This is an overview of features-hamby from 2017. There is a lot of mess and I will try to explain what is going on within the data. There are 3 sets of data within in the data frame and these include: Hamby44, Hamby252, and Cary. Cary has been removed from the data frame because we are only working with Hamby44 and Hamby252. Before I conducted the analysis on the data frame, I did do some minor cleaning to the data. I do not believe this added to the messiness of the data that was there already.

1. read in csv

2. removed columns that were not necessary

3. removed Cary observations

4. Changed the names of certain columns to match and later compare with our current 2019 outputs.

5. Changed the Hamby44 observations to Hamby173

6. Changed the feature names to compare old and new features with no confusion between the two

7. Groups Bullets A, B, C, D,…, Z into a Barrel type “Unknown” and assigned A, B, C, D,…, Z as the Bullets for Barrel “Unknown”

8. I ordered the data frame by Set so all of Hamby173 information come before Hamby252 information

Old:

A screenshot of a social media post

Description automatically generated

New:

A screenshot of a social media post

Description automatically generated

As you can see there is now the same number of Barrels, Bullets, and Lands in both the 2017 and 2019 data. Again, I do not believe performing these changes affected the data besides getting rid of Cary observations. I also checked this by examining the excel file separately.

Now I decided to exam each Barrel by set. So I looked at Set Hamby173 and Hamby252 separately. Here is the notable information that should be mentioned. I will do each Barrel in order.

Barrel1

Looking at Barrel1 and 173 we see there are

* Duplicate rows
* Comparing between Bullets is not symmetric
* Compare\_id is all 2’s
* Duplicate rows have different profile\_ids

Looking at Barrel1 and 252 we see there are



* Duplicate rows
* Compare\_id is all 2’s
* Majority of Comparisons are symmetric
* There is not symmetric comparisons at the end which looks like 173 data
* Duplicate rows have different profile\_ids

In total there is 852 rows with Barrel1

This is the general pattern we see for every single Barrel

A trend is the increasing of duplicate rows for example :



Barrel3 is compared to

Barrel 1 (2 times) and has different profile1\_id

Barrel 2 (2times) and has different profile1\_id

Barrel 3

Barrel 10 (2 times) and has different profile\_id

Barrel 4 is compared to

Barrel 1 (2 times) and has different profile1\_id

Barrel 2 (2 times) and has different profile1\_id

Barrel 3 (2 times) and has different profile1\_id

Barrel 4

Barrel 10 (2 times) and has different profile1\_id

Example of none symmetric data for Barrel 3, Set 173:

A close up of text on a white background

Description automatically generated

* Hamby173 seems to be messier than Hamby252 because there are no symmetric comparisons
* Hamby252 has data input similar to Hamby173 at the end of a Barrel comparison, this data is a duplicate of previous data but is not symmetric like the rest of Hambt252 data
* Different profile\_id has different feature values

When looking at Barrel Unknown, we see the similar pattern we saw in Barrels 1-10 but there are also a lot of NA values.

Looking at Barrel Unknown, we see that there is no Unknown Barrel Comparisons with another Unknown Barrel

Side Note: If the data comparisons are not symmetric it is not a problem because in theory their values should be the same.