Table of contents

Latitude Product Analysis

Scope & Methodology

This report presents a quick analysis of user behavior within the AI Dungeon ecosystem

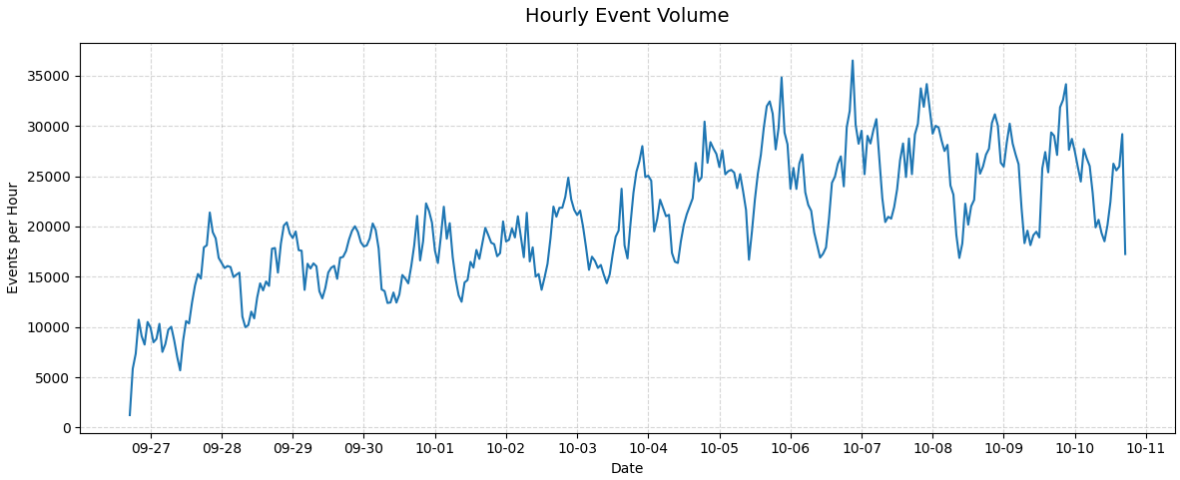
Key Strategic Insights

1. **The "Guest Bias" in Retention:** Our Current Rention Metrics suffer having an unified user_id, which inflates these statistics of interest.
2. **A Sticky Product for Those that Pass Through Initial Introduction To AiDungeon** A bimodal distribution in user engagement reveals two distinct populations: "Casual" (1–5 actions) and "Loyalty" (100+ actions).
3. **Acquisition Efficiency:** The **"Discover"** surface is currently the highest-quality acquisition channel, driving a **78% registration rate**, yet it receives significantly less traffic than the **Homepage Banner**.

created_at datetime[s]	event_type str	event_name str	event_variation str	us i6
2025-10-10 17:39:32.656	”user”	”page_viewed”	null	6(
2025-10-10 17:39:32.462	”user”	”retry_button_pressed”	null	6'
2025-10-10 17:39:32.301	”user”	”too_many_actions_taken_before_registering”	null	6'
2025-10-10 17:39:32.113	”user”	”submit_button_pressed”	null	6'
2025-10-10 17:39:31.338	”user”	”page_viewed”	null	6(

DATA EXPLORATION

created_at u32	event_type u32	event_name u32	event_variation u32	user_id u32	metadata u32	year u32	month u32	day u32	hour u32
0	0	0	6299656	0	0	0	0	0	0



Duplicate Events Table

user_id u32	created_at u32	event_name u32	metadata u32	len u32
3448	3448	3448	3448	3448

Original Rows - Cleaned Rows: 3514
The difference was 66 rows, meaning there was addition duplicates beyond 2 occurances in the

```
Data spans from 2024-10-01 to 2024-10-10
Nothing irregular about the date ranges, seems to be a full 10 days of data
shape: (1, 1)

 created_at
 ---
 datetime[s]

 2025-09-26 17:40:24.032

shape: (1, 1)

 created_at
 ---
 datetime[s]

 2025-10-10 17:39:32.656


The are no invalid user_ids based on length
```

id_length	count
u32	u32
8	6940689

DATA QUALITY IMPROVEMENTS

I would suggest the following enhancements to improve the usability of the dataset:

1. **Adding JSON Metadata Features directly into the Base Schema**
   Promoting key metadata attributes into top-level columns improves analytical accessibility and reduces repeated parsing.
2. **Extracting Session ID from the Metadata Column**
   Session ID should be stored as its own field to enable proper session-level grouping and behavioral analysis.
3. **Extracting Adventure ID from the Metadata Column**
   Adventure ID (scenario identifier) should be separated out for clearer segmentation of user activity.

4. **Adding User Location Data — Demographic Region, Latitude/Longitude**
   Location attributes will support regional analysis, demographic insights, and anomaly detection.

5. **Including Phone Type**
   Surface device information (e.g., iPhone, Android) as a structured field to allow device-level performance and usage analysis.

FUTURE INSTRUMENTATION

1. **Num_Input_Tokens**
   Explains the total number of tokens used within an input prompt.

2. **Num_Output_Tokens**
   Explains the number of tokens generated when returning the output.

3. **Cost_of_Event**
   Represents the associated cost of running the input prompt, based on token usage.

4. **Network_Speed / Latency**
   Measures the time elapsed between the input prompt and the returned output.

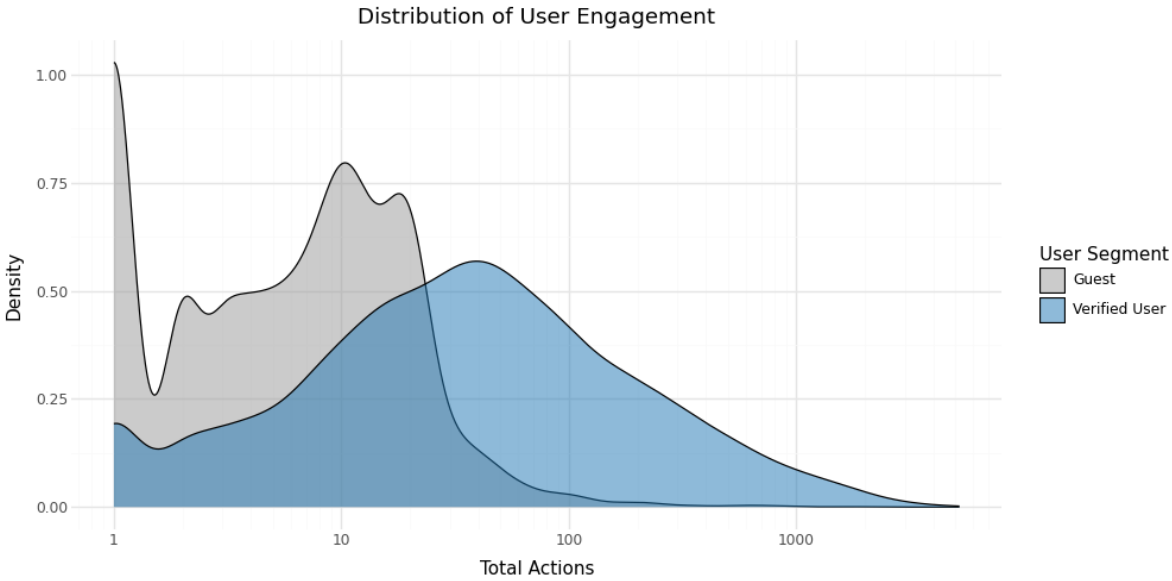
5. **In_Session_Order_Number**
   Provides context on the order in which the screen or event was triggered within the same session.

6. **Out_Session_Process_Number**
   Provides context on when the event was processed relative to other events across or outside the session.

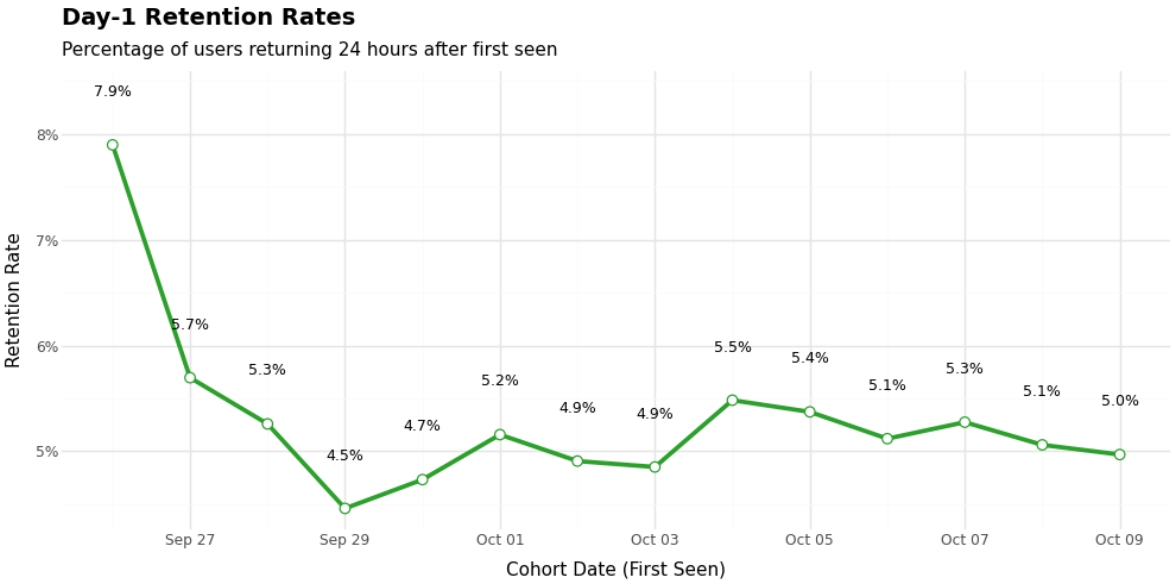
7. **Model_Used**
   Indicates the LLM used within the game for generating the output.

FIRST TIME USER EXPERIENCE

total_users u32	verified_users u32	signup_rate f64
66531	5353	0.080459



retained_users u32	total_users u32	day_1_retention_rate f64
3276	66531	0.04924



Analysis

Reliability:

The provided metrics represent a snapshot of October behavior, not a global population stats. Due to Seasonality effects, retention rates likely differ significantly during holidays or summer months. In fact, assuming that there is an upward trend to the data because this is high growth startup, the data is non-stationary and the stats calculated for this sample may not hold for another time period. Additionally, because the dataset is a sequential slice rather than a randomized study, we must assume Temporal Bias—the specific marketing campaigns or app bugs present during this week heavily influence these numbers.

Guest Bias:

The metrics suffer from having a lack of shared user accounts. The data presents a one-to-many (one user has multiple user_ids) issue because user_id is generated client-side for guests, a single human playing on a phone and then a laptop generates two 'users' with 0% retention. This intrinsically deflates our calculated Retention and Signup Rates, making the product look worse than it actually is. We need a 'Probabilistic Identity Stitching' model (using IP or User Agent) to get the true human retention rate.

Exclusions:

Test/Dev Traffic: Identified by releaseStage != 'production'.

Anomalous High-Frequency Users: Users exceeding humanly possible speeds (e.g., >60 actions per minute), which indicates bot scraping.

Zero-Action Sessions (Bounces): Users who generate session_start but zero submit_button_pressed events. These represent 'Traffic Acquisition' issues, not 'Product Quality' issues, and should be analyzed separately."

PATTERNS THAT DRIVE RECOGNITION

Registration Rate: Currently the data is at the event level, but to track this feature properly, data must be aggregated to user level!

To identify the factors influencing the registration_rate (the dependent variable), I performed a segmentation analysis comparing the populations of Verified Users versus Guest Users. This comparative analysis highlights significant divergences in user behavior prior to conversion.

Surface Exploration (Investment)

People that engaged with the context screen are 84% more likely to register, the context screen is a top priority to explore within a tutorial of how to play the gam effectively.

is_registered i8	user_count u32	avg_actions f64	avg_minutes f64	config_usage_rate f64
0	61178	1.1	128.3	0.06
1	5353	99.9	2474.7	0.843

CONTENT PERFORMANCE AND SELECTION

`MOST POPULAR (Traffic)`

scenario_id str	unique_players u32	total_turns i32	avg_turns_per_player f64
”cj90vvdB14fn”	2422	0	0.0
”8748087”	1927	41133	21.3
”KyMhfQFXO8Bs”	1392	0	0.0
”2503121”	1190	48110	40.4
”Yo_hMuEXJQQI”	943	0	0.0

`MOST ENGAGING (Quality)`

scenario_id str	unique_players u32	total_turns i32	avg_turns_per_player f64
”1828345”	139	11927	85.8
”11482379”	84	4785	57.0
”11507521”	99	5400	54.5
”6231981”	159	7309	46.0
”2503121”	1190	48110	40.4

surface str	unique_users u32	avg_actions f64	reg_rate_pct f64
"Direct/Other"	66508	475.3	77.1
"Search"	9650	287.9	75.6
"Homepage Banner"	6943	72.5	26.3
"Discover"	6906	288.3	78.7

1. **83% of our users only go to quick play, continue, multiplayer, or create scenario, but do not search for new games or click on new banners.**
   This only reinforces that players that are hooked consistently focus on their stories. There is a high barrier to entry, but those invested stay… and that is a sticky product.
2. **With the Discovery tab driving a 78% registration rate (the highest of any surface), we can see that it provides a more streamlined and focused approach to get users to the correct game of interest.**
   I am unsure of the algorithmic approach to displaying the discover tabs and whether it is user personalized or SEO optimized, but for the users that find it, they stick to their game of choice afterwards. There should be further implementation to search deeper into promoting top story creators and testing the social community aspect of the discovery tab.

Algorithm Reccomendation

The Problem:

The deep difference between guest and signed users suggests a bimodal distribution in user engagement: Guests average only 1.1 actions before bouncing. New users are hitting "Writer's Block" or "AI Confusion" immediately.

The Solution:

We should implement an A/B test to serve the feed based on User Maturity rather than a generic "one-size-fits-all" homepage.

1. **Generate Flag on Scenario ID → The 'Easy' Flag**
   We classify content as "Beginner Friendly" if it meets two criteria:
   - **Low Friction:** The Global Retry Rate of the scenario is $< 10\%$ (Indicates simple AI prompts that are easy to follow).

- **Social Proof:** The scenario is in the Top 5% of unique players (Ensures we only show proven, high-quality content).

2. **The Serving Rule**

   - **IF the user is a New Player (user_hours $<$ 24 hours):** The feed ONLY displays scenarios with the 'Easy' Flag.

   - **IF the user is a Returning Player:** The feed displays the standard mixed content.

Result (Hypothesis):

By filtering out complex models and delivering "Quick Wins" fast, we expect to increase the `Activation Rate`: (percentage of users reaching $>$ 5 actions). A smoother first session reduces the "Time-to-Magic," which our data shows is the strongest predictor of downstream Registration Rates.