

# What contributes to player retention in Apex Legends?

## Introduction:

For my Capstone 3 project, I will be focusing on the player experience within Apex legends, a popular free-to-play battle royale developed by Electronic Arts (EA). With frequent seasonal updates, new legends, and evolving gameplay mechanics, the game thrives on keeping its community engaged. As a gamer, game artist, and not aspiring data scientist, I'm very interested in understanding the behavioral patterns that drive player retention or churn. This project will analyze in-game behaviors to identify the key indicators of long-term engagement, using machine learning techniques to classify players and uncover behavioral segments. The goal is to generate business insights specifically around client retention and characterization.

## Problem:

Player retention is one of the most critical success factors in the live-service gaming model. Without strong core active players, free-to-play games like Apex Legends risk declining revenue, ineffective matchmaking, and lower user acquisition through word-of-mouth. The core question I aim to answer is: **What in game behaviors contribute to a player's likelihood to remain engaged or churn?**

By identifying behavioral signals of churn and retention, EA's product teams can better shape gameplay loops, reward systems, and promotional strategies. The primary stakeholders for this analysis include **EA's game designers, product managers, marketing strategists, and monetization teams.**

## Data:

The dataset I will be using for this project will be Kaggle Apex Legends Season 15 Ranked dataset (Raw Match Logs), which includes match-level data such as time timestamps, player IDs, kills, assists, match outcomes, and session duration. This structure makes it ideal for retention analysis, allowing me to define churn based on recent activity and engineer features like win rate, aggression, and session consistency.

## Approach:

### 1. *Retention Classification:*

I'll start by building a model to predict player churn based on recent gameplay activity. Churn will be defined by session recency, and I'll engineer features like session frequency, match placement, kill/death ratio, ranked vs. casual mode usage, and legend variety. I plan to test logistic regression, Random Forest, and XGBoost, using cross-validation and

hyperparameter tuning. SHAP values will help highlight which features most influence retention, making the results easy to interpret and actionable.

2. *Behavioral Segmentation (Unsupervised Learning):*

I'll also cluster players based on their gameplay pattern to identify different player types like daily grinders, casuals, or experimental players. I'll use features like match count, engagement consistency, and legend switching, then apply K-means for clustering and PCA will be used for visualizations. This segmentation will give a clearer picture of who's playing and how their behaviors relate to retention risk.

**Deliverables:**

I'll deliver a complete churn prediction model, behavioral player segments, and a concise report with actionable recommendations. Final outputs will include a well-documented Jupyter notebook (EDA, modeling, insights), summary report, and visuals to support strategic decisions for stakeholder teams at EA.