Homework Thirteen

Chelsea Hughes

Question 19.1

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.

19.1.1, Gather Data:

I think the main problem to be encountered in this homework is gathering useful data followed swiftly by confounding factors. Much of the raw data that we will obtain will not be useful, so **DOE** is going to be a part of gathering data.

Data gathering

(1) Blocking: Splitting the data isn't the primary objective, we would do this as a part of gathering the data to conclude with three distinct batches of data: original sales, sales after increase, sales after decrease. It is important to do this to determine seasonality in products. Best to be done **heuristically** (decreasing and increasing) as we would also want a control group of product that doesn't fluctuate categorized in sections where their complimentary products don't influence sales (I.e. personal care, home décor, grocery) *all of which is assuming* that we have enough store locations and stockers

(2) Factorial: Splitting experiments by variability of income geographically and median income. I live in Cambridge, Massachusetts which is "jokingly" considered to be the 'republic of Boston'. We have higher incomes, higher housing prices and way more organic food stores than Southie. If you are acting off the presumption that this big box retailer has multiple stores, the surrounding neighborhoods would play much a role in what inventory is significant and successful.

Optimization and design

If we consider the ramifications of blocking, changing shelf space can be costly, you have to have stockers, more managers maintaining and ordering inventory and more people like us determining what the data is actually saying. The idea is that you could build an optimization model to reduce this cost by preemptively observing what product space could be modified considering seasonality constraints, storage availability, cost of increasing inventory or space, and contractual import policies.

19.1.2, Model:

19.1.2.1, Testing significance:

Given Data

- Product adjacency
- Fluctuation in sales
- Amount of sales
- Previous sales
- Marginal profit
- Relative profit (DOE)

Use Model

ANOVA testing

To Result

Determine significance of sale fluctuation of relative profit

If assumptions are made in the distribution samples, you could use a non-parametric hypothesis test: **Wilcoxon Median** or **Kruskal-Wallis.**

19.1.2.2, Classification:

Given Data

- Number of sales, sale price, profit
- Stock availability, shelf size, change in shelf size
- Neighborhood household income
- Time of year

Use Model

Logistic regression (probability of change) and **random forest** (analyze significant data)

To Result

Predict increase/decrease in shelf space

You could also use **SVM** and **KNN** to classify whether a change has happened or not. If you could do so accurately, this may be the better way to go so you could see the relationship between the sales. You could probably make stronger predictions going this route.