

Predicting Customer Order Patterns to Reduce Out-of-Stock (OOS) Risk

Department of Industrial Systems Engineering and Management

IE3100R Systems Design Project - Group 20A

SDP Supervisor: Assistant Professor Qin Hanzhang

Procter & Gamble Supervisor: Mr Benjamin Yeo

Group 20 Members: Daniel Chin Yong Jie | Edward Sanjaya | Joel Lo Liang Ze | Sidharth Rajesh



1 PROBLEM OVERVIEW

PROBLEM DESCRIPTION

Procter & Gamble (P&G), as a leading FMCG company, relies on accurate demand forecasting to anticipate fluctuations in customer orders. **Effective prediction models** are essential to mitigate the risks associated with **demand surges and declines**.

P&G is continuously enhancing its **forecasting** capabilities to improve order pattern predictions. By creating a **predictive** model, the company is strengthening its ability to **anticipate market fluctuations** and **demand variability**. These efforts help proactively identify and mitigate potential out-of-stock (OOS) risks.

This forecasting gap can result in **inventory mismatches**, which are either **stockouts** that leads to lost sales and service failures, or **overstocks** that increases holding costs. Given the fast-moving nature of P&G's products, **improving predictive capabilities** is crucial to maintaining supply chain efficiency and responsiveness.



2 OBJECTIVES

KEY OBJECTIVES



Trends and Seasonality Analysis

Identify demand patterns and seasonal fluctuations to anticipate order spikes and declines, improving order forecasting.



Machine Learning Model Development

Leverage advanced machine learning techniques to build a more accurate forecasting model based on past customer orders.



Improve Prediction Accuracy

Refine forecasting capabilities to achieve a prediction accuracy of at least 95%, reducing the risk of stockouts and overstocks.

3 METHODOLOGY



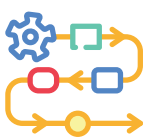
Client Engagement

Communicate in addressing issues and aligning goals with stakeholders



Data Processing

Collect, Clean, and transform raw data into a structured format



Model Development

Design and optimize algorithms based on processed data



Results Analysis

Evaluate model performance to validate accuracy and reliability



Implementation

Deployment of model with seamless integration and effective functionality

4 KEY SKILLSETS

Market Research



Machine Learning



Big Data Analytics



Statistical Techniques



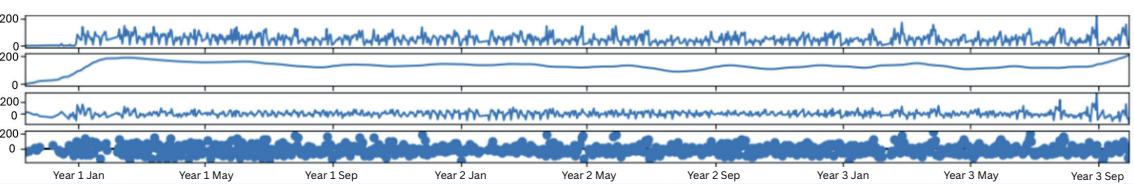
5 PROJECT IMPLEMENTATION

TREND ANALYSIS OF ORDER QUANTITY

Understanding Customer Order Data

Objective: Develop domain expertise by engaging with key business stakeholders to explore the current industry trends, while analyzing data to uncover specific customer behaviour patterns and insights.

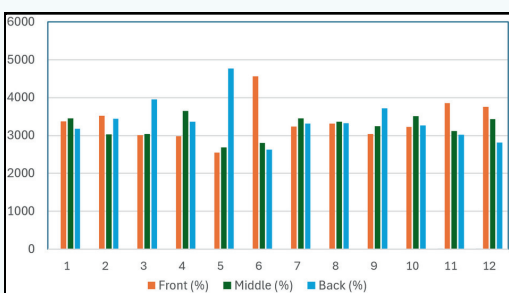
STL Decomposition with X period for 3 years data



STL Decomposition

Analysis reveals monthly sales spikes and a consistent decline in sales during festive seasons at the beginning and middle of the year.

Front-loading, Middle-loading, Back-loading within Each Month



Quarterly Order Trends

Month-over-month analysis shows an even spread in orders with exception in May and June periods.

Further Order Breakdown of a single company

Reporting Customer	Order Year	Order Date	Front (%)	Middle (%)	Back (%)
Company XYZ	Year 1	1	33.72	34.53	31.75
		2	35.21	30.33	34.46
		3	30.13	30.37	39.50
		4	29.80	36.53	33.67
		5	25.44	26.87	47.69
		6	45.65	28.07	26.28
		7	32.33	34.52	33.15
		8	33.13	33.65	33.22
		9	30.37	32.42	37.21
		10	32.22	35.11	32.67
		11	36.56	31.21	30.23
		12	37.54	34.31	28.15

Percentage of Orders

Front Loading: 1-10
Middle Loading: 11-20
Back Loading: 21-30

Color Indication

Pink Shading: Lower %
Blue Shading: Higher %

Analysis of Company XYZ

The results indicate that company XYZ does not exhibit a clear **pattern** in terms of front, middle, or back-loaded orders. However, orders tend to be back-loaded in May and front-loaded in June, possibly due to increased demand during **festive** seasons.

MODEL DEVELOPMENT

Stage	Description
Data Collection	24 months of P&G Customer Order Data
Data Preprocessing	Remove invalid entries and duplicates
Feature Engineering	Use statistically significant metrics
Encode Seasonality	Using Sine and Cosine Functions
Expert Input	Added metrics recommended by P&G Experts

Examples of Key Metrics

- Customer
- Subsector
- Seasons (Months & Quarters)
- Sales Performance Benchmark

Description of Data

- Huge Dataset (>250M Rows)
- Requires Data Cleaning
- Daily Data Extraction
- Exported as CSV Format

Model Development

Training Data: 24 months of data. Testing Data: 1 month.
Performance Indicator = 1-MAPE

Model	Initial Accuracy (1-MAPE)	Limitation
SARIMA	30.2%	Fails to capture the trend of orders
Multiple Linear Regression	45.3%	Non-linear relationship cannot be captured
XGBoost	80.1%	Generally works better

Chosen Model: XGBoost (Initial Accuracy of ~80%)

Model Improvement

Steps Taken	Results (Base Accuracy of 80%)
Removed Sunday Predictions	Prevented model overcompensation , improving accuracy to 90%
Reduced Number of Estimators	Maintained accuracy (80.1%) while reducing training time from >10 min to ≤7 min
Decreased Granularity	Reduced accuracy (64.3%) due to loss of critical parameters affecting key subsectors and customers.

Final Improved Model: XGBoost (Best Accuracy of >90%)

MODEL OUTPUT

Back-Testing for Data Validation

Performed back-testing with at least six months of data to determine the optimal training period for model effectiveness.

Months	Accuracy
24	45%
25	55%
26	34%
27	22%
28	96%
29	92%

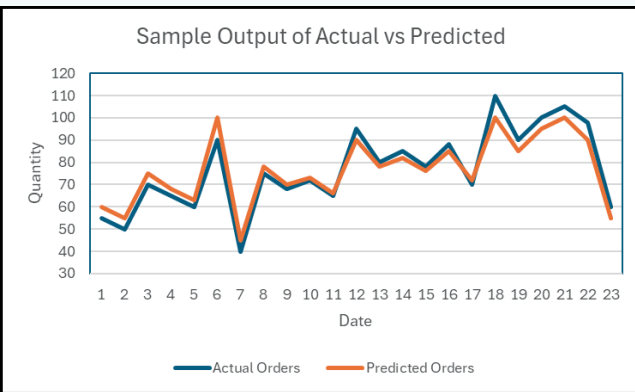
Impact of Reducing Training Set

- Investigated the effect of using a smaller training dataset.
- Assessed whether a reduced training set could maintain consistent accuracy across months
- Findings: Back-testing with six months of data revealed that removing one month significantly decreased sample size and reduced model accuracy (target: ≥28M observations).

Daily Order Prediction Output Table

Date	Predicted Order
Day 11	130
Day 12	110
Day 13	120
Day 14	150
Day 15	110
Day 16	90
Day 17	0

Sample Output of Actual vs Predicted



Analysis of Actual vs Predicted output

- The output represents the final monthly projection.
- Sundays have a predicted value of 0, as there is no order on these days.
- Provides a comparative analysis of actual and forecasted order trends.

6 BENEFITS & ACHIEVEMENT



PROJECT BENEFITS

- Higher customer satisfaction and brand loyalty from consistent product availability.
- Capturing additional sales opportunities by reducing OOS.
- Optimizing inventory levels in warehouses and stores by accurately predicting demand shifts.



KEY ACHIEVEMENTS

- Achieved over 90% accuracy for most months' order predictions.
- Identified key metrics and significant factors influencing orders.
- Successfully predicted demand in a key market using a machine learning model.

7 FUTURE DIRECTION



- Incorporate holiday adjustments that enhance accuracy during festive months, requiring an additional holiday data source.
- Implement a rolling daily prediction model that integrates actual data to improve accuracy over time.
- Explore a combination of other time series and machine learning models to enhance predictive performance.