KAGGLE STORE
: SALES DATA ANALYSIS

INTRODUCTION

Background

- Kaggle currently offers three products: Mugs, Hats, and Stickers.
- These products are sold through two store chains: KaggleMart and KaggleRama.
- Operations span three countries: Finland, Norway, and Sweden

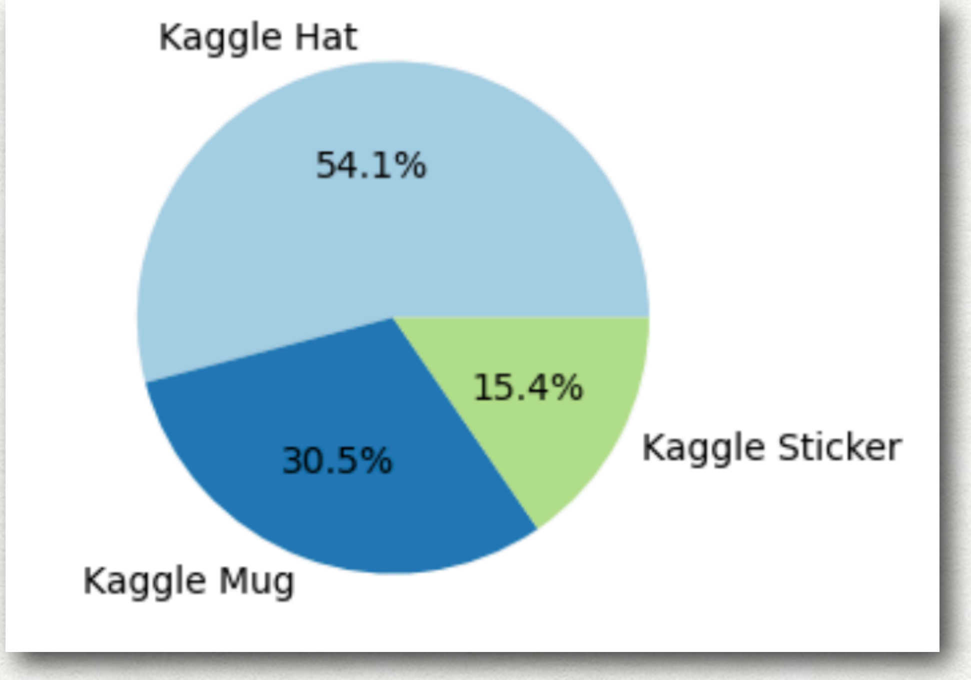
My Goal

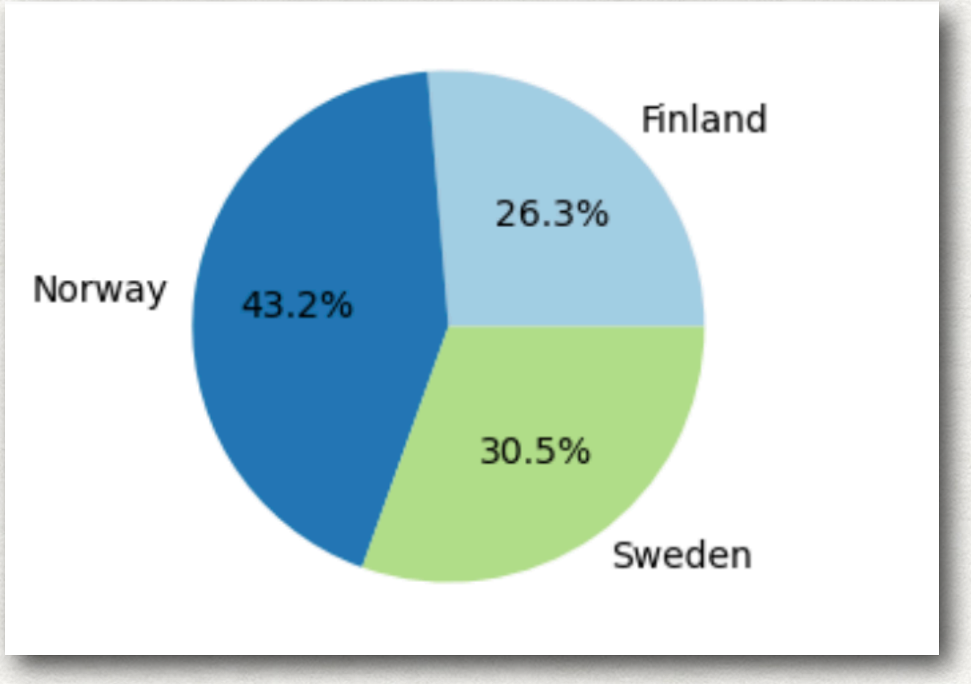
- Determine the optimal official outlet for Kaggle products by developing a predictive sales model and evaluating store chain performance.

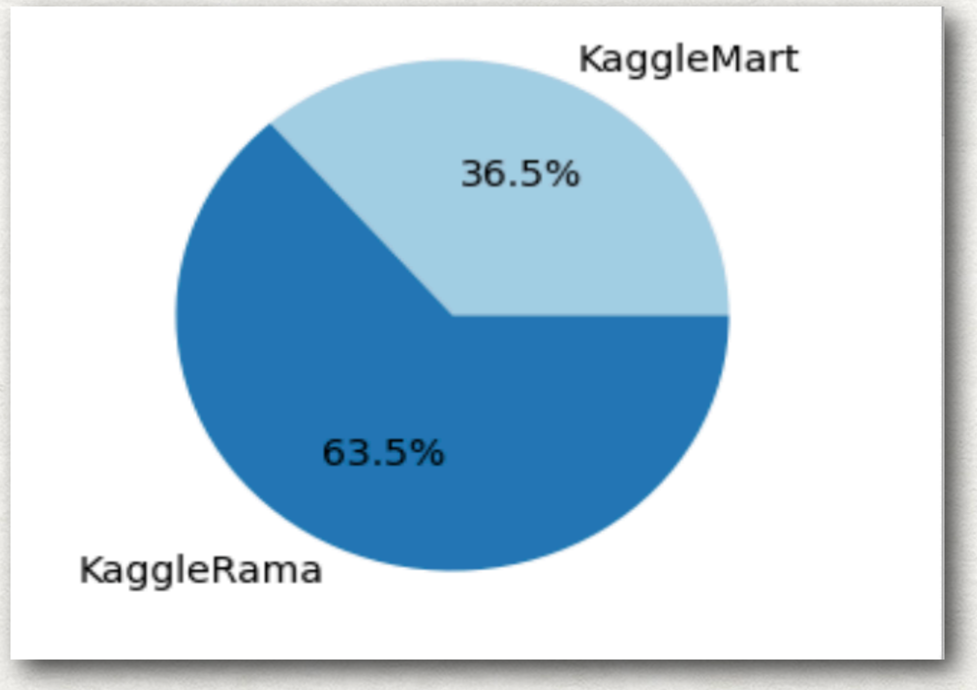
DATA

- The dataset was sourced from Kaggle and comprises time series data.
- It records the sales of three distinct Kaggle products from January 01, 2015, to December 31, 2018.
- The raw data consists of six columns: row_id, date, country, product, store, and num_sold.
- In total, the dataset contains 26,298 rows.

EXPLORATORY DATA ANALYSIS







Product Sales:

- Kaggle Hats emerged as the most popular product, while Kaggle Stickers recorded the lowest sales.

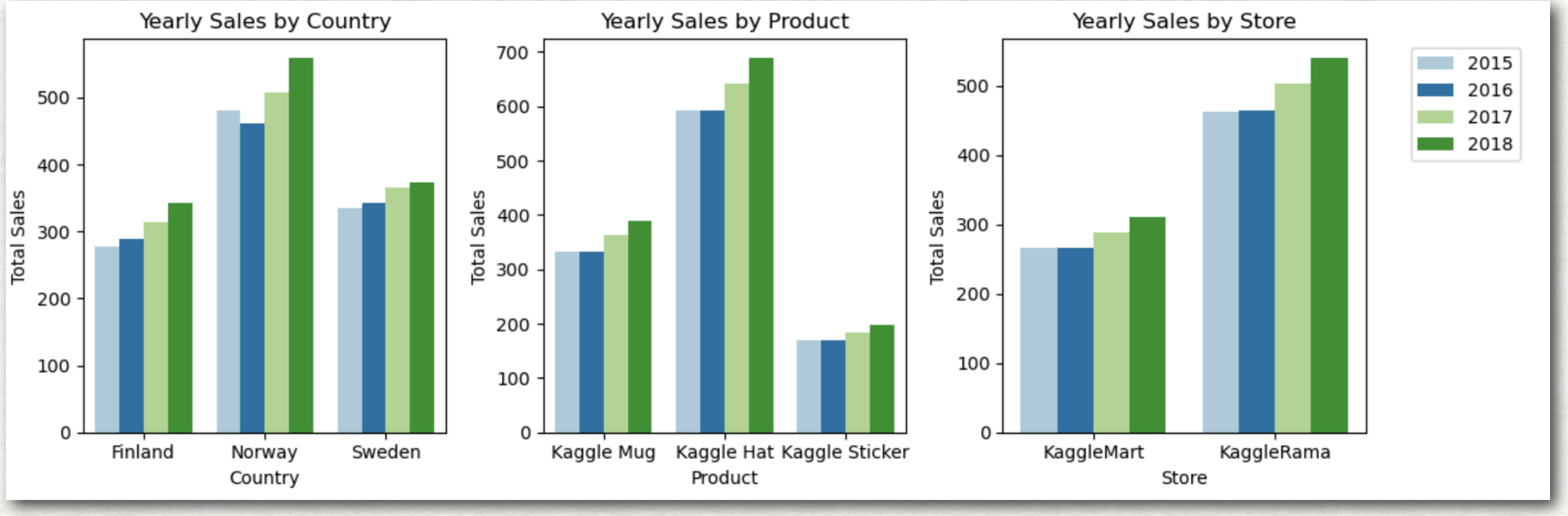
Country Sales:

- Norway stood out with the highest sales among the countries, indicating a robust demand for Kaggle products.
- Sweden and Finland followed, securing the second and third positions in terms of sales.

Store Performance:

- KaggleRama outperformed KaggleMart by selling 1.7 times more products, highlighting a substantial disparity in sales volume between the two store chains.

SALES TRENDS OVER TIME



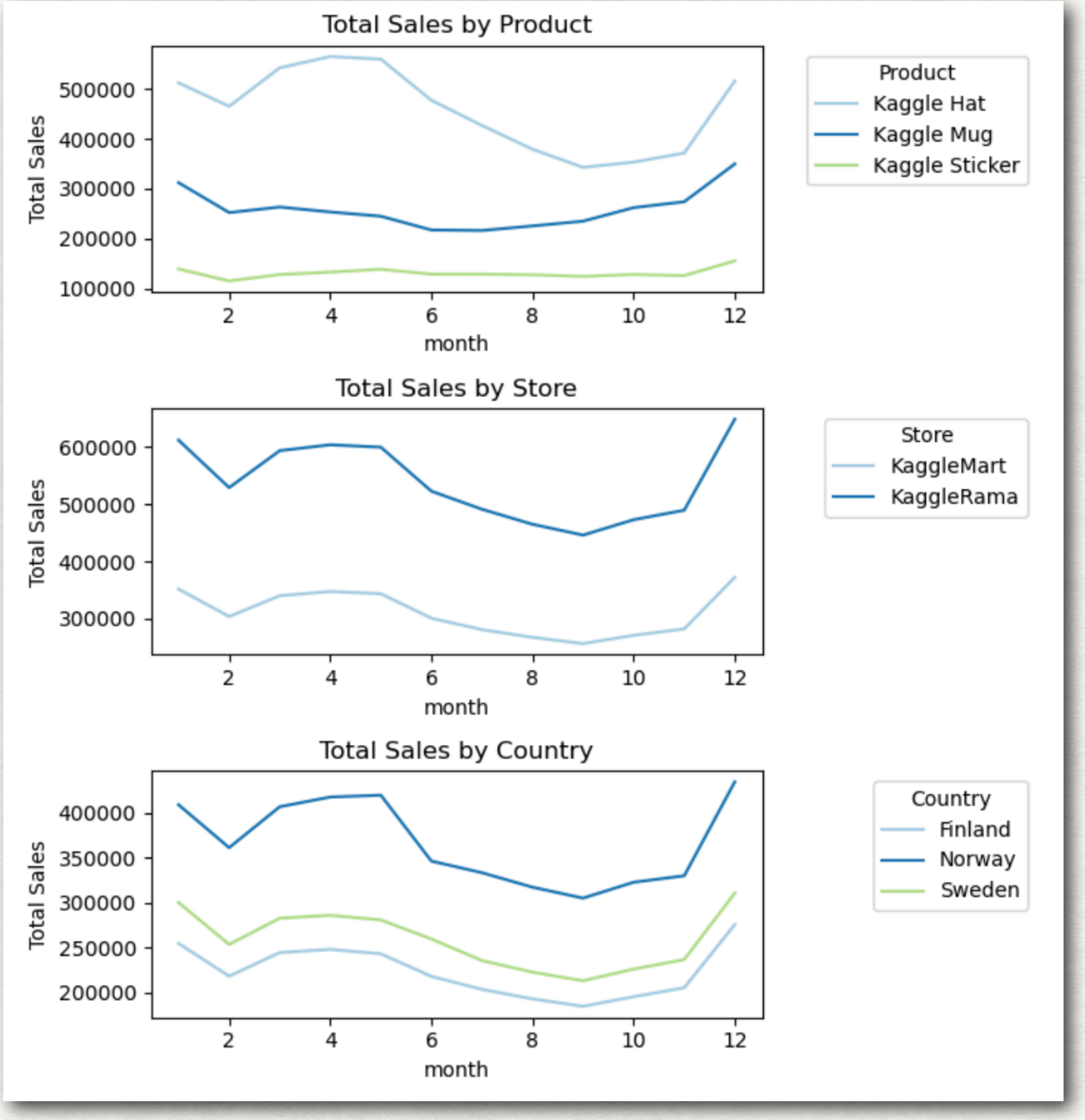
The sales trends over the four-year period

Overall Growth: Across all three countries, there has been a consistent and steady increase in sales, signifying a robust overall growth trend.

Temporary Decline in Norway: Notably, Norway experienced a temporary decline in sales from 2015 to 2016. However, this dip did not disrupt the overarching upward trajectory in sales.

Consistency in Other Countries: In contrast, most other countries witnessed a continuous rise in sales during this period, contributing to the overall positive trend.

SEASONAL SALES PATTERNS



Seasonal Peaks:

- Highest sales occurred during the spring (March to May) and winter (December and January) seasons, reflecting synchronized demand trends.

Consistent Monthly Patterns:

- Monthly sales plots displayed consistent patterns across all stores and countries, indicating a synchronized demand trend.

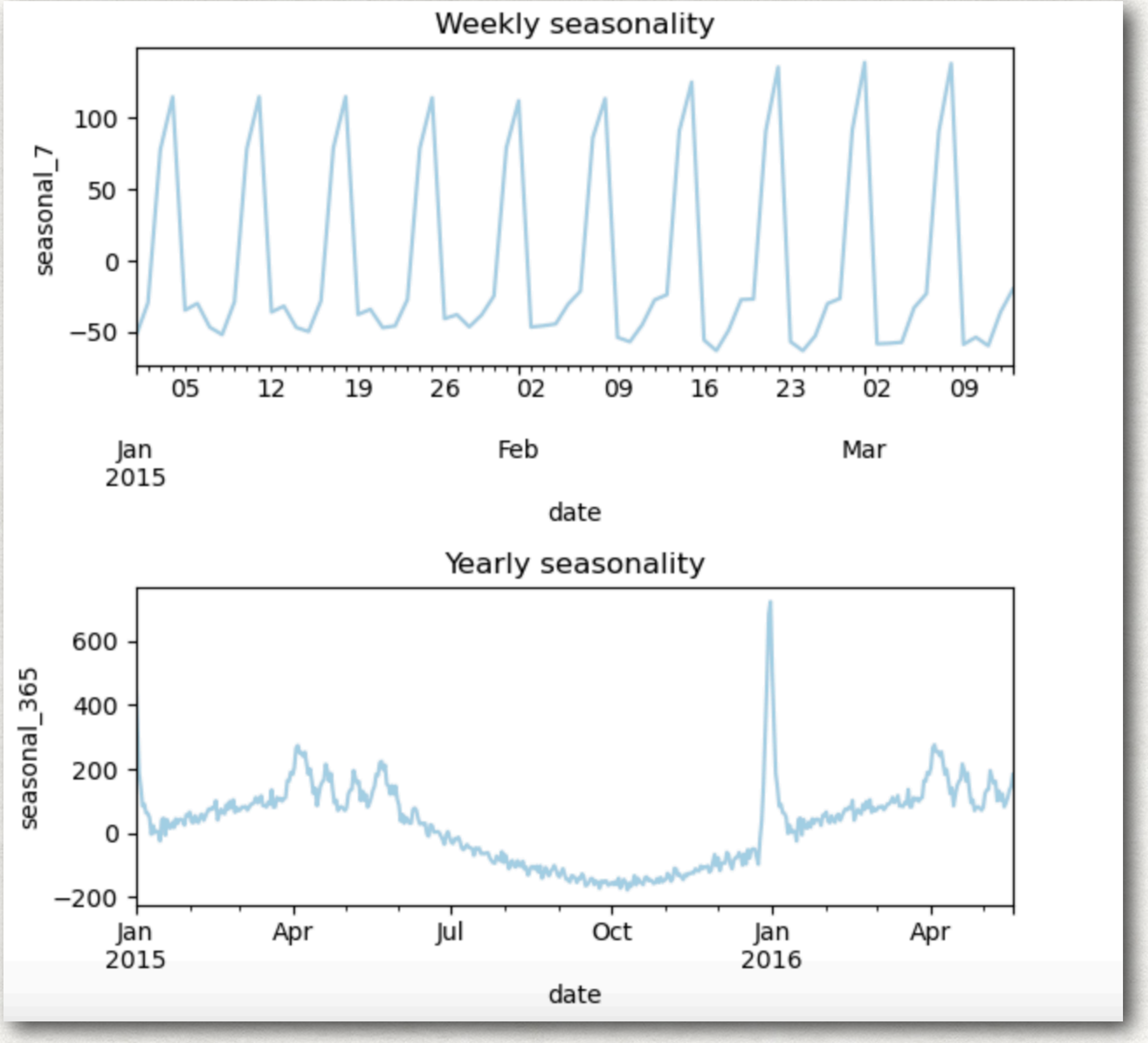
Unique Product Patterns: Each product exhibited distinct seasonal patterns:

- Kaggle Hat: Peaks in April and December but dips in September and October.
- Kaggle Mug: Peak sales in December but declines in July and August.
- Kaggle Stickers: Steady sales throughout the year, irrespective of store or country.

TIME SERIES DECOMPOSITION WITH MSTL

Why Decompose Time Series?

- Gain insights into long-term patterns.
- Identify and analyze seasonal effects.
- Examine residuals for anomalies.
- Improve forecasting and modeling accuracy.



MSTL Decomposition

- Seasonal_7: Captures weekly seasonality with lower weekday sales and weekend peaks.
- Seasonal_365: Represents seasonality that changes over the months, confirming the yearly sales pattern identified earlier in our exploratory data analysis

MODELING

I evaluated the performance of four forecasting models

- Naive Seasonal, SARIMA, BATS, and Prophet.

Performance Metric - MAPE (Mean Absolute Percentage Error):

- MAPE measures the average percentage difference between predicted and actual values, offering a relative indicator of forecasting accuracy.

Baseline Model:

- Naive Seasonal (Baseline) achieved a MAPE of 0.131, serving as a reference for model comparison.

MODEL PERFORMANCE EVALUATION

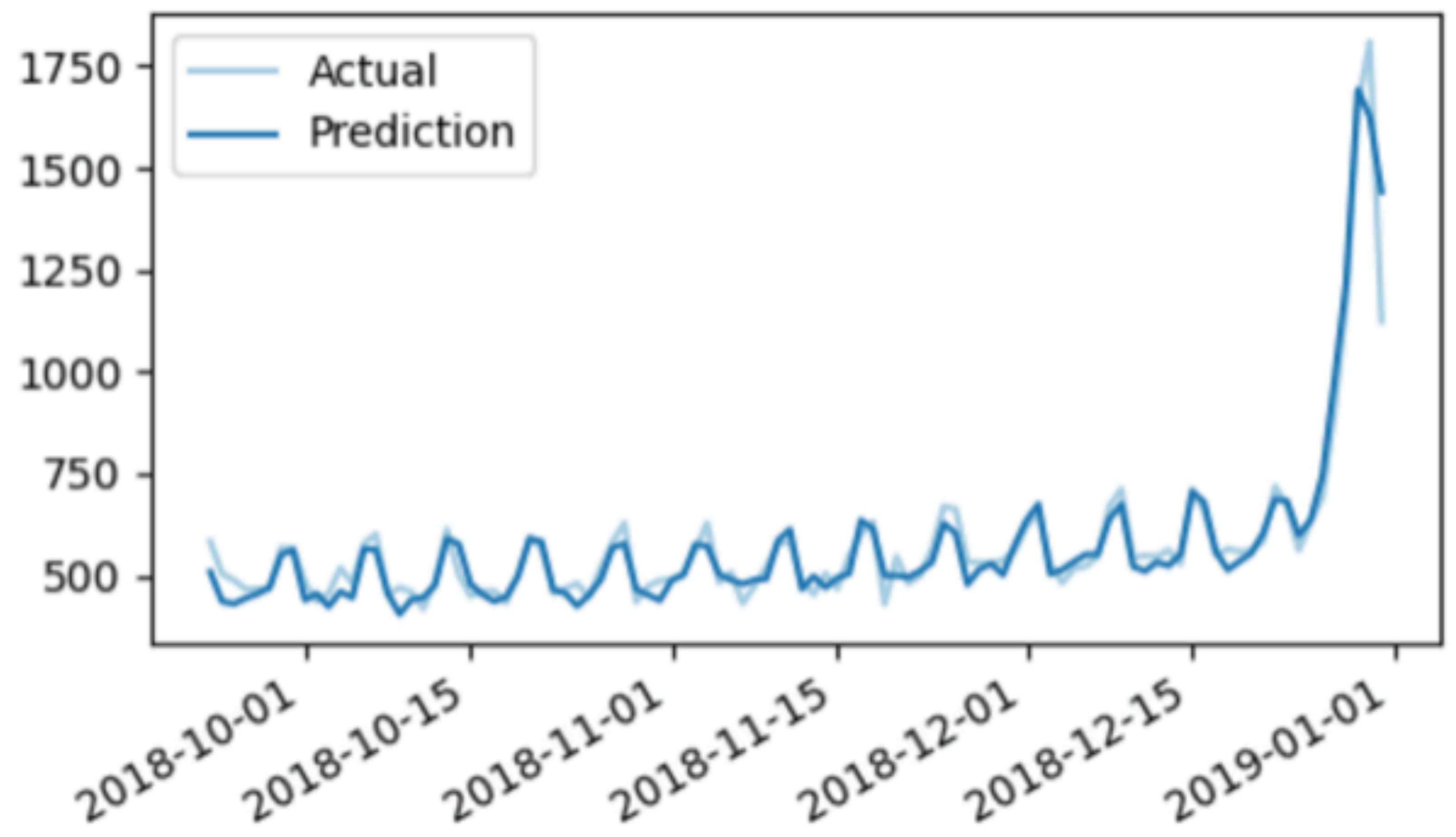
SARIMA Model

- SARIMA outperformed the baseline, yielding a lower MAPE of 0.103.
- SARIMA incorporates autoregressive, moving average, and seasonal components, enhancing forecasting accuracy.

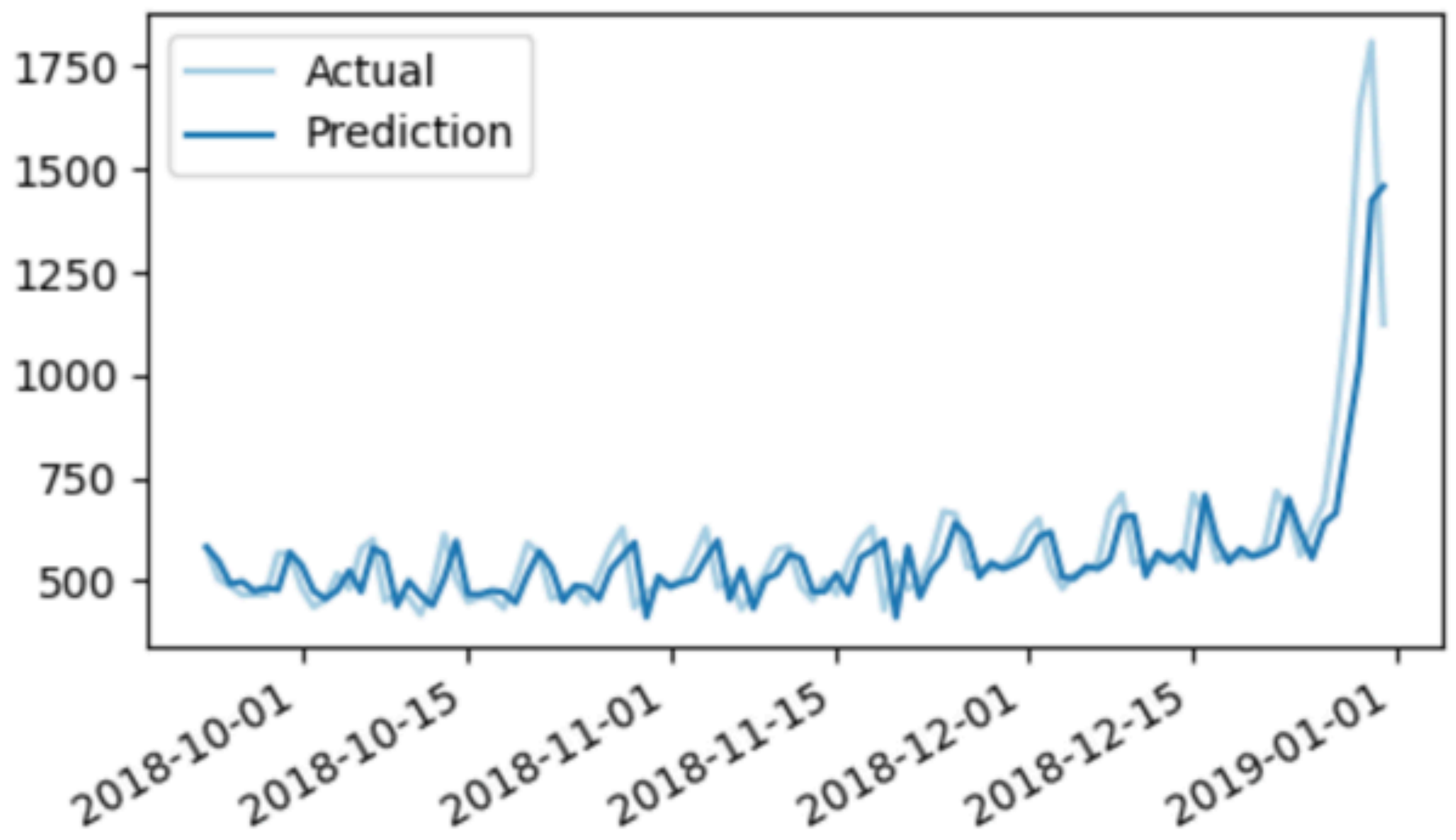
BATS Model:

- BATS, utilizing exponential smoothing, ARIMA, residuals, and Box-Cox transformation, outperformed all models with the lowest MAPE of 0.053.
- BATS can handle data with multiple seasonal patterns effectively.

CLASSICAL APPROACH

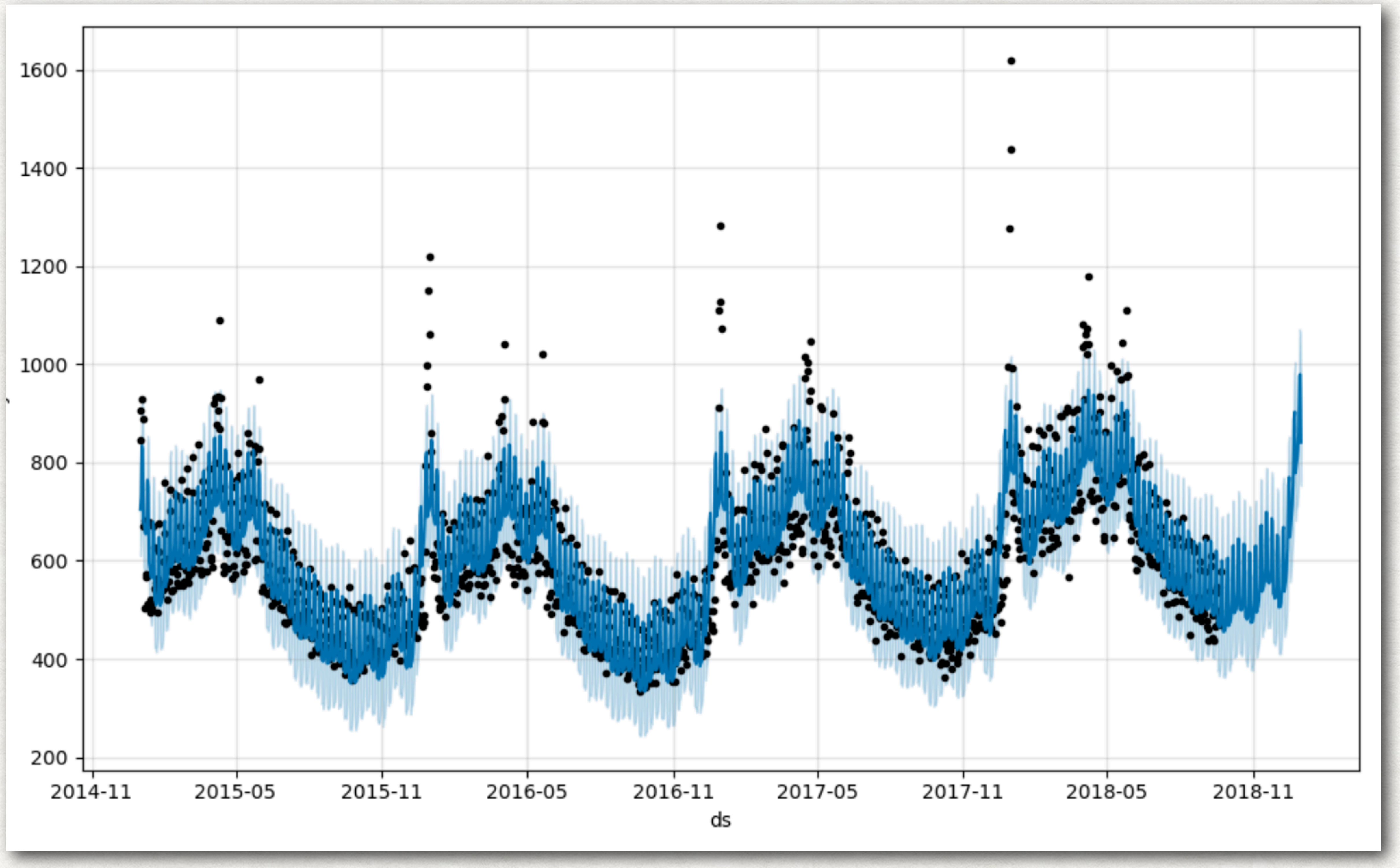


BATS



SARIMA

THE PROPHET MODEL



Despite incorporating holiday features, Prophet achieved a MAPE of 0.264 and did not outperform the other models in this analysis

FUTURE WORK

1. Model Fine-Tuning:
   - Consider fine-tuning and optimizing models, particularly by adjusting the configuration of the Prophet model, to better match the data's characteristics.
2. Feature Augmentation:
   - Incorporate additional relevant features, such as economic indicators or country-specific holiday features, to capture underlying relationships and improve forecast accuracy.
3. Advanced Time Series Models:
   - Explore advanced models like Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks, known for effectively capturing complex temporal patterns. These models can provide improved forecasting capabilities.