# **Retailer Case**

# Problem Statement

**The goal of this project** is to optimize shelf spacing. We are basing our models on three assumptions/hypotheses:

* The more shelf space given to a product, the more it will sell.
* The more of a product type is sold, the more a complementary product will be sold.
* If the final products are next to each other, the complementary effects will be even greater.

**Our first objective** is to confirm the validity of these assumptions. This will be a bit trickier than it may look, so it will be a big chunk of the analytical process.

**Out second objective** is to figure out what model can we use to decide how much shelf should be allotted for each product to maximize sales? Potentially, we could refine our model to optimize our model for profit and not for sales, as it is ultimately profit that the company wants to maximize.

There are a few restrictions that are imposed upon us:

* Minimum and maximum amount of shelf space for each product
* Total shelf space is a constant amount

And we are told to ignore:

* Promotions for certain products or product types
* The fact that some companies pay for more space

The difficulty in optimizing for sales is how to accurate predict the sales of a product/product type based on the shelf space it is given.

Now that we have the problem defined, let’s dive in!

# Objective 1 – Validate the hypotheses

Fortunately, the second and third hypotheses are easy to test.

**Second hypothesis** – Previous data that was collected (providing exact sales data of each product) provided the information to test which products sold well together and which didn’t. In essence, a correlation test could be done on each product pair and if the correlation is strong enough, then a determination could be made that a relationship exists between the sales of two products. One might be inclined to think that just because there is a correlation, does not necessarily indicate causation. In this context, meaning that just because a certain product sells higher, doesn’t mean that it will drive sales for the other product, vice-versa. However, as the grocery store does not need to prove causation but only needs to prove that there is indeed a relationship between certain types of products, a hypothesis test will suffice. The p-value can of course be altered. Increasing it may increase the number of incorrect relationships, where a low p-value might ignore potential relationships. If one wanted to go really deep, then one could pick a p value that would optimize for the cost of ignoring a potential relationship between products and validating incorrect relationships. Leaving the p-value as is at 0.05 may also be acceptable as well.

Using a regression model would also enable us to measure the magnitude of the complementary effects. It is important for our regression model to model non-linear relationships as we should not assume that increasing sales of one product will correspond with a linear increase in the other.

**The third hypothesis** is a bit more tricky. There was relatively limited information on specific product locations from the collected data which makes it more difficult to prove the third hypothesis. But using the limited data according to Professor Sokol should be enough to prove the third hypothesis for many of the relationships detected in our validation of the first hypothesis.

**To prove the first hypothesis**, it requires us to prove that a marginal increase in shelf space will result in a marginal increase in sales. Intuitively, this would not be a linear relationship but might resemble a logarithmic function which I have shown below. That is a small increase in shelf space from 0 to 10 square feet might result in huge sales and at some point, increasing shelf space will not increase the sales by much.

To confirm this, an A/B test could be performed. Seeing whether different shelf space allotted for the same item results in what corresponding change. There are two issues I foresee with this. First, we would have to control for the variable of different stores. That is, can we perform a test so that the variance in change in sales can be 90 to 100 percent predicted by the shelf space. This seems like a difficult thing to do, and also it is difficult to confirm we controlled for other variables successfully.

We could also increase shelf space gradually and track the change in sales. The disadvantage of this is that we would have to take space away from other products as we increase shelf space for the product we are testing. It also may take some time to see the results if we perform this approach. Seasonality and trend may also be a concern as Joel mentioned. And the solutions to these concerns were not feasible.

Ultimately, a modified version of the first approach could be used to determine the impact on sales from shelf space. We could account for the other effects that was of concern in the first approach. The issue still remained however that proving causation – increasing shelf space results in an increase in the sales – was still elusive.