

Jimmy Tran

Ahmadzadeh

CSC 4470

2 November 2017

Mini-Project Milestone 2

Note: Due to last week's discussion, I have changed the data I am working with to the gaming statistics of a popular MMO I used to play: Path of Exile.

The data I have chosen to work with contains information on every character created in Path of Exile up to October 26, 2017. (At least I believe so. That was the latest update date for the .csv file itself. It's recent in any case.) This data includes the rank of the character (ladder-dependent), status (alive or dead), online status (as of when the data was mined), name of the character, level, class, character id, total experience, account name, number of challenges completed, twitch account (if applicable; null if not), and ladder. The only other bits of significant information missing would be skill trees (but that would be overly complex and irritating to add for various reasons - one of them being near-complete randomness in certain parts) and the actual stats of the characters (again annoying due to sheer randomness and far too many variables to account for). Interestingly, Twitch accounts are included here too as is experience (this one being odd as all players cap at the same amount and level would reply similar information). I'm actually not certain what challenges are as of the time of this writing so I may need to look that up again to make use of that attribute.

The Twitch attribute actually uses null values in it for players who lack a Twitch account. If I extract data from the .csv file at all, I need to consider whether to outright ignore that attribute or separate by players who do or do not have a Twitch account. In addition, the data is

organized by rank which can be a bit misleading to look at considering four people have 1st place (and any rank actually). It may be more intuitive to reorder the data by ladder or account name for that reason. Lastly, the id and online status attributes are kind of useless. The online status data pretty much went out-of-date the moment the data was extracted. The id is just a unique identifier for every character and serves no other purpose. It is rather useless anyway as characters must have unique names in the first place. I may just remove the id attribute entirely.

I mentioned that the online status attribute was worthless but assuming I actually wanted to generalize data found at a specific point of time, I could use it to calculate how many players actually are active at a given time on the servers. The caveat would be checking the players as opposed to their characters. I am not sure if the online status is for the account or character which could be bad especially if a player logs in from multiple computers simultaneously on the same account. Ignoring that though, I could more easily find where/when most players seem to stop or quit playing the game by plotting player levels in clusters (assuming I figure out how to plot clusters in R anyway). A game making new endgame content wouldn't be very good if most players never reach it after all. I could also plot which players are in which ladders to see what ladder is most popular (by account preferably over character). In particular for the hardcore ladders, I could calculate how many characters actually survive to the end of it which I could assume to be a character at level 100 (This is much harder than it sounds when you only get one life). I can also see which classes are most popular by counting or plotting the number of characters of each class in the game (It's gonna be necromancer. Best class. Calling it now). There's probably more I could figure out but that's is all I have for now at a glance of the data.