

Problem:

A retailer wants to optimize shelf space to maximize sales.

Assumptions:

1. The more shelf space they give a product they more of it they sell (This is also a problem, because this assumption has not been tested.)
2. The more of a product type they sell the more they'll sell complementary products.
3. If two complimentary products are next to each other the effect from above is even greater.

Given:

Point of Sales (POS) data, sales only, product proximities are not part of this data set.

Constraints:

1. Minimum/maximum shelf space
2. Finite amount of shelf space per store

Discussion:

Although assumption number 2 can be easily tested from the POS data assumptions 1 and 3 are not. Furthermore, there is a disconnect between store management and headquarters. Store management believes they are reacting to customer demand and providing more shelf space to popular products. Headquarters believes that the more shelf space is driving demand.

I think it is important to flush out each of these assumptions through hypothesis testing.

Approach:

- Use pairwise association mining to determine assumption 2
- Find a unicorn product
- Use factorial design to determine if assumptions 1 and 3 are correct
- Use results from factorial design to determine effect size
- Use POS, factorial results, and effect results in an optimization model to maximize sales

Pairwise Association Mining:

Although we did not go over pairwise association mining in this course I am taking CSE 6040 concurrently and it was one of the first exercises we did. A brief explanation on the process is using the POS data and create a hash table where pairs of products are keys and the value is the count of how many times they are purchased together. You can find how strong an association two items have by dividing by the total amount an item is purchased. For example: (rice, beans) = 95, rice = 100. So beans are bought together with rice 95% of the time.

Given POS data we can use a pairwise association model to determine how strongly associated products are with each other.

Find a Unicorn Product

I am under the assumption that this is a big box store with many locations, some of the locations are similar in size and sales volume. I hope there is a product that has minimal seasonal variation sold at multiple stores.

Given a list of products with its pairwise associations that are equivalent at multiple stores, take a subset of the data with products that have a minimum of 5 associations with a threshold of 20% or higher. Meaning a product must be associated with 5 other products that are purchased together at least 20% of the time.

Given this subset of products use an exponential smoothing model to determine if there is a product(s) with minimal variation throughout the year.

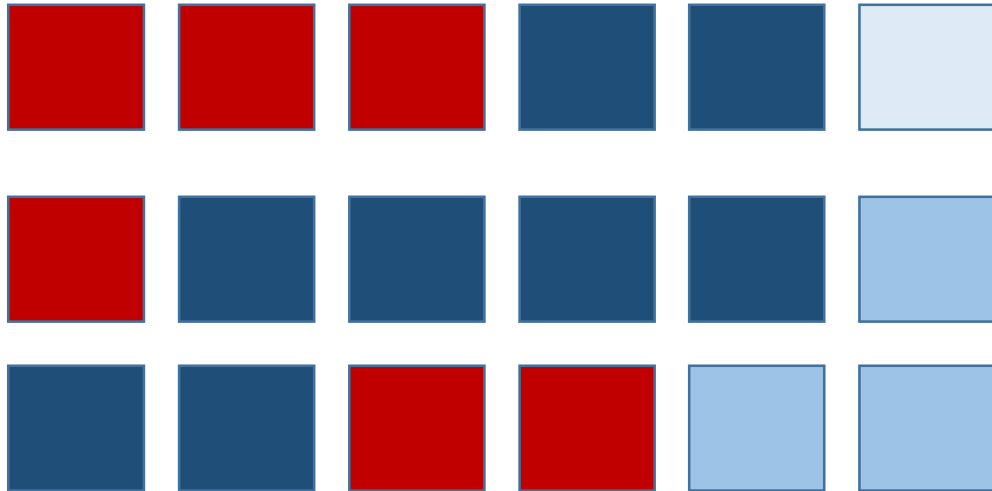
Factorial Design

I want a product with multiple associations to be able to determine if the strength of the products association and the distance away is correlated. In the example below the red is the product and the blue are its associated products with decreasing association, the white box represents an item without any association.



Given a product and its associated products I will use factorial design to test if two complimentary products are next to each other the effect from above is even greater. Distance will not be a linear measure, but a distance of n products away.

Given a product and its associated products I will use factorial design to test the hypothesis more shelf space increases a product's sales. Shelf space will be measured in n products. I can do this with multiple product placement scenarios as shown below.



I am making a big assumption that multiple stores will have the same product(s) and can test at the same time to get the data quickly, but I think this gives a lot of valuable data on top of testing the original hypothesis, such as:

- A products association is inversely correlated with the distance effect
- A grouping of products sales can be maximized by a specific ordering
- An unassociated product sales will increase if placed within a group of associated products.

Effect

Given the results from the factorial design and POS data during that period we can use OLS regression to measure the effect

- **x: product association, product distance y: sales to measure the effect of product association and distance.**
- **x: product shelf space y: sales**

Optimization

Given the POS data, a product's margin, and the previously measured effects an optimization model can be used to maximize sales.

Constraints:

1. Minimum/maximum shelf space
2. Finite amount of shelf space per store

Discussion:

We used pairwise association mining to determine a product's association and if the strength and proximity of products are correlated with an increase in sales. We used a factorial design to test multiple hypotheses regarding shelf space and proximity of grouped items. We used the results and point of sales data from the factorial design to measure the effects of association and effect on sales using regression. All of these results and data were then fed into an optimization model to maximize profits.

I think my current model is heavily reliant on finding a product that can be used to determine the given hypotheses and effects, but big box stores have many related items and have the same offerings throughout the country.

My weakest point may be the factorial design. I think I would have to set it up by rating the association as $1, 2, \dots, n$ where 1 is the strongest and n is the weakest effect, when the associations are not truly factors, but continuous values. I do think the output would add a lot of value to the optimization model, but I would need to really think about how to utilize all of it. I like my approach because this can be used to maximize profit not just sales. For example a given product in the grouping may have the lowest sales, but the highest margin and be only loosely associated with the products. Optimization will help determine how much shelf space will be used to