FORMAL METHODS IN CS

Process Mining Assignment

Steam Retention Analytics

Reconstructing the Customer Journey from Purchase to Loyalty

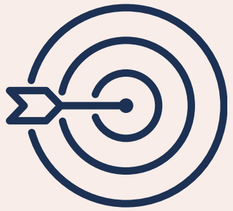
A.Y. 2025/2026

GitHub Repository

Yuri Tateo – 869303
Francesca Palumbo – 869311

01 Introduction & Objectives

This project applies Process Mining techniques to analyze user behavior on the **Steam** gaming platform, adopting an E-commerce perspective.



Our Goal is to reconstruct the *"Customer Journey"* from the initial purchase to gameplay, identifying churn or loyalty.

Primary Objectives:

- **Identify Patterns**: Map the flow between buying and playing.
- **Analyze Retention**: Detect churn points, such as users playing less than 2 hours.
- **Optimize Monetization**: Discover pathways leading to high engagement and DLC purchases

02

Data Source and Structural Limitations

The analysis uses the **"Steam Video Games Dataset"** from *Kaggle*, containing User IDs, Game Titles, and Playtime values.

Our challenge:
The original dataset was static and lacked timestamps, which are mandatory for control-flow discovery in Process Mining.

Our solution:
A logical simulation strategy was implemented to generate a synthetic, yet realistic, Event Log to enable time-based analysis.

03 Simulation Rules for Event Log Generation

We used **Python** and **Pandas** to transform static data into a time-aware log using specific business rules.

Timeline Logic:

Start
“Purchase Event" is assigned as time T_0

Play
"Play Event" occurs at a random interval after purchase ($T_0 + \Delta t$)

03 Simulation Rules for Event Log Generation

Business Logic Tags:

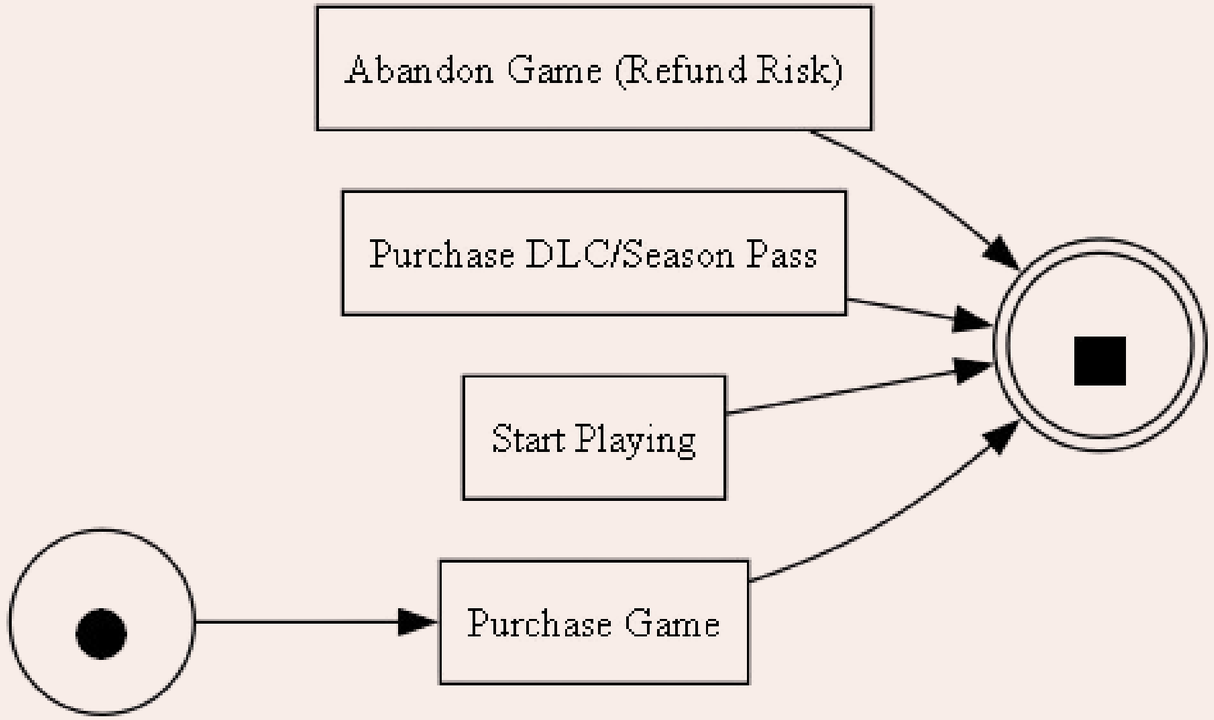
Refund/Churn
If playtime is < 2.0 hours (Steam refund policy), a termination event is triggered.

Loyalty
If playtime is > 50 hours, a "Purchase DLC" event is simulated to model high-value customers.

04

Process Discovery Algorithms

Comparing Mining Techniques using PM4Py library to apply three distinct algorithms.

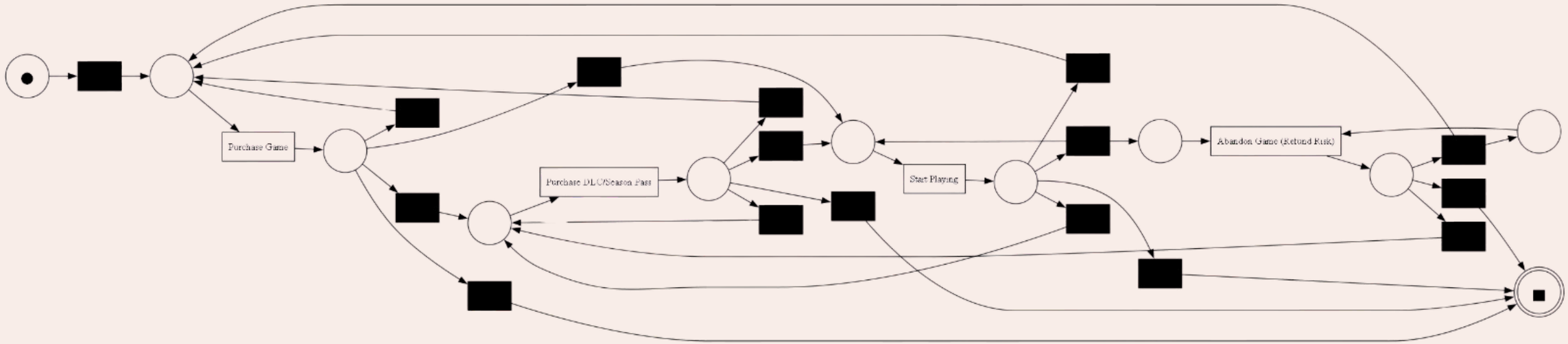


Alpha Miner (Baseline)

Provided a preliminary view but struggled.

Rejected due to poor Fitness (0.132), failing to capture gaming loops.

04 Process Discovery Algorithms

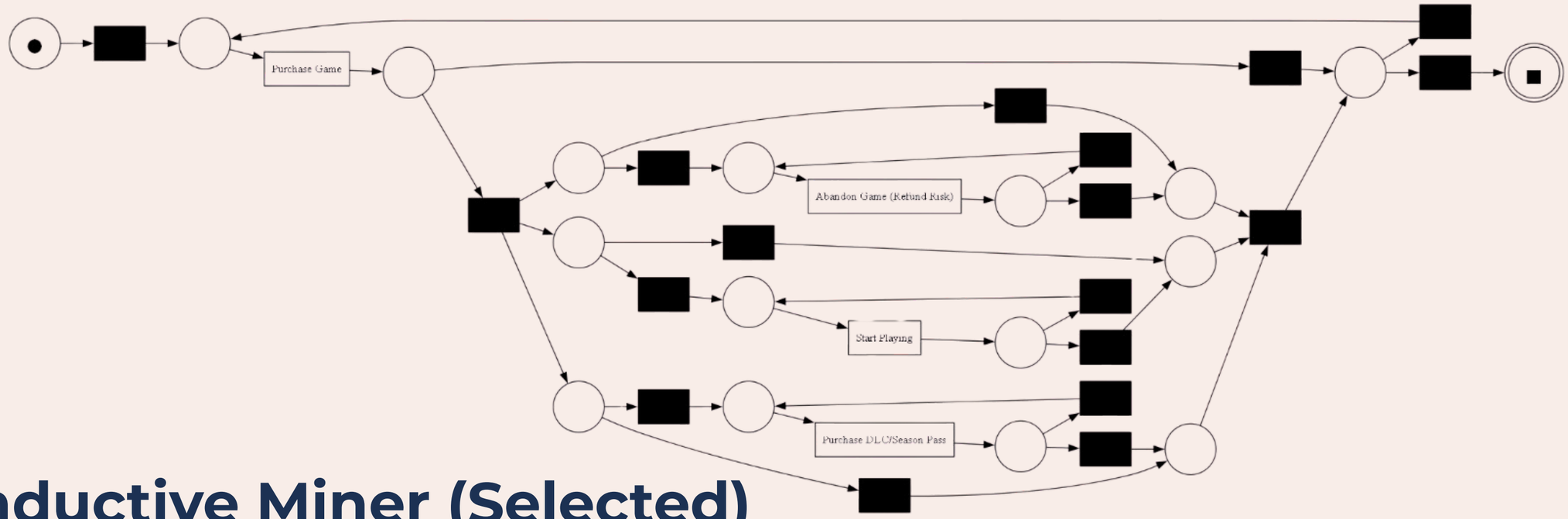


Heuristic Miner

Effectively handled noise and highlighted the "Happy Path" (most frequent behavior).

Achieved high Precision (0.519) but produced a complex graph difficult for stakeholders to interpret.

04 Process Discovery Algorithms



Inductive Miner (Selected)

Chosen for the final analysis because it guarantees a sound model free of deadlocks.

Achieved a perfect Fitness score of 1.000. Perfect fitness ensures 100% of user behaviors, including the "Long Tail" of retention, are represented.

05

Visual Analysis & Preliminary Insights

1. The "Backlog" Path:
The model visualizes a silent transition where users skip gameplay entirely, corresponding to the "Pile of Shame" (buying without playing).

2. The "Refund Risk" Branch:
A distinct path splits after "Start Playing" leading to "Abandon Game," validating the structural difference between churners and loyal players.

3. The "Loyalty" Loop:
Recursive loops around "Start Playing" and "Purchase DLC" confirm that high-value users engage in repeated gameplay sessions.

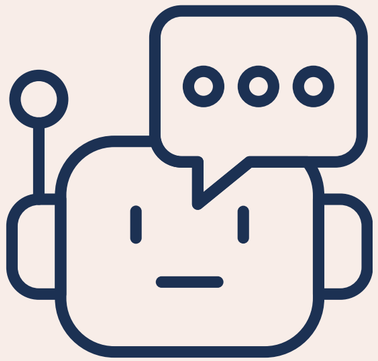
06 AI-Driven Reasoning

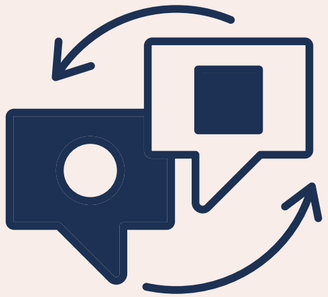
Innovation
An AI Analyst Chatbot was integrated directly into the **Streamlit Dashboard** to fulfill the requirement of using LLMs for reasoning.

Function
The AI interprets the **Inductive Miner model**, translating technical Petri Net patterns into actionable business insights.

Outcome
The system successfully correlated structural model data with business strategies like the **Steam Refund Policy**.

Streamlit



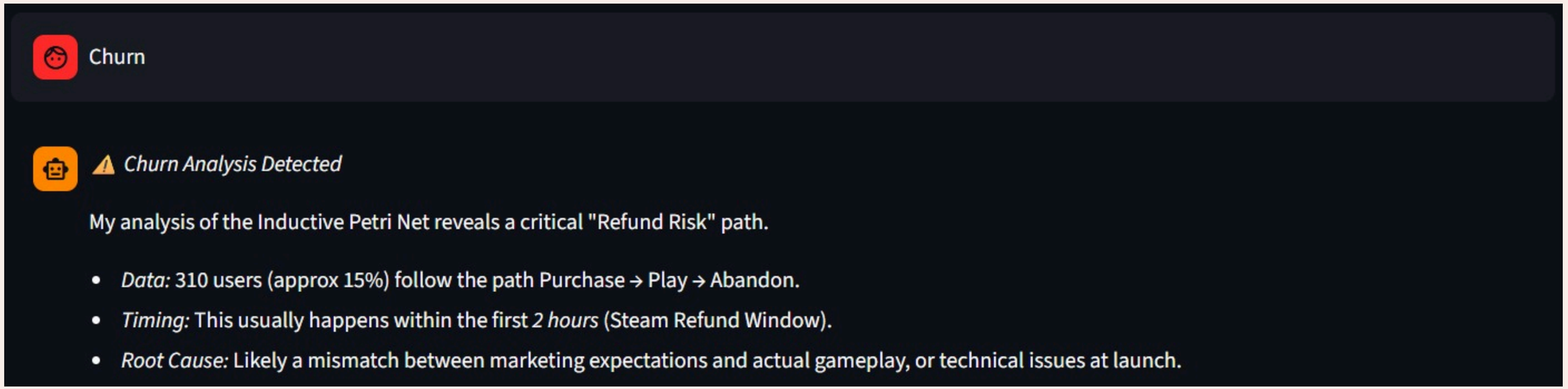


06

AI-Driven Reasoning – Anomaly Detection

1. The "Churn" Path

Anomaly Detection regarding **Refund Risks**



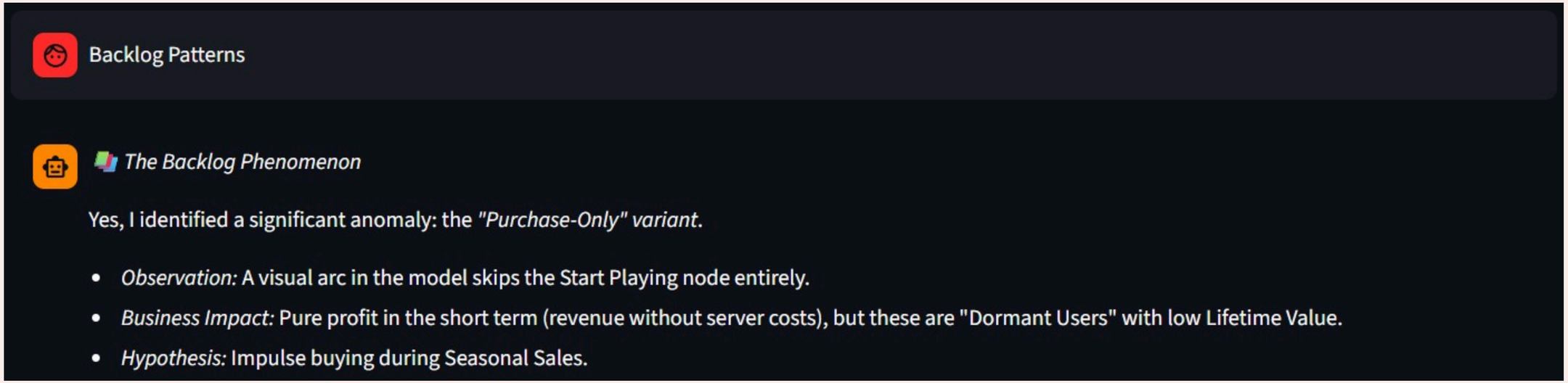
The AI identified a critical **"Refund Risk"** path where users **Purchase → Play → Abandon**. Approximately 15% of users (310 cases) follow this path. This abandonment occurs strictly within the first **2 hours**, aligning with the Steam Refund Window. The AI suggests a mismatch between marketing expectations and actual gameplay or technical launch issues.

06

AI-Driven Reasoning – Anomaly Detection

2. The “Backlog” Phenomenon

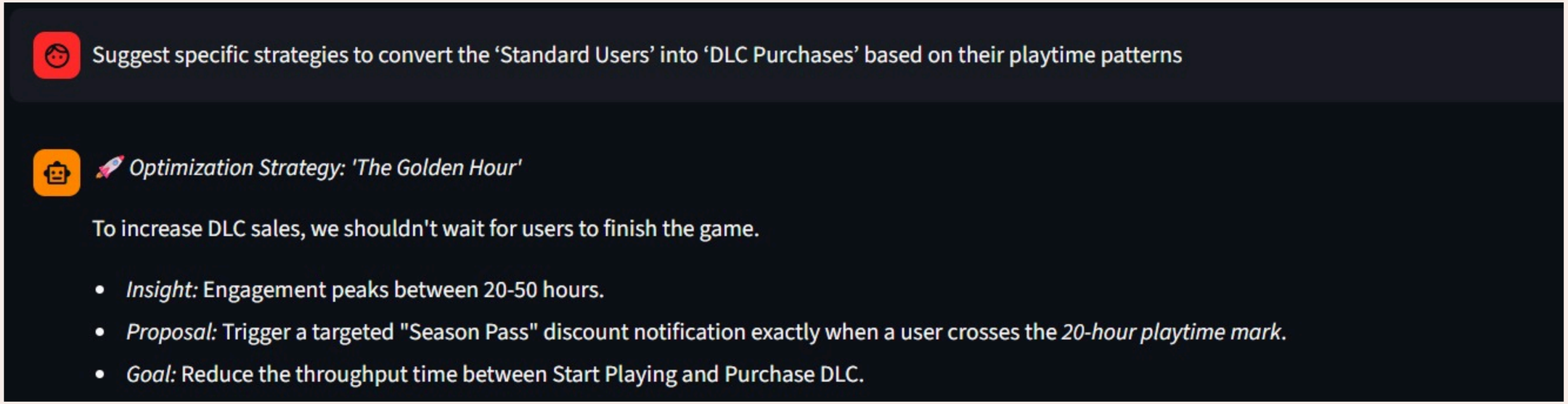
Identifying **"Dormant Users"**



The AI detected a **"Purchase-Only"** variant where the **"Start Playing" node is skipped** entirely via a visual arc. This behavior is likely driven by impulse buying during seasonal sales. While this represents short-term pure profit (revenue without server costs), these are **"Dormant Users"** with low Lifetime Value.

06 AI-Driven Reasoning – Process Optimization

3. “The Golden Hour" Strategy



Convert standard users into DLC purchasers. **Reduce the throughput time** between **"Start Playing" and "Purchase DLC"**. Engagement peaks between 20-50 hours. The AI suggests triggering a targeted "Season Pass" discount exactly when a user crosses the 20-hour playtime mark. Moves from descriptive analysis to prescriptive action.

07

Conclusions & Future Work

The project successfully transformed a static dataset into a **dynamic Event Log** to reconstruct the *Steam User Journey*.

The **Inductive Miner** provided a sound model, while AI reasoning *validated* **"Refund Risks"** and *proposed* the **"Golden Hour"** monetization strategy.

Future Work:

- *Integrate* **real-time API data** to *analyze seasonal trends* (e.g. Summer Sales).
- *Implement* the **dashboard** as a *live monitoring tool* for game developers.

FORMAL METHODS IN CS

Process Mining Assignment

Thank You for the attention!

GitHub Repository

Yuri Tateo – 869303
Francesca Palumbo – 869311