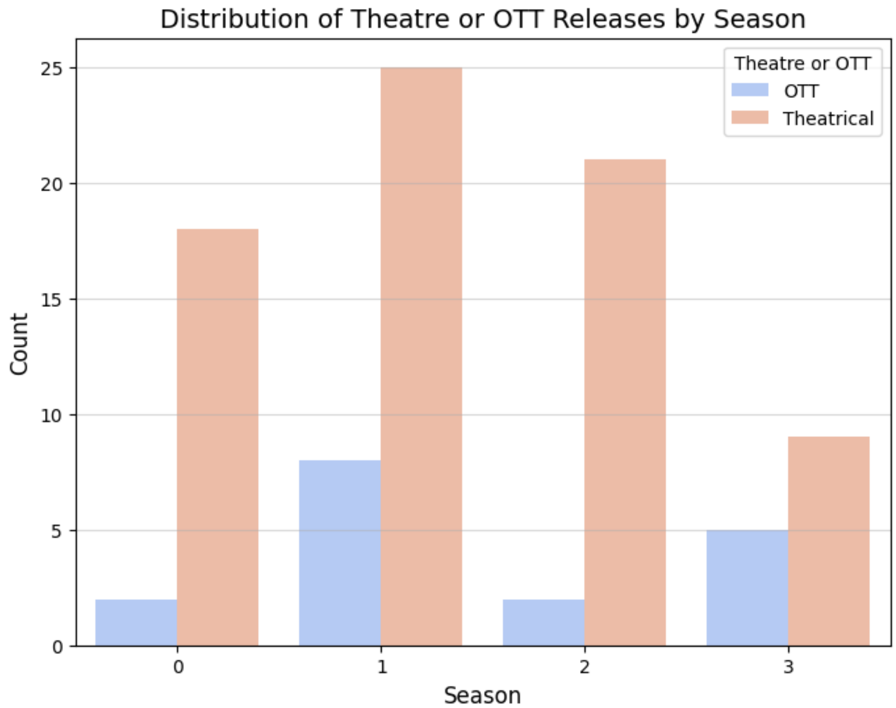


Fig.12: Current Distribution of Theatre or OTT Releases by Season

0: Winter (e.g., December, January, February)

1: Spring (e.g., March, April, May)

2: Summer (e.g., June, July, August)

3: Fall/Autumn (e.g., September, October, November)

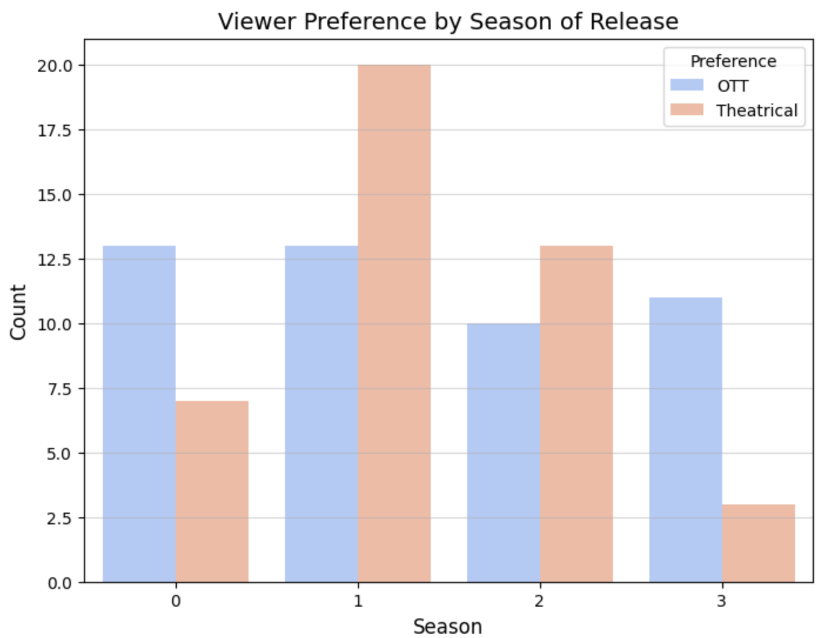


Fig. 13: Viewer Preference by Season of Release

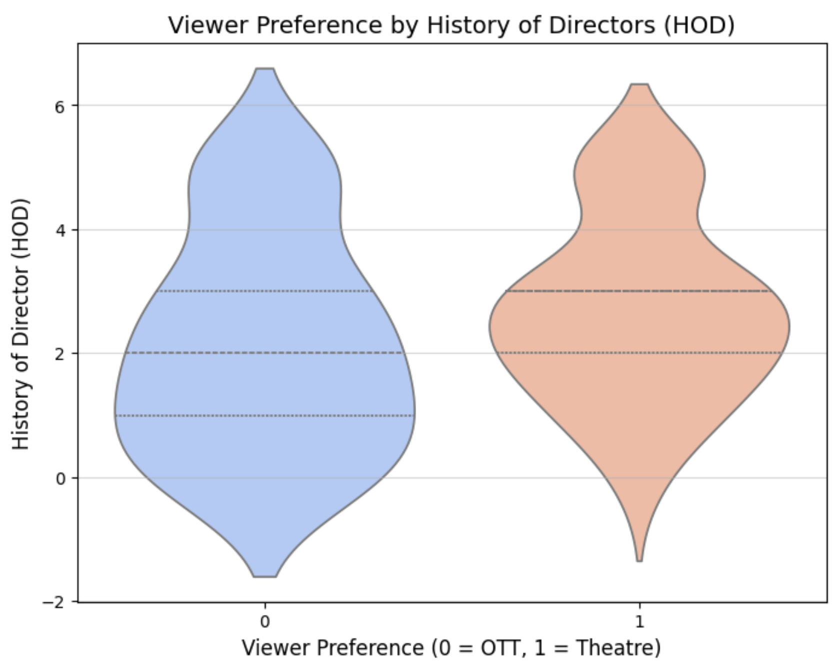
Season 1 (summer) sees the most theatrical releases, as producers target high audience availability. Seasons 0 (spring) and 2 (fall) also have notable theatrical releases, often tied to festivals or family-oriented films. Viewer preferences show theaters dominate in Season 1, aligning with the rule: Season_1 → PREFERENCE_1 (confidence: 60.6%, lift: 1.27). Season 0 has a balanced split between OTT and theaters, while Season 3 (winter) leans toward OTT, likely due to colder weather and a preference for home viewing. These trends show that release timing and content type should match seasonal audience habits.

```
Frequent Itemsets:
    support                     itemsets
0  0.555556                 (HOD_Low_HOD)
2  0.522222                (Preference_0)
3  0.477778                (Preference_1)
1  0.444444                (HOD_High_HOD)
4  0.322222   (HOD_Low_HOD, Preference_0)
7  0.244444  (Preference_1, HOD_High_HOD)
5  0.233333   (Preference_1, HOD_Low_HOD)
6  0.200000  (Preference_0, HOD_High_HOD)

Association Rules for HOD and Preference:
      antecedents     consequents  antecedent support  consequent support  \
0   (HOD_Low_HOD)  (Preference_0)            0.555556            0.522222
7  (HOD_High_HOD)  (Preference_1)            0.444444            0.477778
5  (HOD_High_HOD)  (Preference_0)            0.444444            0.522222
3   (HOD_Low_HOD)  (Preference_1)            0.555556            0.477778

    support  confidence      lift  representativity  leverage  conviction  \
0  0.322222        0.58  1.110638               1.0  0.032099    1.137566
7  0.244444        0.55  1.151163               1.0  0.032099    1.160494
5  0.200000        0.45  0.861702               1.0 -0.032099    0.868687
3  0.233333        0.42  0.879070               1.0 -0.032099    0.900383

   zhangs_metric   jaccard  certainty  kulczynski
0       0.224138  0.426471   0.120930    0.598511
7       0.236364  0.360656   0.138298    0.530814
5      -0.224138  0.260870  -0.151163    0.416489
3      -0.236364  0.291667  -0.110638    0.454186
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarni
  and should_run_async(code)
```



Caption: Fig 14: Violin Plot of Viewer Preferences by HOD

Directors with lower success (HOD_Low) have a moderate link to OTT preferences (58% confidence, 1.11 lift), showing OTT platforms are more suited for content with lower production risks. In contrast, successful directors (HOD_High) strongly influence theatrical preferences (55% confidence, 1.15 lift), reflecting the trust audiences place in their work.

The weak association of HOD_High with OTT (45% confidence, 0.86 lift) and HOD_Low with theaters (42% confidence) highlights that less successful directors struggle in theaters, while OTT provides a platform for their content. These insights emphasize the need to match directors' profiles with the right platform to maximize audience engagement.

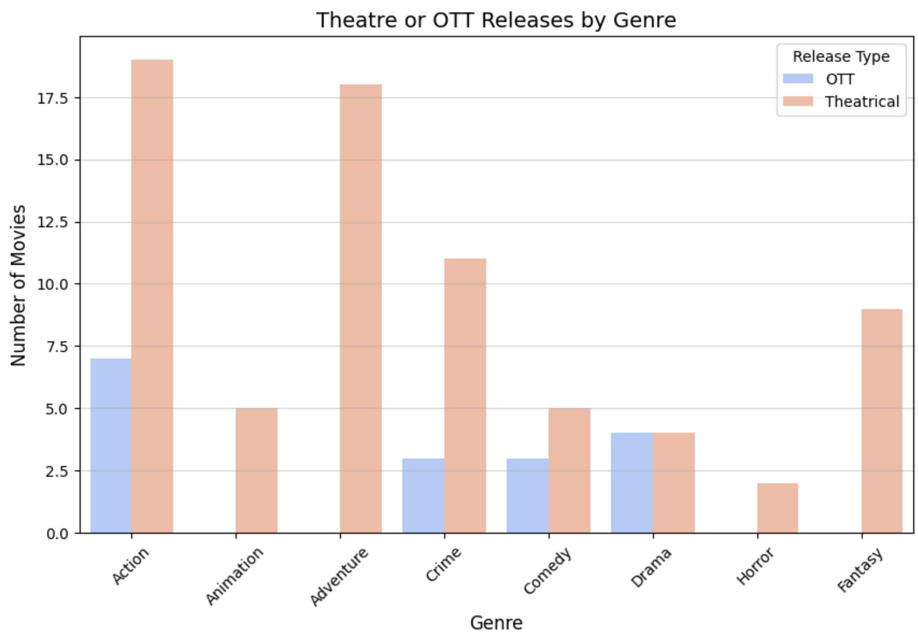
```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: Deprec
  and should_run_async(code)
Frequent Itemsets:
     support                 itemsets
4   0.522222          (PREFERENCE_0)
5   0.477778          (PREFERENCE_1)
0   0.288889               (Genre_0)
1   0.200000               (Genre_2)
7   0.188889  (PREFERENCE_1, Genre_0)
2   0.155556               (Genre_3)
10  0.122222  (PREFERENCE_0, Genre_3)
3   0.100000               (Genre_7)
6   0.100000  (PREFERENCE_0, Genre_0)
8   0.100000  (Genre_2, PREFERENCE_0)
9   0.100000  (Genre_2, PREFERENCE_1)

Association Rules for Genre and Preference (Theatre or OTT):
  antecedents consequents   support  confidence      lift
9       Crime         OTT  0.122222    0.785714  1.504559
3      Action     Theatre  0.188889    0.653846  1.368515
6   Adventure     Theatre  0.100000    0.500000  1.046512
4   Adventure         OTT  0.100000    0.500000  0.957447
1      Action         OTT  0.100000    0.346154  0.662848
```

Crime movies (Genre_3) work best on OTT platforms (78.6% confidence, 1.50 lift) due to their engaging, story-driven nature. Action movies (Genre_0) are perfect for theaters (65.4% confidence, 1.37 lift) because of their exciting visuals and immersive feel.

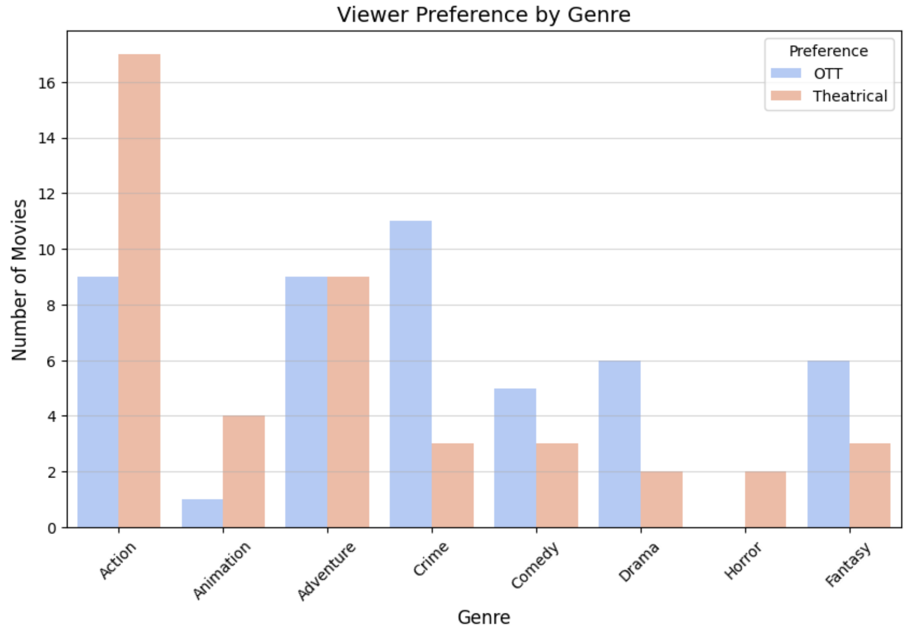
Adventure movies (Genre_2) don't have a clear platform preference, doing fairly well on both OTT and in theaters, making them good for dual releases. The weak connection between Action movies and OTT (34.6% confidence, 0.66 lift) confirms their stronger fit for theaters.

These insights show the value of planning releases based on what works best for each genre.



Caption: Fig 15: Distribution of Theatre or OTT Releases by Genre

Genres like Action, Adventure, and Fantasy are mostly released in theaters, because of their strong visual appeal. On the other hand, Animation and Comedy have a more even split, working well for both theaters and OTT platforms.



Caption: Fig 16: Viewer Preferences by Genre

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarni
  and should_run_async(code)
Frequent Itemsets:
    support                                         itemsets
3  0.515152                                  (PREFERENCE_T)
2  0.484848                                  (PREFERENCE_O)
1  0.121212               (HOA_Budget_Genre_True_True_Action)
4  0.111111  (HOA_Budget_Genre_True_True_Action, PREFERENCE_T)
0  0.101010              (HOA_Budget_Genre_False_True_Action)

Association Rules with Consequent 'PREFERENCE_T':
                          antecedents     consequents  antecedent support  \
0  (HOA_Budget_Genre_True_True_Action)  (PREFERENCE_T)            0.121212

   consequent support   support  confidence      lift  representativity  \
0            0.515152  0.111111    0.916667  1.779412               1.0

   leverage  conviction  zhangs_metric   jaccard  certainty  kulczynski
0  0.048669    5.818182       0.498433  0.211538   0.828125    0.566176
```

The algorithm finds that the combination of Budget = High, HOA = High, and Genre = Action is the strongest predictor for theater preference, with a confidence of 91.6% and a lift of 1.779. Metrics like conviction (5.818) reinforce the reliability of this rule, highlighting how Apriori captures complex patterns that single variables can’t.

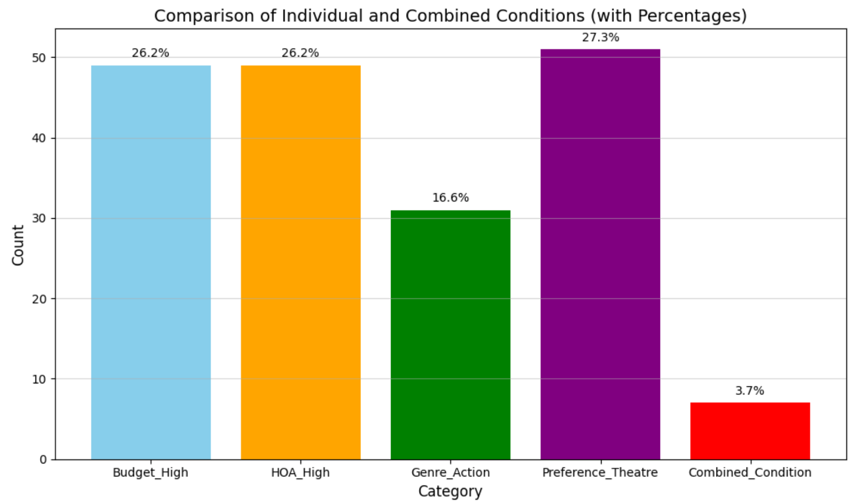


Fig 17 : Comparison of individual and combined conditions(with percentages)

While individual factors like Budget_High or HOA_High are more common (26.2% each), the combined condition (Budget = High, HOA = High, Genre = Action) is less frequent but has a stronger influence. This shows Apriori's strength in identifying niche yet impactful patterns for targeted campaigns.

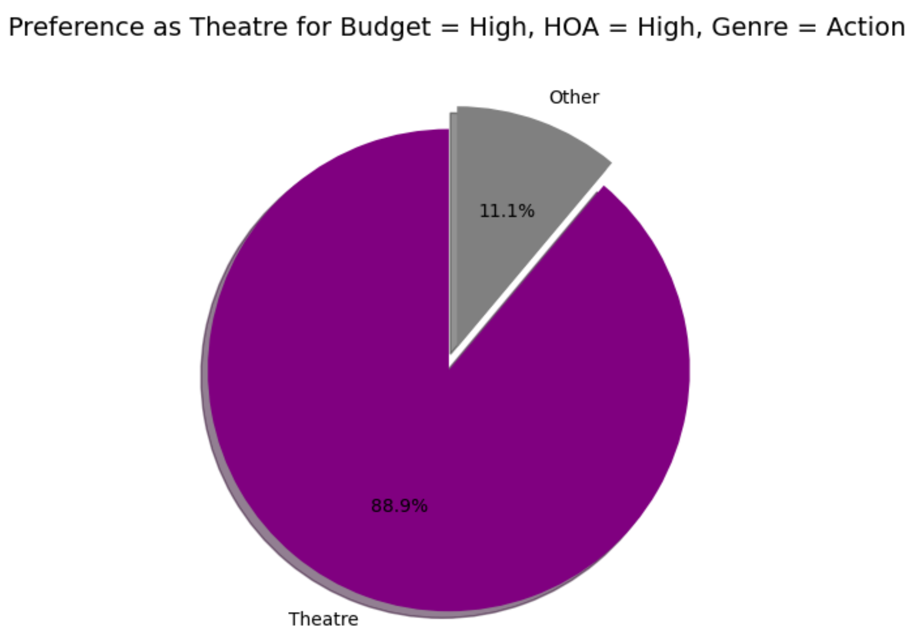


Fig.18: Viewer Preference for Theatres when Budget = High, HOA = High, and Genre = Action.

```
Frequent Itemsets:
     support                                       itemsets
0   0.505051                                   (Low_Budget)
1   0.505051                                      (Low_HOA)
2   0.141414                                 (Genre_Comedy)
3   0.484848                                (Preference_OTT)
4   0.282828                          (Low_Budget, Low_HOA)
5   0.131313                     (Genre_Comedy, Low_Budget)
6   0.313131                   (Low_Budget, Preference_OTT)
7   0.353535                      (Preference_OTT, Low_HOA)
8   0.111111                 (Genre_Comedy, Preference_OTT)
9   0.232323          (Low_Budget, Preference_OTT, Low_HOA)
10  0.111111  (Genre_Comedy, Low_Budget, Preference_OTT)

Association Rules with Consequent 'Preference_OTT':
                 antecedents                     consequents  \
4               (Low_Budget)                (Preference_OTT)
7                  (Low_HOA)                (Preference_OTT)
8             (Genre_Comedy)                (Preference_OTT)
11     (Low_Budget, Low_HOA)                (Preference_OTT)
13              (Low_Budget)       (Preference_OTT, Low_HOA)
15                 (Low_HOA)    (Low_Budget, Preference_OTT)
16  (Genre_Comedy, Low_Budget)              (Preference_OTT)
19            (Genre_Comedy)    (Low_Budget, Preference_OTT)
20              (Low_Budget)  (Genre_Comedy, Preference_OTT)

    antecedent support  consequent support   support  confidence      lift  \
4             0.505051            0.484848  0.313131    0.620000  1.278750
7             0.505051            0.484848  0.353535    0.700000  1.443750
8             0.141414            0.484848  0.111111    0.785714  1.620536
11            0.282828            0.484848  0.232323    0.821429  1.694196
13            0.505051            0.353535  0.232323    0.460000  1.301143
15            0.505051            0.313131  0.232323    0.460000  1.469032
16            0.131313            0.484848  0.111111    0.846154  1.745192
19            0.141414            0.313131  0.111111    0.785714  2.509217
20            0.505051            0.111111  0.111111    0.220000  1.980000

    representativity  leverage  conviction  zhangs_metric   jaccard  \
4                1.0  0.068258    1.355662       0.440421  0.462687
7                1.0  0.108662    1.717172       0.620991  0.555556
8                1.0  0.042547    2.404040       0.445989  0.215686
11               1.0  0.095194    2.884848       0.571341  0.433962
13               1.0  0.053770    1.197157       0.467613  0.370968
15               1.0  0.074176    1.271979       0.645075  0.396552
16               1.0  0.047444    3.348485       0.491543  0.220000
19               1.0  0.066830    3.205387       0.700535  0.323529
20               1.0  0.054994    1.139601       1.000000  0.220000
```

The analysis highlights that low-budget films with less experienced actors are well-suited for OTT platforms, catering to audiences seeking cost-efficient, non-mainstream content. Comedy stands out as the top genre for OTT (78.6% confidence, 1.62 lift), showing its strong appeal in the digital space. Low-budget comedies or productions with low HOA strengthen this preference further (84.6% confidence, 1.74 lift). These insights show how OTT platforms excel at promoting emerging talent and lighter, accessible content, making them a strategic choice for targeted distribution and marketing.

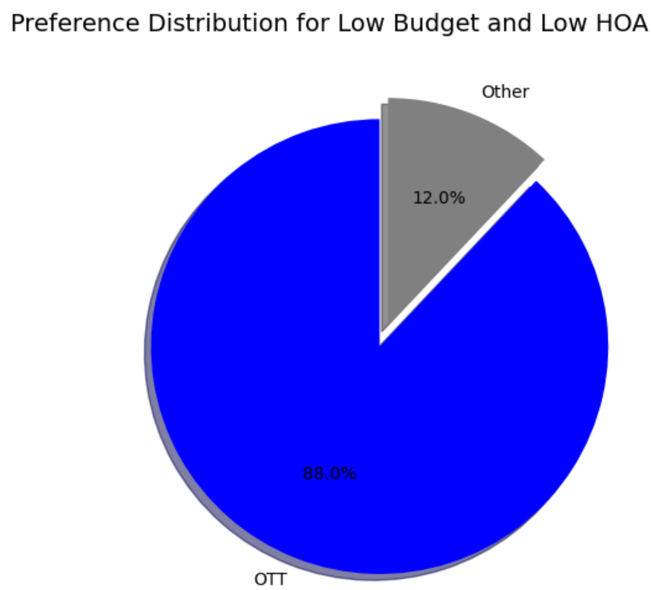


Fig.19: Preference Distribution for low budget and low HOA

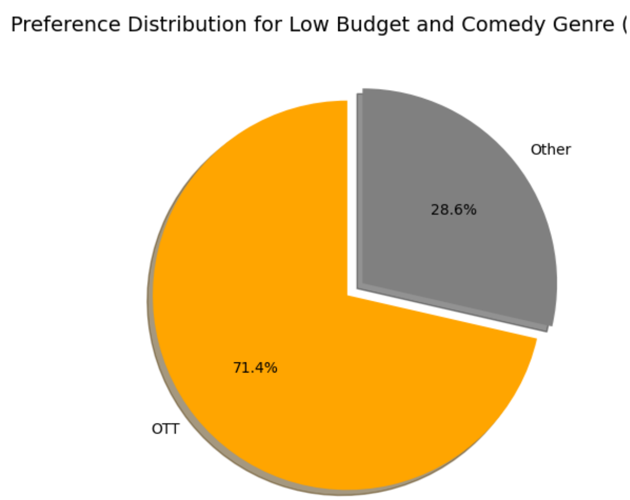


Fig. 20: Preference Distribution for low budget and comedy genre

```
Frequent Itemsets:
     support                                          itemsets
1   0.525253                                         (Low_HOD)
0   0.505051                                         (Low_HOA)
3   0.484848                                  (Preference_OTT)
2   0.414141                                  (High_MetaScore)
6   0.353535                         (Preference_OTT, Low_HOA)
8   0.303030                         (Low_HOD, Preference_OTT)
4   0.292929                                (Low_HOD, Low_HOA)
5   0.232323                         (High_MetaScore, Low_HOA)
11  0.232323                (Low_HOD, Preference_OTT, Low_HOA)
9   0.212121                  (High_MetaScore, Preference_OTT)
7   0.181818                         (Low_HOD, High_MetaScore)
12  0.161616         (High_MetaScore, Preference_OTT, Low_HOA)
13  0.131313         (Low_HOD, High_MetaScore, Preference_OTT)
10  0.121212                (Low_HOD, High_MetaScore, Low_HOA)
14  0.111111  (Low_HOD, High_MetaScore, Preference_OTT, Low_...

Association Rules with Consequent 'Preference_OTT':
                           antecedents                consequents  \
12  (Low_HOD, High_MetaScore, Low_HOA)           (Preference_OTT)
5                   (Low_HOD, Low_HOA)           (Preference_OTT)
9            (Low_HOD, High_MetaScore)           (Preference_OTT)
1                            (Low_HOA)           (Preference_OTT)
8            (High_MetaScore, Low_HOA)           (Preference_OTT)
14           (Low_HOD, High_MetaScore)  (Preference_OTT, Low_HOA)

    antecedent support  consequent support   support  confidence      lift  \
12            0.121212            0.484848  0.111111    0.916667  1.890625
5             0.292929            0.484848  0.232323    0.793103  1.635776
9             0.181818            0.484848  0.131313    0.722222  1.489583
1             0.505051            0.484848  0.353535    0.700000  1.443750
8             0.232323            0.484848  0.161616    0.695652  1.434783
14            0.181818            0.353535  0.111111    0.611111  1.728571

    representativity  leverage  conviction  zhangs_metric   jaccard  \
12               1.0  0.052342    6.181818       0.536050  0.224490
5                1.0  0.090297    2.489899       0.549689  0.425926
9                1.0  0.043159    1.854545       0.401709  0.245283
1                1.0  0.108662    1.717172       0.620991  0.555556
8                1.0  0.048975    1.692641       0.394737  0.290909
14               1.0  0.046832    1.662338       0.515152  0.261905

    certainty  kulczynski
12   0.838235    0.572917
5    0.598377    0.636135
9    0.460784    0.496528
1    0.417647    0.714583
```

The analysis highlights key patterns linking low director history (Low_HOD), high critical acclaim (High_MetaScore), and OTT preferences. Movies with less-established directors and actors but strong reviews are highly favored on OTT platforms, as seen in the rule (Low_HOD, High_MetaScore, Low_HOA) => Preference_OTT (91.67% confidence, 1.89 lift). Similarly, (Low_HOD, Low_HOA) => Preference_OTT (79.31% confidence, 1.63 lift) shows that lower-profile productions align well with streaming services.Additionally, (High_MetaScore, Low_HOA) => Preference_OTT (69.57% confidence, 1.43 lift) highlights the power of strong reviews in driving OTT engagement, even for movies with less-known actors. These insights emphasize that OTT platforms thrive on critically acclaimed content, regardless of production profiles, offering flexibility and reach for diverse films.

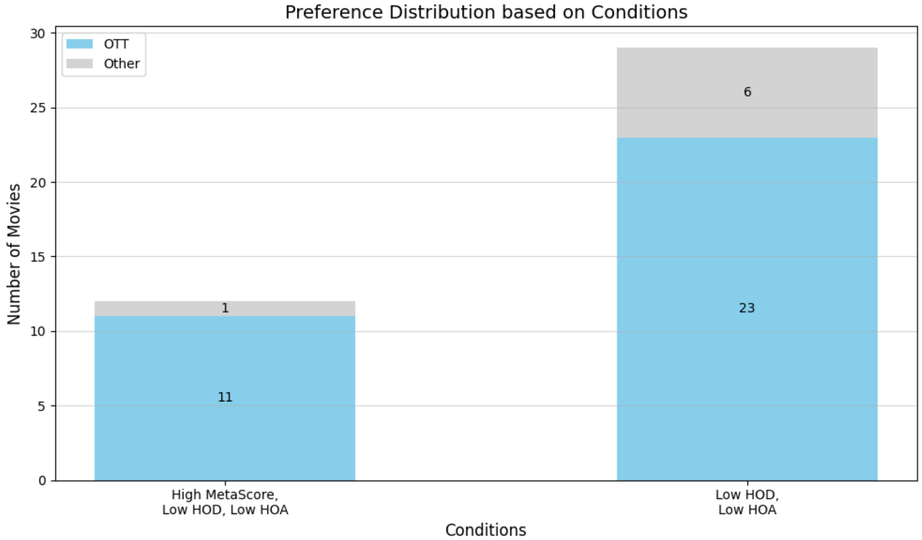


Fig. 21: Preference distribution based on conditions

NOVELTY AND SIGNIFICANCE

The rapid growth of OTT platforms has changed the entertainment industry, but it has also made it harder for producers and directors to predict audience preferences. This project uses the Apriori algorithm to uncover patterns in audience preferences based on factors.

The analysis offers practical insights: low-budget comedy films are popular on OTT platforms, while high-budget action blockbusters do well in theaters during peak seasons like summer. Adding demographic and regional data could improve strategies, and predictive models could estimate a film's success on different platforms. It could also consider factors like OTT subscription trends or social media sentiment to guide decision-making.

References

Mittal, M., Pareek, S. and Agarwal, R. (2014). Efficient Ordering Policy for Imperfect Quality Items Using Association Rule Mining. *Encyclopedia of Information Science and Technology, Third Edition*, pp.773–786. doi:https://doi.org/10.4018/978-1-4666-5888-2.ch074 .

Ormax Media Pvt. Ltd. (2024). *Sizing the cinema: India's theatre-going population in numbers*. [online] Available at: https://www.ormaxmedia.com/insights/stories/sizing-the-cinema-indias-theatre-going-population-in-numbers.htm l [Accessed 22 Dec. 2024].

Stanley, M. (2023). *M BluePaper*. [online] Available at: https://www.oliverwyman.de/content/dam/oliver-wyman/v2/publications/2023/october/Oliver_Wyman_Morgan_Stanley_Global_Wealth_and_Asset_Management_report_2023_The_Generative_AI_Tipping%20Point1.pdf [Accessed 22 Dec. 2024].

Statista (2022). *OTT Video - Worldwide | Statista Market Forecast*. [online] Statista. Available at: https://www.statista.com/outlook/amo/media/tv-video/ott-video/worldwide [Accessed 22 Dec. 2024].