Journal of Universal Computer Science, vol. 29, no. 11 (2023), 1385-1403
submitted: 19/5/2023, accepted: 30/9/2023, appeared: 28/11/2023

Retail Indicators Forecasting and Planning

Nelson Baloian
(Computer Science Department, Universidad de Chile, Santiago, Chile
https://orcid.org/0000-0003-1608-6454, nbaloian@dcc.uchile.cl)

Jonathan Frez
(Computer Science Department, Universidad Diego Portales, Santiago, Chile
https://orcid.org/0009-0003-4812-4483, jonathan.frez@mail.udp.cl)

José A. Pino
(Computer Science Department, Universidad de Chile, Santiago, Chile
https://orcid.org/0000-0002-5835-988X, jpino@dcc.uchile.cl)

Cristóbal Fuenzalida
(Computer Science Department, Universidad Diego Portales, Santiago, Chile
https://orcid.org/0000-0000-0000-0000, c.fuenza6597@hotmail.com)

Sergio Peñafiel
(Fundación Arturo Lopez Perez (FALP), Santiago, Chile
https://orcid.org/0000-0002-0025-7805, sergio.penafiel@falp.org)

Belisario Panay
(Allm Inc., Tokyo, Japan
https://orcid.org/0000-0002-1440-8192, belisariops@gmail.com)

Gustavo Zurita
(Management Control & Information Systems, Universidad de Chile, Santiago, Chile
https://orcid.org/0000-0003-0757-1247, gzurita@fen.uchile.cl)

Horacio Sanson
(Allm Inc., Tokyo, Japan
https://orcid.org/0000-0002-1440-8192, horacio@allm.inc)

Abstract: We present a methodology to handle the problem of planning sales goals. The methodology supports the retail manager to carry out simulations to find the most plausible goals for the future. One of the novel aspects of this methodology is that the analysis is based not on current sales levels, as most previous works do, but on those in the future, making a more precise and accurate analysis of the situation. The work presents the solution for a scenario using three sales performance indicators: foot traffic, conversion rate and ticket mean value for sales, but it explains how it can be generalized to more indicators. The contribution of this work is in the first place a framework, which consists of a methodology for performing sales planning, then, an algorithm, which finds the best prediction model for a particular store, and finally, a tool, which helps sales planners to set realistic sales goals based on the predicted sales. First we present the method to choose the best indicator prediction model for each retail store and then we present a tool which allows the retail manager estimate the improvements on the indicators in order to attain a desired sales goal level; the managers may then perform several simulations for various scenarios in a fast and efficient way. The developed tool implementing this methodology was validated by experts in the subject of administration of retail stores yielding good results.

Keywords: Sales and Operation Planning, Sales Goal Planning, Machine Learning
Categories: H.3.1, H.3.2, H.3.3, H.3.7, H.5.1
DOI: 10.3897/jucs.112556

1 Introduction

A key activity in retail store management is to make a good guess of store sales in the coming days, weeks, months or even years. This is commonly known as Sales and Operation Planning (S&OP) [Thomé, 2012] and is one of the most important tasks managers of retail stores must perform in order to increase store performance. Its results are then used to efficiently plan the provision of resources, such as adequate staffing and enough goods to be sold. According to [Kreuter, 2022] the research in S&OP has been booming in recent years. Sales goal planning is a closely related task and refers to the process of setting targets and objectives for a sales team or individual sales representatives. It involves determining the specific sales objectives, establishing measurable targets, and outlining strategies and actions to achieve those goals. Big data and machine learning based method are increasingly being used to support this task [Aversa, 2021], [Huber, 2020] [Tsoumakas, 2019].

A convenient planning scenario is one in which the manager assumes a target in terms of the expected number of sales. Then, she obtains the values of the store performance indicator allowing her to achieve this target from a certain prediction machine.

There are many typical indicators of store performance (they are reviewed in subsection 2.3). To determine sales and operations planning, managers often use prediction models for these indicators in order to plan their sales goals [Pavlyuchenko, 2021]. The manager evaluates the obtained values of these indicators and may consider them inconvenient or unfeasible. In such a case, she can start asking the machine again with a new sales target as input. The scenario, therefore, implies a simulation cycle, which ends when the administrator is satisfied with the obtained indicator values. Then she can make the decisions necessary to achieve those values. For example, she may set up a promotion and advertising campaign to hopefully achieve the indicator values. If these values actually occur later on and the prediction machine was correct, then the expected sales target would be satisfied.

In this article, we propose a framework, which consists of a methodology for performing sales planning, an algorithm, which finds the best prediction model for a particular store, and a tool, which helps sales planners to perform simulations like the described above, in order to set realistic sales goals based on the predicted sales numbers. In order to validate the framework, we present a real case considering three of the most common performance indicators for retail sales: foot traffic, conversion rate, and average purchase amount per ticket. Therefore, in the mentioned scenario, the output will be a combination of values for these three indicators. These indicators were chosen because they are easy to monitor, yet they convey valuable information.

The reported research is based on using large collections of data and therefore it satisfies one of the topics of this J.UCS special issue.

The rest of the paper is organized as follows: Section 2 presents related work. Section 3 describes our proposal for the Sales Planning Model. Section 4 includes the model implementation and results. Finally, Section 5 presents the conclusions.

2 Related Work

2.1 Sales and Operation Planning

Estimating the outcome of sales promotion has always been a goal for a store chain manager [Boulden, 1957]. Accurate sale estimations allow precise product stocks to be held as well as adequate personnel to be provided. Furthermore, clients being able to purchase the promoted items generate customer satisfaction. On the financial side, accurate sale estimations provide control on cash flow, meet total sales goals and thus obtain expected profits. In short, accurate sales estimation is a core activity for a retail company [Hastings, 1994].

There have been numerous studies on retail indicators forecasting and planning, which involve predicting and preparing for changes in retail sales and supply chain management. These studies have explored various methods including statistical models, judgmental approaches, hybrid methods, data mining techniques, artificial neural networks, support vector machines, random forests, demand forecasting, inventory management, and production planning.

In particular, we are working in an S&OP process, which is a business process used to effectively align organizational demand and supply [Kreuter, 2022]. It involves creating a single plan that integrates sales and marketing, production, and financial goals of an organization.

In S&OP, there is the concept called "demand-driven S&OP", which emphasizes the importance of considering customer demand when developing the S&OP plan [Cassivi, 2007]. Demand-driven S&OP seeks to align the organizational demand and supply by using customer demand as the driving force behind production and inventory plans. This approach is based on the idea that customer demand is the most important factor in determining organizational production and supply chain activities.

The most widely used approaches for demand-driven S&OP correspond to both qualitative and statistical analysis. In qualitative analyses [Garcia-Villareal, 2018], [Danese, 2018], [Hulthén, 2016], experts analyze the behavior of certain key events from which conclusions are obtained about the applied measures. In the statistical and optimization analysis [Nemati, 2013], [Lim, 2014], the goal is to measure the variation of key indicators that are observed when certain measures are applied. This is done through the analysis of historical data or through simulations seeking to replicate the context of the process. It is reported whether the applied measures have a positive impact on the indicators and whether or not this impact is statistically significant to call the set of measures an improvement.

In order to support and improve the demand-driven operation, the use of advanced analytical techniques can be used, such as machine learning and artificial intelligence [Glackin, 2022]. These techniques can be used to analyze historical data and make accurate demand forecasts, which can help organizations better align their demand and supply. In addition, these techniques can be used to identify patterns and trends in demand and supply data, which can help organizations optimize their S&OP processes and make informed decisions.

2.2 Sales Goal Planning

Setting effective sales goals requires a structured approach taking into account the organization sales history, market trends, customer needs, and competitive landscape. Sales goal planning typically includes at least the following steps [Juran, 1992]:

– **Define Objectives:** Start by clearly defining the objectives you want to achieve. These objectives may include revenue targets, market share growth, customer acquisition, or sales volume increase.

– **Analyze Historical Data:** Review past sales performance data to understand trends, identify strengths, weaknesses, opportunities, and threats (SWOT analysis). This analysis supports the setting of realistically achievable goals.

– **Set Specific and Measurable Targets:** Establish specific, measurable, attainable, relevant, and time-bound (SMART) targets. For example, a goal could be to increase sales by 10 percent.

– **Break Down Goals:** Divide the overall sales goal into smaller milestones or benchmarks. This helps create a sense of progress and provides a clear road-map for achieving the final objective.

– **Identify Strategies and Tactics:** Determine the strategies and tactics required to reach the sales goals. This may involve targeting specific customer segments, launching promotional campaigns, improving sales processes, or enhancing product features.

– **Allocate Resources:** Allocate resources such as budget, personnel, training, and technology to support the sales team in achieving their goals. Ensure that the necessary tools and resources are available to maximize their effectiveness.

– **Establish Metrics and KPIs:** Define key performance indicators (KPIs) and metrics that will be used to track progress and measure success. Common sales metrics include revenue, conversion rates, average deal size, customer acquisition cost, and customer retention rate.

– **Monitor and Evaluate Progress:** Continuously monitor sales performance against the set goals and KPIs. Regularly review progress and identify any areas where adjustments or improvements are needed.

Sales goal planning is an iterative process that requires ongoing monitoring, evaluation, and adaptation to ensure continued success.

Several sales goal planning frameworks have been proposed to help organizations set and achieve their sales targets. One popular framework is the Objective and Key Results (OKR) framework [Al Thinyan, 2022]. The OKR framework involves setting ambitious but achievable objectives and defining measurable key results to track progress towards those objectives. Another popular framework is the Sales Pipeline Management Framework [Park, 2020], which involves identifying and tracking the various stages of the sales process, from lead generation to closing the deal. The Sales Pipeline Management Framework helps organizations to identify bottlenecks in the sales process and optimize their sales strategies to improve conversion rates and sales speed.

2.3 Retail Sales Performance Indicators

When it comes to measuring and evaluating the performance of a retail sales operation, there are several key performance indicators (KPIs) that can provide valuable insights [Anand, 2015]. Some of them are the following:

- **Sales Revenue:** The total amount of revenue generated from retail sales within a specific period, such as a day, week, month, or quarter.
- **Sales Growth:** The percentage increase or decrease in sales revenue compared to a previous period. It indicates the overall growth rate of retail sales.
- **Average Transaction Value (ATV):** The average amount of money spent by customers in a single transaction. Calculated by dividing the total sales revenue by the number of transactions.
- **Conversion Rate:** The percentage of potential customers who make a purchase. It is calculated by dividing the number of transactions by the number of visitors or footfall in a store.
- **Customer Traffic:** Also known as foot traffic, it refers to the total number of visitors or footfall in a retail store during a given period. It helps assess the store popularity and customer attraction.
- **Customer Acquisition Cost (CAC):** The average cost associated with acquiring a new customer. It includes marketing and advertising expenses, sales commissions, and other related costs.
- **Gross Margin:** The difference between the total sales revenue and the cost of goods sold (COGS). It represents the profitability of the products being sold.
- **Inventory Turnover:** The number of times inventory is sold or replaced during a specific period. It indicates how efficiently a retailer is managing inventory and can help identify slow-moving or obsolete items.
- **Return on Investment (ROI):** The ratio of the net profit generated from retail sales to the total investment made. It helps evaluate the profitability and efficiency of the retail operation.
- **Customer Satisfaction:** Measuring customer satisfaction through surveys, feedback, or ratings can provide insights into the overall customer experience and loyalty. Satisfied customers are more likely to repeat purchases and recommend the store to others.

These are just a few examples of the many performance indicators that can be utilized in retail sales. The specific chosen KPIs will depend on the nature of the retail business, its goals, and the available data. It is important to regularly track and analyze these indicators to identify areas of improvement and make informed decisions to optimize sales performance.

2.4 Sales Goal Planning and Machine Learning

In the literature, we can find success stories in which the use of machine learning has helped improve retail indicators. For example, one of the simplest but widely used methods is ARIMA, which is a model based on moving averages which is capable of interpolating and extrapolating the behavior of a time series. Lam et al. [Lam, 2020] use ARIMA to forecast store foot-traffic; with this information, they propose a strategy to find the optimal retail sales force.

McIntyre et al. [McIntyre, 1993] propose an early use of Artificial Intelligence to the problem of estimating the outcome of promotional sales. They mention ten factors explaining why making estimation based just on human experience is a very difficult and error-prone process. Therefore, they propose case-based reasoning as a computer-supported tool intended to incorporate the human expertise within the firm. It is an example of using an Organizational Memory [Guerrero, 2001]. They choose ten factors to be included in the system, such as season of the promotion, number of television spots and percentage of price discounts. Guo et al. [Guo, 2013] developed a multivariate intelligent decision making model. The corresponding system is composed of three modules: a data preparation and preprocessing module, a harmony search-wrapper-based variable selection module, and a multivariate intelligent forecaster module. The forecasting precision of the model was tested as better than other proposals at that time (2013). Machine learning appears as an interesting approach to forecast sales. For example, the manager does not need to select the factors to be included in the prediction system, as the case-based reasoning tool mentioned above [McIntyre, 1993] required. This occurs because the machine learning system chooses the best factors. Bendle et al. [Bendle, 2021] have reviewed the issues and perspectives of machine learning in the retailing area.

Another example is the one presented by Panay et al. [Panay, 2021] in which a model is presented that can predict the behavior of foot traffic, conversion rate and sales amount in a range of time in the future. The method uses the historical data of these indicators to train a custom regression model. The model achieves high accuracy in all variables, with an average RMSE of 0.07. In addition, the model is capable of indicating the most relevant variables at the time of making the forecast. However, although this model allows obtaining precise results on the behavior of sales, it does not provide optimal strategy planning to maximize sales.

Conflicting variables involved in promotional planning imply visualizing these effects. Stewart and Gallen [Stewart, 1998] proposed a matrix to visualize such conflicts in budget allocations.

2.5 Proposed Approach and its Relation to Previous Work

We propose a framework consisting of a model and a concrete tool based on it to support the first steps of a sales goal planning process. Therefore, we use the general context described in subsection 2.1. The basic steps for the process are assumed to be the ones listed in subsection 2.2. The KPIs (subsection 2.3) we choose are foot traffic, conversion rate, and average transaction value, which are well-known indicators. For the prediction of the indicators, we use the model by Panay et al. [Panay, 2021] (subsection 2.4). Nevertheless, we extend the applicability of this model, by providing automatic strategy planning to maximize sales. Finally, planning results must be presented in a suitable way; we inspired in Stewart and Gallen [Stewart, 1998] work (subsection 2.4) to offer visual presentation of the results.

3 The Sales Goal Planning Model

Our proposal for supporting the sales goal-planning process is a tool, which sales managers can use to simulate various scenarios. With this tool, they can receive suggestions about the necessary changes that they have to achieve in the values of the three sales performance indicators chosen for this work (foot traffic, conversion rate, and average transaction value) in order to attain a particular total sales amount. In this way, they can explore various scenarios for deciding which are the most convenient actions to take for achieving the suggested values. For this purpose, managers can organize increasing advertising and/or sales promotion campaigns. Unlike most procedures reported in the literature for sales goal planning, this proposal does not use current values for the three KPI's as starting values; instead, it uses values obtained by a prediction model for the period being simulated. This strategy avoids the effects of stationary variations and/or increasing or decreasing trends. The next subsection describes the prediction model and the following one, the way how they are used in order to explore recommendations for the managers to modify the values of foot traffic, conversion rate and average transaction value to achieve the desired sales goal.

3.1 Prediction and Model Selection Process

In a previous work [Panay, 2021], we realized that a predictive model could perform well for certain stores in particular, which was complemented with hyper-parameterization of the model and search for the most appropriate hyperparameters for each store. However, through experimentation, we discovered that some models perform better than others with the correct hyperparameters. Therefore, in this work we designed an algorithm that explores the use of various models with a constrained set of hyperparameters, choosing the one that provides the best results and then improving the selection of the hyperparameters of the chosen model.

The selection process tests the following estimator algorithms:

- **Gradient Boosting Decision Trees (GBDT)** is an algorithm that combines multiple decision tree models to create a more powerful model for prediction. It works by training a decision tree model on the data and then iteratively improving the model by adding new decision trees that focus on the mistakes made by the previous trees. Each new tree is trained using the residual errors from the previous tree, which helps to correct the mistakes and improve the overall accuracy of the model.

Support Vector Regression (SVR) is an algorithm used to predict continuous numerical values. It works by finding the optimal hyperplane in a higher-dimensional space that maximally separates the data into two classes. The hyperplane is chosen such that the distance between it and the closest data points is maximized, which is known as the margin. SVR then uses this hyperplane to make predictions for new data points. SVR can handle both linear and nonlinear relationships between the predictors and the target variable.

XGBoost Regression (XGBR) is an algorithm that is used to predict continuous numerical values. It is a type of Gradient Boosting algorithm which is specifically designed for regression problems. XGBR works by training a series of decision tree models and using the predictions of these trees to improve the overall accuracy of the model. It is known for its ability to handle large datasets, high-dimensional data, and missing values.

Random Forest is an algorithm that is used to predict both continuous numerical values (regression) and discrete categories (classification). It works by training multiple decision tree models on random subsets of the data and then aggregating the predictions of these trees to make a final prediction. Each tree is trained using a different subset of the data and a different subset of the features, which helps to reduce overfitting and improve the overall accuracy of the model. Random Forest can handle both linear and nonlinear relationships between the predictors and the target variable.

Weighted Evidence Regression Model (WEVREG) is an interpretable regression method where the prediction of a new observation can be easily followed [Panay, 2020]. This model is based on a previous one that was proposed by Petit-Renaud and Denœux [Petit-Renaud, 2004] It uses the Dempster–Shafer Theory together with a variation of K-Nearest Neighbors (KNN) to produce an evidence-based regression model. WEVREG extends this model by adding weights to each dimension of the attributes, making certain parameters have more importance than others when deciding the most similar instances of the KNN process. In addition, this change allows measuring the importance of each attribute in the prediction.

Calibrating the hyperparameters of an estimator algorithm is necessary to achieve the best possible fit to the data. However, because each store may have unique characteristics, it is not feasible to use the same estimator with the same hyperparameters for all stores. Instead, we must calibrate the hyperparameters for each store separately to ensure optimal performance.

Finding the optimal hyperparameters for an estimator algorithm and a store can be a time-consuming task, as it involves testing all possible combinations of these parameters. To reduce the amount of time required, we can limit the search by defining ranges and the number of values to test for each parameter. For example, if we want to find the best value for a parameter i, we might define a range of [0-1000] and test 10 values within that range (0, 100, 200, …, 1000). However, most estimators have more than one parameter, so we might also need to define ranges and values for additional parameters, such as j. This process of testing all possible combinations of multiple parameters is known as Exhaustive GridSearch.

We can follow these steps to make the GridSearch process efficient:

- Conduct GridSearch for each estimator and store, using the same ranges and values for all stores. Save the selected points from each grid.

- Count the number of times each point is selected across all stores and estimators. Select the K most common points, creating a new grid with these K combinations of hyperparameters. This process will result in a set of 5 grids, each containing the K most common hyperparameter combinations that performed best during the estimator training process.

The goal at this stage is to choose the estimator that will give us the lowest margin of error. We can achieve this goal by training and testing a model for each estimator using the most common hyperparameter from the corresponding MaxGrid. Alternatively, we can also use the N most common hyperparameters from each MaxGrid to increase our chances of selecting the best one.

After we have the margin of error for each estimator, we select the one that performs the best and evaluate the full MaxGrid with it. Afterwards, we can choose the optimal set of hyperparameters and use it to train our final model with the selected estimator.

The outcome of this training and algorithm selection process enables us to reach a significantly higher number of stores on the platform, with much more accurate predictions than those generated manually by analysts and resource planners. This results in the ability to utilize these insights for the development of effective commercial strategies and resource planning. The main steps of the algorithm are described in Algorithm 1.

Algorithm 1 Hyperparameter Calibration for Store-wise Estimator Algorithm (details)

Input: estimators, stores, ranges, values, K, N **Output:** Optimal hyperparameters and estimator

Initialize empty list *MaxGrids*
for each estimator *e* and store *s* **do**
    Conduct GridSearch with ranges and values
    Save selected points from grid in list *selected*
    for each point *p* in *selected* **do**
        Increment count of *p* in a dictionary *counts*
    end for
end for
for each dictionary *counts* **do**
    Select the K most common points, add to a new grid
    Append the new grid to the list *MaxGrids*
end for
for each estimator *e* and grid *g* in *MaxGrids* **do**
    Train and test a model with the most common hyperparameter from *g*
    Record the margin of error
end for
Choose the estimator with the lowest margin of error
Evaluate the full MaxGrid with the chosen estimator
Choose the optimal set of hyperparameters and train the final model with the chosen estimator

---

3.2 Sales Goal Setting Model

Once the model is working and giving results with a desirable precision, the next step is to bring it together with available historical data from stores. This procedure will allow us to build a "sales-goal planning tool" on the user-end, which allows managers to see both the predictions for their store indicators, and the increment of those values needed to achieve their sales goals.

To carry out the aforementioned purpose, we created a model that takes the predicted indicators as an input and gives the associated variations needed as outputs in function of the coefficients of variation of indicators obtained from the historical data; the latter being representative of how easily the indicators can vary. To simplify their use, we work with their reciprocals, which we will call weights for the rest of the explanation, denoted as WX. The model is defined as follows:

- We define the amount of sales *S* as the product between the number of enters (foot traffic) *E*, the conversion rate *C* and the average purchase amount per ticket *T*.

$$S = E \times C \times T \tag{1}$$

- The predicted indicators (inputs) will have the subscript *p*, and the indicator goals (outputs) will have the subscript *g* (note that S_g is the sales goal selected by the user and is therefore a constant). Thus we have:

$$S_p = E_p \times C_p \times T_p \tag{2}$$
$$S_g = E_g \times C_g \times T_g \tag{3}$$

- Since the percentage change of these indicators are the focus of the algorithm, we use the following definitions to proceed:

$$V_S = \frac{S_g}{S_p} - 1 \quad V_E = \frac{E_g}{E_p} - 1 \quad V_C = \frac{C_g}{C_p} - 1 \quad V_T = \frac{T_g}{T_p} - 1 \tag{4}$$

- Replacing those definitions in (3), we get:

$$S_p(1 + V_S) = E_p(1 + V_E) \times C_p(1 + V_C) \times T_p(1 + V_T)$$
$$\Rightarrow [E_p \times C_p \times T_p](1 + V_S) = [E_p \times C_p \times T_p](1 + V_E)(1 + V_C)(1 + V_T)$$
$$\Rightarrow (1 + V_S) = (1 + V_E)(1 + V_C)(1 + V_T) \tag{5}$$

- As mentioned before, the weights of indicators represent the reciprocal quantities of how easily they can vary, which means variations and weights must be inversely proportional to each other:

$$V_E \cdot W_E = V_C \cdot W_C = V_T \cdot W_T = x \qquad \text{(proportion ratio)} \tag{6}$$

- Then, without loss of generality, we will start to approach the problem focusing on one of the variables, because of the symmetry of the product that defines the equation. We will use the equation (5) and will leave it as a function of V_E; taking V_S and the weights as constants (because their values are fixed for the problem).

$$\Rightarrow (1+V_S) = (1+V_E)\left(1+V_E * \frac{W_E}{W_C}\right)\left(1+V_E * \frac{W_E}{W_T}\right) \tag{7}$$

- When rearranging the equation and amplifying it by $(W_E \cdot W_C \cdot W_T)$, we get the following third degree polynomial equation:

$$\begin{gathered}(V_E \cdot W_E)^3 + (W_E + W_C + W_T)(V_E \cdot W_E)^2 + \\ \left(\frac{1}{W_E} + \frac{1}{W_C} + \frac{1}{W_T}\right)(W_E \cdot W_C \cdot W_T)(W_E \cdot W_E) + \\ (W_E \cdot W_C \cdot W_T) \cdot (-W_S) = 0\end{gathered} \tag{8}$$

- To synthesize the notation, we define the following constants in function of the weights:

$$c_1 = W_E + W_C + W_T$$

$$c_2 = \frac{1}{W_E} + \frac{1}{W_C} + \frac{1}{W_T}$$

$$c_3 = W_E \cdot W_C \cdot W_T$$

$$c_4 = -V_S$$

- Then, we replace those values in equation (8) and use the proportion $x = V_E \cdot W_E$ from (6), obtaining the following cubic polynomial equation:

$$x^3 + c_1 \cdot x^2 + c_2\, c_3 \cdot x + c_3\, c_4 = 0 \tag{9}$$

- The previous equation no longer depends specifically on the indicator and weight associated with foot traffic (*E*), but now depends entirely on the proportion met by the three indicators and the constants c_k. It can be easily seen that the problem is generalized for the three indicators in the same way, since the original equation is symmetric. Therefore, the whole problem boils down to solving the cubic equation as a function of x, for which the general formula of Gerolamo Cardano is used.

Finally, having the proportion ratio x, we can figure the variations using (6) and from there the indicator goals needed for the desired sales goal using the definitions from (4), concluding the algorithm.

Note that similarly, if the user had chosen to fix one or two of the indicators as constants, the problem would turn into second and first order polynomial equations respectively, being easier to solve in the same way.

3.3 Multi Variable Optimization

The mathematical model provides a framework to determine the necessary adjustments required to reach a sales target given these three variables in our system. However, there are other variables outside the system that can be used in combination with the predicted Foot traffic, Conversion rate and Sales. For example, Staffing, Inventory and Customer Satisfaction rates. Assuming that equation (1) can be expanded to include more variables, we can incorporate them into the optimization process to achieve the sales goal.

For example, in (10), V is the average value of other relevant variable(s)

$$S = E \times C \times T \times V \tag{10}$$

In this particular scenario (10), we introduce a single variable, which is directly proportional to sales. An example of such a variable could be staffing. However, in order to allow for the flexibility of modifying the sales modeling function with more complex options that incorporate additional variables and alternative forms of relationships, an objective function will be employed. This objective function aims to determine the optimal values, regardless of the specific form of the sales function being used. The objective function measures the deviation between the actual sales (S) and the sales goal (G). This function can be defined as the squared difference between S and G (11), and with this function the optimization process can find the optimal values for the variables E, C, T, and V in order to achieve the desired sales goal.

$$\text{diff} = (S - G)^2 \tag{11}$$

We employed Bayesian optimization to obtain the optimized values for each variable. This optimizer systematically evaluates the objective function by sampling various points in the search space. Its aim is to minimize the squared difference between the actual sales and the sales goal. By exploring various combinations of E, C, T, and V values, it searches for the optimal solution.

For instance, consider having a foot traffic of 218, an average ticket value of $140,238, a conversion rate of 20.18%, resulting in $6,170,189 of sales. Our sales goal is set at $7,500,000. Additionally, let us introduce another variable called staffing, which has a value of 1.

Through Bayesian optimization, we can find the best combination of E, C, T, V, and staffing values that minimize the difference between the achieved sales and the desired sales goal. This approach allows us effectively explore the search space and identify the optimal configuration for maximizing sales performance.

Based on the optimization results in this example, it is suggested that the foot traffic should be increased to 221, the conversion rate should be improved to 21%, and the average ticket value should be set to $145,854. However, it is important to note that there is no need to hire additional staff based on the optimization outcome. By making these adjustments to the variables, it is expected that the sales goal can be achieved without the need for additional staff. This information provides valuable insights on how to optimize the sales process and drive revenue growth effectively.

Also, it is important to note that the equation (10) and the optimization function (11) in combination with a Bayesian optimization, allows us add as many variables as we need to the model in a simple and easy to understand way. A general overview of the method is described in Algorithm 2.

Algorithm 2 running the Bayesian optimization over the sales function

Input: Ep, Cp, Tp, and V_avg[]**, Output:** optimized values of S, E, C, T, and V[]

1: define the **sales_function**(Ep, Cp, Tp, V_avg[]) as a function returning an estimation of sales based on the predicted values of E, C, T and a varying set of average values.
2: define the **sales_goal_objective** that calculates the squared difference between the sales amount and the sales goal using the **sales_function**
3: Perform the Bayesian optimization using the **sales_goal_objective** function
4: **Calculate** the optimized sales amount So using the optimized values Eo, Co, To, Vo_avg[])
5: return **So, Eo, Co, To, Vo_avg[]**

4 Implementation and Results

The implementation of the sales-goal-planning tool has been met with success in both user acceptance and performance since its introduction for companies' monthly goal planning. It currently boasts a prediction accuracy of over 90%, however, we need to compare these results with the previous human performance accuracy.

To facilitate the user-end implementation of the planning tool, we developed a web-app which collects and displays all results. The app includes various fields for user input, such as desired sales goals and optional indicators. Users can also select the desired date range for predictions, with the default being two months in advance. The resulting goals and percentage change needed to achieve them are displayed prominently on the app.

As seen in the accompanying images, the flexibility of the tool is demonstrated through two examples. In the first, the user only inputs their desired sales goal (see Figure 1). In the second, they also set the foot traffic value as a constant for the algorithm (see Figure 2). It is worth noting that fixing the foot traffic near the predicted value results in larger variations for other indicators, making them harder to achieve.

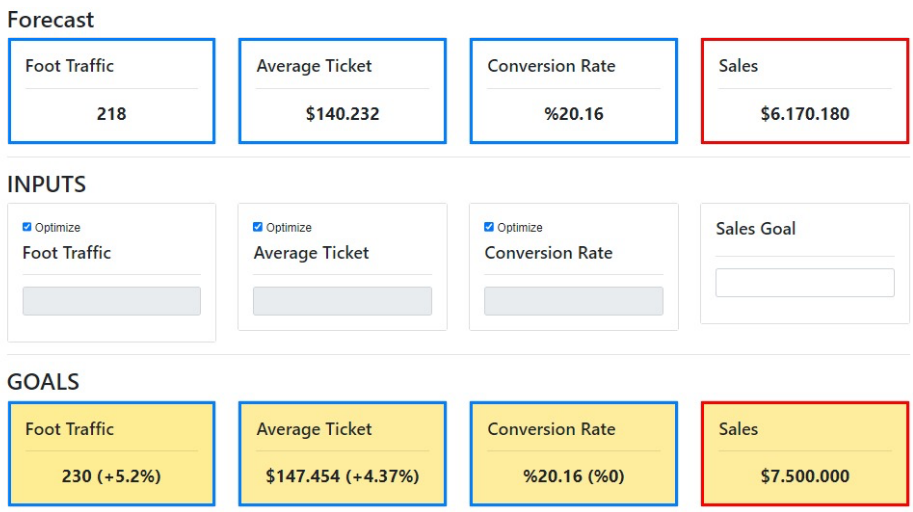

Figure 1: Sales-goal-planning tool results with only Sales Goal as input

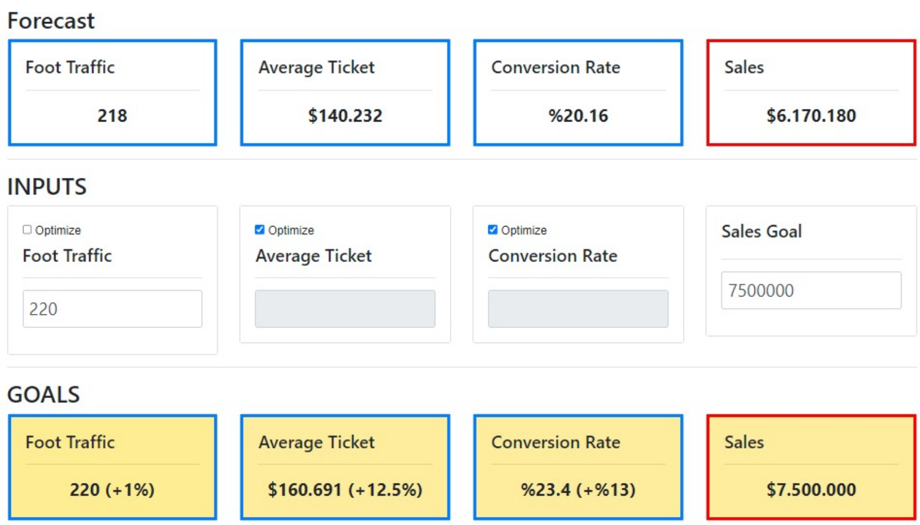

Figure 2: Sales-goal-planning tool results with Sales Goal as input

In order to assess the effectiveness of the tool, we compared its performance against the previously used estimations. These estimations were conducted by a team of experts employed by the holding company, tasked with analyzing the performance of each store across various countries. This intricate process involves collating information from store managers and integrating it with data available in the system. As a result, the outcome represents a collaborative endeavor of the analytics unit. This sum of efforts is referred to as "Human Performance" (HP). A set of 30 stores estimations were selected for this comparison. A team of two to three experts worked on the estimation for each store. They were the people responsible for this task up to now. This allowed us to gauge the accuracy of the tool and its efficiency compared to human performance. The results of a comparison between Human Performance labelled as "HP" and the tool labeled as "Algorithm", are shown in Table 1. The "average" row represents the average error value achieved by each method. We can see that the HP method achieved an average value of -7.73% and the Algorithm method achieved an average error value of -4.44%.

	HP	**Algorithm**		**HP**	**Algorithm**
store1	-14%	-23%	store16	-17%	-8%
store2	-12%	-20%	store17	-21%	-8%
store3	-19%	-20%	store18	-9%	-8%
store4	-26%	-19%	store19	-18%	-7%
store5	-13%	-17%	store20	5%	-5%
store6	-22%	-16%	store21	-20%	-4%
store7	-7%	-16%	store22	-1%	-3%
store8	-16%	-14%	store23	-3%	-2%
store9	-25%	-13%	store24	-9%	-1%
store10	-12%	-13%	store25	-9%	-1%
store11	-15%	-12%	store26	4%	2%
store12	-3%	-11%	store27	4%	2%
store13	-28%	-11%	store28	-12%	2%
store14	-11%	-10%	store29	0%	2%
store15	-8%	-8%	store30	5%	6%
			HP	**Algorithm**	
		average	-7.73%	-4.44%	
		st. deviation	0.516	0,507	
		min error	-1%	-1%	
		max error	-28%	-23%	

Table 1: Monthly sales estimation error for each store and statistics. Negative and Positive values refer to sub estimations or over estimations of foot traffic

The “min error” row represents the lowest error value achieved by each method, which was the same: -1%. The “max error” row represents the highest error value achieved by each method; the HP method achieved a maximum value of -28% and the Algorithm method achieved a maximum value of -23%.
The “st. deviation” row represents the standard deviation of the results for each method, which is a measure of the variability or dispersion of the data. It can be seen that the standard deviation for the HP method is 0.516 and for Algorithm method is 0.507.

In certain instances, it is relevant to highlight that the disparity between HP and the Algorithm is more pronounced than in others. This discrepancy can be attributed to the fact that predictions are not uniformly derived from the same dataset. The experts' additional insights about the environment, such as improvements in window displays or enhanced local management of the establishment, are not necessarily reflected in the data. For example, in the cases of stores 1, 2, 7 and 12 the experts' estimation (HP) yields clearly superior outcomes. Conversely, for stores 9,16, 17, 19, 21, 24, 25 and 28 the algorithm demonstrated a heightened level of precision. We attribute this latter contrast to an underestimation of the chain's marketing campaign, which can be observed through the consistent number of negative predictions across nearly all stores.

The results show the Algorithm method has similar results in terms of min, max, avg and st. deviation (slightly better for the Algorithm).

Moreover, as the reader may have already noticed, the times the algorithm outperforms HP notably are more prevalent—4 instances compared to 7. This underscores the algorithm's value as an option that excels at conducting a comprehensive analysis across a large number of stores using shared metrics when a broad assessment is required. However, as we pointed out before, the experts may have deep knowledge about few particular stores and therefore, their predictions for these stores may be very precise.

As the model was being developed, an implementation of it was created in the form of a web-app to enhance the approach used by multiple retailers and business intelligence companies in providing feedback to their clients, many of whom are among the largest retail companies in Chile and Colombia. After utilizing this implementation for over a year, we sought to gather their perspectives on its usefulness, the reasons for seeking out a new solution, and their future projections for its use.

The positive news is that they continue to utilize the tool, consistently incorporating dozens of additional stores into the system every month. They have found it to be a valuable asset in enhancing their ability to accurately predict sales and set goals for each store.

5 Conclusions

This work presents a methodology to support retail managers while performing sales and operations planning activities. The system consists of two parts: an algorithm for forecasting store performance indicators (foot traffic, conversion rate, and average transaction value) using evidential regression, and a method for planning sales goals based on these predictions. The algorithm is designed to provide a clear interpretation of the prediction process by identifying the most important features for the forecast. The method for planning sales goals uses the predictions and historical data on the variation of the indicators to find the optimal daily variations that are most likely to achieve the sales goal and associated risk according to the users (planners). Also, the quantitative analysis indicates that the Algorithm method has slightly better prediction performance compared to the experts' estimations. However, the results of Algorithm method are similarly consistent with the experts' results.

The described model and tool have been implemented in the real world. They are being actually used in stores chains in Chile and Colombia for over a year with positive results.

The reported research should be of interest to managers in charge of the Sales and Operation Planning of brick and mortar stores. As we mentioned, this planning is key for the performance of those stores. It may be even more relevant at present time in which the brick and mortar retail business has a strong challenge from online retail commerce.

Future research can be done in several areas. One of them is to try to make the Algorithm method results more consistent than they are now. Another area for research is to explore other performance indicators which may make easier and/or performing better the manager's task than it is now.

References

[Al Thinyan, 2022] Al Thinyan, K., Ghawji, H., Al Shehri, A.: What are OKRs and KPIs and can they Coexist within an Organization. International Journal of Innovative Science and Research Technology 7(8), 176-179, 2022.

[Anand, 2015] Anand, N., Grover, N.: Measuring retail supply chain performance: Theoretical model using key performance indicators (KIPs). Benchmarking: An International Journal 22(2), 290-308, 2015.

[Aversa, 2021] J. Aversa, J., Hernandez, T., Doherty, S.: Incorporating big data within retail organizations: A case study approach. Journal of Retailing and Consumer Services 60,102447, 2021.

[Bendle, 2021] Bendle, N., Wang, P.X., Ryoo, J.: The role of machine learning analytics and metrics in retailing research. Journal of Retailing 97, 658–675, 2021.

[Boulden, 1957] Boulden, J.B.: Fitting the sales forecast to your firm. Business Horizons 1, 65–72, 1957.

[Cassivi, 2007] Cassivi, H.P.: The role of joint collaboration planning actions in a demand-driven supply chain. Industrial Management & Data Systems 107(7), 954-78, 2007.

[Danese, 2018] Danese, P., Molinaro, M., Romano, P.: Managing evolutionary paths in sales and operations planning: key dimensions and sequences of implementation. International Journal of Production Research 56(5), 2036–2053, 2018.

[Garcia-Villareal, 2018] Garcia-Villarreal, E., Bhamra, R., Schoenheit, M.: A framework for technology selection to support sales and operations planning in German medical technology organisations. Advances in Transdisciplinary Engineering 285, 2018.

[Glackin, 2022] Glackin, C.E., Adivar, M.: Using the power of machine learning in sales research: process and potential. Journal of Personal Selling & Sales Management 26,1-7, 2022.

[Guerrero, 2001] L.A. Guerrero and J.A. Pino. Understanding Organizational Memory. SCCC 2001: 21st International Conference of the Chilean Computer Science Society, 124-132, IEEE, 2001.

[Guo, 2013] Guo, Z.X., Wong, W.K., Li, M: A multivariate intelligent decision making model for retail sales forecasting. Decision Support Systems 55, 247–255, 2013.

[Hastings, 1994] Hastings R. Fildes, R.: The organization and improvement of marketing forecasting. The Journal of Operational Research Society 45, 1–16, 1994.

[Huber, 2020] Huber J., Stuckenschmidt, H.: Daily retail demand forecasting using machine learning with emphasis on calendric special days. International Journal of Forecasting 36(4), 1420–1438, 2020.

[Hulthén, 2016] Hulthén, H., Näslund, D., Norrman, A.: Framework for measuring performance of the sales and operations planning process. International Journal of Physical Distribution & Logistics Management 46(9), 809-835. 2016.

[Juran, 1992] Juran, J.M.: Juran on quality by design: the new steps for planning quality into goods and services. Simon and Schuster, 1992.

[Kreuter, 2022] Kreuter, T., Scavarda, L.F., Thomé, A.M.T., Hellingrath, B., Seeling, M.X.: Empirical and theoretical perspectives in sales and operations planning. Review of Managerial Science 16(2), 319–354, 2022.

[Lam, 1998] Lam, S., Vandenbosch, M., Pearce, M.: Retail sales force scheduling based on store traffic forecasting. Journal of Retailing 74(1), 61–88, 1998.

[Lim, 2014] Lim, L., Alpan, G., Penz, B.: Reconciling sales and operations management with distant suppliers in the automotive industry: A simulation approach. International Journal of Production Economics 151, 20–36, 2014.

[McIntyre, 1993] McIntyre, S., Achabal, D., Miller, C.: Applying case-based reasoning to forecasting retail sales. Journal of Retailing 69(4), 372–398, 1993.

[Nemati, 2017] Nemati, Y., Madhoshi, M., Ghadikolaei, A.S.: The effect of sales and operations planning (S&OP) on supply chain's total performance: A case study in an Iranian dairy company. Computers & Chemical Engineering 104, 323–338, 2017.

[Panay, 2020] Panay, B., Baloian, N., Pino, J.A., Peñafiel, S., Sanson, H., Bersano, N.: Feature selection for health care costs prediction using weighted evidential regression. Sensors 20(16), 4392, 2020.

[Panay, 2021] Panay, B., Baloian, N., Pino, J.A., Peñafiel, S., Frez, J., Fuenzalida, C., Sanson, H., Zurita, G.: Forecasting key retail performance indicators using interpretable regression. Sensors 21(5), 1874, 2021.

[Pavlyuchenko, 2021] K. Pavlyuchenko, K., Panfilov, P.: Application of predictive analytics to sales planning business process of fmcg company. In Procs. of the 13th International Conference on Management of Digital EcoSystems, 167–170, 2021.

[Park, 2020] Park, K., Lee, G., Kim, C., Kim, J., Rhie, K., Lee, W.B.: Comprehensive framework for underground pipeline management with reliability and cost factors using Monte Carlo simulation. Journal of Loss Prevention in the Process Industries 63,104035, 2020.

[Petit-Renaud, 2004] S. Petit-Renaud S., Denœux, T.: Nonparametric regression analysis of uncertain and imprecise data using belief functions. International Journal of Approximate Reasoning, 35(1), 1–28, 2004.

[Stewart, 1998] Stewart D., Gallen, B.: The promotional planning process and its impact on consumer franchise building: the case of fastmoving goods companies in New Zealand. Journal of Product Brand Management 7, 557–567, 1998.

[Thomé, 2012] Thomé, A.M.T., Scavarda, L.F., Fernandez, N.S., Scavarda, A.J.: Sales and operations planning: A research synthesis. International Journal of Production Economics 138(1), 1–13, 2012.

[Tsoumakas, 2019] Tsoumakas, G.: A survey of machine learning techniques for food sales prediction. Artificial Intelligence Review 52(1), 441–447, 2019.