

Product Segmentation of XYZ Products

In the segmentation of XYZ products we wanted to focus on two aspects initially, risk and opportunity. Our thinking is that our priority when looking at the list of products offered by XYZ is, first and foremost, identifying any products that are inherently risky. In engineering our first priority is the elimination of risk, so if there are products that are risky, finding a way to allay their risk should be our first priority.

After risk is eliminated we can focus on opportunities. This is where segmentation can help us to make decisions on prioritizing negotiations and pursuing brand partners.

Risk

Three behaviours jumped out immediately as being potential sources of risk:

- products with small unit profits pose a risk as external costs could increase
- products with small percent margins pose a risk as they are costly to obtain compared to what they provide in return

We calculated the average unit profit by product and sorted the results. This did not produce any alarming results:

	unit_profit
product_name	
Hazelnuts Natural (13-15mm) 12.5kg Bag	4.640544
Peanut Blanched Splits (61/71) (Subject to availability) 25kg Bag	4.928365
Gluten Free Oats [Ingredient & Zanna's] 25kg Bag	5.681739
Cashew LWP (Large White Pieces) 20kg Bag	6.399066
Cranberries 11.34kg Bag	7.017329
...	...
Dried Blueberries - Sugar infused 12.5kg Bag	18.549984
Sultanas Medium Choice 12.5kg Bag	19.047055
Pineapple Cubes 20kg Bag	19.438636
Buckwheat 25kg Bag	20.218304
Dried Blueberries - Apple Juice Infused 10 kg Bag	20.714167

Similar analysis was done to calculate the average percent margin per product, this also did not produce any alarming results:

	unit_profit	unit_price	percentage_profit
product_name			
White Sugar 25kg Bag	7.670000	21.930000	34.974920
Gluten Free Oats [Ingredient & Zanna's] 25kg Bag	5.681739	16.244782	34.975512
Coconut Sugar 25kg Bag	9.677194	27.659078	34.987398
Plasctic Square Tub 102 Tubs per box PPWTUB	12.087604	34.546866	34.989164
Walnuts Light Halves (Chandler Variety) 15kg Bag	12.284700	35.109432	34.989454
...
Buckwheat 25kg Bag	20.218304	44.932161	44.997694
Macadamia Style 6 Raw 11.34kg Bag	12.919496	28.709267	45.001102
Raisins Medium Choice 12.5kg Bag	17.236858	38.299180	45.005849
Figs - Turkish Large 10kg Bag	12.065258	26.808075	45.006387
Oats Flour 25kg Bag	15.719600	34.919800	45.016486

Opportunity

There were two questions lingering from the warehouse analysis that felt important in terms of segmentation:

- Where should new stock be stored?
- When should new stock be acquired?

These questions presented two segments:

- regional/ universal: products that sell more within a particular region of the United States versus those that don't
- seasonal/ staple: products that sell more given the time of year versus those that don't

Our approach in answering this problem was to use a simple chi-square test to determine whether or not the null hypothesis in each case is true (i.e. the product does *not* sell any better in particular states, the product does *not* sell any better in any particular month). However we ran into some technical issues and were unable to complete the analysis. We still believe that these segments have value and are worth pursuing.

Future Work

The segmentation work is clearly not finished. The risk framework does not provide much value at this point in time (beyond the reassurance that obvious risks are not present) and the opportunity segments are not working. We believe it is most important in this light to pursue the existing opportunity segments.

After the questions of regionality and seasonality have been addressed, the team would like to continue the segmentation process by submitting the dataset to clustering algorithms, namely k-means clustering, which may provide segments the team may otherwise not have thought of. The work to answer the existing questions of regionality/seasonality will only enrich the clustering approach, as they may provide new labels for data (e.g. season as opposed to month) and additional data cleaning will be valuable before clustering.