Machine Learning Final

To begin segmenting our customers, the first question we must ask from a business perspective is: "Who are our most valuable customers?". Isolating and identifying this customer base will allow our company to run promotions more efficiently. This analysis is exploratory in nature, with the goal of finding our most valuable customers using cluster analysis algorithms. The steps of the analysis are as follows:

1. Clean the data
2. Prep the data
3. Analyze

Step 1 & 2:

The analysis starts by first removing records where gender is unidentified. Since the subset of data is focused on bath soap products, and bath soap products are generally gender specific, it makes sense to remove the records of data where gender cannot be identified, as unidentified genders could represent noise in the data. Upon further analysis of the data, 96.1% of customers represented in the data were female. Due to the extremely high concentration of female customers, the male customers were also removed from the data prior to analysis, as these records could also represent noise in the data and alter the final results of the analysis. Therefore, for accuracy purposes, these records were removed.

Step 3:

The analysis was performed in 3 separate ways:

1. Purchase Behavior/Brand Loyalty
2. Basis of Purchasing (Promotion based behavior)
3. Purchase Behavior AND Basis of Purchase

1.) Purchase Behavior:
   a. This analysis was done by grouping customers based on the frequency in which they purchase products from the same brand, the number of times they purchase from the same brand in a row, and the customers brand loyalty factor
   b. Brand loyalty factor is estimated by using the max percentage of purchase volume for a single brand, across all possible bands.

2.) Basis of Purchasing:
   a. This analysis was done by grouping customers based on their adoption of promotions ran, and the price paid for the products

3.) Purchase Behavior AND Basis of Purchase:
   a. This analysis was done by grouping customer based on the criteria of both analysis 1 & 2.

Results:

The results of the analysis were able to isolate the most valuable customer segment, with the 3rd analysis (grouping by Purchase Behavior AND Basis of Purchasing) showing the most favorable results. Here, customer value was determined by average value per transaction.

Analysis 1 Results:

- The first analysis was able to isolate the most valuable customer group, shown below:

	BehavClus	Value
*	<int>	<dbl>
1	1	46.9
2	2	44.1
3	3	97.6
4	4	37.0
5	5	39.2

- This was the most valuable customer group (group 3) identified by any of the analyses done, except it represented the smallest percentage of customers identified by any of the analyses done as well:

	BehavClus	Percentage
*	<int>	<dbl>
1	1	2.74
2	2	26.0
3	3	6.07
4	4	15.9
5	5	49.3

- The problem with this, is that the target promotions will only reach a small customer base

Analysis 2 Results:

- Analysis 2 failed to identify any high value customer segments, but had a most even distribution across customer segmentation:

	BasisClus	Value
*	<int>	<dbl>
1	1	53.7
2	2	38.2
3	3	40.9
4	4	37.3
5	5	47.5

	BasisClus	Percentage
*	<int>	<dbl>
1	1	8.02
2	2	30.1
3	3	8.61
4	4	11.7
5	5	41.5

- This fails to deliver in targeting high value customers

Analysis 3 Results:

- By combining the criteria of the first two analyses, we were able to identify high value customers with a more even distribution:

	BothClus	Value
*	<int>	<dbl>
1	1	44.8
2	2	59.6
3	3	60.3
4	4	37.7
5	5	36.9

	BothClus	Percentage
*	<int>	<dbl>
1	1	19.6
2	2	2.54
3	3	19.4
4	4	11.2
5	5	47.4

- Here we see that customer segments 2 & 3 have a much higher average transaction value than the other two segments, and these two segments combined represent over a fifth of the total customer base (21.94%)

Conclusion:

Using targeted promotions on customer segments 1 & 2 will result in more efficient marketing efforts by driving high value transactions.