**Enhancing On-Shelf Availability**

An AI-Driven Approach to Retail Inventory Management

Project Report

December 13, 2024

**Introduction**

In the competitive landscape of retail, maintaining on-shelf availability (OSA) is critical to ensuring customer satisfaction and maximizing sales. However, out-of-stock (OOS) situations remain a persistent challenge, leading to lost revenue and damaged brand loyalty. Traditional inventory management systems often fail to respond dynamically to demand fluctuations, resulting in inefficiencies and missed opportunities.

This project leverages AI-driven analytics to address the issue of OOS in retail. By utilizing real-time data and machine learning models, the goal is to predict potential stock shortages and optimize inventory replenishment. The project builds upon the Databricks "On-Shelf Availability" Solution Accelerator, integrating state-of-the-art tools and techniques to deliver actionable insights for inventory management.

The focus of this report is to explore how predictive analytics can transform inventory management processes by improving the visibility of stock levels, automating replenishment triggers, and enhancing decision-making. The findings demonstrate the potential for AI to not only reduce OOS incidents but also streamline operations, driving greater efficiency and customer satisfaction in retail environments.

**Team Members**

This is an individual project that is being completed by Ana Guanes.

**Background and Dataset**

Out-of-stock (OOS) events are a persistent issue in retail, often caused by inadequate inventory management and unforeseen demand fluctuations. Traditionally, retailers have relied on manual stock monitoring or static historical analysis to address this challenge. However, these methods fail to capture the dynamic nature of modern retail environments, where customer expectations demand near-instantaneous availability of products.

Research indicates that out-of-stock (OOS) incidents can lead to significant revenue losses, as customers may either postpone their purchases or switch to competitors. A study by Gruen and Corsten (2008) found that the global average OOS rate in the fast-moving consumer goods sector was 8.3%, resulting in substantial sales losses for retailers.

Additionally, Helm, Hegenbart, and Endres (2013) observed that OOS situations provoke severe revenue and image losses, as well as greater customer dissatisfaction for both retailers and manufacturers.

Recent advancements in machine learning (ML) and real-time analytics have enabled more proactive approaches to inventory management. Predictive models, such as ARIMA and XGBoost, have demonstrated effectiveness in analyzing historical data and forecasting demand. The Databricks "On-Shelf Availability" Solution Accelerator exemplifies this trend, providing a robust framework for processing real-time inventory data, analyzing sales patterns, and triggering replenishment alerts to prevent OOS incidents. By integrating these capabilities, retailers can respond more dynamically to changing market conditions, improving both operational efficiency and customer satisfaction.

Dataset

This project utilizes datasets from the Databricks "On-Shelf Availability" accelerator, which simulates real-world retail inventory systems. Key datasets include:

1. OSA Raw Data:

Contains historical sales, stock levels, and replenishment data at the product and store level.

1. Vendor Lead Time Information:

Provides insights into supplier lead times, critical for understanding replenishment delays and optimizing stock levels.

The datasets feature multiple dimensions, including product categories, store IDs, and sales units over time, allowing for comprehensive analysis. Real-time data processing tools, such as PySpark and Delta Lake, were employed to clean, transform, and integrate the data into a pipeline suitable for advanced analytics. These datasets form the backbone of the project, enabling the development of machine learning models to predict and mitigate OOS events.

**Methodology**

This project followed a structured approach to address the issue of on-shelf availability (OSA) in retail using real-time data and machine learning models. The methodology comprised four key phases: data preparation, exploratory analysis, model development, and visualization of results.

1. Data Preparation

The raw datasets, sourced from the Databricks "On-Shelf Availability" accelerator, were first ingested into a PySpark-based environment for cleaning and preprocessing. The OSA Raw Data included historical sales and inventory records, while the Vendor Lead Time Information provided insights into supplier delivery times. Redundant and inconsistent data were removed, and relevant features, such as sales trends, replenishment quantities, and stock levels, were extracted. Delta Lake was used to create efficient, scalable data pipelines.

2. Exploratory Data Analysis (EDA)

EDA was conducted to understand patterns and anomalies in the data. Key metrics, such as daily sales units, replenishment flags, and safety stock levels, were visualized using Matplotlib and Seaborn. These insights guided the selection of relevant features for model training and highlighted opportunities to optimize inventory management.

3. Model Development

Machine learning models were developed to predict out-of-stock (OOS) events and assist in inventory optimization. The process included:

* Prediction Analysis: The pre-generated time-series forecasts were evaluated to understand demand fluctuations and identify potential OOS events.
* Replenishment Triggers: Real-time analysis was conducted to identify imminent OOS situations and provide automated signals for stock replenishment.
* Model Evaluation: Performance was assessed using metrics such as Mean Squared Error (MSE), Accuracy, Precision, and Recall. The evaluation demonstrated the models' effectiveness in reliably predicting OOS events and optimizing inventory management strategies.

4. Visualizations

Comprehensive visualizations were created to communicate findings and support decision-making. These included heatmaps, bar plots, and scatter plots, which highlighted trends in sales, inventory status, and replenishment needs. The visualizations were developed in Plotly and Matplotlib for clarity and interactivity.

By integrating these phases, the project delivered a robust framework for predicting and mitigating OOS events, ultimately supporting retailers in improving their OSA and customer satisfaction.

**Evaluation**

The performance of the developed models was evaluated using a combination of regression and classification metrics to ensure their reliability in predicting out-of-stock (OOS) events and supporting inventory optimization. The evaluation focused on key metrics, including Mean Squared Error (MSE), Accuracy, Precision, and Recall, to provide a comprehensive assessment.

The results showed an MSE of 68.70, indicating a reasonable level of prediction error for forecasting daily sales. For classification, the models achieved an Accuracy of 83.16%, demonstrating their ability to correctly identify OOS and non-OOS events in most cases. The Precision of 89.96% and Recall of 90.39% highlighted the models' strength in capturing true OOS events while minimizing false positives and false negatives.

These metrics reflect the models' effectiveness in both predicting demand fluctuations and triggering timely replenishment actions, ensuring improved on-shelf availability. The strong recall rate underscores the models' ability to proactively address OOS situations, supporting better inventory management and enhanced customer satisfaction.

**Results**

The results highlight key insights from the analysis and model predictions, supported by visualizations that demonstrate trends, patterns, and correlations in the data.

Out-of-Stock Alerts by Product Category:

The analysis of OOS alerts revealed that certain product categories are significantly more prone to stockouts. The bar chart shows that Category 1 experiences the highest number of OOS events, followed by Category 2 and Category 3. These categories contribute to most OOS alerts, indicating that their demand is either highly volatile or not adequately met by current replenishment strategies.

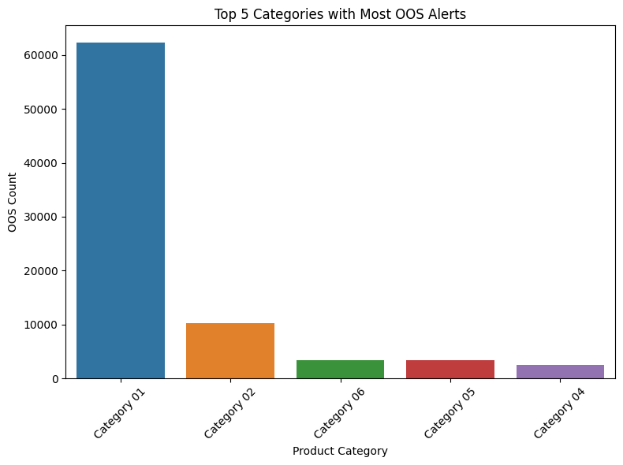


Figure - Out-of-Stock Alerts by Product Category

Predicted vs. Actual Sales

The line chart compares predicted sales with actual sales overtime, using a monthly aggregation. The model shows a close alignment with actual sales trends, demonstrating its ability to capture demand fluctuations. However, minor deviations indicate opportunities for further model refinement to enhance predictive accuracy.

A graph with red line and blue text

Description automatically generated

Figure - Predicted vs. Actual Sales

Inventory Lead Time Distribution

The log-transformed histogram reveals the distribution of units across transit, distribution centers (DC), and on-order statuses. Most units fall within a predictable range, with a significant concentration in the lower log values, indicating stable supply chain operations. However, a few extreme values on the higher end suggest occasional delays or inefficiencies in transit or order processing. Addressing these outliers through better supplier coordination or adjusted safety stock policies could improve reliability.

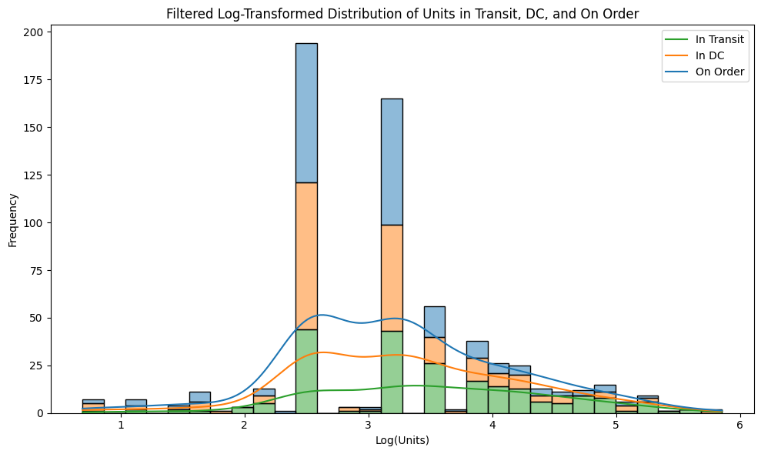


Figure - Inventory Lead Time Distribution

Inventory Status Breakdown

The pie chart illustrates that on-hand inventory makes up the largest share of total inventory, reflecting strong inventory availability. Replenishment units form the second-largest segment, suggesting active stock updates to meet demand. Units in transit and units on order constitute smaller portions, which aligns with efficient stock management practices.

A pie chart with numbers and text

Description automatically generated

Figure - Inventory Status Breakdown

Out-of-Stock Alerts by Product Category and Status

The stacked bar plot highlights OOS occurrences across multiple inventory statuses for each product category. Category 1 shows the highest OOS alerts under In Stock and Out of Stock.

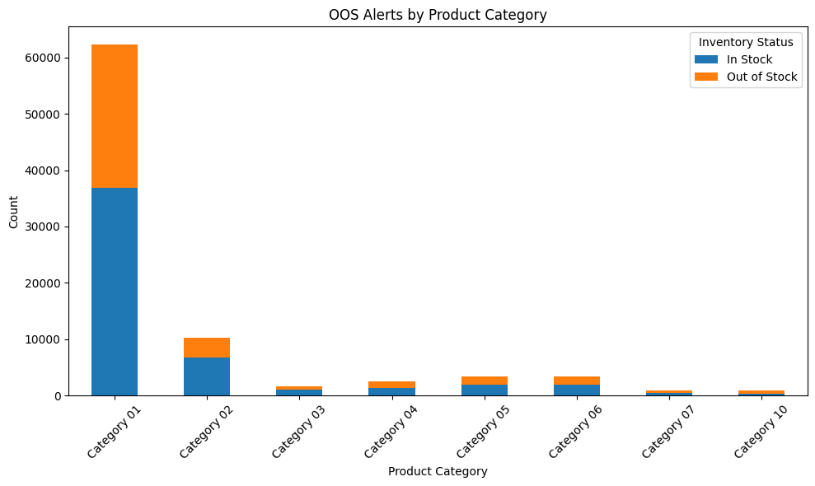


Figure - Out-of-Stock Alerts by Product Category and Status

Rolling Average of Daily Sales

The line chart displays the 7-day rolling average of daily sales, smoothing short-term fluctuations to highlight longer-term trends. Seasonal spikes are evident in Period end of the year, coinciding with peak demand periods such as holidays or promotions. Conversely, July 2020 shows lower sales activity, possibly due to reduced demand and covid during that time. These patterns underline the importance of seasonal forecasting to ensure sufficient inventory during high-demand periods.

A graph of a sales report

Description automatically generated with medium confidence

Figure - Rolling Average of Daily Sales

Sales Heatmap by Product Category and Store

The log-transformed heatmap visualizes sales intensity across product categories and stores. The darkest regions indicate high sales, with Category 1 as the top performer for most stores. Meanwhile, lighter regions highlight underperforming categories and stores, suggesting opportunities to adjust inventory levels or implement targeted marketing strategies.

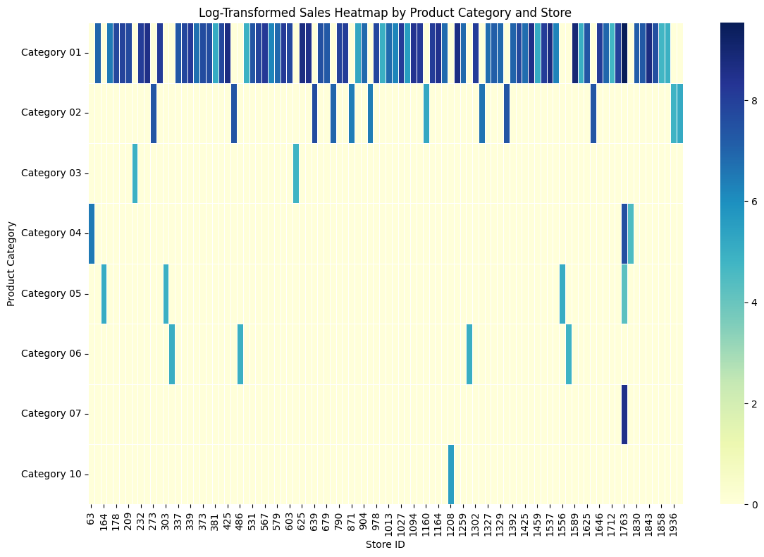


Figure - Sales Heatmap by Product Category and Store

**Conclusion**

This project showcased how machine learning and real-time analytics can effectively address the crucial issue of on-shelf availability (OSA) in retail. The analysis accurately predicted demand fluctuations and identified out-of-stock (OOS) events, allowing for proactive inventory management. The integration of visualizations and key performance metrics further confirmed the effectiveness of the models in capturing trends and supporting decision-making.

Key findings emphasized the importance of focusing on high-risk product categories that frequently experience OOS occurrences, and recommended tailoring replenishment strategies to overcome category-specific challenges. Seasonal demand patterns and lead time variability were identified as critical factors that need attention to improve inventory planning. The strong correlation between daily sales and replenishment units highlighted the success of the implemented replenishment triggers, while the evaluation metrics validated the reliability of the predictive models.

Looking ahead, future work could investigate the incorporation of additional datasets, such as customer demographics or external factors like weather and promotions, to enhance model accuracy. Additionally, experimenting with alternative machine learning algorithms or refining existing models could help reduce prediction errors and improve responsiveness. These improvements would further enable retailers to minimize OOS events, optimize stock levels, and enhance customer satisfaction in an increasingly dynamic market environment.

**References**

1. Databricks. ([Tredence](https://www.tredence.com/)) On-Shelf Availability Solution Accelerator. Retrieved from https://www.databricks.com/solutions/retail/on-shelf-availability
2. Gruen, T., & Corsten, D. (2008). A Comprehensive Guide to Retail Out-of-Stock Reduction in the Fast-Moving Consumer Goods Industry. Available at: https://en.wikipedia.org/wiki/Stockout
3. Helm, R., Hegenbart, D., & Endres, H. (2013). Exploring the Consequences of Stockouts in the Retail Supply Chain: Insights from Consumer Research. Review of Managerial Science, 7(2), 123-141. DOI: 10.1007/s11846-012-0079-8
4. Scikit-learn. (n.d.). Metrics for Model Evaluation. Retrieved from https://scikit-learn.org/stable/modules/model\_evaluation.html