Applying the new digital frontier. Optimisation of sales promotions using heuristic techniques and machine learning.

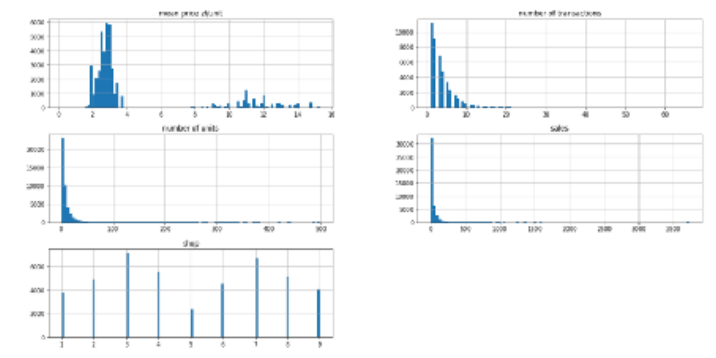
Krzysztof Szlaski, University of Warsaw alumnus, Big Data engineer at P&G

At the verge of the 4th industrial revolution companies must adapt to a new digital environment that is mostly dependent on data and analytics. Companies protect their data more than ever as data are considered as the most valuable resource. This condition is due to the fact of creation of new data analytics tools which allow gaining meaningful business knowledge from data increase profit by several or over a dozen per cent. As a "tool" we can consider models and algorithms which allow extraction of knowledge from data and also – advancement in computation powers of processing units which together allowed this progress. This type of innovation is essential to empowering companies to gain a market advantage in a resource-constrained world.

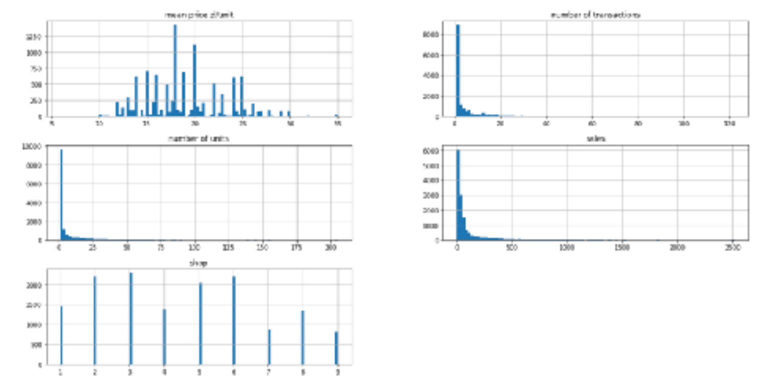
Due to that, the **primary goal of this paper is to answer the following research question: Does the heuristic methods and out of the box machine learning can find empirical patterns and valuable business insight in the small dataset containing local sales data?** Part of this approach is to make analyses that prove that the data provided can give some insight into customers and their behaviour. Therefore, no additional data or attributes will be included in the paper, although, according to the author, they can significantly improve the models it will not be part of this research. To accomplish that, we will be using custom heuristic techniques and machine learning methods that can allow retail analysts or store account managers to make data-driven pricing planning choices without accessing multi-feature big data. The problem we will try to solve is how to optimise price promotions with a relatively small set of data. Finally, we will also show how important it is to choose the right machine learning tools to perform the analysis.

The dataset is provided by the Centrum Monitorowania Rynku Sp. z o.o. for the purposes of this paper. The company is an independent research facility as well as a professional data provider for companies. The provided datasets are actual **retail data from stores across Poland collected from December 31, 2018. until February 29, 2020.** The dataset does not cover a later period to avoid any disruptions during COVID-19. The institute focuses on collecting retail consumers data, and our two datasets that we will analyse are retail food data for coffee and beer. The method of obtaining this data is not known to the authors of this article, however, we assume that this data is complete and correct. The first dataset contains daily sales data for 12 different (Stock Keeping Unit) SKUs that were sold in 9 different store types. We also have information of total coffee sales and total sales for each store. The second dataset contains daily sales data for 20 different SKUs of beer that were sold in 9 different stores. Also includes total beer sales and overall total sales for each store. In total, we have **18,043 observations in the coffee dataset and 50,922 observations in the beer dataset**. The observations do not evenly represent the stores. The sales data for both datasets has the following characteristics: Shop number – describes in which store the sale was made. Date of sales – the date on which a specific product was sold in a specific store, SKU/category – specifies the product name and EAN number, Value of sales – the sum of the daily sales of a specific product in each store. Number of the units sold – a unit sold each day in a specific store. Number of transactions – in how many transactions were total sales, Average price for unit.

For each dataset we have following information about the stores: Index number of the store. Area of the store – this metric is show in square meters. The stores 1-3 are large-area stores and 4-9 are small area stores, Voivodeship, Type of the town – this informs us on the segment of population of the town in which the store is located. We have 5 types of towns: city above 500 000 population, city between 100 000 and 500 000 population, city between 50 000 and 100 000 population, below 50 000 population and countryside.



Plot 1 Summary distribution plot for the complete beer dataset. Source: own analysis.



Plot 2 Summary distribution plot for the complete coffee dataset. Source: own analysis.

The data was pre-processed for the search od missing data and outliers. For the purpose of the analysis most of the columns were dropped e.g.

For our tool to be applicable, some assumption needs to be met:
- Retailers can control price independently – manufactures cannot impose price levels.
- The retailer can fulfil the demand for the product.
- The future demand for the product is based on past prices.

Predicting the optimal period of the year in which the price promotion affects the sales value

1. We take each month of our analysis and check which day of the month it is. We set a separate column in our dataset with those days.
2. Then we use one-hot encoding to split our column in multiple features.
3. We create an algorithm which will mark our days as promotional. The algorithms are using a moving average of the value of sales. The default way we are doing it is per store average, but to make more market-based conclusions in some situations, we might calculate a rolling average based on all stores. Moving average counts ten days before the day which we study at that point. Furthermore, the average is computed using weights that are the sales values for each store to balance the metric. The weight is proportional to the sum amount of sales in one store for a specific product compared to the sales of the same product in the other stores. This procedure we call the promotional moving average, and we will be using it in each of the heuristic's methods shown below.
4. At this point, we have in our dataset all required information. Each day of the month is flagged, whether it is promotional or not, and we also have the value of sales for each day (observation).
5. We create the list where each day of the month is the particular day of the promotion.
6. Now we use linear regression to check the impact of each day of the month and the fact if the day was significant or not to the value of sales. We will have one more feature: is_promotion, which only determines whether each day is a promotion day or not.

The second approach is much simpler. We proceed as follows:

1. We aggregate the sales of the first ten days of each month than the second ten days and third nine to eleven days.
2. Then we aggregate the sales by this day, and we calculate average sales per day for each period.
3. Finally, we also calculate averages for each of the aggregated periods to find out what the sales trend is for the entire dataset for each store.

Suggesting how many days the promotion should cover to maximize the sales volume

1. We create a list of promotion campaigns using promotional rolling average to flag the appropriate days and then add them to the lists. We create lists of campaigns by checking if the next day is also promotional. The first day which is not promotional brakes the chain and the next promotional day starts a new promotion campaign.
2. We count how many promotion days we have each campaign and create a separate column for that.
3. Then we count a weighted average of sales per day for campaign per store.
4. Because we do not want to show promotions which are too short, as it might influence the scores, we also reject promotion campaigns which are shorter than three days.
5. Then we count an average of days for the five best and five worst results for each store.

Suggesting what the highest prices should be in the promotion to make it useful

1. We use our promotional rolling average which is weighted by the number of sales the store has done in proportion to the all other stores during whole observed period to define if each observation(day) is promotion. We can call it *average price or baseline.*
2. Then we create a metric (1)

$$\frac{Sales\ Value}{Average\ price - promotional\ price} \tag{1}$$

3. The *promotional price* is the price each day which we marked as a promotion and *Sales Value* is the sales which are the sum of sales of all units of product each day for each store.
4. What is important to mention: we want to maximize the whole metric.
5. Based on this metric, we create a list ranked in descending order for each store.

Identification of decision-making rules in the case of parallel application of promotion

We use our promotional rolling average to define when we have a promotion day (observation) or not. To calculate the base price, we choose stores from each segment (large-area, small-area). Doing that we can compare the price in each store by using graphs. Actual price from each store as one graph and baseline price, which we calculated using all stores from the segment. We calculate when the price is below the baseline price, and we mark it as the promotional price.

We can picture that on the graph to show in which period of the year situation described

above occurs.

Identification of dependencies related to the maximization of sales during the promotion

1. First, we mark our observations (days) as promotions with use our promotional moving average.
2. Then we make multiple lists where we count promotion campaigns days. The algorithm checks if the day (observation) for each store was a promotional day. If so, we add it to the list, if not, it searches for the next first day of the promotion and creates a new list after finding it.
3. Each of the consecutive days gets indexes of each day of the promotion starting from 1.
4. Then we take each of the consecutive days of the promotion and count the average for that, for example, we take all first days from all lists, and we count average then we count each average for each of the following days.
5. Then we sort the list by the descending order.

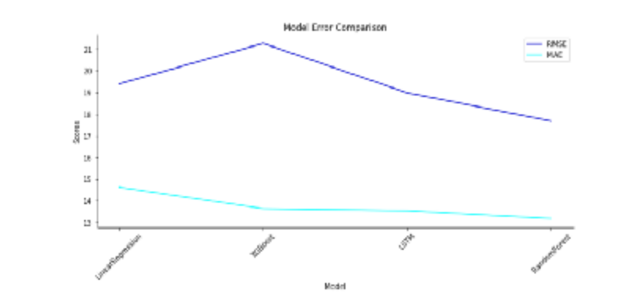
- The tools we will use are: Linear Regression,Random Forrest, XGBoost, LSTM
- All those tools were parametrized.

Data preparation for our analysis consists of a few steps:

1. First of all, based on the correlation and partial correlation we decide on adding lags to the data set. Based on this, we decide to choose our look-back period to be seven days, which means that sales correlate every week with the highest sales during the weekend.
2. Then we split our dataset into training and test. We include twelve months in the training data and two months in the test data.
3. We also create a function to get back with the transformation to show it in the graph.
4. Next, we declare the MEA and RMSE scores to be calculated based on the predicted values.

model	MAE	RMSE
RandomForest	13,19	17,69
LSTM	13,55	18,98
XGBoost	13,65	21,27
LinearReggression	14,64	19,40

Table 11 Model error comparison – summary results, label 'Tyskie Gronie Alk 5.5 light beer, 500ml bottle grb, ean 5901359000367' and store number 7. Source: own analysis



Plot 7 Model error comparison, label: 'Tyskie Gronie Alk 5.5 light beer, 500ml bottle grb, ean 5901359000367' and store number 7. Source: own analysis.

Plot 9 LSTM sales forecasting prediction for the label: 'Tyskie Gronie Alk 5.5 light beer, 500ml bottle grb, ean 5901359000367' and store number 7. Source: own analysis.

We managed to answer our question stated in the introduction to the article positively. Using heuristic methods based on simple statistics and use of out of the box machine learning methods might bring valuable business insight and we also were able to get information about empirical patterns within the analysed data.

Analysing the data using limited or small dataset is a challenging task as the knowledge which we get directly from the data needs to be completed by the subject matter expertise. This is, of course, the case in any data related matter. That might be an obvious truth, but as a data scientist, even the best machine learning or econometrics methods will not be sufficient to make a good analysis if we ignore the context of the data. The food product sales data always needs to be analysed in the context of these products. **In this paper, we showed that taking an insight into the local sales data on the small sample with some addition of knowledge of the fast-selling consumer goods might give as practical business intelligence and we might use this to improve the profit of the companies.**

In this paper, **we spot the difference between the categories of products and the different behaviour of the customers regarding these products. The insights mentioned above were about the length of the promotions and best-selling day and the maximization of the price which all is part of the question: "When, for how long and for how much?" each retailer planning promotion is asking.**

We also concluded that beer sales are much more seasonal than coffee sales. The two primary data stamps drive sales of beer: weekends and the summer period from June to September. Also, coffee sales are much more dependent on the sales on the promotions. The tools used in this paper are showing how easily we can verify some sales hypothesis about the behaviours of the consumers.

All the fundamental questions about the promotion analytics can be answered with the use of the heuristic data mining technics. As retailers collect more data each year, it becomes possible to make our algorithms more robust and accurate. It would also be suitable both for researchers and society if the large global retail chains would be sharing some parts of the data which they collect. The wars for data[17] between those companies cause increasing closure of access to open data. The possibility of using it for something else than for gaining an even more significant competitive advantage in an already very monopolized market decreased in recent years. As a last remark, we can say that, to overcome this situation a smaller local retailer should be using opportunities created for them by technology and collect their data not to get left behind on the new digital frontier.