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.

1. Before any sort of testing happens to figure out the best shelving of items, I thought it would make sense to make sure that all the original hypotheses were correct, and if they were correct what is the effect of each of them. The data that came to mind to use for this original test was what is the past sales history of each product, what do the shelves currently look like/are filled with, what products tend to sell with others (for example milk and cereal), and what products are next to each other on the shelves. For this part, it sounds like I must predict responses of certain attributes, I believed a regression model made the most sense here.

   Confirming the hypothesis and estimating the effects:

   Given {previous sales history, current shelf space allotments, complementary product relationships, and neighboring products on shelves}, use {a regression model} to {find the impact of shelf space, validate the store's hypothesis on complementary effects, and find the impact on sales of neighboring products}.

2. Assuming the hypotheses were correct, I would now have values for each product's impact on sales depending on how much shelf space they use and how much neighboring products account for sales of the product. I now need to find the sales for each product. This not only depends on how much of that product is on a shelf, but it also depends on the shelf space of surrounding products. A regression model made sense to me again for finding these values. I am predicting the sales amount of a product based on the attributes above.

   Estimating sales based on shelf space:

   Given {each product's sales change per unit of shelf space and each product's neighboring product's effects}, use {a regression model} to {predict future sales for each product placement and shelf space allotments}.

3. At this point, I would have the numbers for estimated sales based on shelf space for each item and their surrounding products, the effects of complementary products, and the current constraints of the store. With these values, I can find the best way to fill the shelves to promote maximum sales. Originally, this sounded like a clear and cut optimization problem to me, but then I realized it would most likely be an integer optimization problem. Even though heuristics can be used to solve them, I wanted to think of a way for the model to just do most the work instead of needing to use a heuristic. I came up with using a greedy algorithm (forward regression, backwards

elimination, stepwise regression) to solve this problem. One product by one product, this would add/remove shelf space to products based on their marginal benefit (for example their profit per unit of space). Each product type could be allocated space until the marginal benefit drops below a threshold or until space constraints are reached.

Finding best shelf space allotment:

Given {shelf space constraints (minimum amount of shelf space required, and a maximum amount that can be devoted and the physical size of the store), sales and neighboring product impact estimates, and complementary product relationships}, use {a greedy algorithm} to {find a near optimal solution for shelf space}.

4. As with most models we use, validating them is important so that we know the numbers are working with the real data instead of the random data. A/B testing here would allow the store to arrange the shelves in two different ways (the way our model says and a different way) and see which one promotes better sales, use the better one and test it against a different arrangement, and so on until it is clear which is the best way. If the model predicted the best way after the A/B testing, then the model is likely validated and can be used.

   Validation of the model:

   Given {the modeled results}, use {A/B testing} to {see the actual outcomes of the model's predictions}