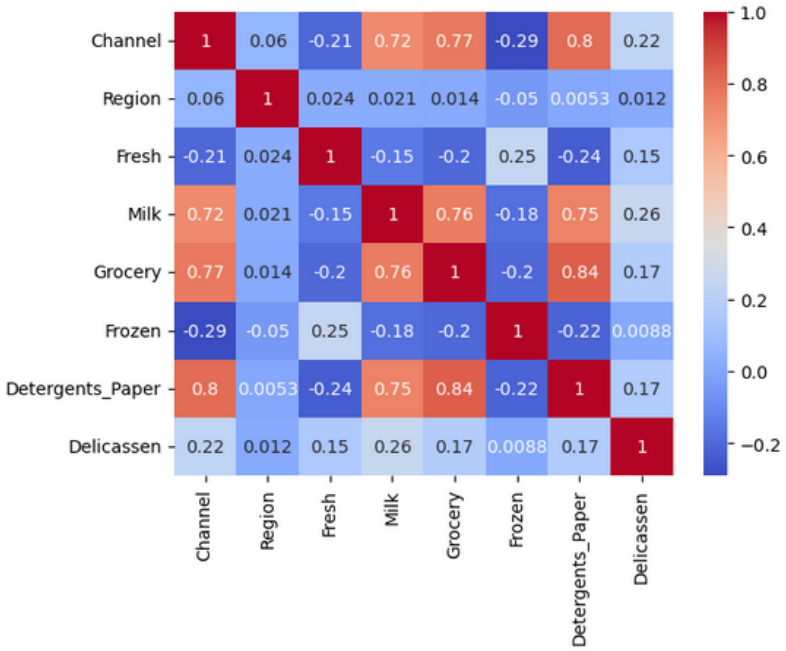
Wholesale Data Modelling
Andrew Elliott

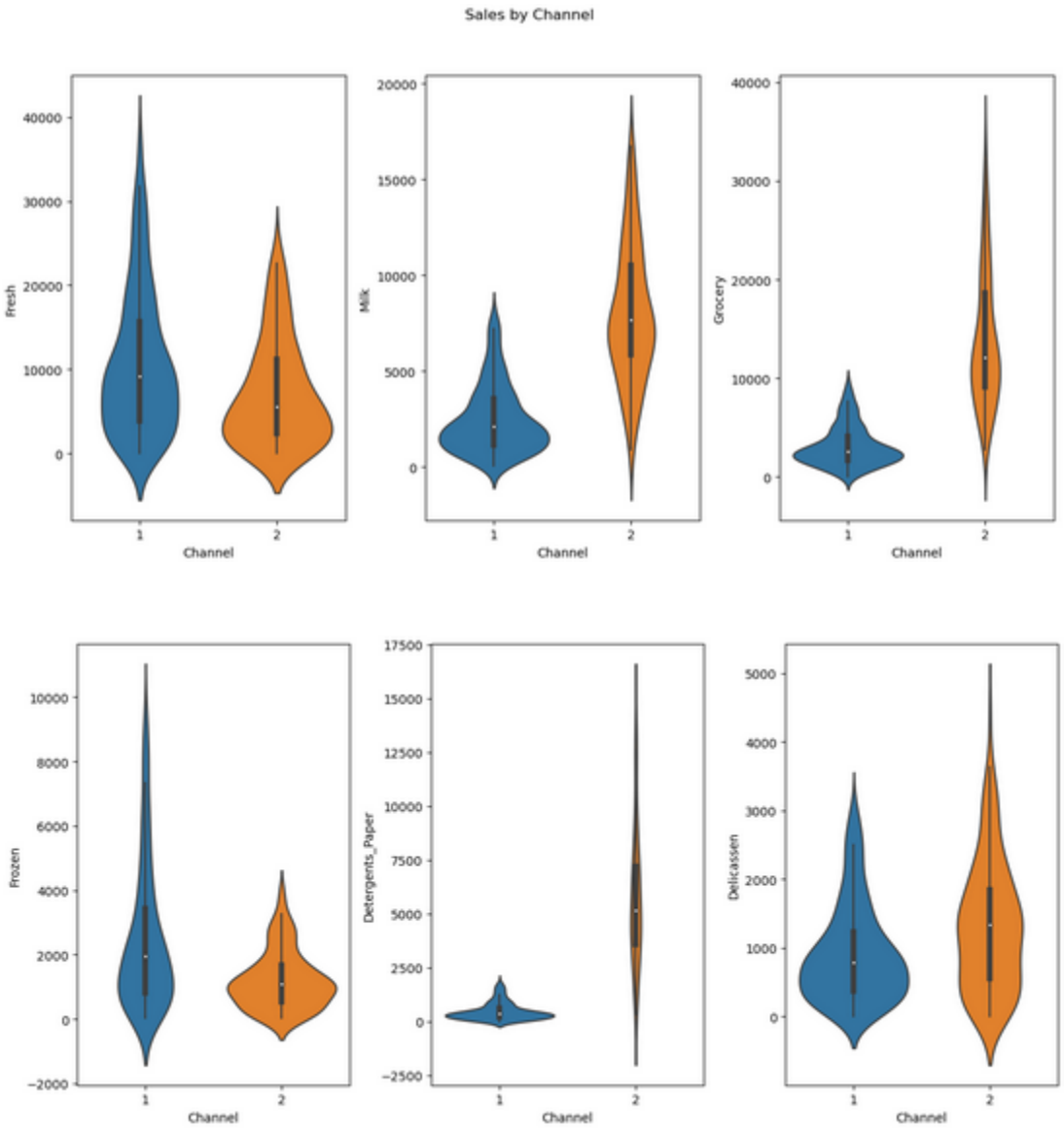
While working with the data available, which involves wholesale product clients (i.e. who we provide products to), I was able to work out a few core facts that could be used to refine marketing strategies and increase sales.

- Correlations:
    - Businesses that buy detergents & paper products also tend to buy a lot of grocery and milk products, and a similarly strong correlation exists for milk and grocery buyers sans detergents & paper products.
    - Similarly, there is a mild correlation between fresh and frozen food sales, as well as delicatessen and milk sales.
    - We have a mild negative correlation between fresh and detergents/paper products, which suggests that clients are less likely to buy both together. There are similarly mild negative correlations for frozen and grocery products, as well as for frozen and detergents & paper products.



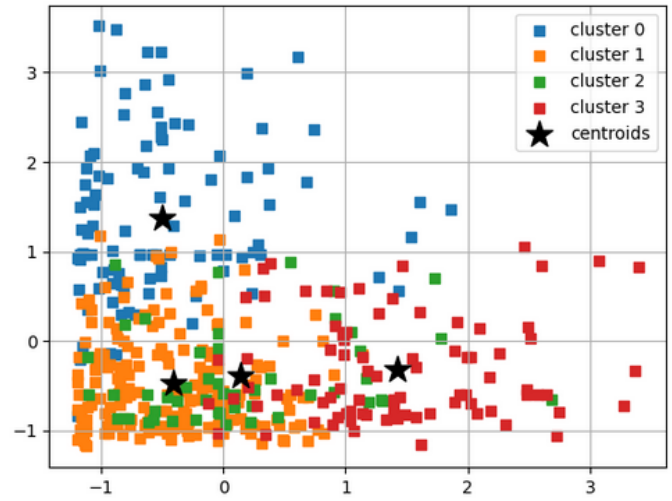
Depicted: Strong correlations in red, weak and negative correlations in blue.

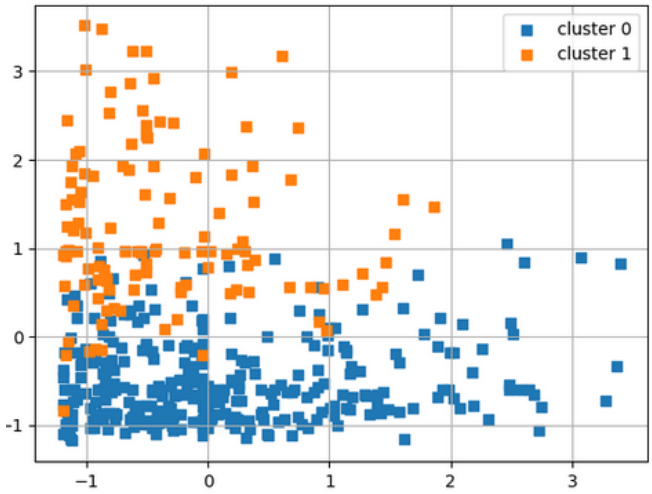
- There are no substantial differences in the purchasing habits of wholesale customers between different regions, meaning that any strategies employed to improve sales should be applicable in any of the three identified regions or region categories. However, there are differences between different channels (Ho/Re/Ca and Retail) as noted in the above chart.



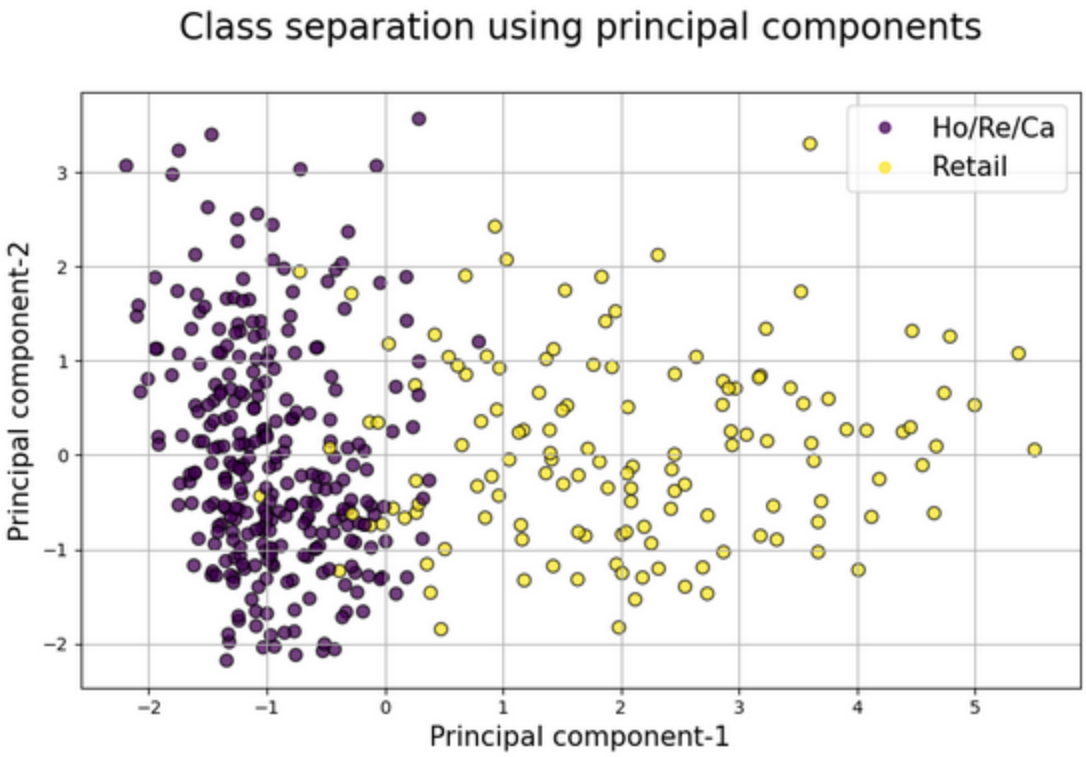
Depicted: Differences in sales to clients based on how we internally classify them. A wider “violin” means a greater number of records fall near a datapoint (with narrower violins indicating the opposite), while taller violins indicate a greater range of m.u. across product categories.

- Depending on the clustering method used, we can identify 2-4 distinct business clusters from the data. This could, for example, help strengthen the employer’s marketing or sales strategies to maximize their business operations and resulting profits.

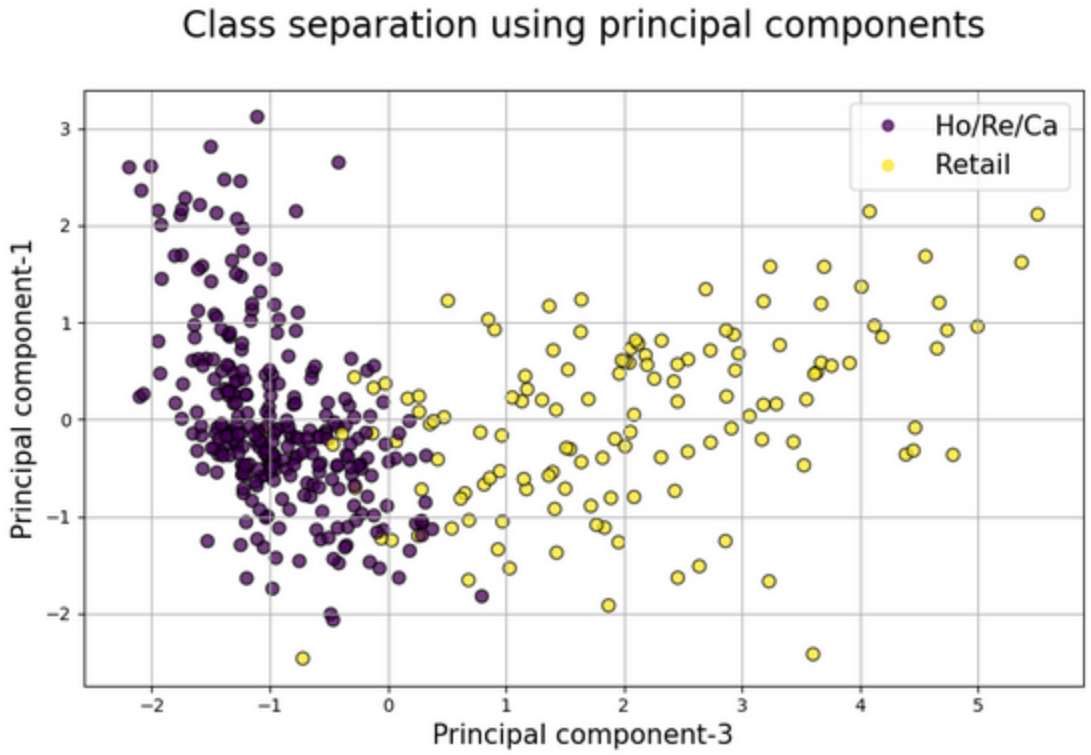




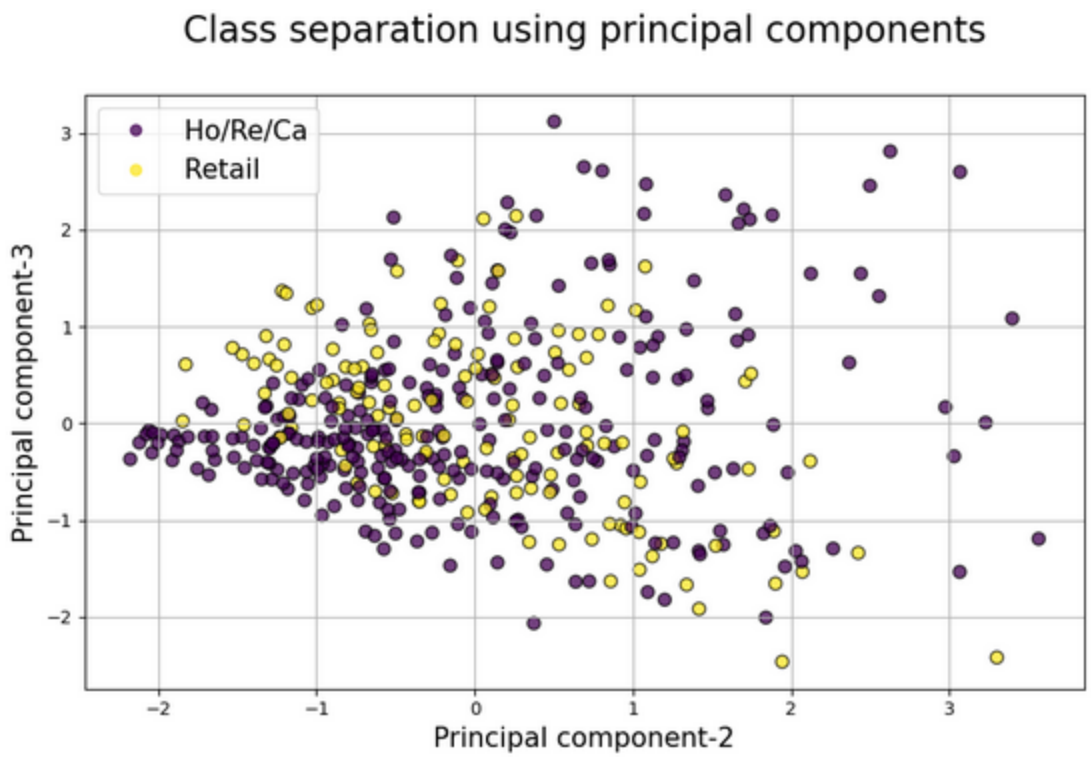
- There are some areas where our two Channels, Ho/Re/Ca and Retail, differ substantially in terms of their buying habits and what they need to conduct business. Specifically, Fresh and Milk products. This could be used to conduct further analyses to figure out exactly how to proceed, but these variances exist and are important to note.



Both channels are substantially different from each other when comparing these two principal components, having only a little overlap approaching PC-1 = 0.



There is a pretty clear distinction between both channels here; Retails is largely PC-3 positive while Ho/Re/Ca is largely PC-1 negative. They also diverge further the greater PC-1 gets, on average.



Even though it is the third highest principle component, we can see that there is not much variance between our two Channels (Ho/Re/Ca and Retail) when it comes to our third principal component; the clearest distinction can be made in the lower portions of PC-2, with Ho/Re/Ca having a clear origin point in the negatives of PC-3 and Retail having a clear origin point in the positives of PC-3. However, both spread outward quite a bit from that point and the distinction diminishes.

Overall, I would say that we need to conduct further analysis to determine the best strategies for improving our business operations. The plus side is that, using the techniques I have briefly outlined, I have determined that we have a good overview for how to proceed and can begin the process of developing and deploying new tactics to improve