# Question:

**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.

# Approach

1. **Ask:** Correlation of shelf space to the sales of the product

How to achieve?

* To understand correlation of shelf space and Sales for tens of thousands of products can be challenging. I will start with simple unsupervised clustering model to categorize the products to products group. For each group, use non parametric Wilcoxon signed test to test if the change in sales is significant.
* Given Information of existing products group, use the historical sales records to understand the correlation of Sales and product groups to obtain probabilities using Logistic regression.

1. Given smaller subset from the results of high Sales product group and its respective median, Establish Design of Experiments using Fractional Factorial to study the impact of complementary products and see if they are placed next to each other will benefit business.
2. Build Optimization model to determine the shelf space of the products. When determining the products Sales vs. Shelf space , below considerations should be accounted

* Will increasing space for one product decrease shelf space of another?
* Do we increase cost by adding additional shelf? (This can be ignored per requirements)
* Shelf space dependent on seasonality?

1. **Clustering:**

In this use case, the key issue will be collecting the data, more specifically collecting the correct set of products to use for design of experiments. Before suggesting the best optimization for shelf space for products, it is important to understand the current trend of the products and its contribution to Sales & Profit. Using this information, we associate the allotted shelf size for the highly functional product clusters due to high volume of individual products across the retailer shop. Hence clustering becomes the first step in focusing on specific target group for our downstream process

1. **Design of experiments:**

For this specific use case, though collection is data is achievable. Designing a robust experiment accounting into seasonality, an individual products relation with its complementary products and subsequently the complimentary products relationship with its own set of complementary products could run into scenarios of parent child hierarchy or a loop structure. Hence to avoid overwhelming ourselves and the model, design of experiments can setup a balanced designs of Fractional Factorial on the product groups as result of clustering on high impact factors ONLY. The set of highly impactful factors should be worked with the business owners of the retailer to determine a balanced set.

Given the data collected through DOE, Logistic regression can be used to verify the Sales as response to understand the probability of getting a Sales and the interaction of the two of more complementary products to determine the placement option

1. **Optimization:**

Given the downsized product groups and results of DOE that provides combination of factors that work better together, Optimization model can be setup to provide suggested shelf size for each product group. My suggestion will be to categorize the product groups based on the sales pattern and target the highest sales pattern group in the first, followed by the other group. Basically an extension of Multi arm bandit approach that will give us opportunity to exploit the learning of our exploration model.

This will not be one time design and done effort. As Sales of the products will be frequently changing based on seasonality, cyclic/seasonality trends should be modeled as part of scenario based modeling to understand the high impact product group and focus on the complementary products associated to these groups.