

Netflix Top 10 Performance and Data Anomaly Analysis Report

Author: Manus AI **Date:** December 26, 2025

1. Analysis of 'Films (Non-English)' Category Performance

The primary objective of this section was to identify the top-performing film in the 'Films (Non-English)' category based on its cumulative weeks in the Top 10 and to assess the methodology for estimating its total viewership.

1.1. Film with the Most Weeks in the Top 10

Analysis of the `NFLX Top 10` dataset revealed that the film with the highest number of cumulative weeks in the Top 10 within the 'Films (Non-English)' category is **"Through My Window"**.

Film Title	Category	Maximum Cumulative Weeks in Top 10
Through My Window	Films (Non-English)	13
Gangubai Kathiawadi	Films (Non-English)	6
Black Crab	Films (Non-English)	6
RRR (Hindi)	Films (Non-English)	6
The Takedown	Films (Non-English)	6

1.2. Estimation of Viewership for "Through My Window"

To estimate the number of times "Through My Window" was watched, the total viewing hours are normalized by the film's runtime. This method converts the volume of time spent watching into an equivalent number of full views.

Data Points:

- **Total Weekly Hours Viewed (in dataset):** 10,380,000 hours
- **Film Runtime (from `Runtime` sheet):** 116 minutes

Calculation:

1. Convert runtime to hours: $116 \text{ minutes} / 60 \text{ minutes/hour} \approx 1.933 \text{ hours}$
2. Apply the estimation formula:
$$\text{Estimated Views} = \frac{\text{Total Weekly Hours Viewed}}{\text{Film Runtime (in hours)}} \times \text{Estimated Views} = \frac{10,380,000 \text{ hours}}{1.933 \text{ hours/view}} \approx 5,369,891 \text{ views}$$

The estimated number of views for "Through My Window" based on the provided data is approximately **5.37 million**.

1.3. Assumptions and Risks in Viewership Estimation

The estimation process relies on several key assumptions, each carrying inherent risks that must be acknowledged for accurate data interpretation.

Assumption	Description	Risk Involved
Full Watch Time	Assumes that every hour viewed contributes to a single, complete watch of the film.	Overestimation: The calculation is based on total hours, not unique users. A single user re-watching the film multiple times, or users not finishing the film, can inflate the "views" count relative to the actual number of unique viewers.
Consistent Runtime	Assumes the runtime of 116 minutes is a fixed and accurate duration for all viewing sessions across all regions.	Inaccuracy: Minor variations in runtime (e.g., due to regional cuts, or the exclusion of credits) could slightly skew the final estimate.
Data Completeness	Assumes the provided <code>weekly_hours_viewed</code> data captures all relevant viewing activity.	Underestimation: The data only includes viewing hours for weeks where the film was in the Top 10. Viewing hours from weeks before or after this period, or from regions not included in the Top 10 reporting, are excluded, leading to an underestimation of the film's true total views.
View-to-User Proxy	Assumes that the "Estimated Views" figure is a reasonable	Overestimation: The most significant risk is that this metric does not represent

	proxy for the number of unique users.	unique users, but rather the total number of times the film was consumed. The number of unique users will be lower than the estimated views.
--	---------------------------------------	--

2. Risks of Ignoring Outage Data from the Week of May 22nd

The data for the week of **May 22nd, 2022** exhibits a critical anomaly: the `weekly_hours_viewed` for all 40 titles in the Top 10 (10 in each of the four categories) are recorded as **0**. This pattern is a strong indicator of a data collection or reporting outage. Ignoring this week's data when calculating performance metrics introduces significant risks to the integrity of the analysis.

2.1. Impact on Key Performance Metrics

Metric	Risk of Ignoring Outage Data	Impact on Metric
Total Weekly Hours Viewed	Inaccurate Summation: The total viewing hours for any film that was in the Top 10 during this week (e.g., <i>Gangubai Kathiawadi</i> , <i>RRR (Hindi)</i> , <i>The Takedown</i>) will be artificially lower than the true value.	Systematically Understated: The total viewing hours will be understated by the actual number of hours viewed during the outage week.
Estimated Views	Skewed Viewership Estimate: As the total hours are understated, the resulting estimate for the number of views will also be artificially low.	Understated: The estimated number of views will be understated, leading to a false impression of lower popularity and engagement.
Average Weekly Hours	Severe Skew: While the <code>cumulative_weeks_in_top_10</code> count is correctly incremented in the provided data, any metric relying on the <i>average</i> weekly hours (e.g., average hours per week in Top 10) will be severely skewed downwards due to	Skewed Downwards: The average performance of affected titles will be misrepresented as significantly lower than reality.

	the inclusion of a zero-value week.	
Comparative Analysis	Misleading Performance Comparison: Films that were in the Top 10 during the outage week will appear to have significantly lower average weekly hours than comparable films that were not in the Top 10 during that specific week.	Unreliable: Comparisons between titles will be unreliable, leading to incorrect business conclusions about relative performance and content value.

2.2. Mitigation Strategy

To maintain data integrity, the preferred strategy is **data imputation**. The missing `weekly_hours_viewed` data for the week of May 22nd should be estimated rather than ignored. A robust imputation method would involve:

1. **Averaging:** Using the average of the viewing hours from the week immediately preceding and the week immediately following the outage for each affected title.
2. **Exclusion from Averages:** If imputation is not feasible, the zero-value week should be explicitly excluded from any calculation of *average* weekly hours to prevent the systematic downward skewing of performance metrics.

References

[1] Netflix. (2022). NFLX Top 10, IMDB Rating, and Runtime Data [Excel Spreadsheet]. Provided by User.

[2] Netflix. (2022). Data Support Specialist NFLX Exercise Instructions [Document]. Linked by User.