Final Project Data Science

Jihad Akbar

📞 +62 813 3232 6785

in https://www.linkedin.com/in/jihadakbr

✉ jihadakbr@gmail.com

TABLE OF CONTENTS

01

Introduction

Jihad Akbar

linkedin.com/in/jihadakbr

"Transforming complexity into **clarity**, uncovering **insights**, and turning data into actionable strategies through **collaboration**."

Experience

Now–Oct 2024	**Machine Learning Annotator** PT Nomura Research Institute Indonesia
Sep–Feb 2024	**Data Annotator** CVAT.ai Corporation

Education

| Now–Oct 2024	**Data Science Bootcamp** dibimbing.id
2020–2016	**Bachelor of Science in Physics** Universitas Sebelas Maret

02

Previous projects

What I Have Learned in the Bootcamp

Churn Analysis (Feb-Mar 2025)

- Predicted customer churn using feature importance analysis and optimized recall (91%) for imbalanced data via PR AUC and threshold tuning, achieving top XGBoost performance (Precision: 70%, F1: 79% at a threshold of 0.45).
- Highlighted total transactions as the top churn predictor and deployed a Streamlit app for real-time predictions, delivering actionable insights for stakeholders.

github.com/jihadakbr/churn-analysis
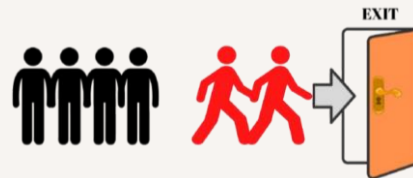
New York City Taxi Trip Duration (Oct-Dec 2024)

- Predicted NYC taxi trip durations using 2016 data by defining business objectives, preprocessing data, engineering features like distance and speed, and addressing NYC-specific location constraints.
- Built scalable machine learning workflows using reusable transformer components with OOP principles in scikit-learn pipelines, achieving an RMSLE score of 0.55 with a LightGBM Regressor.

github.com/jihadakbr/new-york-city-taxi-trip-duration

FutureSight: Revolutionizing E-Commerce
with Data-Driven Forecasting

03

Main Project

MAIN PROJECT OUTLINE

Project Background

Olist, one of Brazil's largest e-commerce platforms, faced critical challenges despite its growing sales:

⚠️ Unpredictable Order Volumes: Sudden fluctuations make it challenging to manage resources effectively.

⚠️ Revenue Volatility: Shifting revenue trends are hard to track, hindering strategic planning.

Daily Orders Over Time

Ghibli–style Data Scientist Illustration created by ChatGPT

Disclaimer: *The following story is fictional and created solely to illustrate the challenges addressed by this project.*

Business Objective
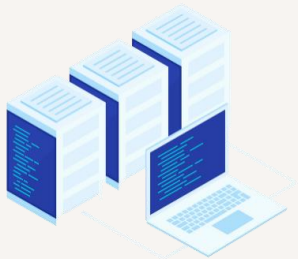
The Mission: Predict the Future, Optimize the Present

The primary goal of this project was to build a state-of-the-art time series forecasting system that accurately predicts key business metrics. This mission includes:

- 📈 **Accurately forecast order volumes** to optimize inventory and resource management, minimizing stockouts and overstocking.
- 📈 **Forecast revenue trends** to provide a clear financial roadmap, supporting strategic planning and long-term growth.

This project, **FutureSight**, delivered an AI-driven forecasting engine that leverages advanced analytics to address these business-critical need.

*Powered by Olist's real transaction data, comprising over 100,000 orders across Brazil's diverse regions, collected from [Kaggle](#).
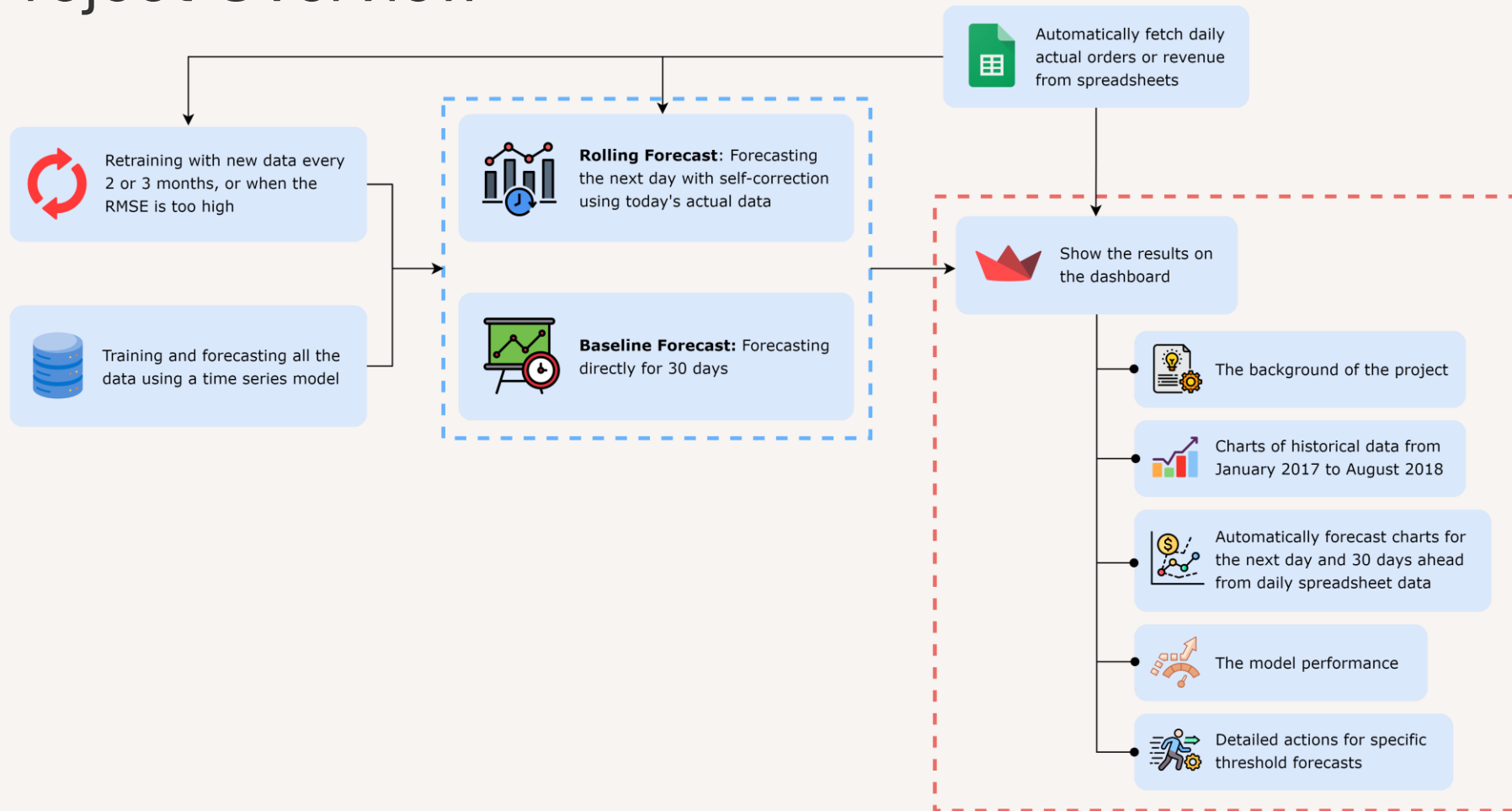
Data Understanding

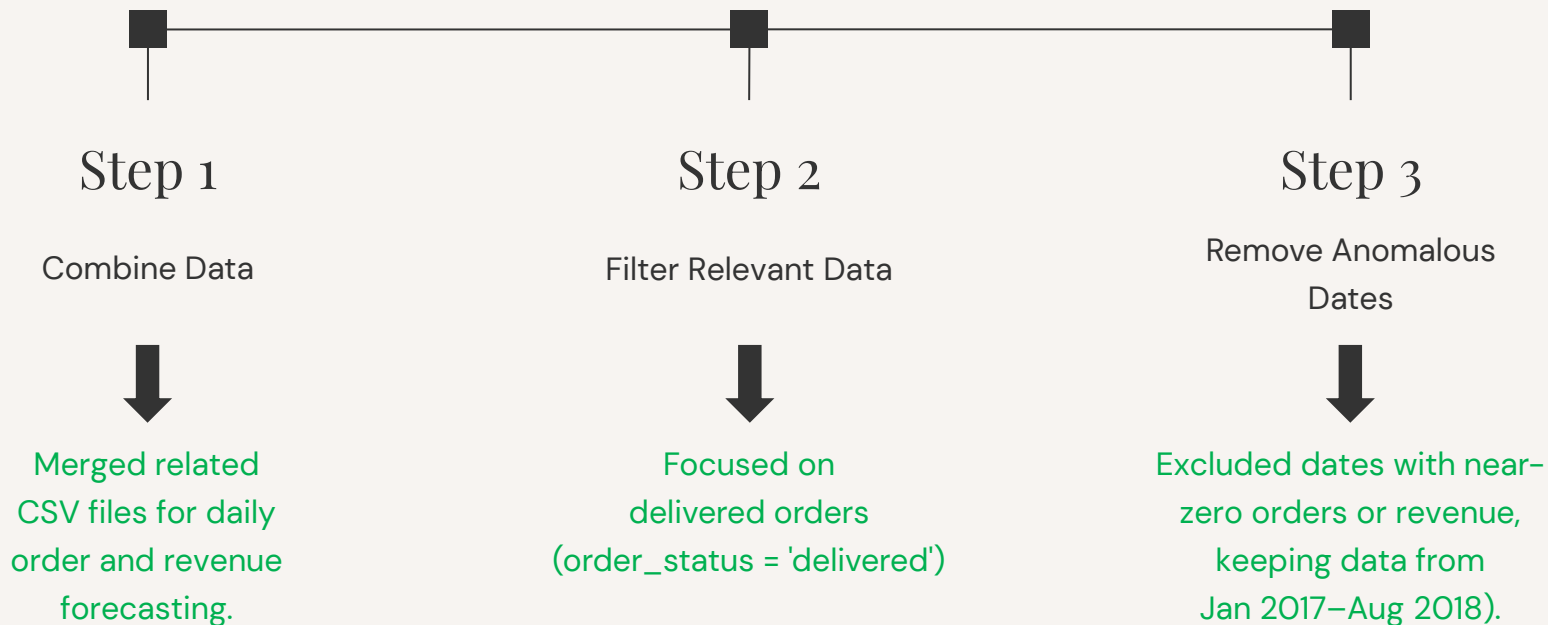Collected from Kaggle – Brazilian E-Commerce Public Dataset by Olist.

Approximately 100,000 orders from Oct 2016 to Sep 2018 across Brazil.

Includes order details, payments, customer and seller information, product data, shipping logistics, and customer reviews.

Project Overview

Automatically fetch daily actual orders or revenue from spreadsheets

Retraining with new data every 2 or 3 months, or when the RMSE is too high

Training and forecasting all the data using a time series model

Rolling Forecast: Forecasting the next day with self-correction using today's actual data

Baseline Forecast: Forecasting directly for 30 days

Show the results on the dashboard

The background of the project

Charts of historical data from January 2017 to August 2018

Automatically forecast charts for the next day and 30 days ahead from daily spreadsheet data

The model performance

Detailed actions for specific threshold forecasts

Data Preprocessing

Step 1

Combine Data

⬇

Merged related CSV files for daily order and revenue forecasting.

Step 2

Filter Relevant Data

⬇

Focused on delivered orders (order_status = 'delivered')

Step 3

Remove Anomalous Dates

⬇

Excluded dates with near-zero orders or revenue, keeping data from Jan 2017–Aug 2018).

Data Preprocessing

Step 4

Exclude Outlier Dates

Removed end-of-period dates with unusually low activity, likely due to in-transit orders.

Step 5

Convert Data Types

Reformatted 'date' to date type

Step 6

Address Outliers

Identified Black Friday 2017 as an outlier. Handled as an exogenous variable in the model.

Data Preprocessing

Step 7

Analyze Seasonality

Detected weekly and yearly patterns through ACF/PACF analysis.

Step 8

Assess Stationarity

Data approached non-stationarity (p-value ~0.5).

Step 9

Save Clean Data

Exported a clean CSV dataset for use in the dashboard.

Data Preprocessing Result

~100,000
Transaction Data

Sep 2016 – Sep 2018
Timeframe

Data Preprocessing

~600
Unique Transaction Data

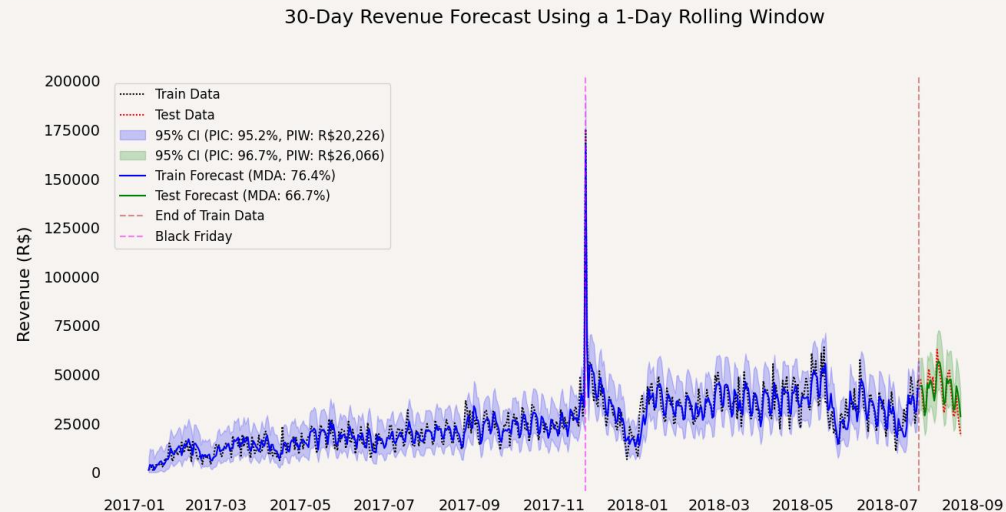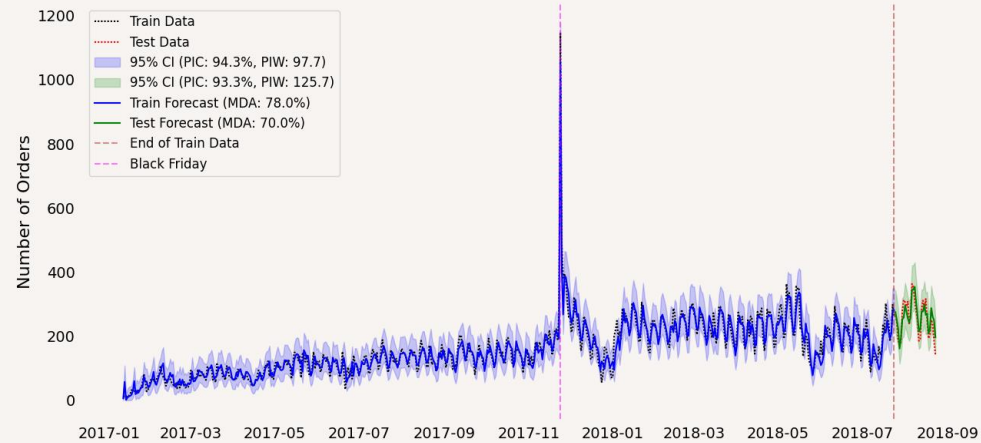**Jan 2017 – Aug 2018**
Timeframe

Findings and Results

Insight

- The actual and predicted daily orders and revenue exhibit a similar pattern throughout the year, with a notable peak around the Black Friday event (2017-11-24).
- The average prediction error (RMSE) is 30 for daily orders and R$5,500 for daily revenue.
- The prediction interval width (PIW) ranges approximately 126 orders (e.g., 100–226) and R$26,000 for daily revenue (e.g., R$10,000–36,000).
- The prediction coverage is 93% for orders and 97% for revenue, indicating that 93% of actual daily orders and 97% of actual daily revenue fall within their respective predicted ranges.
- 70% of trend predictions (up/down compared to the previous day) are correct for daily orders, while 67% of trend predictions are correct for daily revenue.

	Daily Order Volume	Daily Revenue
Avg. Prediction Error (RMSE)	30	R$5,458

30-Day Order Volume Forecast Using a 1-Day Rolling Window

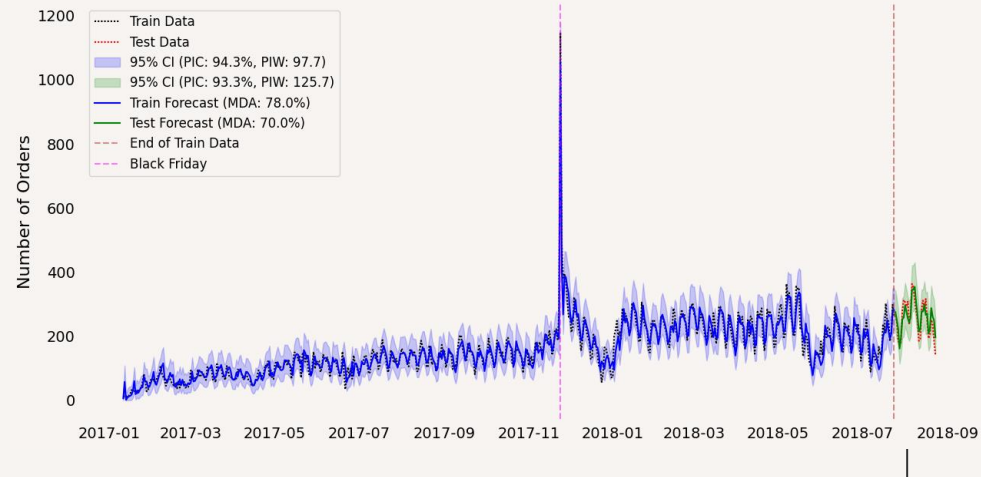### 30-Day Revenue Forecast Using a 1-Day Rolling Window
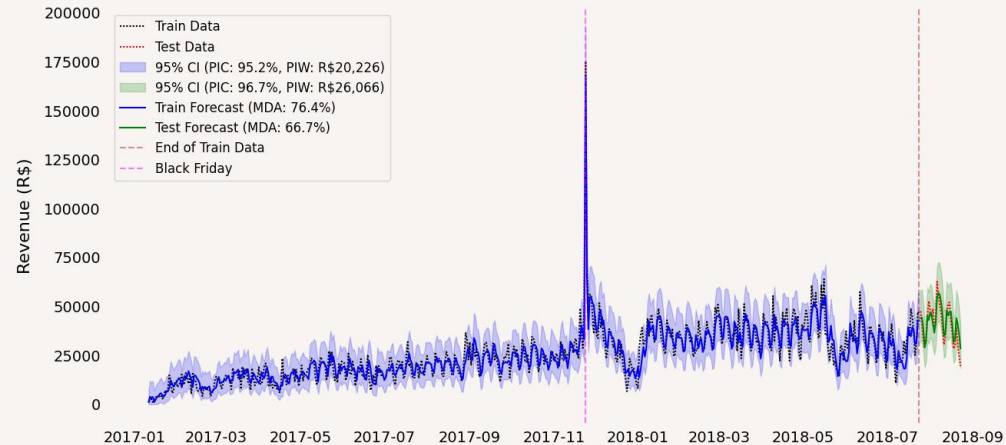
Findings and Results

Insight (2)

- The seasonal ups and downs in this dataset were caused by two main patterns: weekly seasonality and yearly seasonality.

- **Weekly seasonality** happens because customer behavior changes depending on the day of the week. For example, customers often place orders on Saturdays or Sundays, but since the dataset tracks the "delivered" status, these orders are recorded as deliveries during the weekdays.

- **Yearly seasonality** happens because of major shopping events throughout the year. For instance, the Black Friday event in late November causes a sharp increase in both orders and revenue, followed by a noticeable drop afterward.

	Daily Order Volume	Daily Revenue
Avg. Prediction Error (RMSE)	30	R$5,458

30-Day Order Volume Forecast Using a 1-Day Rolling Window

30-Day Revenue Forecast Using a 1-Day Rolling Window
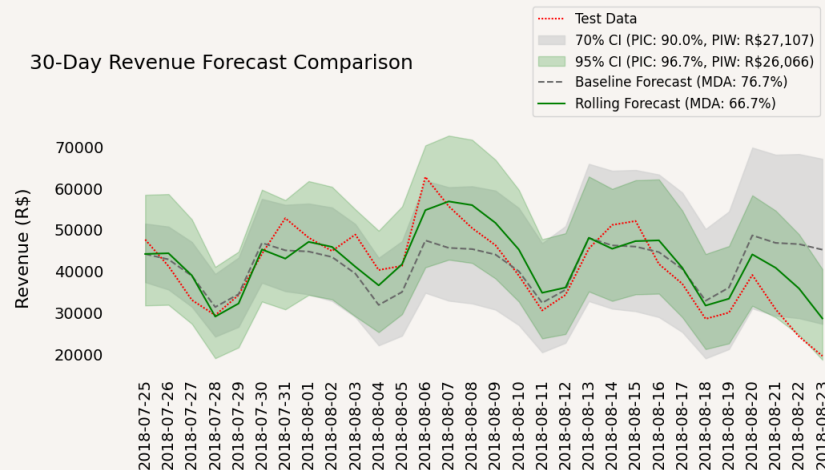
Findings and Results

Insight

- The baseline forecast provides a rough prediction for orders or revenue over the next 30 days, giving a broad overview of future trends. In contrast, the 1–day rolling window forecast leverages today's actual data to predict tomorrow's values, making it significantly more accurate.

- As a result, while the baseline forecast is less precise than the 1–day rolling window forecast, it serves as a useful tool for initial preparation and planning.


30-Day Order Volume Forecast Comparison
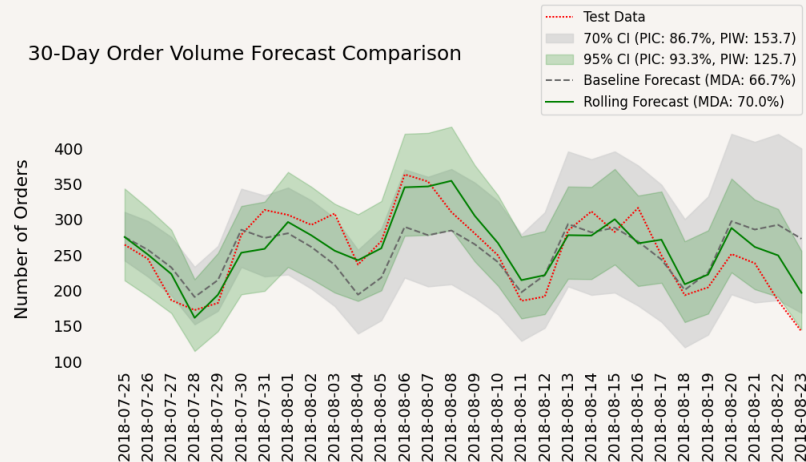
30-Day Revenue Forecast Comparison

Findings and Results

Insight

- This monitoring dashboard helps identify whether the predictions are too high or too low compared to the actual data (normalized to zero).

- The gray line represents the actual number of orders or revenue, serving as a benchmark for comparison.

Daily Olist Orders: Forecast vs Actual (August 2018)

Daily Olist Revenue: Forecast vs Actual (August 2018)
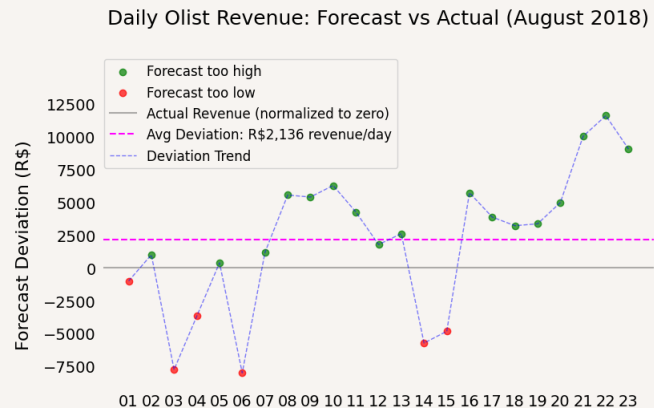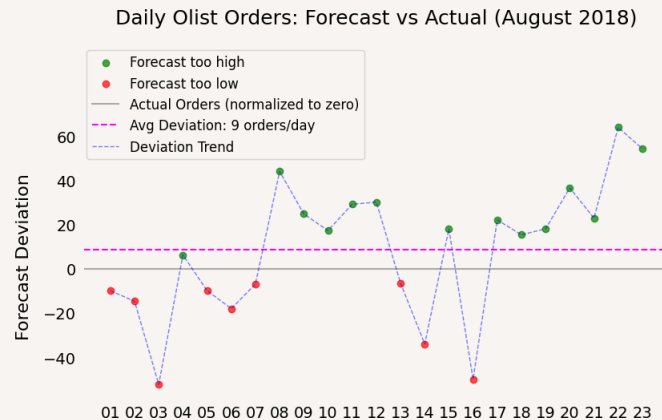
Findings and Results

Daily Order Volume

Business Impact Analysis

- Average Daily Orders: 255
- Average Deviation: 30 orders (11.9% of average)

Insight

- The forecasting model deviates on average by 30 orders per day, which is roughly 12% of the daily volume.
- This level of deviation indicates moderate accuracy but with room for improvement to reduce costs linked to forecast errors.

Findings and Results

Daily Order Volume

Error Cost Analysis (Local Currency R$) [Using hypothetical numbers]

- Under-prediction Cost: R$50 per order (lost profit)
- Over-prediction Cost: R$30 per order (excess inventory)
- Daily Average Cost: R$958
- Monthly Error Cost Impact: R$28,728

Error Cost Analysis (USD $)

- Under-prediction Cost: $9 per order
- Over-prediction Cost: $5 per order
- Daily Average Cost: $165
- Monthly Error Cost Impact: $4,945

Insight

- The forecasting errors translate into a significant daily cost of nearly R$1,000, accumulating to nearly R$29,000 monthly.
- Since under-prediction costs are higher, errors causing stockouts (lost sales) are more expensive than excess inventory costs.

Findings and Results

Daily Order Volume

Return on Investment (ROI) [Using hypothetical numbers]

- Investment cost: R$40,000 (one-time)
- Ongoing monthly cost: R$1,500
- Target error reduction: 25%
- Projected monthly savings: R$7,182
- Annual net gain: R$68,184
- ROI after 1 year: 170.5%
- Breakeven point: ~5.6 months

Insight

- Investing R$40k in improving the forecast model with an ongoing cost of R$1,500/month is highly financially justified.
- Achieving a 25% reduction in forecasting errors would generate substantial savings and lead to a very attractive ROI (>170%) within just over 5 months. This is a strong business case for funding forecasting improvements.

Findings and Results

Daily Revenue

Business Impact Analysis

- Average Daily Revenue: R$40,787
- Average Deviation: R$5,458 revenue (13.4% of average) ($939 per day)

Insight

- The forecast error on revenue is about 13.4% of the daily revenue, meaning the model deviates on average by R$5,458 daily. This is a substantial variation given the revenue magnitude and directly impacts profitability and costs.

Findings and Results

Daily Revenue

Error Cost Analysis (Local Currency R$) [Using hypothetical numbers]

- Under-prediction Cost: R$50 per R$100 error (lost profit)
- Over-prediction Cost: R$30 per R$100 error (excess inventory)
- Daily Average Cost: R$1,680
- Monthly Error Cost Impact: R$50,415

Error Cost Analysis (USD $)

- Under-prediction Cost: $9 per R$100 error
- Over-prediction Cost: $5 per R$100 error
- Daily Average Cost: $289
- Monthly Error Cost Impact: $8,677

Insight

- The costs associated with forecasting errors (lost profit + inventory holding) sum to over R$1,600 daily, leading to a monthly loss exceeding R$50,000. Under-prediction carries a higher penalty, reflecting lost sales/profit impact.

Findings and Results

Daily Revenue

Return on Investment (ROI) [Using hypothetical numbers]

- One-Time Investment: R$75,000 (~$12,909 USD)
- Ongoing Monthly Cost: R$3,000 (~$516 USD)
- Target Error Reduction: 25%
- Projected Monthly Savings: R$12,604 (~$2,169 USD)
- Annual Net Gain: R$115,245 (~$19,836 USD)
- ROI After 1 Year: 153.7%
- Breakeven Period: 6.0 months

Insight

- Investing R$75k in improving the revenue forecast model—with a higher ongoing maintenance cost—is justified financially.
- Achieving a 25% error reduction delivers substantial monthly savings and a strong ROI of over 150% within a year, breaking even in just 6 months.

Recommendations

- Use updated daily order forecasts to **adjust supply and inventory plans**, ensuring products are available without overstocking or understocking.

- **Base inventory and revenue strategies** on daily updates to forecasts, ensuring alignment with actual sales and preventing overproduction or lost opportunities.

- **Regularly monitor forecast accuracy** on the dashboard, adjusting orders and inventory levels when predictions are off to minimize excess or shortage costs.

- **Investing in model improvement** is financially advantageous with strong ROI and quick breakeven.

Dashboard

To monitor daily order volume and revenue trends, I have provided a dashboard for the relevant stakeholders to access.

Link to the Dashboard: [Streamlit](#)

FutureSight: Revolutionizing E-Commerce with Data-Driven Forecasting

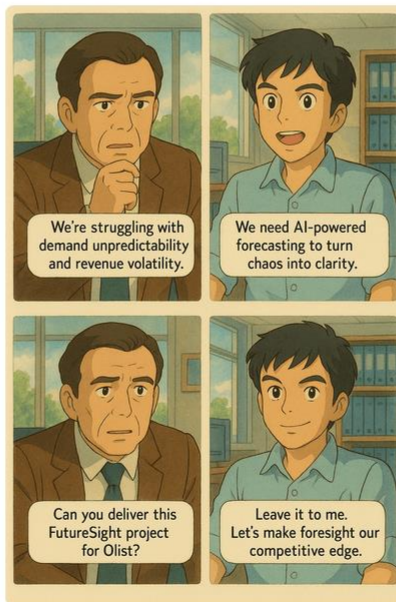📝 **FutureSight Analytics**

Welcome to FutureSight!

Disclaimer: The following story is fictional and created solely to illustrate the challenges addressed by this project.

🎬 The Background: A Business on the Brink

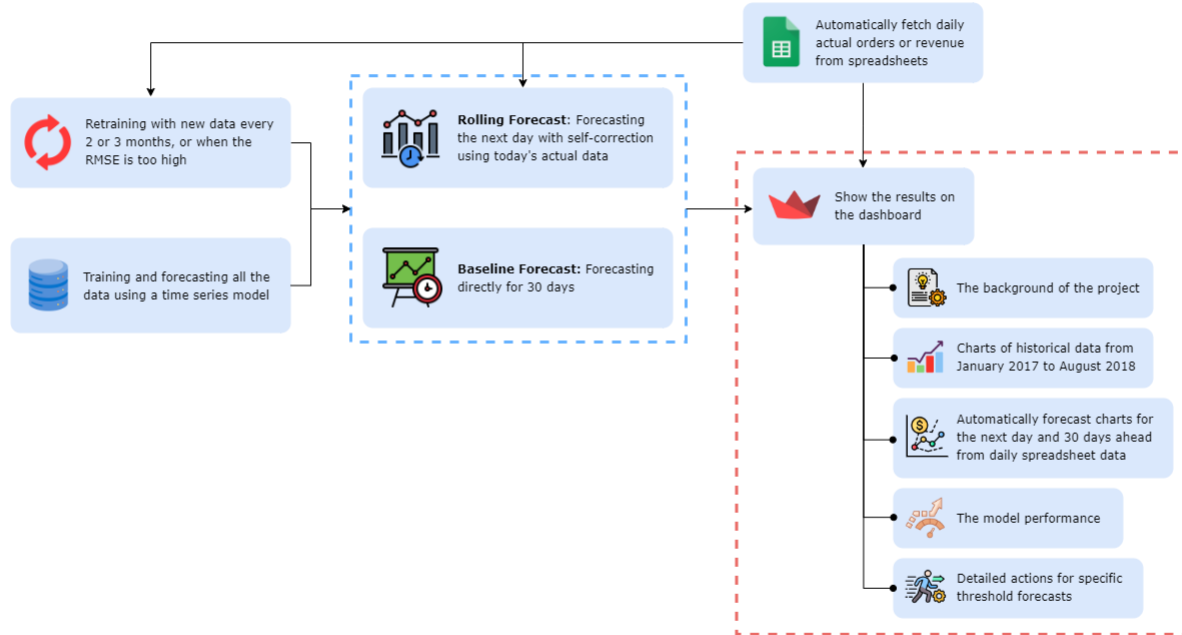Olist, one of Brazil's largest e-commerce platforms, faced critical challenges despite its growing sales:

⚠ **Unpredictable Order Volumes**: Sudden fluctuations make it challenging to manage resources effectively.

⚠ **Revenue Volatility**: Shifting revenue trends are hard to track, hindering strategic planning.

The CEO of Olist issued a bold challenge:

Ghibli-style Data Scientist Illustration created by ChatGPT

🎯 The Mission: Predict the Future, Optimize the Present

🔲 FutureSight Analytics

- 🏠 Home
- 🔍 Overview
- 📊 Dashboard
- 🎯 Performance
- 🛒 Order Volume
- 🏅 Revenue Trend
- 📞 Contact

Made with ❤️ by Jihad Akbar

🔍 Project Overview

Automatically fetch daily actual orders or revenue from spreadsheets

Retraining with new data every 2 or 3 months, or when the RMSE is too high

Rolling Forecast: Forecasting the next day with self-correction using today's actual data

Training and forecasting all the data using a time series model

Baseline Forecast: Forecasting directly for 30 days

Show the results on the dashboard

The background of the project

Charts of historical data from January 2017 to August 2018

Automatically forecast charts for the next day and 30 days ahead from daily spreadsheet data

The model performance

Detailed actions for specific threshold forecasts

Manage app

FutureSight Dashboard

📦 Next Day Orders Forecast

140

↓ -2 vs Today's Actual Orders

⚠️ Next Day Order Forecast Deviation
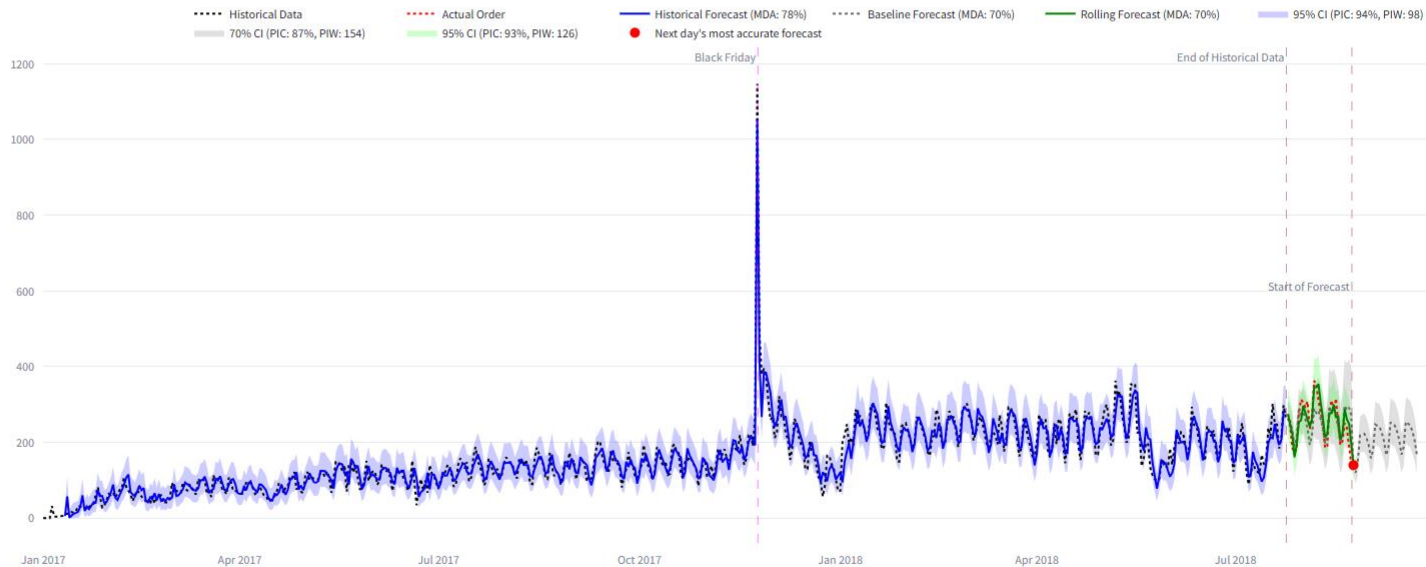
96 — 191

🏅 Next Day Revenue Forecast

$21,120

↑ $1,597 vs Today's Actual Revenue

⚠️ Next Day Revenue Forecast Deviation

$12,604 — $31,539

FutureSight Analytics

- 🏠 Home
- 🔍 Overview
- 📊 Dashboard
- 🎯 Performance
- 🛒 Order Volume
- 🏅 Revenue Trend
- 📞 Contact

Made with ❤️ by Jihad Akbar

30-Day Order Volume Forecast

Legend: Historical Data · Actual Order · Historical Forecast (MDA: 78%) · Baseline Forecast (MDA: 70%) · Rolling Forecast (MDA: 70%) · 95% CI (PIC: 94%, PIW: 98) · 70% CI (PIC: 87%, PIW: 154) · 95% CI (PIC: 93%, PIW: 126) · Next day's most accurate forecast

30-Day Revenue Trend Forecast

Manage app

FutureSight Analytics

- 🏠 Home
- 🔍 Overview
- 📊 Dashboard
- 🎯 Performance
- 🛒 Order Volume
- 🏅 Revenue Trend
- 📞 Contact

Made with ❤️ by Jihad Akbar

Share

🎯 Forecast Performance

Select Month:

August 2018

Daily Orders: Forecast vs Actual (August 2018)

—— Actual Orders (normalized to zero)   ● Forecast too high   ● Forecast too low   ---- Avg. Deviation: 9 orders/day   ---- Deviation Trend
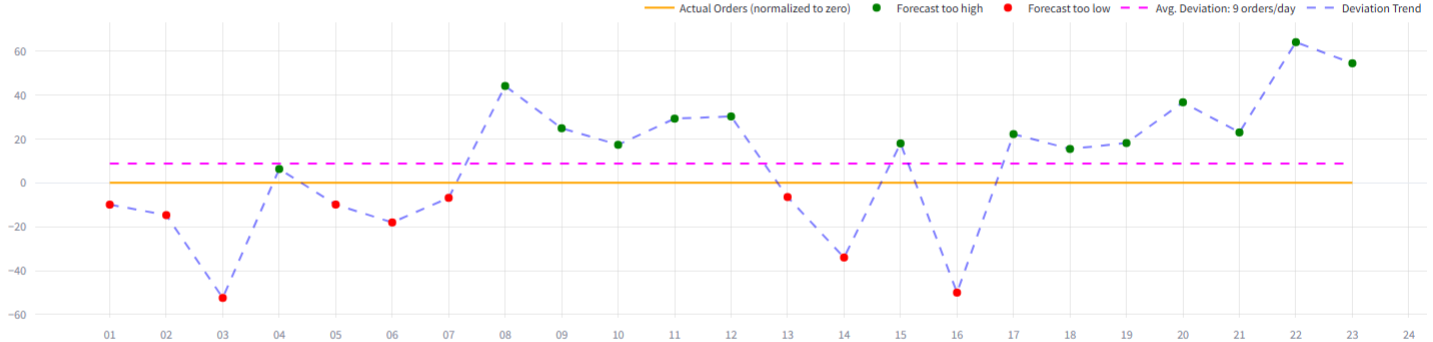
Select Month:

August 2018

Daily Revenue: Forecast vs Actual (August 2018)

enue (normalized to zero)   ● Forecast too high   ● Forecast too low   ---- Avg. Deviation: R$2,136 revenue/day   ---- Deviation Trend

0 1 : 18 : 48

Manage app

FutureSight Analytics

- 🏠 Home
- 🔍 Overview
- 📊 Dashboard
- 🎯 Performance
- 🛒 Order Volume
- 🏅 Revenue Trend
- 📞 Contact

Made with 🧡 by Jihad Akbar
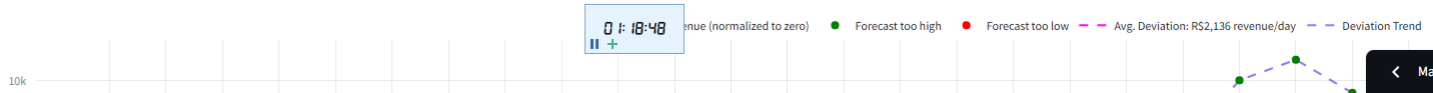
🛒 **Order Volume Forecasting**

30-Day Order Volume Forecast Comparison

Actual Order · · · · · Baseline Forecast (MDA: 70%) — — Rolling Forecast (MDA: 70%) — 70% CI (PIC: 87%, PIW: 154) 95% CI (PIC: 93%, PIW: 126) ● Next day's most accurate forecast
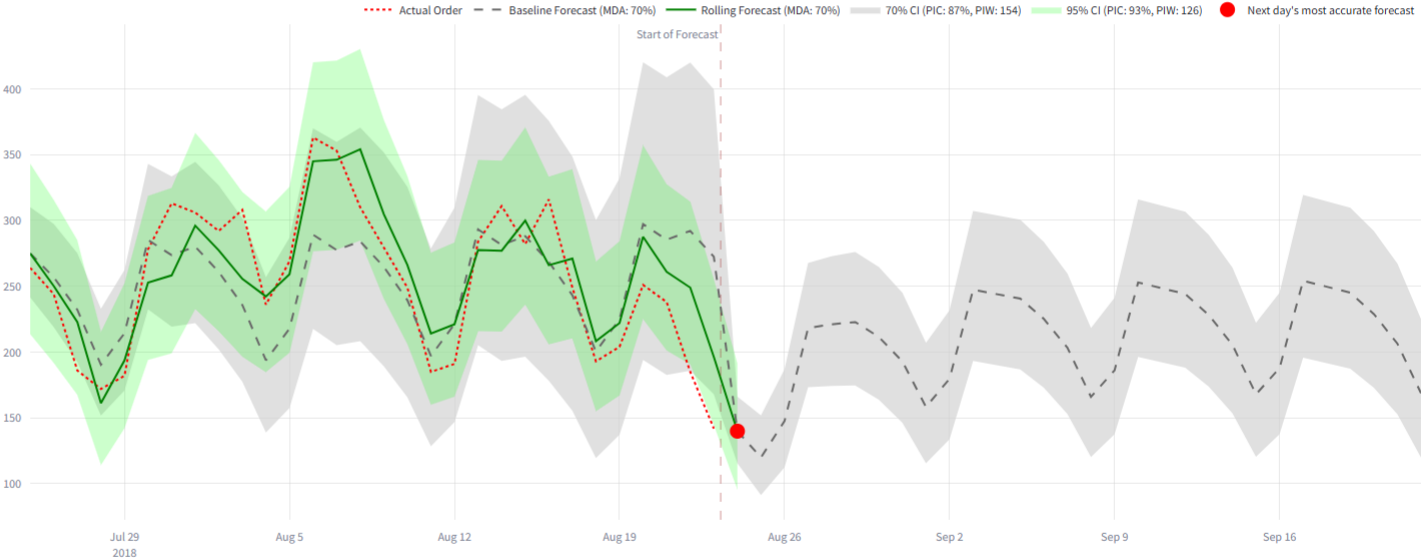
Expected Orders Volume Range

Next 30-day projections with possible variation

Date	Actual Orders	Expected Orders (Rolling)	Worst Case	Best Case

◄ Manage app

| 2018-07-25 | 47,600 | 44,133 | 37,328 | 51,452 |
2018-07-26	41,184	42,824	35,533	50,728
2018-07-27	33,017	38,842	31,376	47,026
2018-07-28	29,350	31,323	24,173	39,305
2018-07-29	34,233	34,430	26,530	43,256
2018-07-30	44,166	46,793	37,170	57,414

A Quick Look at Rolling Forecast:

- Typical expected volume: **R$41,559 revenue/day**
- Normal fluctuation range: **R$29,567–R$55,403 revenue/day**

Recommended Actions

Revenue Range	🗹 Forecast Condition	✅ Priority Action	⚡ Next Step
> R$50,000	High Revenue	Scale marketing campaigns	Upsell premium products/services
R$30,000 - R$50,000	Moderate Revenue	Maintain current ad spend	Run targeted promotions
< R$30,000	Low Revenue	Reduce non-essential costs	Analyze customer churn drivers

Time Series Model Performance

📊 Metric	🎯 Value	📋 Interpretation
Avg. Prediction Error (RMSE)	5,458	On average, predictions are off by R$5,458 revenue/day.
Prediction Width (PIW)	26,066	Prediction ranges are R$26,066 revenue wide (e.g., R$10,000–R$36,066).
Prediction Coverage (PIC)	97	97% of actual sales fall within the predicted range.
Trend Accuracy (MDA)	70	70% of trend predictions (up/down vs. yesterday) are correct.
Forecast Deviation	-1,454	Tends to underpredict by R$1,454 revenue
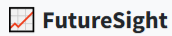
FutureSight Analytics

- 🏠 Home
- 🔍 Overview
- 📊 Dashboard
- 🎯 Performance
- 🛒 Order Volume
- 🏅 Revenue Trend
- 📞 Contact

Made with ❤️ by Jihad Akbar

Contact Information

😶 Jihad Akbar

📍 Indonesia

📱 (+62) 8133 2326 785

▭ jihadakbr@gmail.com

🔗 LinkedIn

💻 GitHub

Feel free to reach out!

04

Appendix

Link

Streamlit: time-series-forecasting-olist-e-commerce-jihadakbar.streamlit.app

GitHub: github.com/jihadakbr/time-series-forecasting-olist-e-commerce

THANK YOU

Do you have any questions?

Jihad Akbar

+62 813 3232 6785

https://www.linkedin.com/in/jihadakbr

jihadakbr@gmail.com