# DATA SCIENCE PORTFOLIO

Presented by:

Rahma Anggana Rarastyasa

# PRESENTATION CONTENTS

## Introduction

Self-Overview

## Project Overview

Previous Project

## Main Project

Smarter Stocking —
ML-Powered Sales Forecasting

# INTRODUCTION

# ABOUT RARA

A highly motivated Industrial Engineering graduate from Sebelas Maret University with 1 year of professional experience in port services and healthcare services. **Strongly interested in Data Science and Data Analytics,** with foundational skills in data processing, statistical analysis, and data visualization to support effective decision-making. Proficient in **Python**, **SQL**, and **Excel**, with experience using tools such as Pandas, Numpy, Matplotlib, and Looker Studio. Known for being **analytical**, **detail-oriented**, and a **fast learner**, with the ability to work both independently and in teams. Eager to contribute to data-driven projects and continuous improvement initiatives.

## CONTACT ME!


www.linkedin.com/in/rahma-anggana-rarastyasa

rahmaanggana04@gmail.com

http://wa.me/6285257896356

# WORKING EXPERIENCE

- PT Revolusi Kesehatan Indonesia

  Operation Team

  Nov, 2024 - Present

- PT Krakatau Bandar Samudera

  Port Area and Warehouse Intern

  Jan, 2023 - Feb, 2023

# EDUCATION BACKGROUD

- Dibimbing.id

  Data Analyst & Data Science Bootcamp (Non-Degree)

  May, 2025 - Present

- Universitas Sebelas Maret Surakarta

  Industrial Engineering (Bachelor of Engineering)

  Aug, 2020 - July, 2024

# PROJECT OVERVIEW

# WHAT I’VE LEARNED

## Time Series Forecasting

- To forecast monthly sales of anti-diabetic drugs in Australia (1991–2008) using time series analysis. The data was analyzed to identify trends, seasonal patterns, and sales fluctuations, followed by building a forecasting model using AutoARIMA
- The model with a 12-month seasonal length was selected as it produced the lowest RMSE of 0.1304 (≈13.94% error), indicating reasonably accurate predictions to support drug inventory and distribution planning.
- Link: https://bit.ly/time_series_forecasting_


## Churn Analysis

- This project aimed to predict customer churn in the telecommunications industry using machine learning. Random Forest performed best with 0.788 accuracy and 0.661 F1-score for churned customers, outperforming SVM, XGBoost, and Decision Tree, helping identify high-risk customers and key factors influencing churn.
- Link: https://bit.ly/https://bit.ly/time_series_forecasting

# 📦 SMARTER STOCKING – ML-POWERED SALES FORECASTING

TIME SERIES: SALES VOLUME FORECASTING

# STAY WITH ME TO UNLOCK SMARTER AND FASTER WAREHOUSE PLANNING!

# SMARTER STOCKING CONTENTS

# PROJECT OVERVIEW

Business understanding : Define project background, objectives, and understand dataset

↓

Data Preprocessing

↓

Create weekly time series for each warehouse

↓

Checking time series characteristics, making time series analysis, find the best model, and evaluate for Bakers Warehouse

↓

Checking time series characteristics, making time series analysis, find the best model, and evaluate for Thompson Warehouse

→

Checking time series characteristics, making time series analysis, find the best model, and evaluate for Nickolson Warehouse

↓

Calculate the bussiness impact using bussiness metric

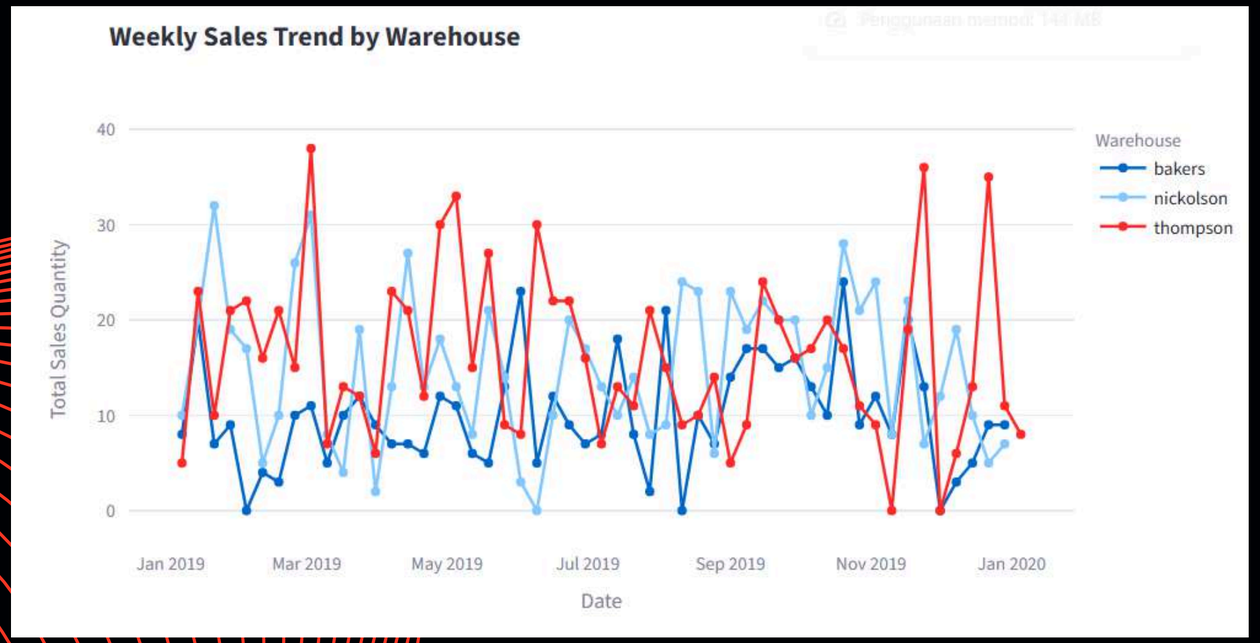
↓

Giving some recomendation based on model result and business metric

↓

Deploy app and dashboarding using Streamlit

# PROJECT BACKGROUND

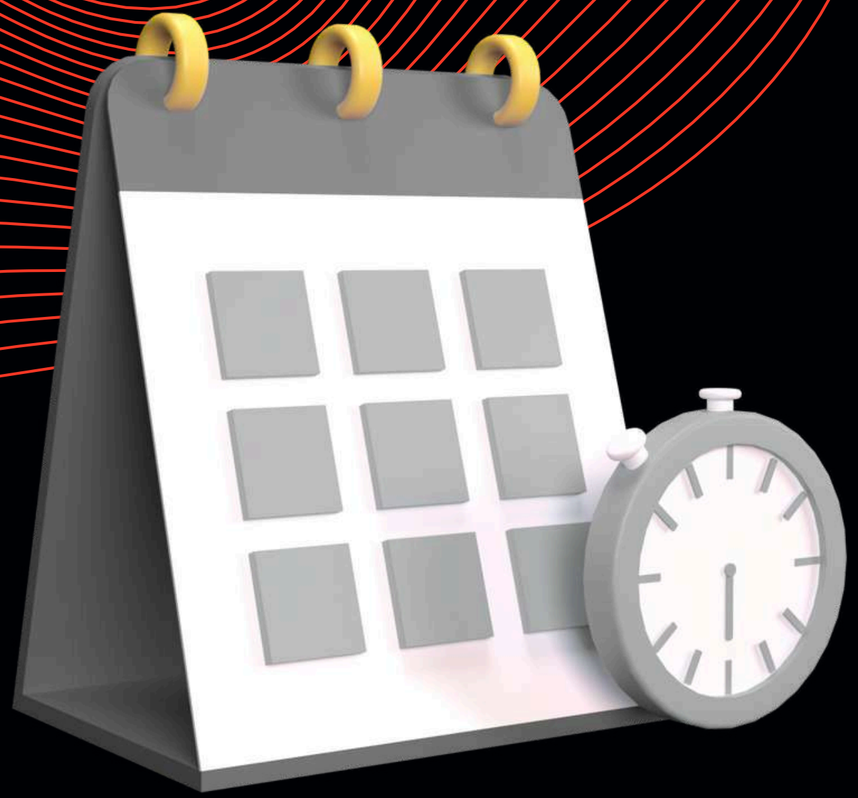



In the electronic retail business, weekly sales are highly volatile — with frequent zero-sales days, sudden spikes from promotions, and occasional supply delays.

When demand forecasting is inaccurate, the impact can be significant:

- **Overstock** → increased storage costs and reduced profit margins.
- **Stockout** → lost sales and decreased customer satisfaction.

Since procurement is managed on a weekly basis, a warehouse-specific weekly forecasting model (for Nickolson, Thompson, and Bakers) was developed.



# OBJECTIVES

- To build an **accurate machine learning / time series–based forecasting model** for predicting weekly demand.
- To use the forecast results to generate **business recommendations** — identifying **when to increase stock, when to hold inventory**, and how to **optimize warehouse costs** effectively.

# DATA UNDERSTANDING

Source : Kaggle - Transactional Retail Dataset of Electronics Store

Data used: Transactional weekly sales of 3 warehouses (Bakers, Thompson, and Nickolson)

Cover transaction from Jan - Dec 2019

Exogenous Features: coupon discount, order price, seasonal event (payday)

# MODEL APPROACH

🧠 Methodology / Modeling Approach

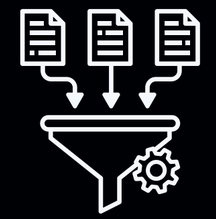
- **SARIMAX** — captures seasonal components and external (exogenous) effects.
- **Prophet** — provides flexibility in modeling weekly trends and special events.
- **Hybrid Models** (Prophet + XGBoost / SARIMAX) — combines statistical and machine learning methods to correct residuals and minimize forecast error.

**Process Overview**:
Data preprocessing → Transformation (Yeo-Johnson) → Hyperparameter tuning → Model evaluation.

# DATA PREPROCESSING

the initial step in develop model aimed at cleaning and preparing the dataset for further exploration.

500 rows
16 columns



Preprocessing

479 rows
17 columns

1. Changing Data Types

**Date (Object) >> Datetime**

2. Extract Quantity

**Extract quantity from ‘shopping_cart’**

3. Checking Duplicate

**No duplicate data found**

4. Missing Value Handling

**20 missing value on ‘date’ and 1 in ‘latest_customer_review’ >> Drop it**

5. Outlier Handling

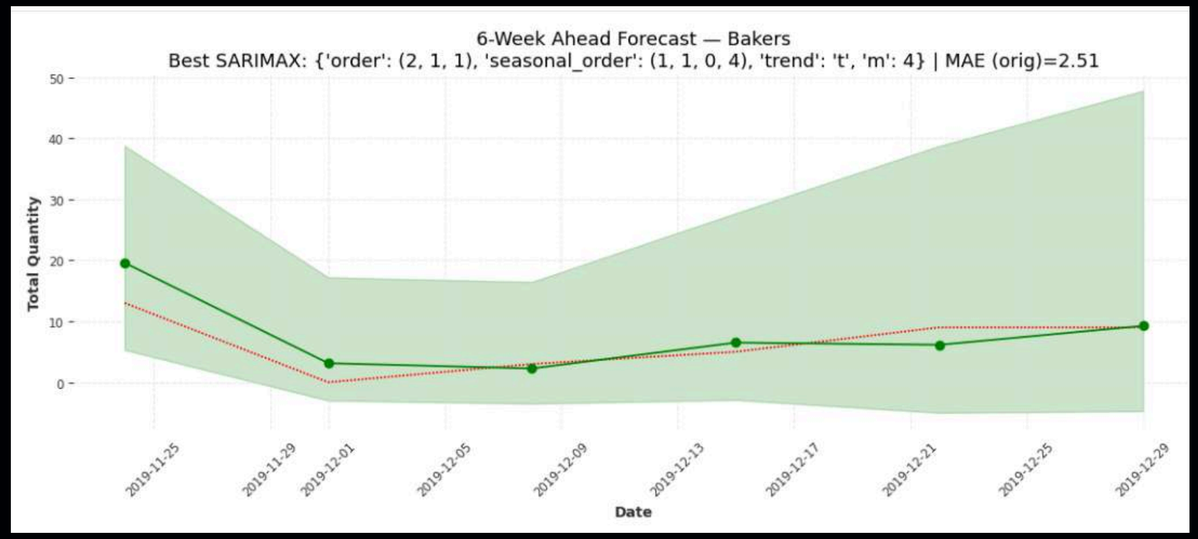
**found in columns ‘order_price’, ‘delivery_charges’, ‘order_total’, and ‘distance_to_nearest_warehouse’ . Handling with Capping IQR method**

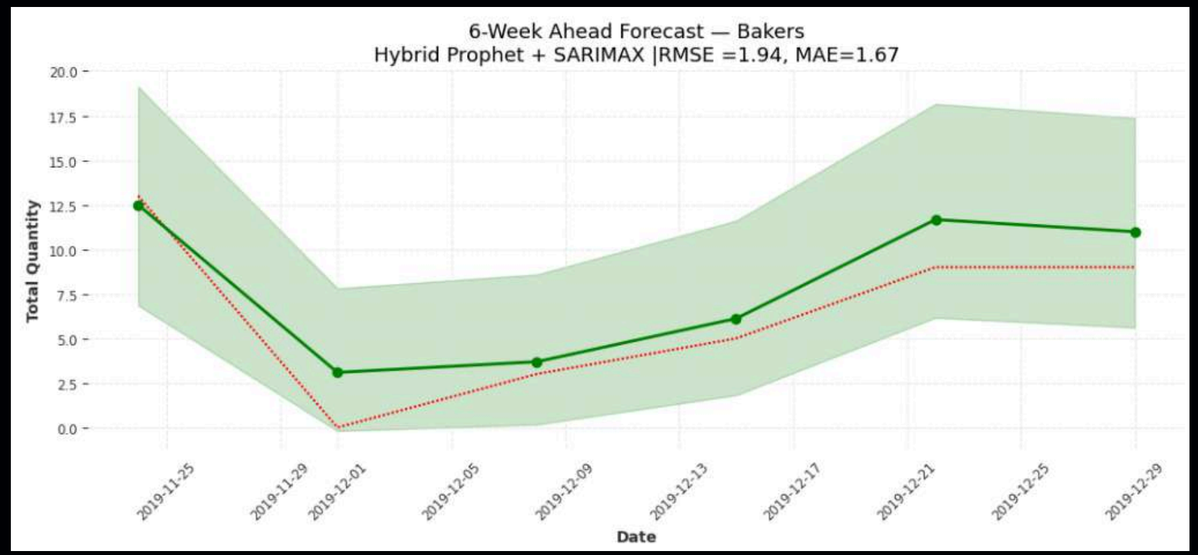
6. Making Weekly Time Series

**Group by ‘date’ with ‘nearest_warehouse’ and making weekly time series for each warehouse**

# MODEL PERFORMANCE

# SARIMAX VS HYBRID PROPHET + SARIMAX ON BAKERS WAREHOUSE




Actual (Test Dataset)

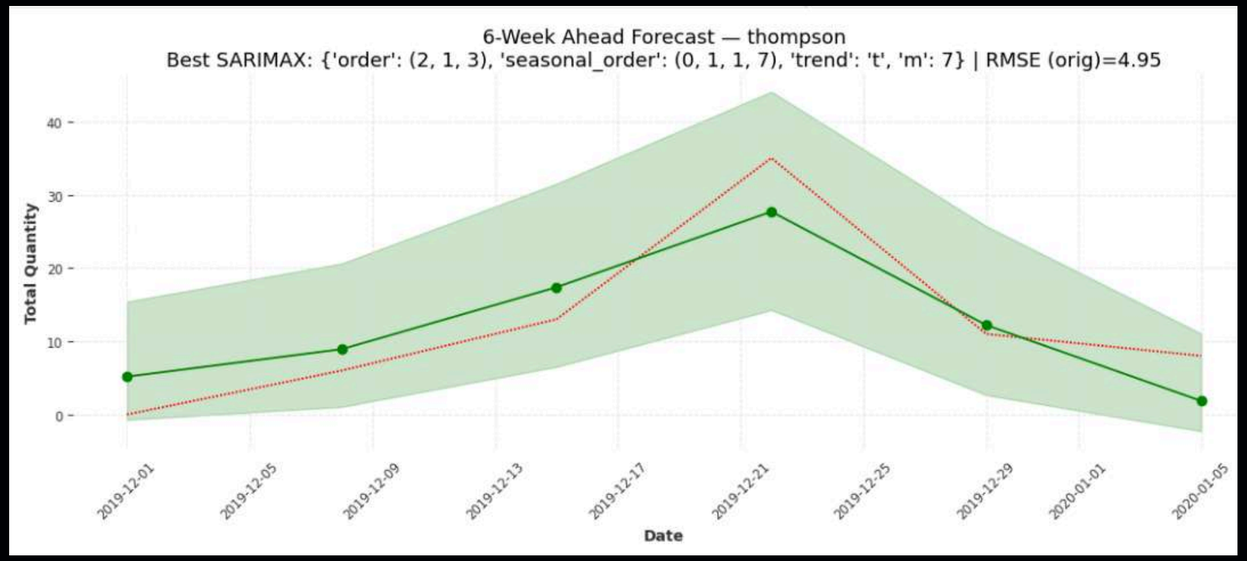
Forecast Result

| Model | MAE | RMSE |
|---|---|---|
| SARIMAX | 2.51 | 3.26 |
| HYBRID PROPHET + SARIMAX | 1.67 | 1.94 |

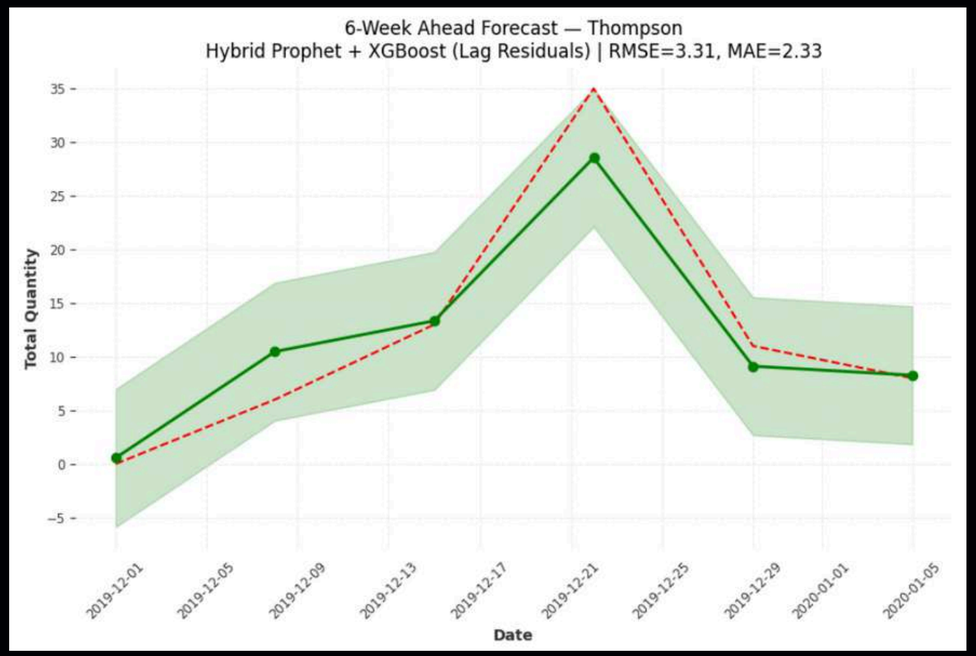
The hybrid model has higher prediction accuracy. Visually, the green line (forecast result) in the hybrid model aligns more closely with the red line (actual data), especially during periods of demand fluctuation.

👉 **Conclusion**: **The Hybrid Prophet + SARIMAX** model is more effective in capturing short-term trends and seasonal patterns, making it **more suitable for demand forecasting at Bakers Warehouse**.

# SARIMAX VS HYBRID PROPHET + XGBOOST (LAG RESIDUAL) ON THOMPSON WAREHOUSE



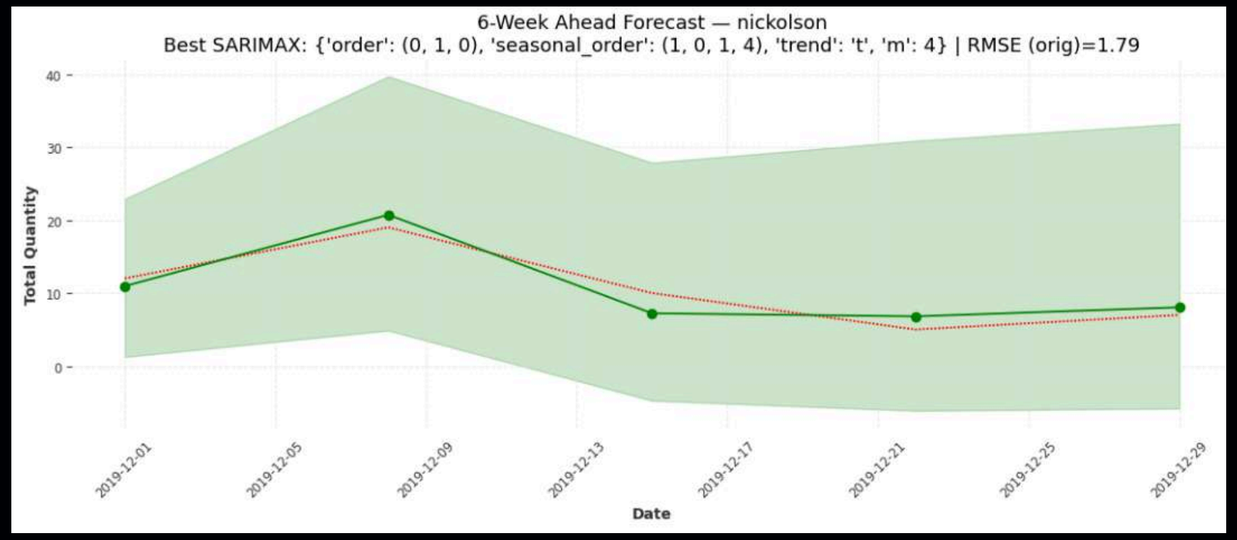

| Model | MAE | RMSE |
|---|---|---|
| SARIMAX | 4.51 | 4.94 |
| HYBRID PROPHET + XGBOOST | 2.33 | 3.31 |

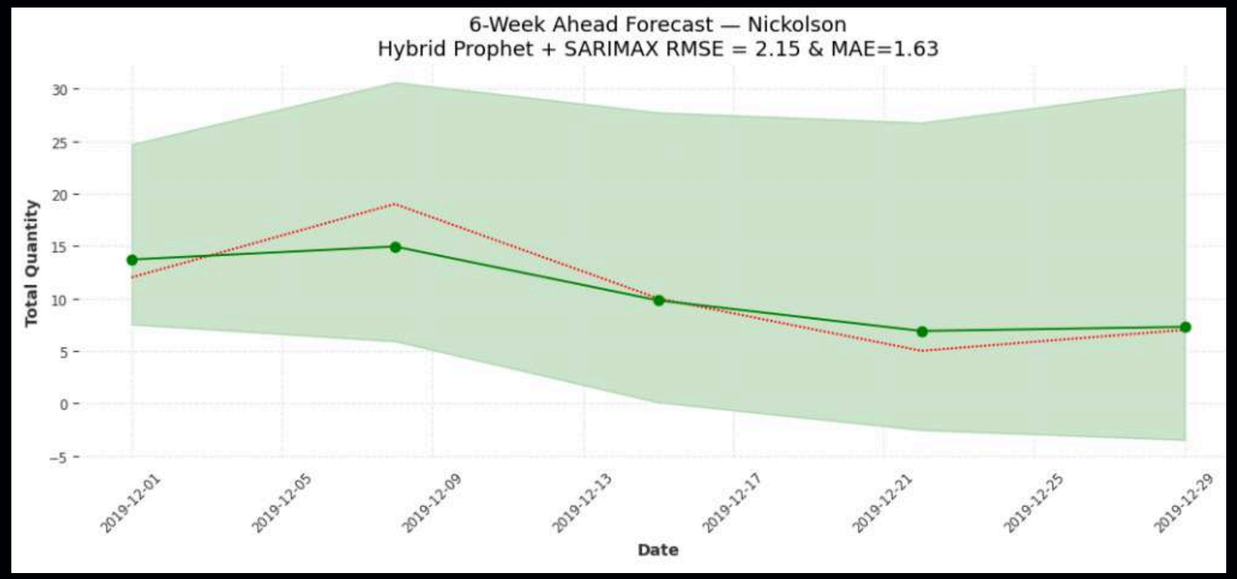



Hybrid Prophet + XGBoost (Lag Residual) model delivers a significant improvement over the standalone SARIMAX model. The hybrid model achieves MAE and RMSE, notably lower than SARIMAX's , indicating stronger predictive accuracy and stability. Visually, the hybrid model's forecast (green line) follows the actual demand trend (red line) more closely, especially around peak demand periods.

👉 Conclusion: **The Hybrid Prophet + XGBoost model effectively captures both trend and nonlinear residual patterns**, making it more reliable for demand forecasting in Thompson Warehouse.

# SARIMAX VS HYBRID PROPHET + SARIMAX ON NICKOLSON WAREHOUSE




Actual (Test Dataset)

Forecast Result

| Model | MAE | RMSE |
|---|---|---|
| SARIMAX | 1.67 | 1.79 |
| HYBRID PROPHET + SARIMAX | 1.63 | 2.15 |

Both models – SARIMAX and Hybrid Prophet + SARIMAX – deliver strong and consistent forecasting performance for Nickolson Warehouse. The hybrid model slightly improves accuracy. Visually, both models follow the actual demand trend closely, with minimal deviation across the forecast horizon.

👉 Conclusion: While both models perform comparably well, **the Hybrid Prophet + SARIMAX provides slightly better precision**, making it suitable for stable demand forecasting in Nickolson Warehouse.

# MODEL EVALUATION

| Warehouse | Model | MAE | RMSE |
|---|---|---|---|
| Bakers | SARIMAX | 2.51 | 3.26 |
| | HYBRID PROPHET + SARIMAX | 1.67 | 1.94 |
| Thompson | SARIMAX | 4.51 | 4.94 |
| | HYBRID PROPHET + XGBOOST | 2.33 | 3.31 |
| Nickolson | SARIMAX | 1.67 | 1.79 |
| | HYBRID PROPHET + SARIMAX | 1.63 | 2.15 |



# BUSINESS IMPACT

# BUSINESS IMPACT ANALYSIS

| Warehouse | Avg. Weekly Order | Avg. Order Value | MAE | Avg. Weekly Cost | Avg. Monthly Cost (Error Impact) | Insights |
|---|---|---|---|---|---|---|
| Bakers | 6.5 | Rp 1,200,000 | 1.67 | Rp 289.952 | Rp 1.246.795 | Model performs well overall but struggles slightly during sudden demand shifts |
| Thompson | 12.17 | Rp 1,500,000 | 2.33 | Rp 1.181.923 | Rp5.082.267 | Reflects high demand volatility and potential inefficiencies in inventory management. |
| Nickolson | 10 | Rp 1,000,000 | 1.63 | Rp408.722 | Rp1.757.506 | Indicates stable demand patterns and high model reliability. |

*Disclaimer: Average Order Value figures are hypothetical and used only for illustrative analysis*
*Under-prediction Cost : 25% of order value (lost sales) Over-prediction Cost : 10% of order value (holding cost)*

# BUSINESS IMPACT USING BUSINESS METRIC

| Warehouse | Forecast Accuracy (%) | FCR (%) | CPO (Rp) | Monthly Cost (Operational) (Rp) | ROI (%) | Insights |
|---|---|---|---|---|---|---|
| Bakers | 83.7 | 2.85 | 28.451 | 1.223.372 | 58.7 | Prioritize model retraining or hybrid parameter adjustment to improve short-term forecast precision. |
| Thompson | 80.9 | 3.35 | 50.268 | 2.629.838 | 126.2 | Model delivers strong business value; fine-tuning forecasts could further reduce cost volatility |
| Nickolson | 74.2 | 4.51 | 54.097 | 1.512.021 | 72.6 | Model performs well with consistent accuracy; focus can shift toward optimizing ROI through inventory cost control |

*Insight: Overall, hybrid models significantly improved forecast-driven ROI across warehouses, especially in Thompson*


# RECOMMENDATION

# RECOMMENDATIONS FOR EACH WAREHOUSE

## BAKERS WAREHOUSE

- Retrain model monthly with updated data to adapt to short-term market changes (e.g., promotions or holiday spikes).
- Implement early-warning thresholds for forecast deviation > 15% to trigger proactive procurement review.
- Integrate forecast output with automated reorder systems to reduce manual lag in restocking decisions.

## THOMPSON WAREHOUSE

- Combine the hybrid forecast with an automated alert system that notifies the inventory team when forecast error exceeds ±3 units.
- Coordinate inter-warehouse delivery schedules, especially with nearby facilities, to balance stock levels during localized demand surges.
- Review and adjust supplier procurement policies to make them more flexible to fluctuating weekly demand volumes.

## NICKOLSON WAREHOUSE

- Maintain the current Prophet+SARIMAX hybrid model, but perform seasonal parameter fine-tuning quarterly to sustain performance stability.
- Optimize inventory costs by reducing safety stock by 5–10% without compromising the service level.
- Use forecast results to automate minimum order thresholds in the ERP system, improving purchasing efficiency and stock planning.

# Cross-Warehouse Recommendation

## CENTRALIZED FORECASTING DASHBOARD

**Develop a centralized forecasting dashboard** (auto-refreshing weekly) to monitor accuracy and cost performance across all warehouses.

## AUTOMATED MODEL SELECTION

**Implement automated model selection**, allowing each warehouse to dynamically adopt the best-performing hybrid configuration (e.g., Prophet+SARIMAX or Prophet+XGBoost).

## INTEGRATE FORECAST OUTPUTS

**Integrate forecast outputs into procurement and distribution scheduling**, aligning supply chain operations with predicted regional demand to minimize overstock and stockout risks.

# DASHBOARD

Share

## Smarter Stocking Starts Here — AI-Powered Sales Forecasting Dashboard

**Smarter Stocking Analytics**

Home

Dashboard

Performance

Business Impact

Contact

Developed by Rahma Anggana Rarastyasa © 2025

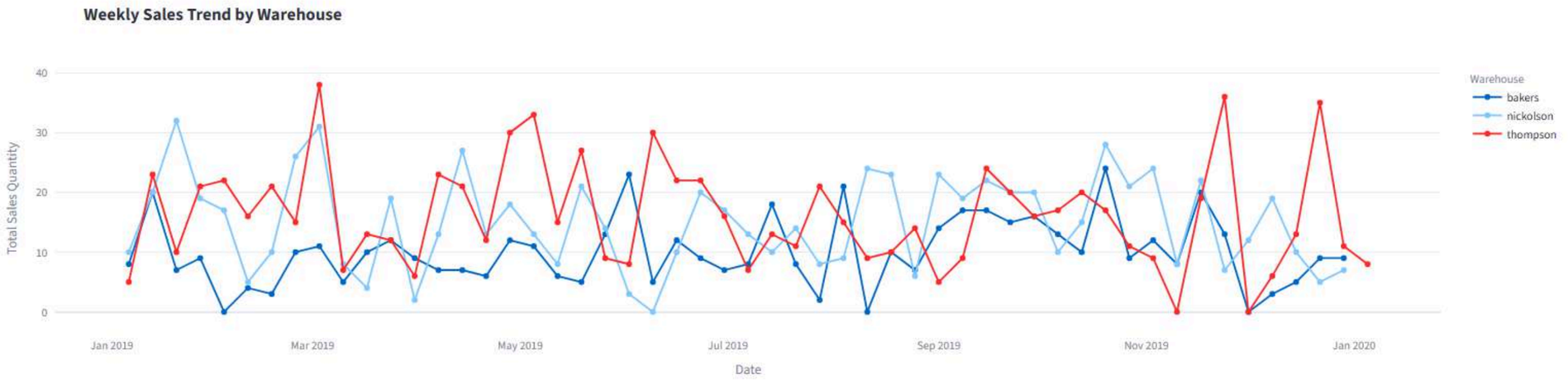
### Company Overview

Our retail electronics company operates through **three main warehouses — Nickolson, Thompson, and Bakers.**
Each warehouse serves different regional markets, playing a critical role in ensuring timely delivery and efficient stock management.

However, fluctuating demand patterns and supply chain inconsistencies across regions have made inventory forecasting increasingly complex.
To address this, the company leverages **AI-driven time series forecasting** to improve accuracy, optimize stock levels, and enhance operational decisions.

> **Data Source:** The dataset used for this project is derived from the *Transactional Retail Dataset of Electronics Store on Kaggle*, which provides detailed transactional records and warehouse information used to simulate real-world retail operations.

### Weekly Sales Trends per Warehouse




*Developed using Strimlit. Access heres: https://time-series-forecasting-retail-electronic.streamlit.app/*

# DOCUMENTATION

**Kaggle Source:**
TransactionalRetailDataset

**Dataset:**
Dataset Retail Electronic

**Colab Notebook:**
Notebook

**Developed App**
Developed App Streamlit


THANK YOU :)