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.

Answer :

Shelf Space Optimization problem: We need to find the optimum shelf space for each product and need to find which pair of products can be placed together to get best profit.

Phase 1 : Data Preparation : If there are any missing values use imputation techniques to estimate the missing data

Given	Required Shelf Space for each product, Count of each product, Cost of each Product, Daily Profit of each product, Customers list of purchased items, Daily Total Sales, Placement Category, Product Inventory, Day of week, Cost of hosting the product, Customer Data
Use	Impute the missing data with Regression & Perturbation
To	Estimate the missing data.

Phase 2: Get the list of products that sell well

Given	Customer purchased list of items (data collected over the years) i.e Products that sell well
Use	Kmeans Clustering and Logistic Regression on each cluster
To	Get the most relevant items that sell well .

Phase 3: Get the list of possible combinations of products that has high probability of getting purchased together (they can be put together in the shelfs)

Given	From the Highest Profitable clusters – product data obtained as a result of the Kmeans Clustering output from the phase 3
Use	Fractional Factorial Method and Bayesian Modeling
To	Find the possible combination of products to be kept together in the shelfs

Phase 4: Find the Best profitable combinations of products from the list generated from Phase 3 using Optimization.

Given	Using the different combinations of products from the phase 3
Use	Stochastic optimization on that data with constraints such as Shelf Space of Product < Max Shelf Space of Product > Min Sum of all product shelf spaces =< Total shelf space in the store. Prob(Products adjacent to the product under consideration to be purchased) > Threshold T1 Prob(Profit made by the combination) > Threshold T2
To	Maximize the Profit.

Phase 5: Testing the Optimum Solution in different stores which hold similar product inventory.
Once the best combination of products that have high probability of selling together is found, we can test this solution with many stores which hold the same product inventory. Depending on the geographic location and the customer base, not all such stores might find this solution optimum but surely there will be few stores who find this prescribed solution to be optimum.
For the stores where the first prescribed solution did not work try the next best optimal solution and repeat the process until all stores get the optimum results.

Conclusion :

1. First the Store's Product data, Revenue Data of the store and Customer Data is collected, the data is prepared for the models by checking for missing values and using imputation using regression and perturbation the missing values are estimated.
2. Then Using the KMeans Algorithm and Logistic Regression on the Customers list of purchased products, products that are sold well are found.
3. Then using Fractional Factorial and Bayesian Modeling on the products that sold well , we find the possible combinations of products that can bring profits when kept together.
4. We then use Stochastic optimization using different constraints to get the optimum solution of the products list that need to be kept next to each other for higher profitability.
5. We test the optimum solution on different similar stores. For the stores where the first prescribed solution did not work try the next best optimal solution and repeat the process until all stores get the optimum results.