**Heroes of Pymoli Summary Data Analysis**

As with most data sets collected, the Heroes of Pymoli in game purchase data showcases elements that are expected, unexpected as well as some that turned out to be rather frustrating to analyze.

The most expected trend within the data was that males make up a huge majority of their purchase base at around 84%. However, they turned out to be the most frugal overall with their purchases as their average item purchase price of $3.02 fell below the overall item purchase price of $3.05. This would mean the other two group (Female and Other) would have to have a much higher item purchase price average to bump the overall average up and they certainly did. Despite female and other only accounting for around 14% and 2% of the player count, they had averages of $3.20 and $3.35. I went into this data set believing the three averages as a hypothesis test that they would not have much difference between them but that does not seem to be the case. Pymoli may benefit from focusing some marketing into these two groups if they have not already.

The age bracket statistics were also a bit unexpected. The three brackets that span the most gamers ranges from 15 to 29 years old and accounts for nearly 77% of all gamers yet they had item purchase price averages either at the overall purchase average or below. 15 to 19 years old averaged $3.04, 20 to 24 years old averaged $3.05 and 25-29 years old averaged just $2.90. Meanwhile, the two brackets with the highest item purchase price averages were 10 years and younger ($3.35) and 35 to 39 years old ($3.60). Despite only accounting for just over 8% of the player base, Pymoli may benefit from focusing some marketing into these two groups due to their higher than average item purchase price.

The most surprising element while analyzing the data was the relatively small total revenue generated from the top 5 spenders. I always follow the Pareto principle for most types of sales data that you usually get most sales from a minority of the clientele. However, the top 5 spenders only accounted for just over 3% of total revenue. This would be a main point that I would love to analyze further to see if revenue grouped by each player would follow a power distribution (as again would be expected with a hypothesis test) or if it follows something more conventional like a normal distribution.

Despite the interesting points found within this data set, there were some parts that proved to be a headache in the analysis. The first point had to do with the total Item ID numbers. After using a value\_counts() on Item ID in the data set and then doing a count() on it, a result of 183 unique Item IDs was given. However, I knew this was incorrect as the Item IDs ranged from 0 to 183 and assumed 184 was the correct total. Sure enough, while looking over the actual data set, Item ID number 36 was not accounted for in the data set. While I could have made a function to determine this rather than looking over the data myself, I feel this would not be something that anyone would bother to check.

The second point of agony came when doing analysis on the top 5 most popular items. Grouping the data set by Item name, counting all the elements in each row for each column and then sorting it, the item Final Critic had the most purchases at 13. However, I soon noticed a huge issue with this. Final Critic has 2 different item IDs (92 and 101) with 2 different item prices ($4.88 and $4.19) each with different purchase counts. I mulled over this until I decide to regroup the original data set by using Item ID as the definitive way to get the total number of unique items.

I went with this route of thinking since Heroes of Pymoli is a fantasy game, there could very easily be items with the same name yet have different levels of elements to them. It would seem Final Critic is not the only item that this applies to as a count of the number of item names yields a total of 179 meaning that there are 4 items in the data set with a unique name but have 2 different Item ID numbers.

Whenever I analyze a data set that has an odd discrepancy within it, I tend to call them boobytraps. They are little elements in the data set that one would usually not check for but will cause issues with the analysis and can be hard to pinpoint. In this case, item total could either be 179, 183, or 184 depending on whether this issue is realized.