



STEAM PLATFORM ANALYSIS

Analysis of Engagement, Popularity, and Market Structure on the
Steam Platform

Sector: Digital Gaming Industry

Group: G-2

Team Members:

Rajat Srivastav 2401010256

Mahir Abdullah 2401010256

Krishiv Gupta 2401010235

Manas Selukar 2401010259

Ankit Singh 2401010075

Kavya Katal 2401020031

Executive Summary

This project analyzes Steam game data to understand how game attributes such as pricing model, user feedback, platform accessibility, and content structure influence player engagement and market outcomes.

The analysis focuses on identifying patterns in player activity, user satisfaction, and catalog concentration using derived metrics, KPIs, and pivot-based analysis implemented entirely in Google Sheets.

Key findings show that the Steam ecosystem is highly skewed, with a very small number of blockbuster titles capturing a disproportionate share of player attention and reviews. Free-to-Play games demonstrate significantly higher average concurrent player activity, while paid games show slightly stronger player satisfaction metrics. Social proof, measured through recommendations, emerges as a major driver of popularity.

Based on these insights, the project provides recommendations around pricing strategy, discovery optimization, and portfolio balancing for game publishers and platform decision-makers.

Sector & Business Context

Steam operates as a large-scale digital marketplace where thousands of games compete for limited player attention. Decision makers on such platforms must balance discoverability, monetization, and long-term engagement while managing an extremely skewed catalog.

The gaming sector faces challenges such as (but not limited to):

- Oversupply of niche titles
- Heavy concentration of engagement in a few top games
- Difficulty in predicting success based on early signals

This problem was chosen to simulate how data driven analysis can support platform strategy, pricing decisions, and content investment planning.

Problem Statement & Objectives

Problem Statement

How do pricing models, content structure, platform accessibility, and user feedback influence game popularity, engagement, and satisfaction on a steam marketplace?

Objectives

- Identify engagement patterns across pricing models
- Measure market concentration using popularity tiers
- Evaluate the role of social proof and accessibility
- Support strategic decision-making using measurable KPIs

Success Criteria

- Clear, defensible KPIs
- Insightful segmentation beyond simple averages
- Actionable recommendations supported by data

Data Description

The dataset used in this project was sourced from **Kaggle** and contains publicly available **Steam game metadata** collected from the Steam platform.

Link: <https://www.kaggle.com/datasets/artermiloff/steam-games-dataset>

Data Size

- **Original dataset:** ~50,000 game records; **Columns:** 47
- **Working dataset (post selection, cleaning):** ~7,000 game records; **Columns:** 42 columns

*Additional analytical columns are introduced later during analysis and are documented separately in the **Data Cleaning & Preparation** section.*

Metric Legend (Used across the report)

Metric	Description	Analytical Use
Peak CCU	Max concurrent players	Popularity & reach
PCT POS Total	Percentage of positive review	Direct measure of player satisfaction
Total Reviews	Review volume	Engagement scale
Weighted Sentiment	Volume-adjusted satisfaction	Credible sentiment
Pricing Type	Free vs Paid	Monetization analysis
Platform Count	OS support count	Accessibility
Popularity Tier	Engagement segmentation	Market structure
Recommendation Band	Social proof level	Discovery & growth

Data Structure

Each row in the dataset represents **one Steam game**.

The dataset is structured across the following major dimensions:

- Game identity (App ID, Name, Release Date)
- Pricing and monetization
- Player engagement and popularity
- User feedback and sentiment
- Platform and accessibility support
- Content classification (genres, categories)
- Social proof and discoverability indicators

Key Columns Used

The analysis primarily focuses on the following columns:

- **Price**: upfront cost of the game
- **Peak CCU**: peak concurrent users, used as the primary popularity indicator
- **Positive & Negative Reviews**: raw user feedback volume
- **PCT POS Total**: overall user satisfaction percentage
- **Recommendations**: proxy for word-of-mouth and social proof
- **Supported Platforms (Windows / Mac / Linux)**: accessibility indicators
- **Genres & Categories**: content classification and breadth
- **Estimated Owners**: long-term reach proxy (where available)

(These columns form the foundation for all KPIs, pivot tables, and dashboard views.)

Data Limitations

- The dataset does **not include revenue or marketing spend data**, limiting direct monetization analysis
- Playtime-related fields are **partially missing** across games
- Analysis limited to the **Steam ecosystem** only.

(These limitations are addressed where possible through proxy metrics and are discussed further in later sections.)

Data Cleaning & Preparation

All cleaning and preparation was performed in **Google Sheets**, as required.

Major Cleaning Steps

- Removed non-analytical text and media columns
- Standardized data types (dates, numbers, currency, percentages)
- Normalized boolean platform indicators
- Converted array-based fields into count metrics
- Standardized naming conventions across columns

Handling Missing Values

- Playtime fields contained significant missing data
- Missing values were filled with column averages for consistency
- These fields were later treated cautiously during analysis

(Detailed cleaning and preparation process can be found in the log files - which can be found in the project's github repository along with everything else)

Link: https://github.com/mahir-m01/SectionC_Group2_SteamGames

Feature Engineering

To enable deeper analysis beyond raw platform metrics, additional analytical columns were derived from the cleaned dataset. These derived fields transform raw attributes into interpretable indicators aligned with the project's business objectives.

Derived Columns Created

- **Total Reviews**

Calculated as the sum of positive and negative reviews to represent overall engagement volume.

- **Weighted Sentiment**

Computed as $\text{pct_pos_total} \times \ln(\text{total_reviews} + 1)$ to adjust sentiment scores by review volume and reduce low-sample bias.

- **Pricing Type**

Categorized games as Free-to-Play or Paid to support monetization and reach comparisons.

- **Platform Count**

Derived by summing Windows, Mac, and Linux indicators to measure platform accessibility.

- **Popularity Tier**

Games were segmented into Blockbuster, Hit, Mid, and Niche

categories based on Peak CCU thresholds to support market structure analysis.

- **Discount Impact**

Calculated as $\text{discount} \times \text{peak_ccu}$ to explore the relationship between discounting and short-term player activity.

- **Price Band**

Grouped games into interpretable pricing ranges (Free, Low, Mid, High) for pricing impact analysis.

- **DLC Band**

Classified games by post-launch content depth to analyze monetization strategies.

- **Recommendation Band**

Grouped games based on recommendation volume to evaluate social proof and discoverability effects.

Weighted Sentiment Deep Dive

The metric multiplies percentage positive reviews by the log of total reviews to balance sentiment quality with engagement credibility. Percentage positive reflects player satisfaction, while review volume indicates reliability of that sentiment. The logarithmic scale (\ln) reduces the dominance of extremely high review counts, and the $+1$ ensures the metric remains defined for games with zero or minimal reviews.

These engineered features form the foundation for KPI construction, pivot table analysis, and dashboard development.

KPI & Metric Framework

KPIs were designed to directly support the project objective and answer core business questions around engagement, satisfaction, market structure, and discovery.

Metrics are grouped into thematic KPI categories and analyzed dynamically using interactive slicers, enabling flexible comparisons across pricing models, recommendation tiers and more.

KPI Categories, Definitions & Relevance

Player Activity & Engagement

KPI: Average Peak CCU

Formula: Average of Peak CCU by Pricing Type

Why it matters: Measures short-term popularity and player attention, enabling comparison of reach across monetization models.

Objective: Identify which pricing model drives higher engagement

Player Satisfaction

KPI: Weighted Sentiment

Formula: $pct_pos_total \times \ln(\text{total_reviews} + 1)$

Why it matters: Combines satisfaction quality with review credibility, reducing bias from low review volumes.

Objective: Compare satisfaction levels between free and paid games

Catalog Composition

KPI: Shares;

Formula:

COUNT(games by pricing type) / TOTAL games

Why it matters: Provides supply-side context for interpreting engagement and popularity trends.

Objective: Understand how the Steam catalog is structured.

Market Structure & Concentration

KPI: Popularity Tier Distribution
(Blockbuster / Hit / Mid / Niche)

Metric Used: COUNTA(appid) by Popularity Tier

Objective: To assess how Steam games are distributed across popularity tiers.

Why it matters: It shows that the Steam catalog is dominated by niche titles, with very few highly popular games.

Discovery & Growth

KPI: Viral Games Count

Formula: COUNT(games) where recommendation_band = Viral (>10k)

Why it matters: Measures organic discovery driven by word-of-mouth and social proof.

Objective: Evaluate growth driven by player recommendations

Exploratory Data Analysis (EDA)

Exploratory analysis was conducted using pivot tables to identify patterns in engagement, pricing, sentiment, and market structure across the Steam catalog.

Distribution Analysis

The Steam catalog is highly skewed. Over 90% of games fall into the *Niche* category, while only a handful qualify as *Blockbuster* or *Hit* titles. This confirms a long-tail structure where most games receive limited visibility and engagement.

Comparison Analysis

Free-to-Play games show **significantly higher average Peak CCU** than Paid games, indicating stronger reach and accessibility. However, Paid games consistently demonstrate **higher user sentiment**, suggesting better satisfaction among players willing to pay upfront.

Engagement & Content Patterns

Games spanning multiple genres tend to achieve **higher Peak CCU** than single-genre titles, indicating broader appeal. At the same time, user sentiment declines slightly as genre count increases, pointing to a trade-off between reach and experience depth.

Discovery & Social Proof

Games in the *Viral* recommendation band exhibit **dramatically higher Peak CCU** compared to other categories. This highlights the importance of community-driven discovery and word-of-mouth in driving player attention.

Pricing & Monetization Effects

Higher-priced games generally show **stronger sentiment**, but do not consistently achieve higher engagement. Free and lower-priced games benefit from reach, while premium pricing aligns more closely with satisfaction.

Accessibility Insights

More popular games tend to support multiple platforms, suggesting accessibility acts as a supporting factor for reach rather than a primary driver of popularity.

EDA Summary

Overall, engagement on Steam is concentrated among a small number of titles. Pricing model, discoverability, and content breadth strongly influence reach, while player satisfaction is more closely tied to paid and premium experiences.

Dashboard Design

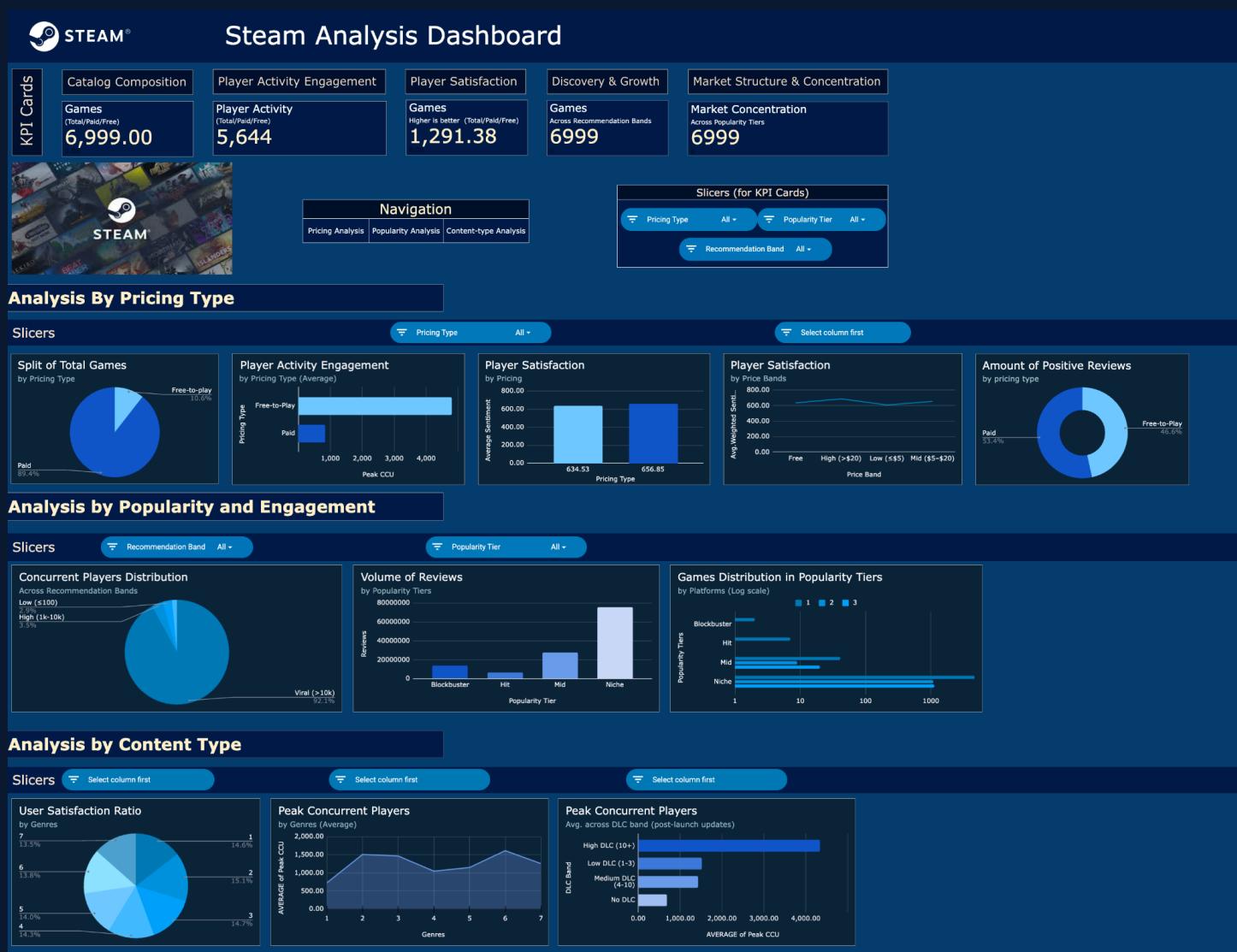
The dashboard brings together the most critical KPIs and pivot-based insights to give a clear view of **engagement, satisfaction, market structure, and discovery on Steam**.

Key metrics are displayed as **scorecards**, supported by comparison and distribution charts derived from the core pivot tables.

Interactive **slicers** (Pricing Type, Popularity Tier, Genre Count, DLC Band, Price Band) allow users to quickly drill down into specific segments while keeping the overall view intact.

The layout is designed for **easy navigation and quick decision-making**, with visuals organized to highlight patterns rather than raw numbers.

(Screenshots added below)



Insights Summary

Free-to-Play titles drive disproportionate player activity:

Free-to-Play games achieve ~5.7× higher average Peak CCU than paid titles (4,807 vs 837), confirming that zero-price entry significantly boosts reach and visibility.

Paid games deliver higher player satisfaction despite lower reach:

Paid titles show stronger sentiment scores (Weighted Sentiment ≈ 657 vs 635 for Free-to-Play), indicating higher perceived quality and satisfaction among paying users.

The Steam catalog is heavily skewed toward low-visibility titles:

Over 93% of games fall into the Niche tier, highlighting a long-tail marketplace where most titles struggle to gain meaningful traction.

Engagement is highly concentrated among very few games:

Just 4 Blockbuster titles generate over 11% of total reviews, demonstrating strong winner-takes-most dynamics across the platform.

Social proof is the strongest driver of popularity:

Games with Viral recommendation levels (>10k) average ~6,665 Peak CCU - far exceeding all other recommendation bands, underscoring the impact of word-of-mouth.

Multi-genre games attract higher engagement up to an optimal point:

Games spanning 2–3 genres achieve the highest average Peak CCU, while excessive genre breadth shows diminishing returns.

Broader platform support aligns with higher popularity tiers:

Blockbuster and Mid-tier games support more platforms on average than Niche titles, suggesting accessibility contributes to scale.

Pricing influences sentiment more than engagement:

Higher-priced games (> \$20) show the strongest sentiment scores, but Free-to-Play models outperform in raw engagement and reach.

Post-launch content depth correlates with engagement:

Games with 10+ DLCs achieve substantially higher average Peak CCU, indicating sustained content updates help retain players.

Discovery dynamics differ sharply by pricing model:

Viral recommendation counts are significantly higher among Paid titles, while Free-to-Play games rely more on scale than intensity of recommendations.

Recommendations

1) Hybrid Monetization (Free + DLC)

- **Evidence:** Free-to-Play avg Peak CCU = **4,807** vs Paid = **837**
- **Insight:** Free pricing maximizes reach; paid games show higher satisfaction
- **Action:** Launch free core gameplay, monetize via DLC / expansions

2) Invest in Early Social Proof

- **Evidence:** Viral games (>10k recs) avg Peak CCU = **6,665**
- **Insight:** Recommendations strongly drive discoverability and engagement
- **Action:** Prioritize community building, reviews, and influencer outreach

3) Target Discounts to Mid-Tier Games

- **Evidence:** Mid-tier = **70 games, 27.8M reviews**, moderate CCU
- **Insight:** These games show upside but lack visibility
- **Action:** Apply limited-time discounts to boost attention and traffic

4) Scale Accessibility Selectively

- **Evidence:** Blockbusters avg platform count = **1.75** vs Niche = **1.48**
- **Insight:** Platform expansion correlates with higher popularity
- **Action:** Expand OS/language support only after early traction

Impact Estimation

Business Impact

- **Cost Reduction:** Higher organic discovery (Viral band >10k) supports **1,145 games**, reducing dependence on paid marketing channels
- **Efficiency:** Targeted discounts and selective platform expansion improve ROI versus blanket strategies, especially for **mid-tier games (70 titles, 27.8M reviews)**
- **Engagement Growth:** Free-to-Play titles achieve **~5.7x higher Peak CCU** (4,807 vs 837), maximizing player reach
- **Revenue Upside:** Higher-priced games show stronger satisfaction (**79–84% positive**) indicating willingness to pay among engaged users
- **Risk Mitigation:** Portfolio-based strategy avoids over-reliance on **0.05% blockbuster titles**, stabilizing performance across **93% niche games**

Limitations

Platform Scope: Analysis is limited to Steam data only. Cross-platform performance and revenue impact cannot be inferred.

Playtime Gaps: Playtime metrics contain significant missing values and were treated cautiously, limiting long-term engagement conclusions.

Causality Limits: Relationships between pricing, discounts, and engagement are correlational, not causal, due to lack of experimental or time-series data.

Future Scope

Revenue Integration: Incorporate sales and revenue data to directly link engagement and sentiment to financial outcomes.

Marketing Attribution: Add marketing spend and promotion timing to assess the true impact of discounts and visibility boosts.

Predictive Modeling: Build tier-based models to forecast engagement and identify potential breakout (future blockbuster) titles.

Conclusion

This project analyzed how pricing models, content structure, platform accessibility, and user feedback influence game popularity, engagement, and satisfaction on the Steam marketplace.

Using a combination of feature engineering, KPI design, and pivot-based analysis, the study revealed a highly skewed market structure where a very small number of games drive a disproportionate share of player activity and reviews. Free-to-Play titles consistently outperform paid games in terms of reach, while paid games tend to show stronger user satisfaction metrics. Social proof, measured through recommendations, emerged as one of the strongest drivers of visibility and engagement.

The analysis moves beyond simple averages by using segmentation (pricing models, popularity tiers, bands) to support more realistic and actionable insights. Overall, the project demonstrates how structured KPI frameworks and exploratory analysis can support strategic decision-making for digital marketplaces.

Appendix

Appendix A: Data Dictionary

A consolidated data dictionary was created and used throughout the analysis to ensure consistent interpretation of raw, cleaned, and derived fields.

This includes definitions for pricing models, popularity tiers, sentiment metrics, and all banded variables used in KPIs and pivot tables.

(Refer to the Data Dictionary or Legend in the 'Data Description' section)

Appendix B: Feature Engineering & Derived Metrics

Key analytical columns were derived to enable deeper analysis, including:

- Total Reviews
- Weighted Sentiment
- Pricing Type
- Platform Count
- Popularity Tier
- Discount Impact
- Price Band, DLC Band, and Recommendation Band

These derived fields form the foundation for KPI calculations and segmentation-based analysis.

Appendix C: Pivot Tables & Dashboard Support

Multiple pivot tables were created to support KPI scorecards and dashboard visualizations.

These pivots enable comparisons across pricing models, popularity tiers, content depth, and discovery levels, and serve as the primary data source for the interactive dashboard.

Contribution Matrix

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Mahir Abdullah							Analysis Lead
Ankit Singh							Project & Strategy Lead
Manas Selukar							Dashboard Lead
Krishiv Gupta							Data Lead
Rajat Srivastav							PPT & Quality Lead
Kavya Katal							Strategy Lead

Declaration:

We confirm that the above contribution details are accurate and verifiable through version history and submitted artifacts.