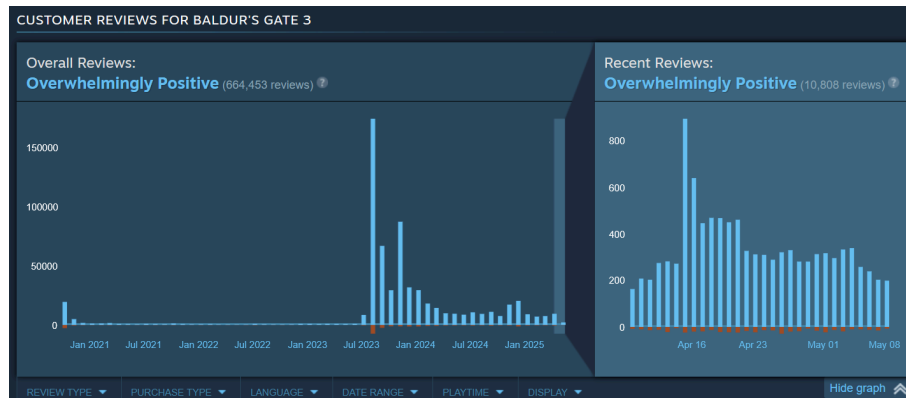# Steam Review Sentiment Analysis

Context

Almost everyone who plays PC games uses Steam: with its massive sales and wide selection of titles, it's one of the most popular online gaming marketplaces today. Gamers can buy games, trade collectible items, set up custom profiles, connect with friends, and, most relevant for this project, review games that they've purchased. Gamers often consult these reviews before making decisions about purchases, looking at factors like graphics, gameplay, artstyle, writing, value, and more. Most importantly, reviewers can either recommend or not recommend a game, which contributes to that game's overall review score, as seen above for the popular RPG *Baldur's Gate 3*.

Motivation

Review scores in general can provide valuable information on the popularity and public view of video games, and Steam reviews are no different. By analyzing text sentiments, we can determine whether reviews are positive or negative, and which game components come up frequently in reviews. Getting sentiment scores and then analyzing the significance of those scores would allow Steam and the game companies that partner with it to sell their products to analyze the performance of their titles, as well as find out which aspects people like and dislike, which can help them adjust their production and marketing strategies to increase sales.

Deliverable

Your goal is to use data scraped from Steam reviews to analyze the extent to which the inferred sentiment of Steam reviews aligns with recommended/not recommended status of the reviews, and which specific game components correlate more with positive review scores. To do this, you will use the scripts and data provided to generate various charts and visualizations demonstrating the correlation (if there exists one) between steam review sentiments and recommendations, as well as between identified gameplay elements in the reviews and recommendations. Throughout this project, you will use sentiment analysis applications in R and Python to derive information from the data, before using statistical tests to determine the significance of your findings.