ACME Textiles.

Data-driven Sales Forecasting
Proposal

Vignesh Nambiar, Rita Orvalho,
Daniel Sebastián, Laufey Sverrisdóttir, Lucas Trenzado

Agenda

1. Problem Definition
2. EDA
3. Models
4. Benefits

Problem Definition

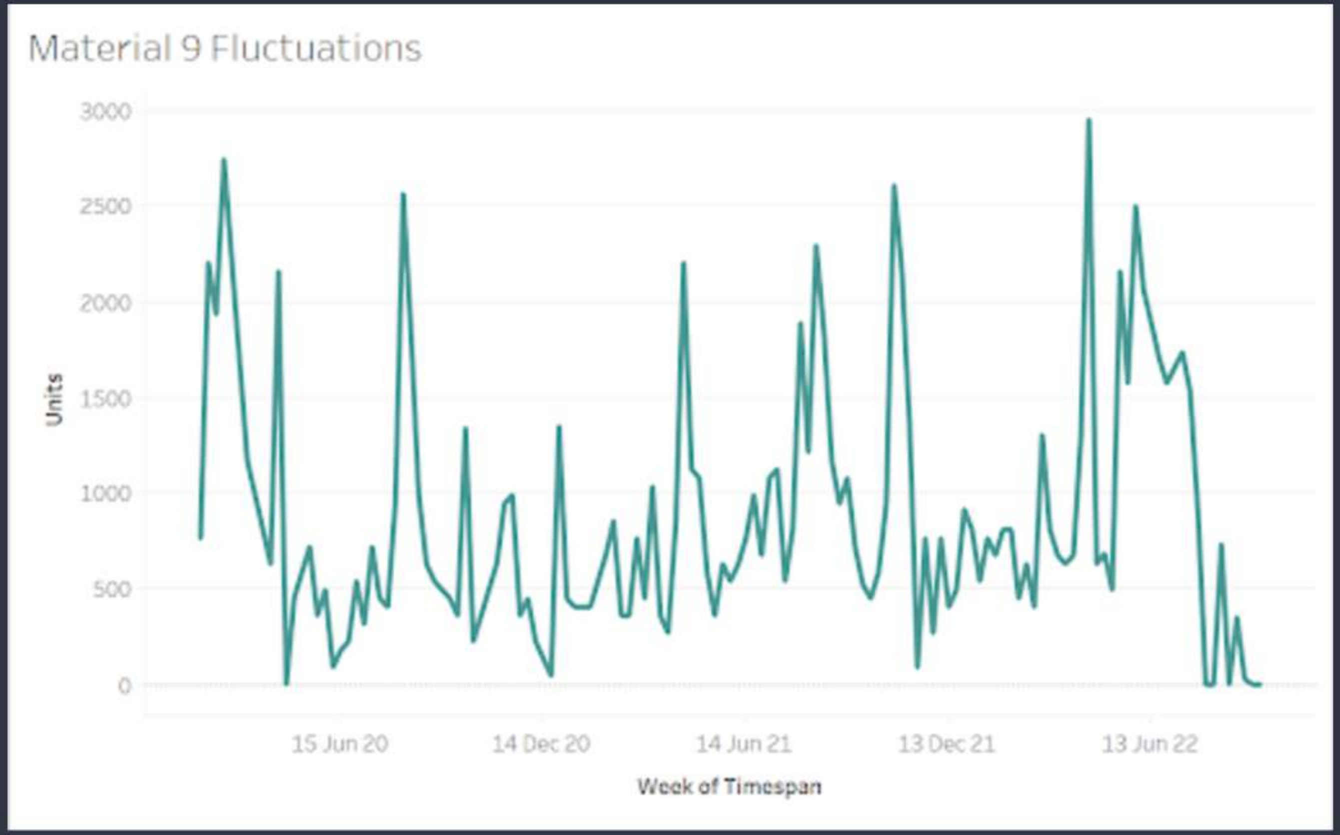
ACME's Sales representatives forecast their end-client product demand. Nevertheless, the error in demand forecast has increased significantly, causing:

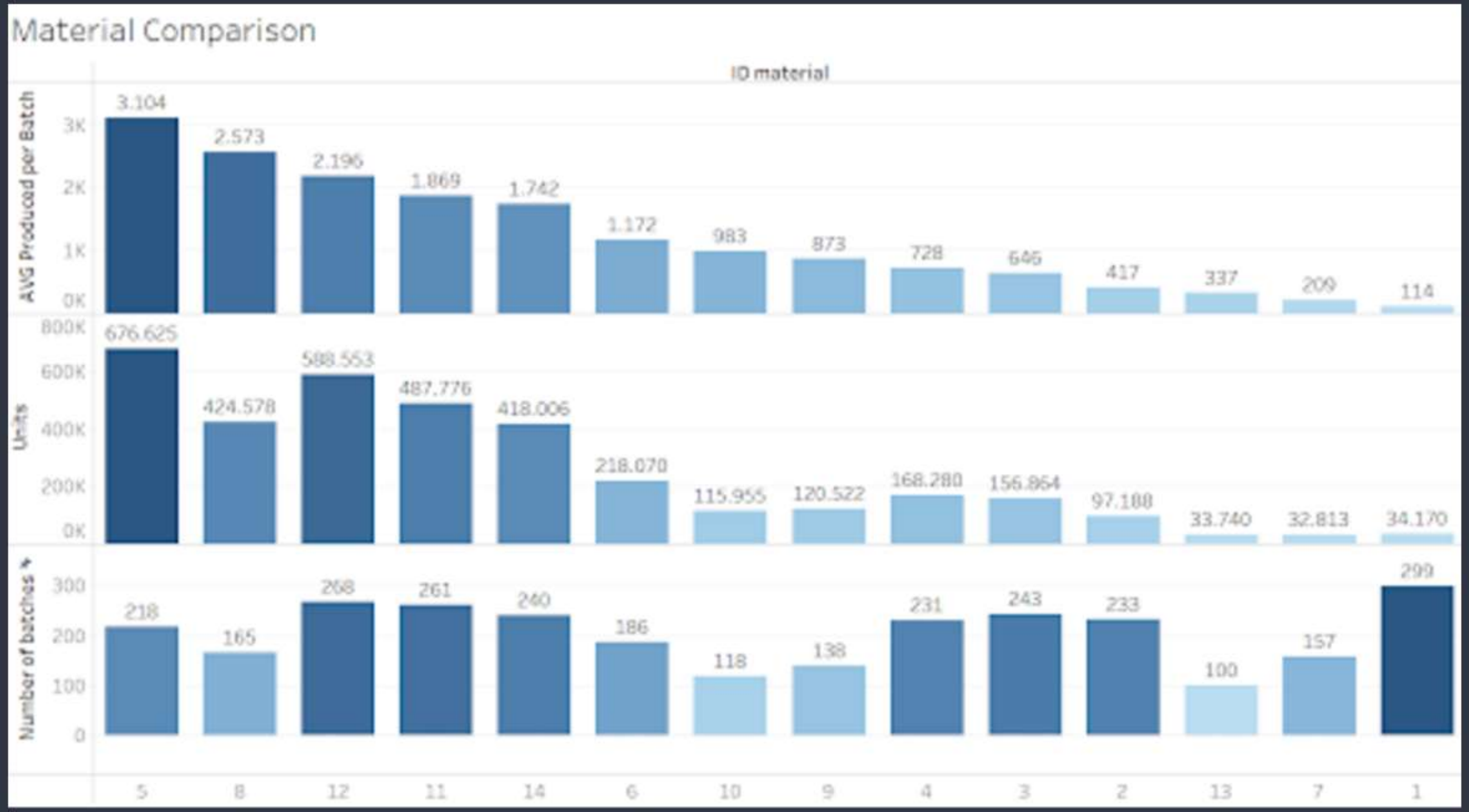
- Stress in production
- Stress in stock
- Overproduction and Underproduction
- Financial Inefficiencies

We aim to enhance optimization by developing a robust analytical model. This project holds the potential to **reshape** ACME TEXTILE's future by leveraging data-driven insights for improved operational efficiency and profitability.

Data-driven Quick Insights

- Visualized the ordering history for better understanding
- Plotted each time series
- No nulls
- Zeros left unaltered





Statistical Testing For Models

Technical EDA

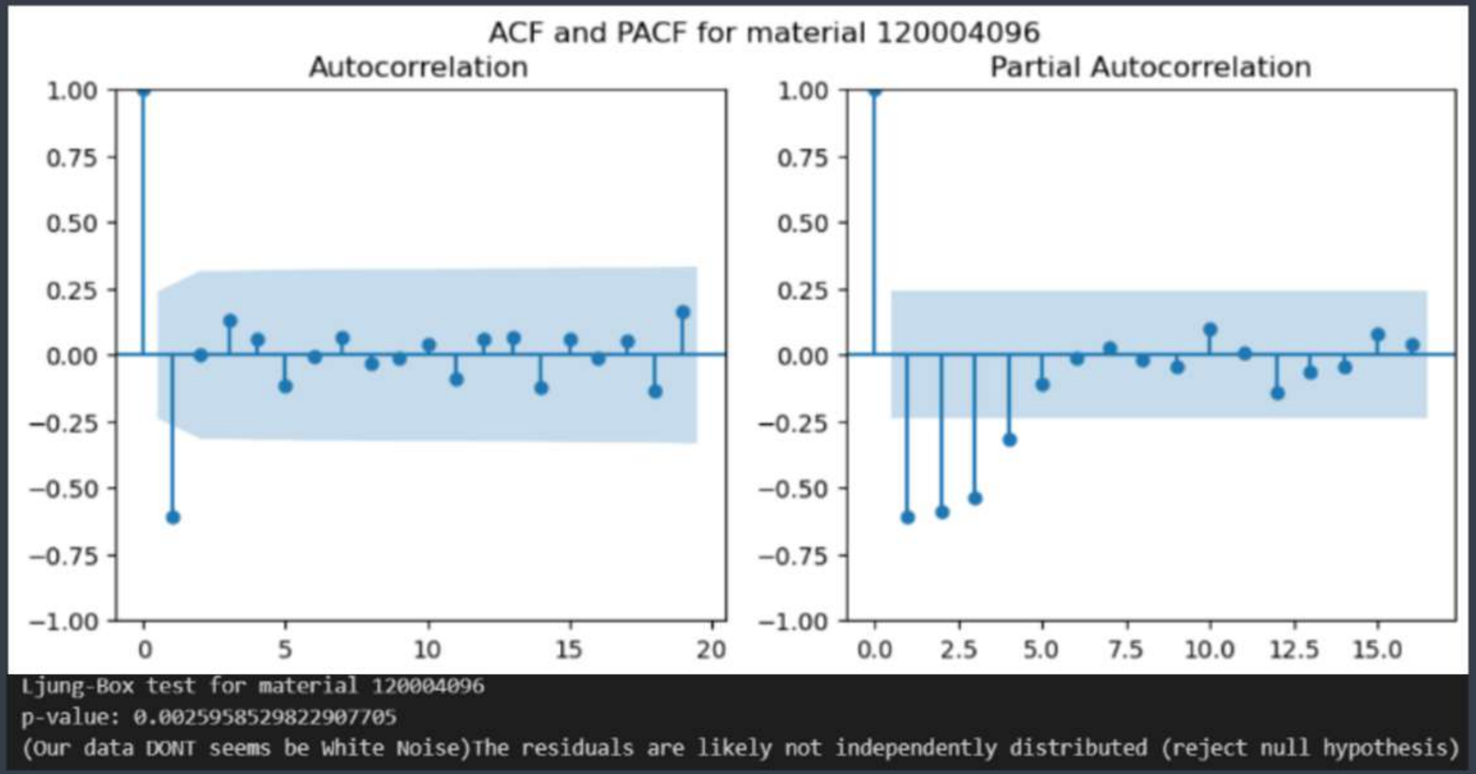
Most time series models assume stationarity and do not work for White Noise series, so we will be following the Box-Jenkins Methodology, for each product we will:

1. Test for stationarity using ADF → We will take a difference [T-(T-1)] to make the series stationary.

2. Plot ACF & PACF and Test for White Noise → If our ACF & PACF do not have significant bars out bounds our series is probably White Noise.

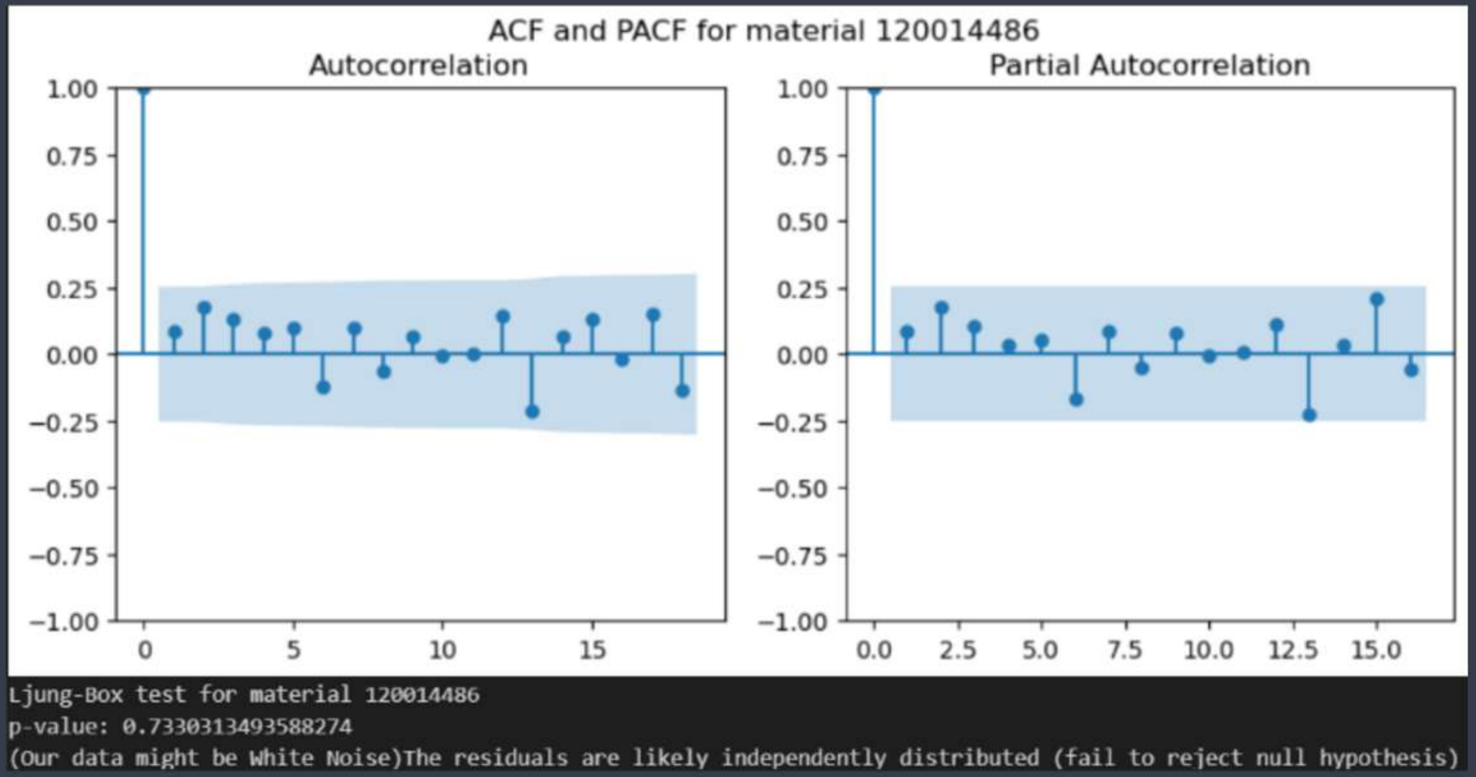
3. Test for normality using Shapiro-Wilk test → If our series is called Gaussian WN, the series is totally unpredictable in mean and variance.

Statistical Testing For Models
Technical EDA



Product 1 has bars out of bounds in the ACF & PACF. The Ljung-Box test p-value is <0.05.

Not White Noise, as such, data is predictable.



Product 3 does not have bars out of bounds in the ACF & PACF. The Ljung-Box test p-value is >0.05.

White Noise, as such, the best prediction is just the series' mean.

Statistical Testing For Models
Technical EDA

Material_ID	Stationary (ADF)	White Noise (Ljung-Box)	Normally Distributed(Shapiro)
120004096	No (Need one Difference)	No	Yes
120014488	Yes	No	No
120014486	Yes	Yes	Yes
120015996	Yes	Yes	Yes
120009816	Yes	Yes	Yes
120010342	Yes	No	No
120009814	Yes	Yes	Yes
120010566	Yes	Yes	Yes
120010970	Yes	Yes	Yes
120011782	No (Need one Difference)	No	No
120011556	Yes	Yes	Yes
120012154	Yes	Yes	No
120012606	Yes	Yes	Yes
120015842	Yes	No	Yes

- 3 Predictable
- 2 Not Stationary but Predictable
- 1 White Noise
- 8 Gaussian White Noise

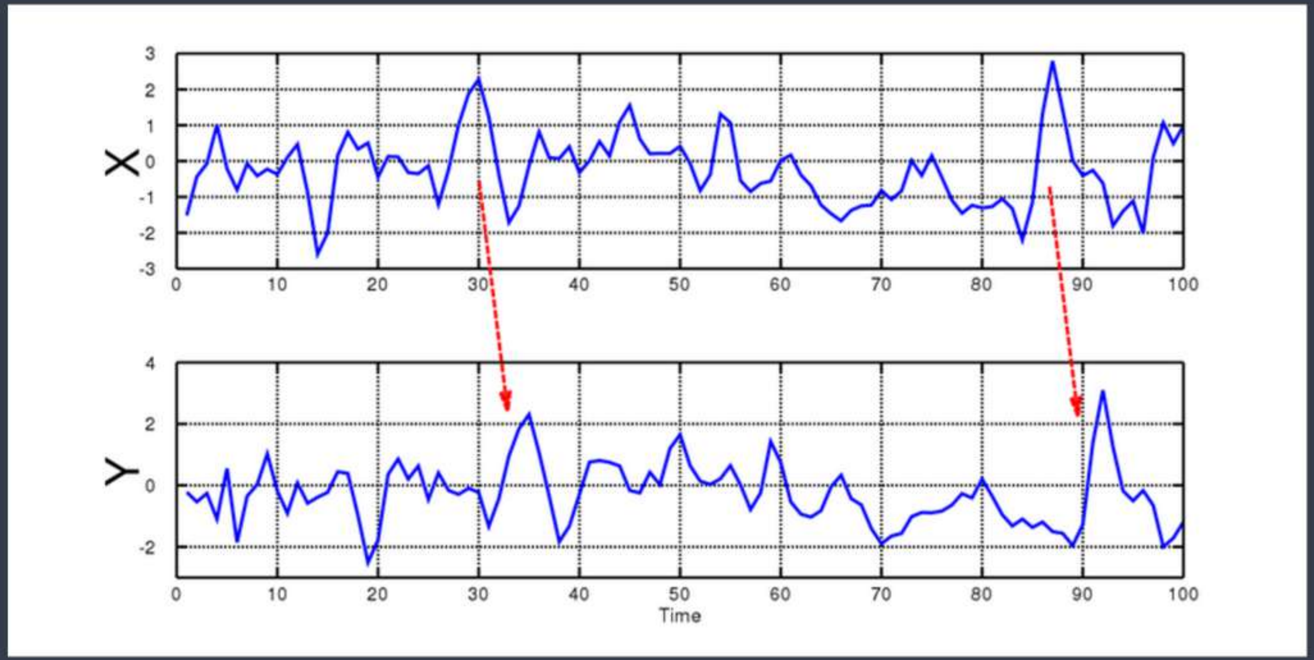
So most series cannot be predicted just by looking at their past, but… What if they are influenced by other series?

Correlation was plotted:

Series 4 vs Series 9, have (0.57) correlation.

Statistical Testing For Models
Technical EDA

Not normally distributed, we test if one influence the other with the Granger Causality Test:



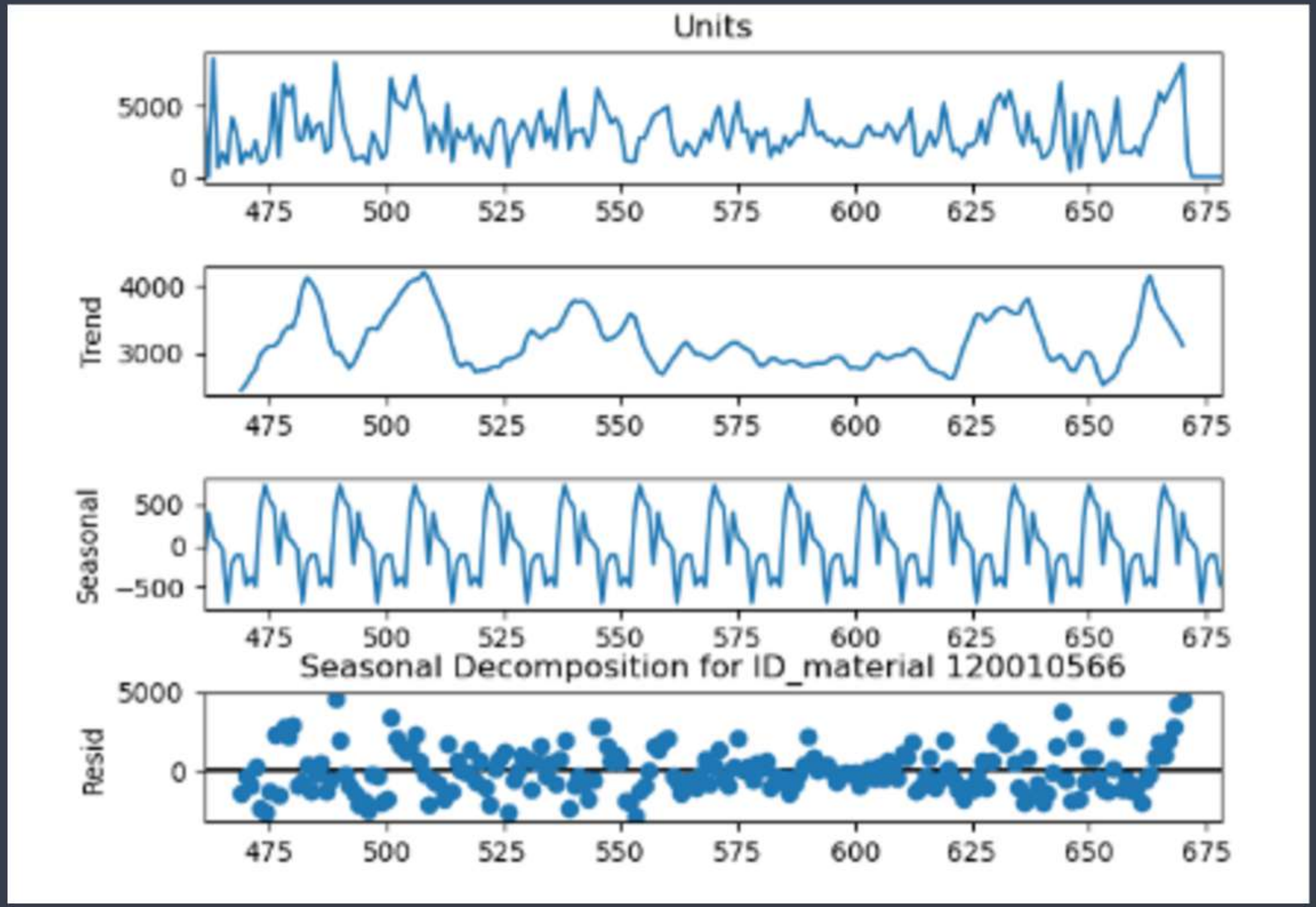
Unable to reject the null hypothesis

Products will no influence each other, which decreases likelihood of overfitting.

Statistical Testing For Models
Technical EDA

It was also important to look at Seasonal Decomposing:

- Trend
- Seasonality
- Error



Lazy / Dummy Models

These simple models are able to predict our series fairly good, with a straightforward approach:

- GlobalMeanGuessing
- GlobalMedianGuessing
- MonthSpecificMedian
- RollingMedianWindow

Models were cross validated to ensure they were robust.

Around 70% Accuracy on unseen data

ARIMA Exponential Smoothing & Prophet

ARIMA-AIC
Balancing the trade-off between model complexity and goodness of fit.

ARIMA-ManualFinding
Model picked that was best on unseen data, some overfitting risk.

Exponential Smoothing
Takes into account similar parameters as ARIMA, more importance given to recent observations.

Prophet
Works well with A LOT of data, with strong seasonality and trends a bit worse results than other models.

Around 80-90% Accuracy on unseen data

On non White Noise data

Machine Learning Models

Why should we look into Machine learning models?

- Enriched columns
- External Features

Some features were added: Season, MonthName, GDP, Holidays in Spain, Unemployment, Time of release.

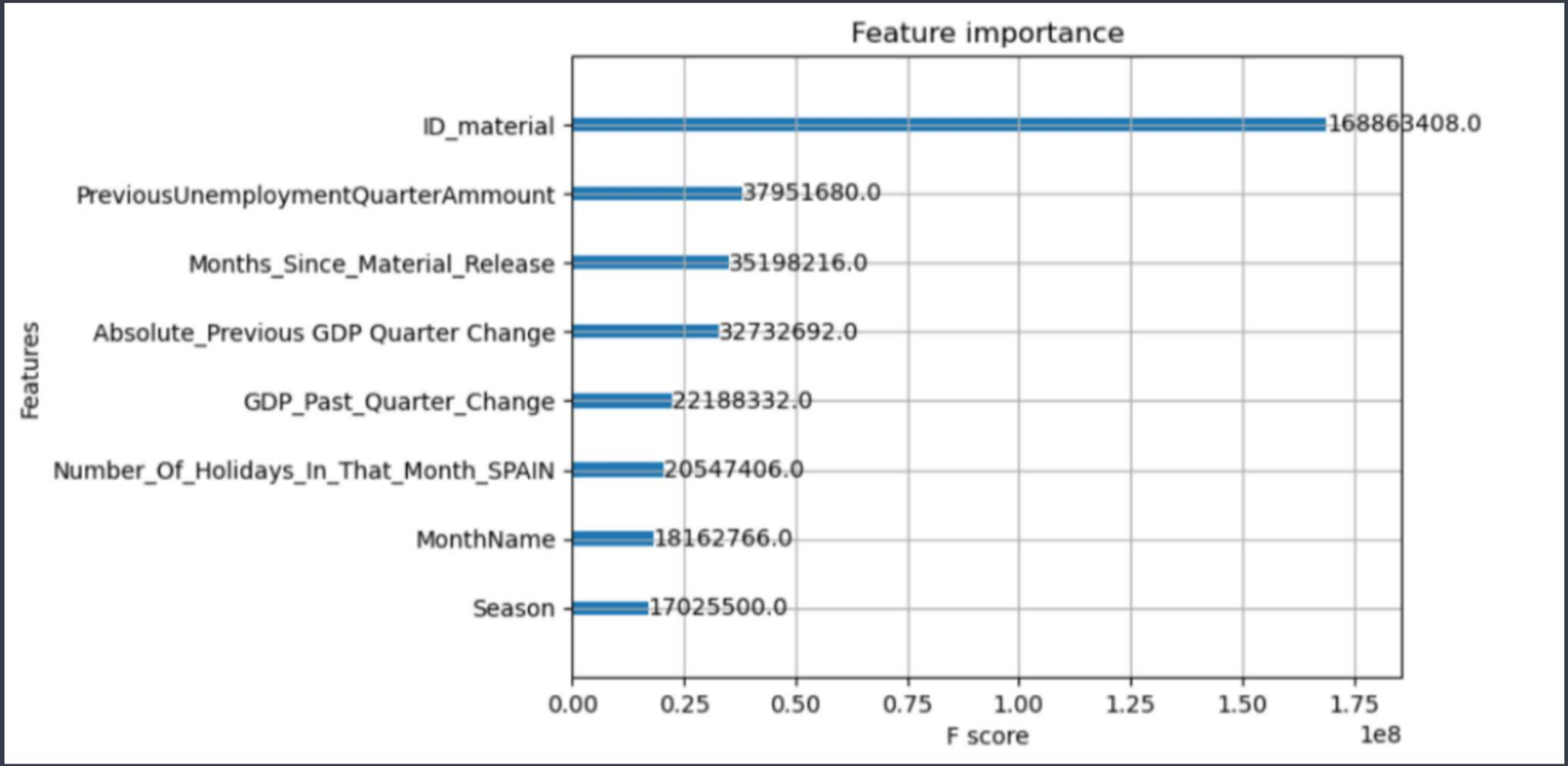
Machine Learning Models

Models tested:

- LightGBM
- KNN
- RandomForest
- SVR
- XGBoost

Outperforms 8 out of the 14 Products

Feature Importance



Machine Learning Models
XGBoost

Why should we look into XGBoost?

- Relationships between Products
- Single Model for all Products

But the accuracy was lower…

- K-Fold Crossvalidation
- Requires more Data
- Overfitting to external Features

80.26% Accuracy on unseen data

Machine Learning Models
DL RNN

Why should we look into RNN?

- Sequential Dependencies
- Feedback Mechanism

RNN Details

LSTM - 64 cells
Activation Function - Relu
Dense - 8 cells

But the accuracy was lower...

- Less robust compared to XGBoost because of it's complexity.
- Prone to overfitting.

85.99% Accuracy on unseen data

General Results

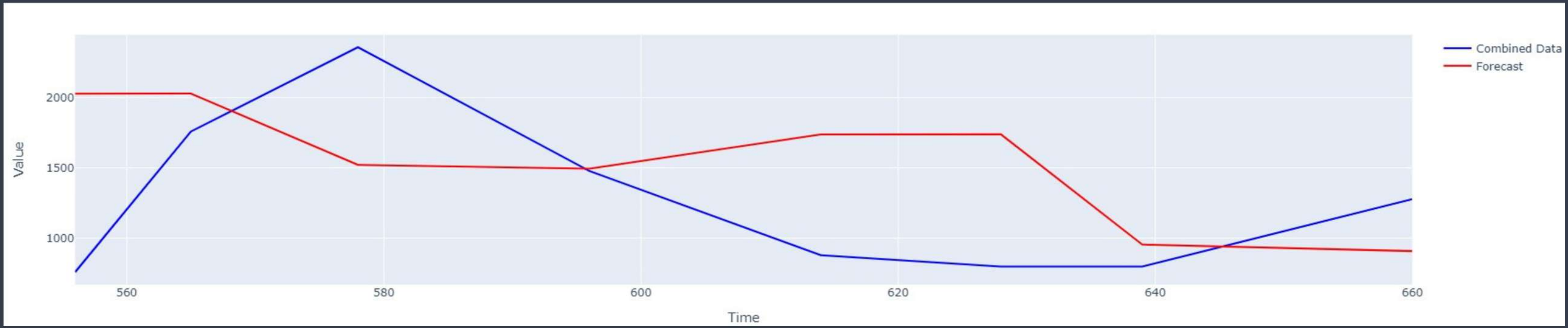
	Model	Accuracy	Notes
01	STATISTICAL MODELS	86.79%	Performed Statistical tests:- ACF, PACF, box-test, and normality tests for data decomposition, checking for seasonality and trends
02	GUESSING MEAN/MEDIAN	76.03%	A simple dummy model by just predicting the mean/median for each product
03	ML MODELS (XGBOOST)	80.26%	Implemented XGBoost after trial with carefully feature engineered columns and hyperparameter tuning
04	DEEP LEARNING MODELS (RNN)	85.99%	RNN being able to use the feedback mechanism from the previous steps enables short and long term dependencies
05	OTHERS PYCARET,PROPHET	--%	Trial of different approaches which proved to not lead to better results

89.44%

Best Accuracy achieved by Combinations of all models for each product.

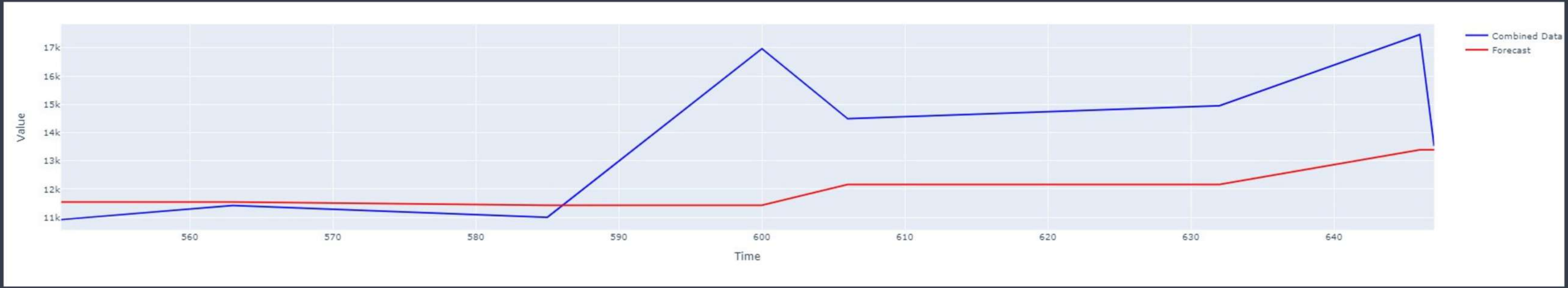
How does our predictions look like for 2022

Exponential Smoothing material 13



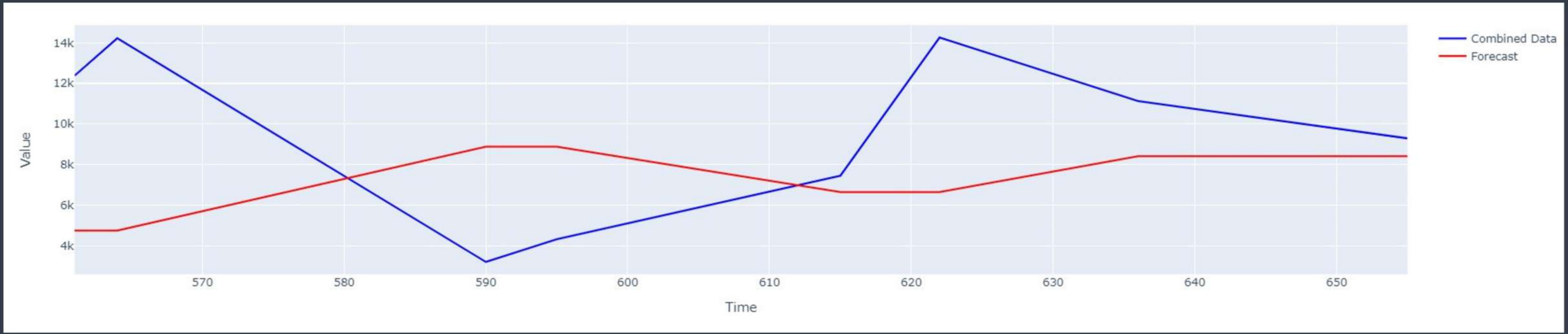
How does our predictions look like for 2022

Exponential Smoothing material 8



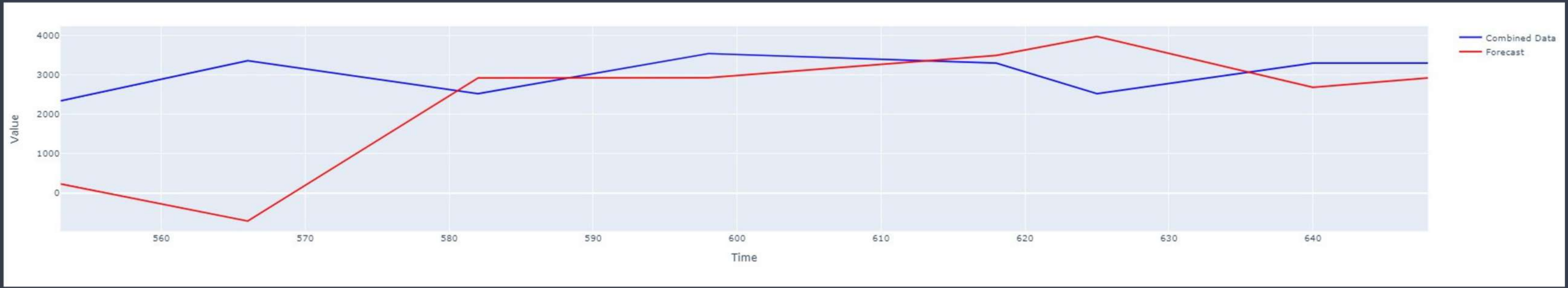
How does our predictions look like for 2022

Exponential Smoothing material 12



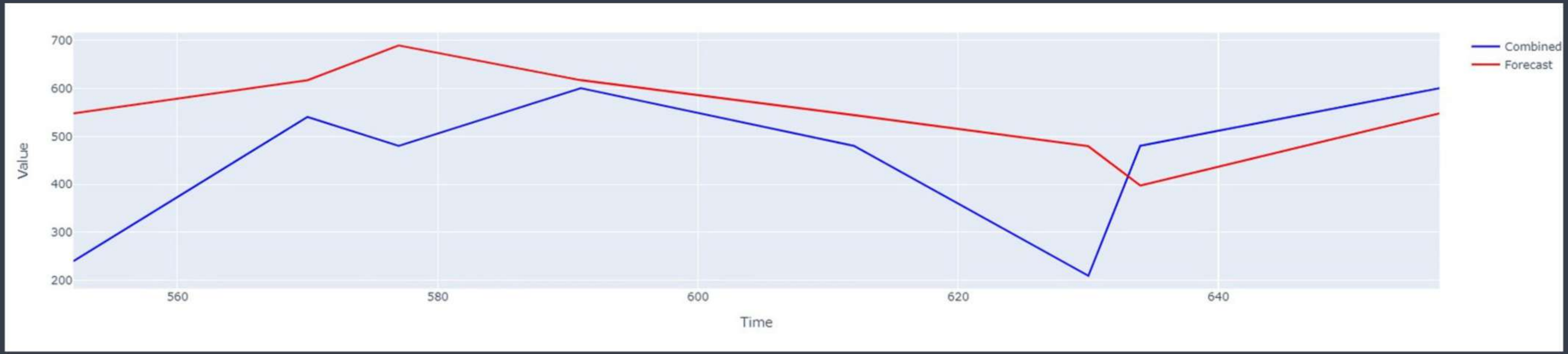
How does our predictions look like for 2022

ARIMA material 4



How does our predictions look like for 2022

ARIMA material 1



Next Steps

How can we further increase performance?

Data Aggregation

Additional Data

MLOps Pipeline

Optimize Operations

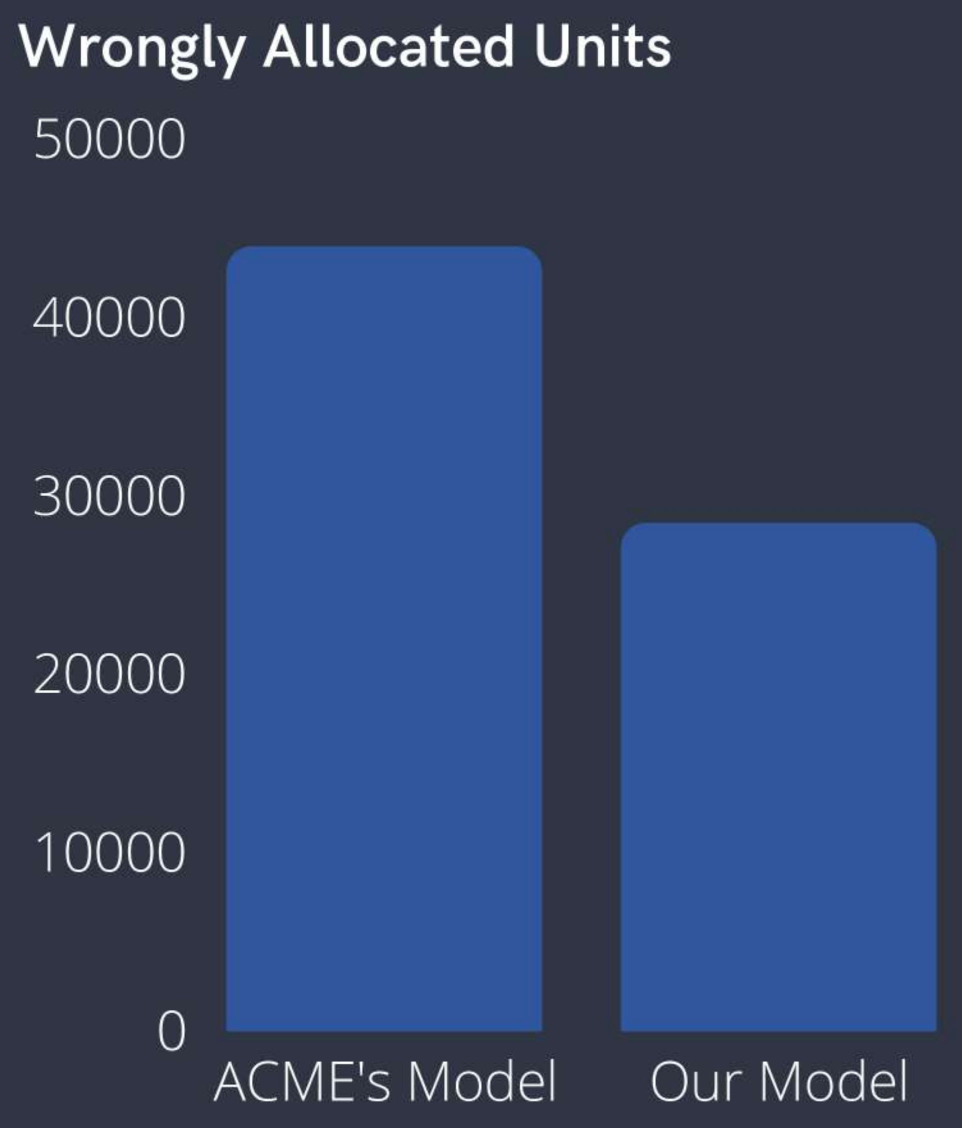
Business Benefits of
The ARIMA Model

Sales forecasting allows businesses to:

1. Fully optimized Inventory, Production, and Supplier Management.

2. Improve their overall profitability

Business Benefits of The ARIMA Model



Accuracy

80% › 87%

Wrongly Allocated Units

44000 › 28500

Over-produced units decrease: 13000

Under-produced units decrease: 2500

Values are rounded for presentation purposes

Business Benefits of The ARIMA Model

With this increase in accuracy, these were the monetary results:

- Average price 359 €
- Average profit margin 20%

Over-produced units decrease: 13000 → 3 600 000 €

Under-produced units decrease: 2500 → 880 000 €

Total amount saved: 4 480 000 €

Values are rounded for presentation purposes

ACME Textiles.

Data-driven Sales Forecasting Proposal

Thank you!

Vignesh Nambiar, Rita Orvalho,
Daniel Sebastián, Laufey Sverrisdóttir, Lucas Trenzado