WordCount Writeup

Jordan Taylor

December 9, 2021

In this project, I produced the top ten most negatively reviewed games on Steam. Hadoop is quite frustrating to deal with when it comes to navigating between the web address to view your file, using filezilla to acquire these files, and running commands on ubunto using your written code. I ran into unexpected issues because of user-created comments in the dataset. I tried running everything in Java but got sick of the dataset not doing what I needed so I ran one program of Python to "clean up" the dataset.
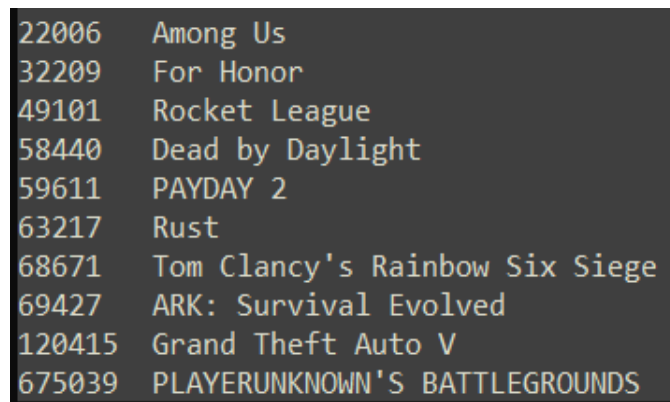
(1) The Map Reduce design was pretty complicated. Initially, I went off of the WordCount code to figure out how to produce the results but I ran into an issue with that. For starters, I tried doing what I felt was simple and stored all of the data into an array. The reason I did this was to grab all of the negative reviews by index, along with the name of the game. Unfortunately, the user's reviews were quite problematic. In these reviews, they had commas that interferred with my delimiter and ASCII art, which had newlines. These factors ultimately made me lose sanity and I ended up created a Python program to only represent the name and reviews. After this, I had to add back the commas that it removed so that I could sort it back out with the highest number of negative reviews (I used the Sort program that was used for the WordCount, with a small modification).

(2) My greatest trouble was going through the thought process on how I can remove the comments from the users. With some advice from my professonal friends, I was able to conclude that Python would be an easy fix for this.

(3) One of the odd issues I had was the program would sometimes confuse entire sentences for being a game, when they were only a review from a user. This was because of the delimiters in the reviews (commas).

(4) It was very interesting and exciting to see what the most negatively reviewed games were. I hear so much about PlayerUnknownBattlegrounds but I would have never imagined it would be the most negatively reviewed game.

As for conclusions, the most negatively reviewed game was PlayerUnknownBattlegrounds. However, as I mentioned before, this game is talked about a lot. In fact, every game on the top ten are popular and that is why I think they should be viewed as the most popular games rather than negatively reviewed games. They receive so many reviews that of course they are going to be the most negatively reviewed. I initially wanted to do this because I was curious on what the "worst game" would be. However, I believe I did the opposite and found the most popular and favored games. It's ironic actually how it all resulted and came to me.

Figure 1: Here is a picture with the top 10 most negatively reviewed Steam games.