There are several demographic factors that have potential to influence a customer’s interaction with the store and product. For this example, we are examining how demographic factors may show a correlation between a customer’s attitude towards products with an organic label, marked by the NATURAL\_ORGANIC\_FLAG. The demographic factors we have chosen to focus on are age and income range, marked by the AGE\_RANGE and INCOME\_RANGE attributes respectively. We ignored null values for these attributes because they could represent any age or income range.

For the purpose of simplicity, we are choosing to examine these attributes separately from one another, first let us look at a customer’s age. Counts of products with a value of ‘Y’ for the organic flag were grouped by age range and then averaged. From the following visual representation of these results, it appears that younger households purchase more organic product.

But it is important to consider that younger households could be larger in size, and generally just tend to purchase more products of all types because there are more people. To normalize this data, the number of all products purchased was all calculated by using the same queries, but without the filter for organic products. The ratio of organic products purchased to all products purchased was then calculated for each age-range.

From this normalized view of the data, we can see that the trend previously observed holds true, younger customers are more likely to purchase organic products.

Now let us examine a similar representation of the data using Income-Range as opposed to Age-Range. First, we observe solely organic purchases by the quantity again:

From this representation without normalization, it appears that those with higher incomes are much more likely to purchase organic items, with those in brackets of 75-99k and 150k+ purchasing around three times the amount of organic products than the lower two income brackets. We also believe it is interesting to display the intermediate data visualization of all products purchased by income range, as it seems there is not much difference in the amount of products purchased based on income:

And now for the final step using these values to normalize the ratio of organic purchases:

As we can see, there is practically no change in the appearance of the bar graph due to the count of all item purchases by income being so similar to one another. This leads to an interesting conclusion that households with higher income don’t actually purchase more, rather they purchase more organic products, again at nearly a 3 to 1 rations between the higher and lower ends of the income range. Organic products are typically more expensive than non-organic products, so households with more disposable income are able to make these purchases whereas lower income households don’t have the ability to. Perhaps marketing of these products should primarily be aimed at those making over $75,000 since that is the range where there is a significant spike in the number of organic purchases.