

Abstract

The video game industry's evolution, with millions of players generating vast and complex data, necessitates data-driven approaches to understand player behavior. Traditional user persona creation relies on methods like questionnaires, which are often limited by subjective data and bias, posing a significant challenge for Games as a Service (GaaS) models that depend on continuous player loyalty. This study presents and validates a data-driven workflow to develop accurate player personas by analyzing objective player activity logs. The methodology involves parsing raw server logs from the sandbox survival game, Project Zomboid, encoding player actions into TF-IDF feature vectors, and applying K-Means clustering to group players with similar behavioral profiles. The analysis yielded four distinct player personas, each characterized by unique playstyles grounded in statistically significant in-game behaviors, such as tendencies toward combat, crafting, support, or mobility. To validate these personas, a study was conducted with returning players from the server. The results demonstrated strong face validity, as 55.6% of participants correctly selected their data-driven persona as their first-choice match, with the cumulative top-two alignment reaching 77.8%. These findings suggest the generated personas are meaningful representations of player behavior. This research presents a validated method for integrating quantitative player analytics with human-centric game development, providing developers with an evidence-based tool to inform content creation and enhance player satisfaction.

Index Terms—Data-Driven Persona, Video Game Analytics, Video Game Log, Clustering, Player-Centric Game Development