

How can we increase revenue from Catch the Pink Flamingo?

Lim Lee Keong, Jeffrey

Problem Statement

Context:

- The Flamingo app tracks game interactions, purchases, and ad clicks from players.
- Different users exhibit varied spending behaviors, affecting revenue potential.

The Problem:

- How can Eglence predict and influence user spending behavior?
- How do engagement patterns correlate with revenue?

Impact & Data Science Story:

- This is a predictive analytics challenge, requiring multiple datasets.
- Session data, ad clicks, purchases, and chat activity together help identify key user behaviors and optimize monetization strategies.

Goal:

- Use graph analytics, clustering, and classification to segment users and drive targeted revenue strategies.

The analysis incorporates multiple datasets, including user sessions, ad clicks, in-app purchases, and chat interactions. These sources provide insight into player behavior, spending habits, and engagement patterns.

By leveraging data science techniques, we can uncover patterns in user engagement and spending habits. This allows Eglence to move beyond surface-level trends and proactively identify opportunities to boost revenue.

User session data shows how long players engage. Ad-click and purchase data reveal spending behavior. Chat interactions provide insights into community influence. Together, these sources enable Eglence to refine monetization strategies—targeting the right users with the right offers

Data Exploration Overview

User Spending Patterns:

- Total revenue from purchases: \$428
- Total amount spent on in-game items: \$21,407
- Most expensive item: \$20, cheapest item: \$1
- iPhone users dominate top spenders, with a hit ratio between 11.6% and 14.5%.

Engagement & Monetization:

- Higher chat activity correlates with spending—users who interact more tend to buy more.
- Most purchased item sold over 750 times, proving strong demand for specific in-game products.

Spending Concentration Among Users:

- Top 10 spenders contributed a significant percentage of total revenue, emphasizing a small group drives major profits.
- Highest spender spent over \$250, showcasing extreme variance in user purchases.

A histogram showing how much money was made from each item:



A histogram showing total amount of money spent by the top ten users (ranked by how much money they spent).



Amount spent buying items	\$21,407	Rank	User Id	Platform	Hit-Ratio (%)
Number of unique items available to be purchased	6	1	2229	iphone	11.60
Company revenue made from the purchased in buy-clicks.csv	\$428	2	12	iphone	13.07
Item cost (max)	\$20	3	471	iphone	14.50
Item cost (min)	\$1				

In exploring the Flamingo app data, I worked with multiple datasets—user sessions, purchases, ad-clicks, and game interactions—to uncover spending patterns. One of the most important findings was that spending behavior is highly segmented, with a small group of high-spenders driving the majority of revenue.

Total in-game purchases amounted to \$21,407, but revenue generated was only \$428, suggesting that purchase pricing and engagement strategies could be optimized. The most expensive item costs \$20, while the cheapest is \$1, with a few items dominating the economy—one was purchased over 750 times.

Interestingly, iPhone users spend significantly more than others, and their hit ratios range from 11.6% to 14.5%, indicating higher engagement. Another major insight is that users with higher chat activity tend to make more purchases, showing the powerful connection between social interaction and monetization.

These findings help us predict spending behaviors, refine engagement strategies, and uncover new revenue opportunities for Englece.

Predicting User Spending: Insights from Decision Tree Classification

Data Preparation & Partitioning

- Categorized users as HighRollers (>\$5 purchases) vs. PennyPinchers (≤\$5 purchases).
- Removed unnecessary attributes (userid, teamLevel) to improve predictive accuracy.
- Train-Test Split (60/40) ensured model validation.

Decision Tree Implementation

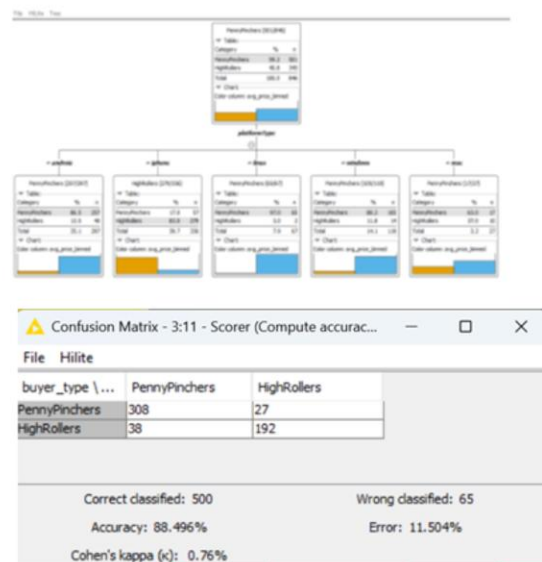
- Used Decision Tree Learner in KNIME to classify users based on spending behaviors.
- Analyzed results using Scorer and Confusion Matrix to measure prediction accuracy.
- Key predictors: Total chat interactions, Platform type, Past purchase history.

Findings from the Decision Tree Analysis

- iPhone users tend to be HighRollers → 83% buy big-ticket items.
- Android, Windows, and Linux users are predominantly PennyPinchers → Over 85% buy only low-cost items.

Business Strategy Implications

- HighRoller Strategy (iPhone Users): Premium bundles, VIP offers, exclusive promotions.
- PennyPincher Conversion Strategy (Android, Windows, Linux Users): Discounted packs, loyalty incentives.



Using Decision Trees, we classified users into HighRollers and PennyPinchers based on their spending behavior.

Feature selection was key in making accurate predictions, ensuring only relevant attributes—chat activity, platform type, and past purchases—were included. After partitioning the dataset into an 60/40 train-test split, we implemented the Decision Tree model in KNIME, which effectively categorized users.

One of the most striking insights was that iPhone users are far more likely to be HighRollers, with 83% purchasing high-value items, while Android, Windows, and Linux users overwhelmingly fall into the PennyPincher category.

For Eglence, these findings shape new revenue strategies—premium bundles for HighRollers, discounted packs for PennyPinchers, ensuring optimized monetization across user groups.

Insights from Clustering Analysis

Why Clustering Was Used

- Segmented users into four distinct clusters based on ad-click behavior and spending patterns.
- Used Elbow Method to determine optimal K=4 clusters.

Major Findings from Cluster Analysis

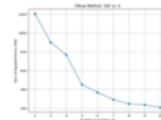
- Cluster 2 (High Rollers) → Users with above-average ad clicks and high spending.
- Cluster 4 (Engaged but Low Spenders) → Highly active in ads but not converting into purchases.
- Cluster 3 (Low Engagement Users) → Minimal ad clicks, low revenue—likely disengaged audience.
- Cluster 1 (Slightly Below Average Users) → Near the mean but less active overall.

Business Implications for Eglence

- Premium Offers for Cluster 2 → Exclusive in-app bundles to boost revenue from High Rollers.
- Conversion Strategies for Cluster 4 → Personalized discounts & loyalty incentives to drive spending.

Attribute	Rationale for Selection
ad_click_count	Represents the total number of ad clicks per user. Directly conveys the overall engagement with ads.
InApp_Purchase_Revenue	Captures the total revenue generated by the user through in-app purchases. Indicates this measure tracks monetary contributions.
Revenue_Per_AdClick	Denotes the average revenue obtained per ad click (i.e., total revenue divided by the total ad clicks). Reflects the conversion efficiency of ad interactions into revenue.

of clusters created: 4
Using Elbow method, K = 4 is chosen as an optimal value for this case.



Cluster #	Center
1	[-0.197 -0.278 -0.172]
2	[0.225 1.792 1.558]
3	[-1.175 -0.720 -0.319]
4	[1.094 0.002 -0.327]

Clustering allowed us to segment users into four distinct groups based on ad clicks, in-app purchases, and revenue per ad click.

The most important findings were in Cluster 2 and Cluster 4—Cluster 2 represents High Rollers, users who convert ad engagement into high spending, while Cluster 4 includes highly engaged users who don't spend much.

This insight allows Eglence to craft targeted monetization strategies—Premium bundles for High Rollers, and discount incentives for Cluster 4, encouraging low spenders to increase their purchases.

By understanding engagement and spending behavior through clustering, Eglence can maximize revenue opportunities while optimizing user experience.

Graph Analytics: Understanding User Conversations & Influence

Graph Model Structure & Why It Matters

- Users (User nodes) and chat messages (ChatItem nodes) form a rich conversational network.
- Relationships like CreateChat, ResponseTo, and Mentioned reveal patterns in engagement and influence.

Most Remarkable Findings from Graph Analysis

- Longest Conversation Chain: Found a 9-message response chain involving 5 unique users. Shows strong engagement dynamics within chat networks.
- Chattiest Users & Teams: User 999 from Team 52 proves high activity in both individual and group chats. Overlap between active users & chatty teams highlights the impact of community-driven conversations.
- Clustering Coefficients & Network Density: Users 209, 668, and 999 had the highest clustering scores, showing tight conversational links. Reveals how connected users drive discussions and shape engagement behavior.

Future Exploration: How Can We Go Deeper?

- Predicting Influencers:
 - Identify users who act as bridges between different chat groups.
 - Detect central figures in key conversations based on response frequency.
- Sentiment & Topic Analysis in Conversations:
 - Categorize chat topics to uncover emerging discussion trends.
 - Analyze positive vs. negative sentiment to optimize engagement strategies.
- Community Detection & Group Interactions:
 - Use graph clustering to discover strong social subgroups within chats.
 - Improve strategies for team collaboration & user retention.

Chattiest Users

Users	Number of Chats
394	115
2067	111
1087	109

Chattiest Teams

Teams	Number of Chats
82	1324
185	1036
112	957

User ID	Cluster Coefficient
209	1
668	1
999	0.958

Our graph analysis uncovered rich conversational patterns, revealing how users engage and interact.

One remarkable finding was the longest response chain, spanning 9 messages with 5 unique participants, showcasing deep discussions. We also identified the chattiest users and teams, where user 999 appeared in Team 52, proving an overlap in individual and group chat engagement.

Network density analysis highlighted users 209, 668, and 999 as having high clustering coefficients, meaning they play central roles in chat interactions.

For further exploration, we can predict conversation influencers, analyze sentiment trends, and refine community detection to enhance engagement strategies. These insights can drive smarter communication models and optimize user retention within Eglence's chat network.

Recommendation: Personalized Engagement Strategies to Maximize Revenue

Action:

Implement AI-driven personalized promotions based on user classification and clustering insights.

Rationale:

- HighRollers (iPhone users) → Offer exclusive premium bundles & VIP perks to boost retention.
- Cluster 4 (Engaged but Low Spenders) → Deploy targeted discounts & incentives to increase conversion.
- Chat activity-driven spending behaviors → Customize offers based on engagement trends, leveraging in-app social interactions.

Why This Strategy Works:

- Maximizes revenue by targeting the right users based on spending behavior and engagement insights.
- Uses clustering analysis to refine promotional strategies, ensuring the right incentives reach the most receptive audience.
- Leverages graph analytics to predict key influencers and amplify community-driven engagement.



Our analysis has shown clear segmentation of user spending behavior and engagement patterns. To improve revenue, Eglence should implement AI-driven personalized promotions, focusing on HighRollers and engaged but low-spending users.

For HighRollers, exclusive premium bundles and VIP perks will enhance retention. Meanwhile, Cluster 4 users, who engage heavily but spend less, can be converted through targeted discounts and loyalty incentives.

Additionally, chat activity strongly correlates with spending, suggesting that in-app social interactions should drive personalized offers.

This strategy optimizes revenue by ensuring the right users receive tailored promotions, increasing overall spending while improving user experience.

Thank you for your attention!

