Quantifying Lost Sales and Optimizing Inventory Availability

Shriyan Singh and Dirk Hoffman

April 12, 2024

1   Introduction

Retail companies depend heavily on inventory management, and Shoprite Holdings Ltd., the biggest grocery operator in Africa, is no different. Shoprite, which has 3,543 locations around the continent as of December 31, 2023, has a difficult task ahead of it: efficiently managing its inventory to maximise sales and guarantee customer happiness. Out-of-stock situations are a major problem in inventory management since they can result in lost sales, a decline in client loyalty, and ultimately a bad effect on the bottom line of the business.

When a product isn't available for purchase when a customer wants to, it's said to be out-of-stock. Numerous factors, including imprecise demand projections, disturbances in the supply chain, and ineffective inventory replenishment procedures, may contribute to this. When things are out of stock, customers have three options: buy something else, put off buying it, or, worse, go with a competitor. The retailer loses money as a result of these situations, which may have long-term effects on consumer happiness and brand loyalty.

The present project aims to accomplish three principal objectives: firstly, to measure the lost sales resulting from out-of-stock scenarios in various product categories, retail locations, and time frames; secondly, to create a predictive model that precisely predicts customer buying patterns and approximates the potential sales in the event that out-of-stock products become available; and thirdly, to optimise inventory decisions by determining the ideal equilibrium between product availability and inventory holding costs. Through the pursuit of these goals, the project seeks to furnish Shoprite with practical perspectives and approaches to alleviate the consequences of stock-outs and enhance the overall efficacy of inventory management.

This project is important because it has the potential to significantly enhance Shoprite's inventory management procedures. Shoprite can optimise its inventory levels, decrease stockouts, and improve customer happiness by using data-driven decision-making to precisely measure missed sales and comprehend the elements that lead to out-of-stock scenarios. Additionally, Shoprite will be able to more accurately predict customer demand and make proactive adjustments to its inventory planning and replenishment methods thanks to the construction of a predictive model. Shoprite will achieve greater operational efficiency and profitability by optimising inventory decisions to assist them strike the correct balance between guaranteeing product availability and reducing inventory holding costs.

In conclusion, this project tackles the crucial problem of out-of-stock expenses within the framework of Shoprite Holdings Ltd., the biggest grocery chain in Africa. The project aims to give Shoprite key insights and strategies to improve its inventory management practices, boost customer satisfaction, and drive long-term business success by utilising data-driven approaches to quantify lost sales, predict customer behaviour, and optimise inventory decisions.

2   Literature Survey

The effects of out-of-stock circumstances on sales and consumer behaviour in the retail sector have been the subject of much research. Stockouts have serious negative effects on a retailer's finances and reputation, as several studies have shown. For example, Corsten and Gruen's 2003 study calculated that retail sales can drop by an average of 4% as a result of out-of-stock conditions. They discovered that 31% of consumers would buy an alternative item, 26% would postpone their purchase, and 9%

would buy the item from a rival when presented with an out-of-stock situation. These results highlight how crucial it is to reduce stockouts in order to save sales and keep customers loyal.

Previous studies have used a variety of approaches to quantify lost revenues caused by out-of-stock circumstances. Utilising previous sales data is a popular method for estimating the potential sales that may have been made if the out-of-stock products had been accessible. For instance, Gruen and Corsten (2007) presented a model that accounts for seasonality and promotions by comparing a product's sales during an out-of-stock period with those during a corresponding in-stock period. Utilising customer surveys to collect information on planned purchases and what customers did in the event of an out-of-stock scenario is another strategy. Using this technique, Van Woensel et al. (2007) estimated the substitution rates and lost sales for several product categories in a retail environment.

Researchers have looked into ways to optimise inventory levels and reduce the likelihood of out-of-stock situations in addition to measuring lost sales. The newsvendor model is one popular method that seeks to ascertain the ideal order quantity that strikes a balance between the expenses of overstocking and understocking (Silver et al., 1998). Other optimisation approaches, notably simulation-based techniques and dynamic programming, have been used to address complicated inventory problems while taking supply chain restrictions, lead times, and demand uncertainty into account (Zipkin, 2000; Axsäter, 2006).

This initiative intends to address the gaps and limitations in current approaches that remain despite the substantial study in this field. Numerous research endeavours concentrate on analysis at the aggregate level, potentially failing to grasp the intricate dynamics of out-of-stock circumstances at the shop or product level. Furthermore, there is still a lot of room for improvement in the integration of machine learning approaches to forecast customer behaviour and optimise inventory decisions in real-time. By utilising comprehensive sales, item, and inventory movement data from Shoprite, this project aims to close these gaps by creating a thorough methodology for estimating lost sales, forecasting customer behaviour, and improving inventory management techniques.

With the use of modern machine learning and analytics techniques, these constraints will be addressed, and the project's goal is to give Shoprite a solid and workable solution to deal with out-of-stock scenarios and enhance its inventory management procedures.

3 Data Description

This project utilizes three main datasets provided by Shoprite Holdings Ltd: sales data, article data, and inventory movements data. These datasets form the foundation for quantifying lost sales, building predictive models, and optimizing inventory decisions.

3.1 Sales Data

The sales data contains information about individual sales transactions at Shoprite stores. It includes the following attributes:

- `date_key`: The date of the sales transaction (format: YYYYMMDD)

- `site_code`: The unique identifier of the store where the sale occurred

- `unique_ticket_id`: A unique identifier for each sales transaction

- `ticket_end_time`: The timestamp of when the sales transaction was completed

- `article_code`: The unique identifier of the article (product) sold

- `sales_qty_alternate_uom`: The quantity of the article sold in an alternate unit of measure

- `alternate_uom`: The alternate unit of measure for the sold quantity

- `sales_qty_base_uom`: The quantity of the article sold in the base unit of measure

- `base_uom`: The base unit of measure for the sold quantity (e.g., "EA" for each)

3.2 Article Data

The article data provides detailed information about each article (product) sold at Shoprite stores. It includes the following attributes:

- `article_code`: The unique identifier of the article

- `article_desc`: A description of the article

- `base_uom`: The base unit of measure for the article

- `super_dept_no`: The number of the super department to which the article belongs

- `super_dept_desc`: The description of the super department

- `dept_no`: The number of the department to which the article belongs

- `dept_desc`: The description of the department

- `category_no`: The number of the category to which the article belongs

- `category_desc`: The description of the category

- `merchandise_category_no`: The number of the merchandise category to which the article belongs

- `merchandise_category_desc`: The description of the merchandise category

- `sub_category_no`: The number of the sub-category to which the article belongs

- `sub_category_desc`: The description of the sub-category

3.3 Inventory Movements Data

The inventory movements data captures the changes in inventory levels for each article at different stores over time. It includes the following attributes:

- `unique_key`: A unique identifier for each inventory movement record

- `posting_date`: The date of the inventory movement (format: YYYY-MM-DD)

- `entry_timestamp`: The timestamp of when the inventory movement was recorded

- `site_code`: The unique identifier of the store where the inventory movement occurred

- `article_code`: The unique identifier of the article involved in the inventory movement

- `movement_type`: The type of inventory movement (e.g., receipt, issue, transfer)

- `movement_type_category`: The category of the inventory movement (e.g., inbound, outbound)

- `quantity_base_uom`: The quantity of the inventory movement in the base unit of measure

- `base_uom`: The base unit of measure for the inventory movement

- `stock_level_before_posting`: The stock level of the article before the inventory movement

Common characteristics like `article_code` and `site_code` link these datasets together. The article data offers more information about each sold article, whereas the sales data records the actual sales transactions. Inventory movements data allows for the discovery of out-of-stock situations and the calculation of lost sales by tracking changes in inventory levels for each article at various retailers.

During the data preprocessing stage, several steps were taken to ensure data quality and consistency:

1. **Missing Values**: The datasets were checked for missing values, and appropriate strategies were applied to handle them. For example, records with missing critical information such as `article_code` or `site_code` were removed, while missing values in non-critical fields were handled through imputation or default value assignment.

2. **Duplicate Records**: The datasets were inspected for duplicate records, and any identified duplicates were removed to ensure data integrity.

3. **Outliers and Anomalies**: The data was analyzed for outliers and anomalies that could potentially skew the analysis. Outliers were investigated and treated appropriately based on their nature and the specific requirements of the analysis.

4. **Data Validation**: The datasets still need to be validated against known business rules and constraints to ensure consistency and accuracy. For example, checks will be performed to ensure that `sales_qty_base_uom` and `sales_qty_alternate_uom` are non-negative values and if there are, find reasons for it.

By thoroughly understanding the structure, relationships, and quality of the datasets, we can effectively leverage them to quantify lost sales, build predictive models, and optimize inventory decisions for Shoprite Holdings Ltd.

4 Methodology

4.1 Data Exploration and Preprocessing

The data will be examined using exploratory data analysis (EDA) approaches to find patterns, trends, and connections. Calculating summary statistics, displaying distributions, and looking at correlations between variables will all be part of this. We will construct visualisations to show product distributions, inventory movements, and sales patterns, including bar charts, scatter plots, time series plots, and heatmaps. These data visualisations will offer insightful information and serve as a guide for the methodology's next steps.

The benefits of conducting thorough data exploration and preprocessing include:

- Ensuring data quality and consistency

- Identifying potential issues or anomalies in the data

- Gaining a deeper understanding of the data characteristics and relationships

- Facilitating the selection of appropriate modeling techniques

4.2 Quantifying Lost Sales

A multi-step method will be created to measure the amount of sales that are lost as a result of out-of-stock circumstances. In order to determine when an out-of-stock situation occurs, the inventory movements data will first be examined to find any occasions in which the stock level prior to posting falls below zero or goes negative. The relevant sales transactions in the sales data will subsequently be mapped to these occurrences.

Subsequently, a time series forecasting model, such ARIMA or Prophet, will be utilised to approximate the anticipated sales quantity for every product that is out of stock by utilising past sales trends. The lost sales during the out-of-stock period will be calculated as the difference between the expected and actual sales quantities.

Moreover, customer behavior patterns will be analyzed to account for potential substitution effects. If a customer purchases a similar product within the same category during an out-of-stock situation, it will be considered as a substitution sale. The lost sales will be adjusted accordingly to avoid overestimation.

The benefits of this approach to quantifying lost sales include:

- Accurately estimating the financial impact of out-of-stock situations

- Identifying the products and periods with the highest lost sales

- Providing a foundation for optimizing inventory decisions

4.3 Predictive Model for Customer Purchase Behavior

Utilising machine learning techniques, sales potential and customer purchase patterns will be predicted, as well as the likelihood that out-of-stock merchandise will be sold. The dataset will be prepared by combining pertinent attributes from the inventory movement, sales, and article data. Among the factors are stock availability, sales history, and product qualities.

Several machine learning techniques, such as gradient boosting, decision trees, and logistic regression, will be evaluated after the dataset has been split into training and testing sets. After the models have been trained on the training set, grid search or random search techniques will be used to optimise the model's performance through hyperparameter tuning.

The trained models will be evaluated on the testing set using appropriate evaluation metrics, such as accuracy, precision, recall, and F1-score, to assess their predictive capabilities. The best-performing model will be selected based on its performance and interpretability.

The benefits of building a predictive model for customer purchase behavior include:

- Understanding the factors influencing customer purchasing decisions

- Estimating the potential sales impact of out-of-stock situations

- Providing insights for improving inventory planning and customer satisfaction

4.4 Optimization Framework for Inventory Decisions

An optimisation framework will be created in order to balance waste costs and availability while making the best inventory decisions. The framework will take into account variables including lead times, ordering costs, holding costs, and product demand.

The optimisation problem will be phrased as a multi-objective optimisation with the goal of minimising the costs associated with waste resulting from excess inventory as well as the missed sales caused by out-of-stock situations. The order amounts and reorder points for each product at each store will be among the decision variables.

To determine the ideal inventory levels, the optimisation framework will apply methods including simulated annealing, genetic algorithms, and linear programming. The particular business needs and cost structures of Shoprite will serve as the basis for defining the objective function and limitations.

The benefits of developing an optimization framework for inventory decisions include:

- Balancing the trade-off between availability and waste costs

- Identifying optimal inventory levels for each product and store

- Improving inventory turnover and reducing holding costs

4.5 Evaluation and Validation

Numerous assessment metrics and validation procedures will be used to evaluate the predictive model's and the optimisation framework's performance. Metrics like accuracy, precision, recall, and F1-score will be computed for the predictive model to assess how well it predicts the purchasing behaviour of customers. We will employ cross-validation approaches, like k-fold cross-validation, to make sure the model is resilient and generalizable.

Metrics like decreased lost sales, increased inventory turnover, and cost savings will be assessed for the optimisation framework. To evaluate the optimised inventory levels' effect on key performance indicators, past data will be contrasted with the current levels. To verify the optimisation outcomes against real sales and inventory data, backtesting will be done.

The benefits of conducting thorough evaluation and validation include:

- Ensuring the reliability and effectiveness of the developed models and frameworks

- Identifying areas for improvement and refinement

- Providing confidence in the implementation of the proposed solutions

5 Experimental Design

The experimental design will involve the following steps:

5.1 Model Development

- To anticipate client purchasing behaviour, a range of machine learning models, including decision trees, logistic regression, and gradient boosting, will be constructed.

- The training dataset will be used to train the models with the proper feature engineering and preprocessing methods.

- To maximise model performance, grid search and random search methods will be used for hyperparameter tuning.

5.2 Model Evaluation

- The trained models will be evaluated on the validation dataset using metrics such as accuracy, precision, recall, and F1-score.

- The best-performing model will be selected based on its performance on the validation set.

- The selected model will be further tested on the testing dataset to assess its generalization capabilities.

5.3 Optimization Framework

- The optimization framework for inventory decisions will be implemented using techniques such as linear programming, genetic algorithms, or simulated annealing.

- The objective function and constraints will be defined based on the business requirements and cost structures of Shoprite.

- The optimization framework will be tested on historical data to evaluate its effectiveness in reducing lost sales and optimizing inventory levels.

5.4 Sensitivity Analysis

- Sensitivity analysis will be conducted to assess the robustness of the optimization framework under different scenarios.

- The impact of varying input parameters, such as demand forecasts, holding costs, and lead times, will be evaluated.

- The results of the sensitivity analysis will provide insights into the stability and reliability of the proposed solutions.

5.5 Iterative Refinement

- Based on the evaluation results and sensitivity analysis, the models and optimization framework will be iteratively refined.

- Adjustments will be made to the feature engineering, hyperparameters, and optimization constraints to improve performance and align with business objectives.

- The refined models and framework will be re-evaluated to ensure their effectiveness and robustness.

The benefits of this experimental design include:

- Ensuring a systematic and structured approach to model development and evaluation

- Allowing for the comparison and selection of the best-performing models

- Providing insights into the sensitivity and robustness of the proposed solutions

- Facilitating iterative refinement and improvement of the models and optimization framework

6 Timeline and Next Steps

The project steps will consist of the following phases. (The timeline and due dates will be confirmed with the supervisor for clarity):

1. Model Development (3-4 weeks):

   - Predict customer purchase behavior.
   - Feature engineering and preprocessing techniques.
   - Evaluate the models using appropriate evaluation metrics.
   - Hyperparameter tuning.

2. Optimization Framework (3 weeks):

   - Optimization framework for inventory decisions.
   - Define the objective function and constraints based on business requirements.
   - Implement the optimization framework using suitable techniques.
   - Test the optimization framework.

3. Sensitivity Analysis (1 week):

   - To assess the robustness of the optimization framework.
   - Evaluate the impact of varying input parameters.
   - Analyze the results and draw insights.

4. Iterative Refinement (2 weeks):

   - Refine the models and optimization framework based on evaluation results and sensitivity analysis.
   - Make necessary adjustments to improve performance and align with business objectives.
   - Re-evaluate the refined models and framework to ensure their effectiveness.

7 Bibliography

1. Aksoy, A., Ozturk, N., & Sucky, E. (2012). A decision support system for demand forecasting in the clothing industry. International Journal of Clothing Science and Technology, 24(4), 221-236.

2. Arunraj, N. S., & Ahrens, D. (2015). A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting. International Journal of Production Economics, 170, 321-335.

3. Huang, T., Fildes, R., & Soopramanien, D. (2014). The value of competitive information in forecasting FMCG retail product sales and the variable selection problem. European Journal of Operational Research, 237(2), 738-748.

4. Huber, J., Gossmann, A., & Stuckenschmidt, H. (2017). Cluster-based hierarchical demand forecasting for perishable goods. Expert Systems with Applications, 76, 140-151.

5. Kilger, C., & Wagner, M. (2008). Demand planning. In H. Stadtler, C. Kilger, & H. Meyr (Eds.), Supply chain management and advanced planning (pp. 133-160). Springer, Berlin, Heidelberg.

6. Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra-and inter-category promotional information. European Journal of Operational Research, 249(1), 245-257.

7. Prak, D., Teunter, R., & Syntetos, A. (2017). On the calculation of safety stocks when demand is forecasted. European Journal of Operational Research, 256(2), 454-461.

8. Ramos, P., Santos, N., & Rebelo, R. (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. Robotics and Computer-Integrated Manufacturing, 34, 151-163.

9. Thomassey, S. (2010). Sales forecasts in clothing industry: The key success factor of the supply chain management. International Journal of Production Economics, 128(2), 470-483.

10. Zhang, Y., Jia, H., Diao, Y., Hai, M., & Li, H. (2016). Research on demand-driven leagile supply chain operation model: a simulation based on system dynamics in system engineering. Systems Engineering Procedia, 5, 308-312.