A journey into Supermarket macrospace

By Laura Varsandan

City University

INM430 – Principles of Data Science

# Introduction

## 1.1 Analysis Domain

The supermarket industry has an abundance of data which could be used to provide guidance towards decision-making. One such key decision relates to ‘macro-space’: how much space should be allocated to each section in each store. Examples of sections are could be Grocery, Produce, Cleaning etc. The ‘micro-space’ then deals with how much space should be allocated to which products within a section. The two different decisions need different approaches. The emphasis of ‘macro-space’ is comparing how sections perform relative to each other, whereas the ‘micro-space focuses on how products within a section perform relative to each other. I came across these terms and concepts during my professional experience at Sainsbury’s, a UK supermarket store, but the concept is used at an industry level (Retail Acumen, 2010).

The key question for this analysis is therefore which sections a supermarket chain needs to increase and which sections it needs to decrease to achieve its sales maximization objective.

## 1.2 Data Sources

In order to develop a solution for the ‘macro-space’ challenge, I am going to use data from Corporacion Favorita, an Ecuador supermarket chain (Kaggle, 2017). They have provided the following data for a Kaggle competition:

* Daily volume of sales by product and store since January 2013 to August 2017
* Daily volume of transactions by store
* Daily prices of oil
* Holiday events
* Store metadata
* Product Metadata.

## 1.3 Key Assumptions

The key limitation of this dataset is that it does not contain how much space each product or section occupies in each store. Therefore, I have decided to use how many unique products were ranged in each section-store combination as a proxy for the size of the section. This important aspect will be further discussed in section [ ].

Since the data does not include any profit information, but only unit sales volumes, I will assume that the chain wants to maximize unit sales. This is not an uncommon objective as a lot of supermarket chains pay close attention to market share, particularly in the UK (BBC, 2017).

Since it can be operationally expensive to change store layout, I will assume that this decision is only made once, during the current month for the following month. Since this may be a centralised or a decentralised decision, I will design the analysis in such a way that the results will be usable at all levels, by providing results by store and section, which can then be aggregated up.

The results will not include performance forecasts of new introductions of products in certain stores. The question of how well a new product is going to perform in a store is a separate analytical question which is outside the scope of this analysis.

## Main Objective

The main objective of this analysis is to provide a tool for decision makers that suggests how many products they should allocate to a particular section (in our case ‘family’) in a particular store for the following month. Since July 2017 is the latest month with complete data, that will be used as our target month for the suggestions.

## Analysis Strategy

In order to achieve our objective, the first step will be to predict sales per unit for July 2017 at a day-store-class level, which will then be summed to sales per unit at a monthly level for each store-class. Predicting sales per unit at the intermediary level between ‘family’(i.e. section) and product has the benefit of taking into account that different classes within sections might perform really good or really bad. Additionally, it takes the focus away from product-level, which is a ‘micro-space’ decision and should not be dealt with in this analysis. Predicting sales at a day level allows us to consider that certain classes might perform well on certain days and it also adds more points of analysis to train our model compared to doing a forecast at a monthly level. Thirdly, by forecasting at a store level, we can firstly capture better the way classes behave in particular stores.

Providing an output at a store level allows decentralised decision making, where store managers can increase their sales by adding more products to the high performing sections in their store.

A base algorithm will be created and several more advanced algorithms for predicting continuous variables will be compared to it.

The second step will be to use the monthly predictions for July at store-class level in an optimization algorithm that will identify how many unique products should be ranged in each store-class combination to maximize sales for that store.

The final step will be to add the results of the optimization into an interactive visualisation tool designed for the decision-makers in the supermarket chain.