

## WEEK 9 Homework

---

Describe analytics models and data that could be used to make good recommendations to the retailer. How much shelf space should the company have, to maximize their sales or their profit?

Of course, there are some restrictions – for each product type, the retailer imposed a minimum amount of shelf space required, and a maximum amount that can be devoted; and of course, the physical size of each store means there's a total amount of shelf space that has to be used. But the key is the division of that shelf space among the product types.

For the purposes of this case, I want you to ignore other factors – for example, don't worry about promotions for certain products, and don't consider the fact that some companies pay stores to get more shelf space. Just think about the basic question asked by the retailer, and how you could use analytics to address it.

As part of your answer, I'd like you to think about how to measure the effects. How will you estimate the extra sales the company might get with different amounts of shelf space – and, for that matter, how will you determine whether the effect really exists at all? Maybe the retailer's hypotheses are not all true – can you use analytics to check?

Think about the problem and your approach. Then talk about it with other learners, and share and combine your ideas. And then, put your approaches up on the discussion forum, and give feedback and suggestions to each other.

You can use the {given, use, to} format to guide the discussions: Given {data}, use {model} to {result}.

One of the key issues in this case will be data – in this case, thinking about the data might be harder than thinking about the models.

---

Assumptions in the problem:

The underlying ideas to optimize shelf spacing:

1. First, that the more shelf space they give to a type of product, the more of it they'll sell.
2. Second, the more of a product type they sell, the more they'll sell complementary products.
3. And third, if two complementary product types are next to each other on the shelves, then the complementary effects will be even greater.

Given:

Sales data (e.g. POS data) for each product of a product type and the complimentary products. This data can be obtained from either one store or from multiple stores.

Getting data from multiple stores across multiple cities will help understand seasonality and trends of sales of a certain product type and its complimentary product. For example, a Walmart in a predominantly Latin population might sell Chips and salsa more during Cinco de Mayo than a location with significantly less Latin ethnicity.

Sometimes, it would also benefit to understand the effect of a nearby retail store on a particular product placement. For example, I live in Peoria (IL), and we have a Menards opposite to a Walmart

Supercenter. Understanding sales data of certain types of products (like lumber – hammer and nails (complimentary)) that are sold in both stores would help understand the effect of product placements.

Another point to consider is time series sales data. This would help understand the seasonality of a certain product type. For example, turkey – Cranberry sauce in November, Chips, beer and soda in a Superbowl game, or a Television set and PlayStation/X-box during Black Friday sale event.

To:

Test the second hypothesis: “Second, the more of a product type they sell, the more they'll sell complementary products.” This ranks the product types (and complimentary products) with highest sales based on POS data. This would rank the product types with high to low correlation.

Using one of the following approaches:

- **Clustering:** Clustering would tell us the clusters of different products sold along with the complimentary products that often sell with a type of product. This would help with understanding correlation between sale of one product with another, and we can categorically dismiss sale of irrelevant products with one type of product. For example, a person buying a pair of socks has nothing to do with the same person buying milk in the same sale. Clustering will help reduce the number of factors causing the sale of a certain type of product, thereby reducing the complexity. I think this can be looked at as dimensionality reduction. We should see an inverse correlation between the sale of a product and the distance to its complimentary product.  
Not only could clustering group similar products sold, it can also be used to rank the groups by revenue over a period of time. For example, over a year, milk and eggs might have brought in more revenue than TV and PlayStation. This would come in handy to make the decision of which product type needs better placement in which neighborhood.
- **A/B Testing:** Or, we can also perform A/B Testing to pair products and sales and rank them by highest to lowest revenue.
- **Factorial Design:** Factorial Design will help find association of a product with its complimentary products by sales. This will help to determine if sale of a product is higher when placed near complimentary products. A comparison of individual sales vs. combined sales will give an idea about product clusters.

As a next step:

Given:

Now that we have established the second hypothesis, we now know what complimentary products sell best with a product type.

To:

Test the first and third hypotheses:

“First, that the more shelf space they give to a type of product, the more of it they'll sell.”, and  
“If two complementary product types are next to each other on the shelves, then the complementary effects will be even greater.”

Using:

- **Change Detection using CUSUM:** This would help detect a significant change in sales over a period of time. An example is how much shelf space a retailer like Walmart needs to allocate for candy. In a regular month, there is usually two aisles of candy. But there are more aisles of candy in the month leading to Halloween. After Halloween, there is going to be a drop in the candy sales. However, people will buy candy up to days leading to Christmas, and one can test the change in candy sales below a certain threshold after Halloween. Varying the number of aisles containing candy across a number of stores will help determine the correlation between number of aisles vs. candy sales in months after Halloween.
- **Original Linear Squares Regression:** This is going to correlate the shelf space to sales of a product. Regression is going to help understand the correlation between product grouping, distance between the products in a group, and shelf space; and help determine the sales (as a function of group, distance between products and shelf space).
- **Optimization and Simulation Model:** Run an optimization model to calculate the optimum shelf space needed in a store for a particular product group. Constraints would be maximum sales, minimum distance between products in a group, and a fixed amount of shelf space in a store. Some additional constraints can be the restocking times to replenish empty shelf spaces. Running a simulation model can help determine the optimal shelf space needed to maximize sales of a product type and its complimentary products.