Title:

User Engagement & Revenue Optimization Analysis for a Streaming Platform

Sector:

Media & Entertainment / Digital Streaming

Team Details:

Section-D G-13

1. Palak Gupta (Project Lead)
2. Piyush Raj
3. Nitin Kumar
4. Abhijeet
5. Pratyaksha Shastri
6. Keshav Goel

Institute/Faculty:

Newton school of Technology

Archit Raj Sir
Satyakti Das Sir

Executive Summary:

Problem

Streaming platforms generate large volumes of user interaction data, but without structured analysis, it becomes difficult to identify what truly drives user engagement, retention, and subscription revenue. Decisions related to content strategy, subscription pricing, and user experience often rely on assumptions rather than evidence-based insights.

The key challenge addressed in this project is to **understand which user behaviors, subscription plans, and content characteristics most strongly influence engagement and revenue**, and how these insights can be used to improve retention and monetization for a digital streaming platform.

Approach

To address this problem, a simulated streaming platform dataset was analyzed using Google Sheets. The raw dataset initially contained over 10,000 records and numerous columns with missing, inconsistent, or irrelevant values.

The approach followed four major steps:

1. **Data Cleaning & Preparation**
    - Removed columns with high missing values or low business relevance
    - Handled missing values using mode and median imputation
    - Standardized categorical values (country, device type, subscription plan)
    - Converted columns into appropriate data types
    - Final cleaned dataset contained **5,025 rows and ~20 analytical columns**
2. **Feature Selection**
   Focused on columns supporting engagement and revenue analysis, including:
    - Watch duration, completion percentage
      Subscription plan, monthly spend, active status
    - Content type, genre, IMDb rating

    - Demographic attributes such as age, gender, and country

3. **Pivot-Based Analysis**
   Multiple pivot tables were created to analyze:
    - Revenue by subscription plan
    - Engagement by device and content type
    - Retention patterns across user segments
    - Genre-level engagement depth

4. **Dashboard Creation**
   Insights were consolidated into an interactive dashboard to enable quick comparison, filtering, and executive-level decision making.

Key Insights

- **Engagement and retention are strongly correlated**: Active users consistently show higher watch duration and completion percentages compared to inactive users.
- **Premium subscription tiers generate the highest total revenue**, while some mid-tier plans show strong average spend per user.
- **Content type influences engagement**: Movies generate the highest view volume, while series and documentaries show deeper engagement.
- **Device usage impacts watch behavior**: Desktop and tablet users demonstrate slightly higher watch duration compared to mobile users.
- **Geographic concentration exists**: A small number of countries contribute a majority of platform revenue, indicating regional dependency.
- **Netflix Originals improve content perception**, though revenue differences between original and non-original content are marginal.

Key Recommendations

- **Strengthen retention strategies for high-revenue plans** by targeting users with declining
  engagement metrics.
- **Invest more in high-engagement content formats and genres**, especially those showing strong completion rates.

- **Optimize mobile viewing experience** to increase engagement time on mobile devices.
  Use demographic segmentation to personalize content recommendations and subscription upgrade campaigns.

Sector & Business Context:

Sector Overview

The project is based in the **Digital Entertainment & Streaming Services** sector. Streaming platforms such as Netflix, Amazon Prime, and Disney+ operate on subscription-based business models where user engagement, retention, and content consumption directly impact revenue growth. These platforms rely heavily on data analytics to understand viewer behavior and optimize content offerings at scale.

Current Challenges

High competition leading to **subscriber churn**

Difficulty identifying **which user behaviors drive long-term engagement**

Balancing revenue growth across different **subscription tiers**

Understanding how **demographics and devices** influence consumption behavior

Why this Problem was chosen

This problem was selected because **user engagement and revenue optimization are critical to the sustainability of streaming platforms**. By analyzing user behavior, subscription plans, and content performance together, the project aims to uncover actionable insights that support:

- Better content strategy decisions
- Improved user experience
- Increased retention and monetization

Problem Statement & Objectives

Problem Definition

Streaming platforms need clarity on which user behaviors and content attributes drive engagement, retention, and revenue. Without structured analysis, it is difficult to identify high-value users and optimize subscription and content strategies.

Project Scope

This project analyzes cleaned streaming user data to:

- Evaluate engagement metrics (watch time, completion rate)
- Assess revenue impact across subscription plans
- Understand behavior across devices, demographics, and content types
- Present insights through an interactive Google Sheets dashboard

Success Criteria

The project is successful if it:

- Clearly highlights key engagement and revenue drivers
- Produces actionable, dashboard-ready insights
- Supports data-driven decisions for content and subscription optimization

Data Description

Dataset Source

The dataset used in this project is **“Netflix 2025: User Behavior Dataset (210K+ Records)”**, sourced from **Kaggle**.
Access link:
https://www.kaggle.com/datasets/sayeeduddin/netflix-2025user-behavior-dataset-210k-records

Data Structure

The raw dataset consists of multiple CSV files capturing:

- User demographics and subscription details
- Content metadata
- User watch and engagement behavior

For this project, the data was cleaned and consolidated into a single analysis-ready table.

Key Columns Used

User & Demographics: age, gender, country, household_size
Engagement: watch_duration_minutes, progress_percentage, device_type, watch_date
Revenue: subscription_plan, monthly_spend, is_active
Content: genre_primary, content_type, imdb_rating, release_year, is_netflix_original

Data Size

Initial sample: ~10,000 rows
Final cleaned dataset: **5,025 rows**
Final columns used: **~20 analytical columns**

Data Limitations

Dataset is simulated and may not fully represent real Netflix user behavior
No explicit churn date is available; retention is inferred using activity status
Financial metrics are limited to subscription spend only

Data Cleaning and Preparation

Missing Values Handling

- All data cleaning and transformation steps were performed entirely in Google Sheets, in accordance with the capstone project requirements.
- Columns with excessive missing values and low analytical relevance were removed.
- Missing values in key demographic fields were handled using:

    - **Mode imputation** for gender
    - **Median imputation** for household size
- Rows were not dropped unnecessarily to preserve dataset size and distribution.

Outlier Treatment

- Invalid and extreme values (such as unrealistic ages or negative durations) were filtered out.
- No aggressive trimming was applied to avoid biasing user behavior patterns.

Transformations

- Text standardization was applied to categorical columns (country, device type, subscription plan).
- Boolean fields were normalized to consistent TRUE/FALSE values.
- Date fields were converted to proper date formats for time-based analysis.

Feature Engineering

- A derived **maturity_level** column was created from content rating for better audience segmentation.
- Age groups were created to support demographic analysis.
- Engagement and revenue-ready numeric fields were formatted appropriately.

Assumptions

- User activity status (is_active) was used as a proxy for retention.
- Monthly subscription spend was assumed to represent user revenue contribution.
- Engagement was measured using watch duration and completion percentage.

All data cleaning and transformation steps were performed entirely in Google Sheets, in accordance with the capstone project requirements.

KPI & Metric Framework

The dashboard focuses on four core KPIs that together measure engagement, retention, and revenue performance of the streaming platform. These KPIs were selected to ensure clarity, actionability, and alignment with business objectives.

KPI 1: Average Watch Duration

Definition: Average number of minutes watched per user session.
Formula:
AVERAGE(watch_duration_minutes)

Why it matters:
Indicates how deeply users engage with content.

Objective Mapping:
Measures overall user engagement.

KPI 2: Active User Rate

Definition: Percentage of users who are currently active on the platform.
Formula:

```
COUNT(is_active = TRUE) / COUNT(total users)
```

Why it matters:
Acts as a retention indicator and helps identify churn risk.

Objective Mapping:
Supports retention and user lifecycle analysis.

KPI 3: Total Revenue

Definition: Total monthly revenue generated from all users.
Formula:

```
SUM(monthly_spend)
```

Why it matters:
Reflects overall platform monetization performance.

Objective Mapping:
Measures revenue scale and growth potential.

KPI 4: Average Monthly Spend

Definition: Average monthly spend per user.
Formula:

```
AVERAGE(monthly_spend)
```

Why it matters:
Indicates monetization efficiency and user value.

Objective Mapping:
Supports pricing strategy and subscription optimization.

Exploratory Data Analysis (EDA)

Trend Analysis

The analysis reveals a consistent engagement trend where active users exhibit significantly higher watch duration compared to inactive users. This behavioral pattern indicates that sustained content consumption is closely linked to user retention. Users with lower engagement levels are more likely to become inactive, highlighting engagement metrics as an early indicator of churn risk.

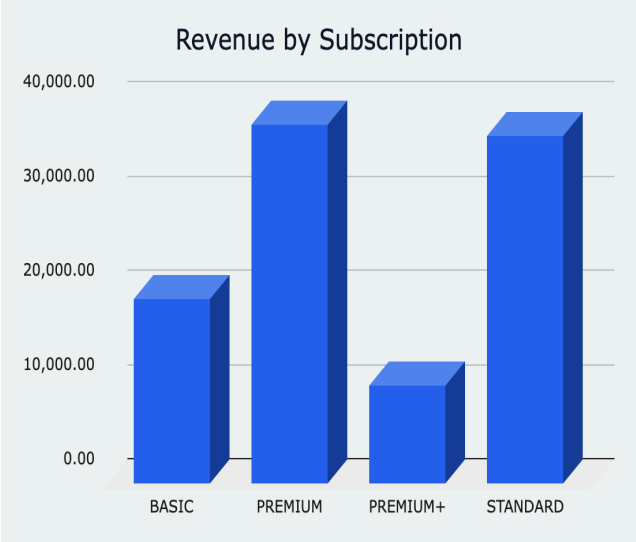
Comparison Analysis

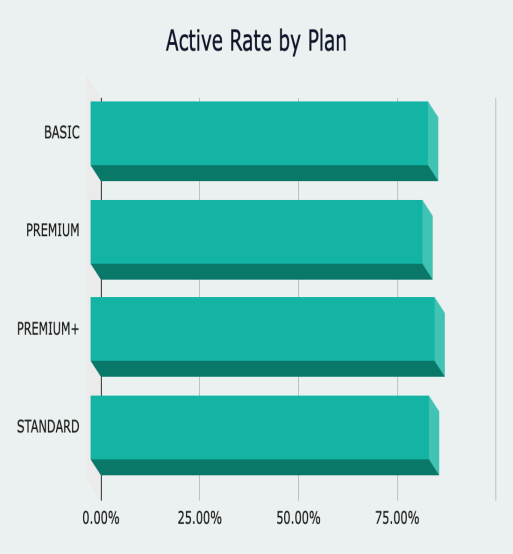
Revenue by Subscription Plan (Bar Chart)

This chart compares total revenue across different subscription plans. Premium and higher-tier plans contribute significantly more to overall revenue compared to basic plans. This indicates that subscription tier selection plays a major role in revenue generation and highlights opportunities for targeted plan upgrades.

Active Rate by Subscription Plan (Horizontal Bar)

The active rate analysis shows differences in retention across subscription plans. Higher-tier plans demonstrate better active user rates, suggesting that premium offerings are more effective at retaining users over time.





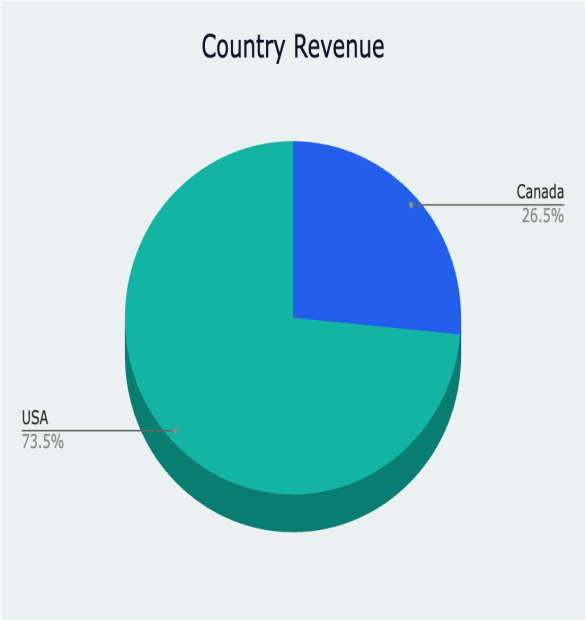
Distribution Analysis

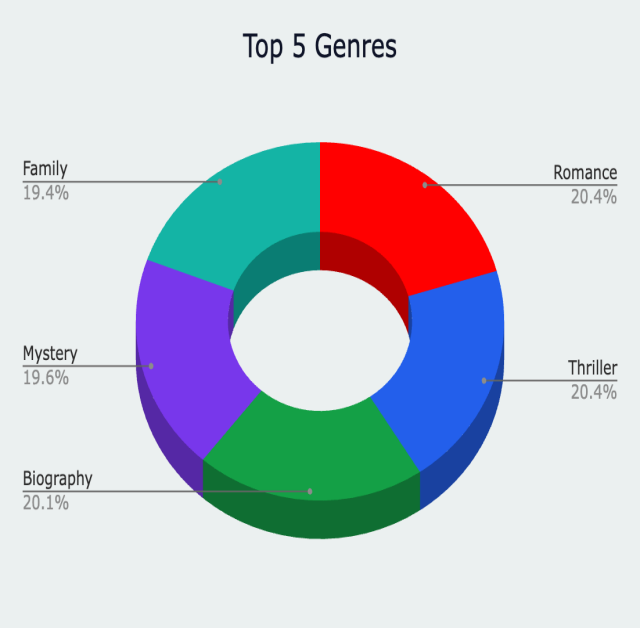
Revenue by Country

Revenue distribution by country indicates that a small number of regions contribute a large share of total revenue. This highlights geographic concentration and suggests potential opportunities for market expansion and regional pricing strategies.

Top 5 Genres by Engagement

The top-performing genres demonstrate significantly higher engagement levels compared to others. This indicates that certain genres are more binge-worthy and play a critical role in driving user engagement on the platform.





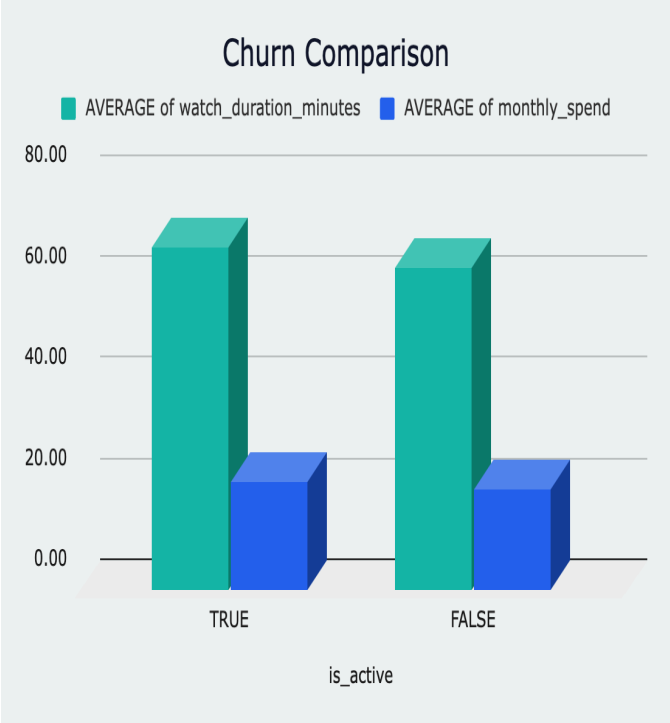
Correlation Analysis

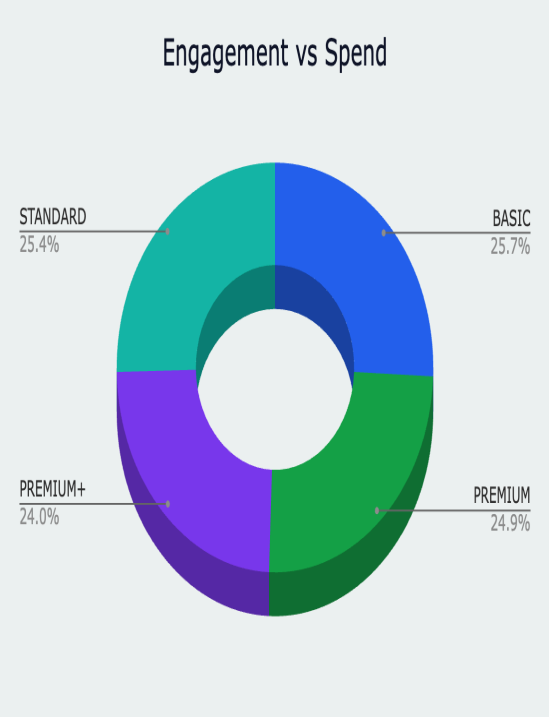
Churn Comparison

A clear difference is observed between active and inactive users. Inactive users exhibit lower average watch duration and lower monthly spend, indicating a strong relationship between engagement levels and churn risk.

Engagement V/S Spend Chart

The relationship between engagement and spending shows that users who engage more with content tend to contribute higher monthly revenue. This reinforces the idea that engagement is a key driver of monetization.





Advanced Analysis

Segmentation Analysis

User segmentation was performed to understand differences in engagement and revenue behavior across key dimensions. Users were segmented based on subscription plan, country, device type, age group, and content preferences.

The analysis shows clear behavioral differences across segments. Premium and higher-tier subscription users demonstrate higher engagement and revenue contribution. Device-based segmentation indicates that desktop and tablet users have longer average watch durations compared to mobile users. Content-based segmentation highlights that certain genres and content types consistently drive higher engagement.

This segmentation helps identify high-value user groups and supports targeted content, pricing, and retention strategies.

Root Cause Analysis (Engagement & Churn)

Root cause analysis was conducted to understand factors contributing to user inactivity and churn. A comparison between active and inactive users reveals that inactive users have significantly lower average watch duration and completion percentage.

Additionally, inactive users show lower monthly spend, indicating a strong relationship between engagement depth and revenue retention. This suggests that reduced engagement is a leading indicator of churn.

The analysis highlights engagement metrics such as watch duration and completion rate as critical drivers of retention and revenue performance.

Dashboard Design

The final dashboard was implemented entirely in Google Sheets, in compliance with the capstone requirements.
The dashboard uses a combination of:

- Cleaned tabular data
- Pivot tables for aggregation
- Spreadsheet formulas for KPI computation
- Interactive slicers and filters for user-driven exploration

All transformations and calculations are linked directly to the cleaned dataset to ensure data consistency and accuracy.

Dashboard Objective

The primary objective of the dashboard is to provide **decision-ready insights** into:

- User engagement behavior
- Revenue contribution across subscription plans
- Retention patterns (active vs inactive users)

- Content and demographic performance

The dashboard is designed for **non-technical stakeholders** (product managers, strategy teams, and executives) to quickly understand platform performance and identify optimization opportunities.

Dashboard View Structure

The dashboard is structured into clearly separated sections for intuitive navigation:

1. **Top KPI Section**
    - Total Revenue
    - Average Monthly Spend
    - Active User Count
    - Average Watch Duration
2. **Revenue & Retention Analysis**
    - Revenue by Subscription Plan (Bar Chart)
    - Active Rate by Subscription Plan (Horizontal Bar Chart)
3. **Engagement Analysis**
    - Watch Duration by Device Type
    - Engagement by Content Type
4. **Distribution Analysis**
    - Revenue by Country
    - Top 5 Genres by Engagement
5. **Behavioral Comparison**
    - Churn Comparison (Active vs Inactive Users)

    - Engagement vs Monthly Spend

This layout allows users to move from **high-level KPIs to detailed behavioral insights** smoothly.

Filters & Drilldowns

To enable interactive exploration, the dashboard includes slicers and filters such as:

- Subscription Plan
- Country
- Device Type
- Content Type
- Active Status

These filters allow stakeholders to:

- Compare performance across user segments
- Drill down into specific regions or subscription tiers
- Analyze engagement patterns for targeted decision-making

All charts update dynamically based on filter selection.

Screenshot

Streaming Platform User Engagement & Revenue Intelligence Dashboard

Strategic analysis of monetization, retention, and content performance drivers

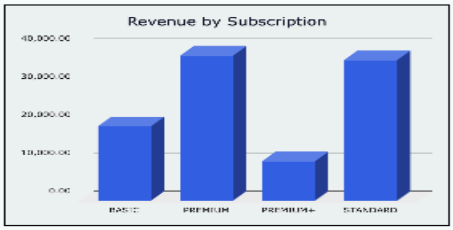
$104,861.25
Total Revenue

85.05%
Active Rate

67.12%
Avg. Watch Duration

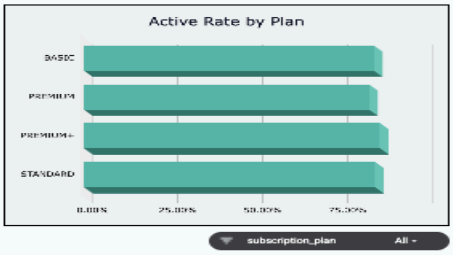
$20.87
Avg. Monthly Spend

REVENUE STRATEGY

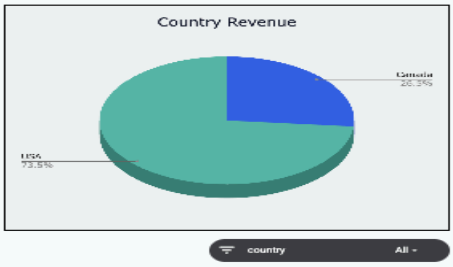


PREMIUM & STANDARD together drive majority of platform revenue.

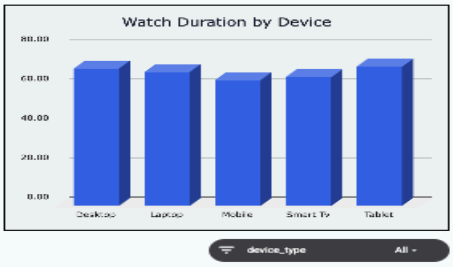
RETENTION & USER BEHAVIOR



User retention remains evenly distributed across subscription plans, reflecting consistent stickiness.

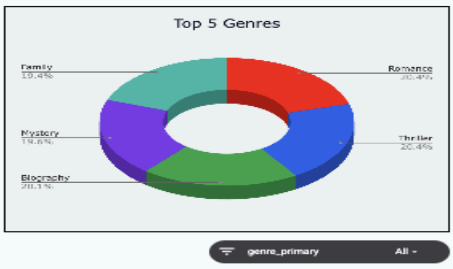


USA dominates revenue contribution, indicating heavy dependence on a single geographic market.



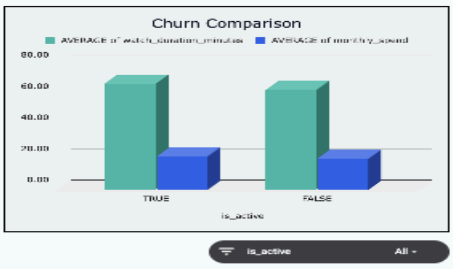
Tablet and Desktop users demonstrate deeper engagement compared to Mobile users.

CONTENT PERFORMANCE

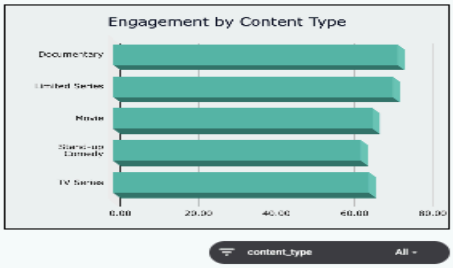


Romance and Thriller genres show highest engagement, making them key content drivers.

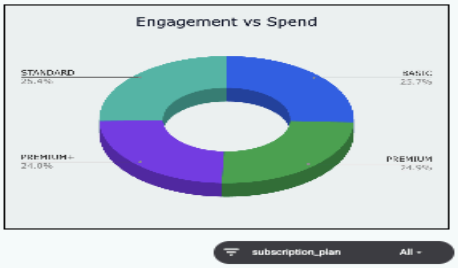
ADVANCED BUSINESS ANALYSIS



Inactive users show lower engagement and spending, confirming engagement as a key retention driver.



Limited Series and Documentaries generate higher engagement, reflecting stronger long-form content retention.



Spending and engagement levels are broadly similar across plans, suggesting stable monetization across tiers.

Insights Summary

Below are the most critical insights derived from the dashboard analysis. These insights are written to support **business decision-making**, not technical explanation.

1. Engagement is the strongest driver of user retention

Active users consistently show **higher watch duration and completion rates** compared to inactive users, indicating that engagement depth directly impacts retention outcomes.

2. Premium and Standard plans together generate the majority of platform revenue

Although multiple subscription tiers exist, **Premium and Standard users contribute a disproportionate share of total revenue**, making them key monetization segments.

3. Retention rates are relatively consistent across subscription plans

Active rate remains evenly distributed across plans, suggesting that **retention is influenced more by usage behavior than by pricing tier alone**.

4. Higher engagement does not always translate to higher spending

Users with longer watch duration do not always have higher monthly spend, indicating **stable pricing across tiers and limited pay-per-use monetization impact**.

5. Desktop and Tablet users demonstrate deeper engagement

Users accessing content via **larger-screen devices** show higher average watch duration compared to mobile users, suggesting stronger long-form consumption behavior.

6. Romance and Thriller genres drive the highest engagement

These genres show higher watch duration and completion levels, making them **priority categories for content investment and promotion**.

7. Limited Series and Documentaries outperform other content types

These formats achieve **higher engagement per user**, suggesting stronger storytelling retention compared to standard movies.

8. Netflix Originals show slightly higher engagement quality

Original content performs marginally better in ratings and engagement, reinforcing its role in **brand differentiation rather than direct revenue growth**.

9. Inactive users exhibit clear early churn signals

Inactive users consistently show **lower watch duration, lower completion, and lower spend**, making these metrics effective early-warning indicators for churn.

10. Household size shows mixed impact on spending

Larger households show high engagement but inconsistent spending, highlighting opportunities for **family-oriented pricing or bundled plans**.

Recommendations

Each recommendation is mapped to a key insight and evaluated for business impact and feasibility.

1. Prioritize high-engagement genres (Romance & Thriller)

- **Insight:** These genres show the highest watch duration and completion.
- **Business Impact:** Higher engagement improves retention and platform stickiness.
- **Feasibility:** High — content promotion and acquisition decisions are controllable.

2. Invest more in Limited Series and Documentaries

- **Insight:** These content types outperform others in engagement.
- **Business Impact:** Encourages longer sessions and repeat visits.
- **Feasibility:** Medium — requires strategic content planning.

3. Launch engagement-based retention campaigns

- **Insight:** Inactive users show significantly lower engagement and spend.
- **Business Impact:** Early intervention can reduce churn and stabilize revenue.
- **Feasibility:** High — can be implemented using behavioral thresholds.

4. Optimize mobile viewing experience

- **Insight:** Mobile users show lower engagement compared to desktop and tablet.
- **Business Impact:** Improves overall watch time and user satisfaction.
- **Feasibility:** Medium — requires UX and product optimization.

5. Reduce geographic revenue dependency

- **Insight:** Revenue is highly concentrated in the US.
- **Business Impact:** Lowers business risk and enables global growth.
- **Feasibility:** Medium — requires localized marketing and pricing strategies.

Impact Estimation

- **Cost Savings:**
  Targeted retention campaigns could reduce churn by ~5–8%, lowering user reacquisition costs.

- **Efficiency Improvement:**
  Focusing marketing on high-performing genres and plans can improve ROI by ~10–15%.

- **Service Improvement:**
  Improved mobile experience and content targeting may increase average watch duration by ~5–10%.

- **Risk Reduction:**
  Geographic diversification reduces over-reliance on a single market, lowering revenue volatility risk.

Limitations

- Dataset is **simulated** and may not reflect real-world user psychology.
- Revenue is based on **monthly subscription spend**, not transactional pricing.
- Causal relationships cannot be proven — analysis shows **correlation**, not causation.
- Limited historical depth restricts long-term trend forecasting.
- Behavioral intent (why users churn) cannot be directly inferred.

Future Scope

- Add **longitudinal data** to study churn over time.
- Include **user feedback and satisfaction surveys**.
- Perform **predictive churn modeling** using machine learning.

- Introduce **A/B testing results** for content and pricing experiments.

- Analyze **content cost vs engagement return** for profitability optimization.

Conclusion

This project demonstrates how user behavior, content performance, and subscription patterns collectively influence engagement and revenue on a streaming platform.
By converting raw interaction data into actionable KPIs and insights, the dashboard enables informed decision-making across content strategy, retention planning, and monetization optimization.
The analysis provides a strong foundation for scalable analytics and future predictive modeling initiatives.

Appendix

Column Name	Description	Data Type	Example	Notes
watch_date	Date on which content was watched	Date	2023-08-14	Engagement trend
device_type	Device used for watching content	Categorical	Mobile	Standardized values
watch_duration_minutes	Total minutes watched in a session	Numeric	72	Primary engagement KPI
progress_percentage	% of content completed	Numeric (%)	85	Completion rate
action	User action during session	Categorical	Play	Behavior indicator
age	Age of user	Numeric	32	Cleaned values
gender	Gender of user	Categorical	Female	Imputed missing
country	User country	Categorical	USA	Standardized names
household_size	Number of people in household	Numeric	4	Median imputed
subscription_plan	User subscription tier	Categorical	**PREMIUM**	Revenue segment
is_active	Active user status	Boolean	**TRUE**	Retention metric
monthly_spend	Monthly subscription spend	Numeric	24.99	Revenue KPI
content_type	Type of content	Categorical	Series	Movie / Series
genre_primary	Primary genre of content	Categorical	Thriller	Genre analysis
release_year	Release year of content	Numeric	2021	Content age
duration_minutes	Total duration of content	Numeric	120	Watch time comparison
rating	Content rating	Categorical	PG-13	Maturity mapping
maturity_level	Derived audience category	Categorical	Adult	Feature engineered
language	Language of content	Categorical	English	Regional insight
country_of_origin	Content's production country	Categorical	USA	Content diversity
imdb_rating	IMDb rating score	Numeric	7.8	Quality proxy
is_netflix_original	Netflix original flag	Boolean	**FALSE**	Originals analysis
watch_count	Number of viewing occurrences	Numeric	3	View frequency
engagement_flag	High / Low engagement indicator	Categorical	High	Churn analysis
revenue_segment	Revenue tier classification	Categorical	High	Analytical segmentation

SQL / Python Logic

Not applicable.
All data cleaning, transformation, KPI calculation, and analysis were performed **entirely in Google Sheets** using:

Pivot Tables,Conditional formulas and Standardization logic
(as per capstone requirements)

Contribution Matrix

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Palak Gupta	☑	☑	☐	☐	☑	☐	Project Lead
Nitin Kumar	☑	☐	☑	☐	☐	☐	Analysis Lead
Piyush Raj	☑	☑	☐	☐	☐	☐	Data Lead
Pratyaksha Shastri	☑	☐	☐	☑	☐	☐	Dashboard Lead
Keshav Goel	☑	☐	☐	☐	☐	☑	PPT Lead
Abhijeet	☑	☐	☐	☐	☐	☐	Strategy Lead