Kevin Yao

11/17/17

GW Homework 4 - Pandas Data Challenge 1

Three observable trends:

1. There is a correlation between the most purchased items and the most profitable items. This should be no surprise, as the most profitable items are calculated using the amount purchased. For dataset 2, the same items appear on both the top five lists of most purchased and most profitable items. However, for the first dataset, this is not the case. There was only one item (the Retribution Axe) that was on both top five lists. This could be because all the other most purchased items cost less than $2.50 and were likely purchased more often due to the lower price.
2. Overall, there are vastly more male players than female players, around 81% male versus 17.6% female with the remainder made up of other/non-disclosed players. This explains the greater overall purchase counts by males. However, in both datasets there are also more repeat purchases by male players. Perhaps the genre of the game appeals more to males or perhaps the marketing heavily targeted them.
3. For both datasets, there appears to be an approximate normal distribution for player age on purchase count. There are big spikes at the 20-24 age range with counts in all the surrounding age ranges. This age range did not stand apart from the others in average purchasing prices. Their increased profitability came solely due to their large number of purchases. It could be that people in this age range tend to be more attracted to this genre of games or to gaming in general. It could also be that marketing that heavily targeted this age group was effective.