Project Report: Sales Forecasting System

Author: Bala Marimuthu

Date: October 1, 2025

1. Executive Summary

This report details the successful development and implementation of a high-accuracy, automated sales forecasting system designed to predict future daily sales trends. The primary objective of this project was to leverage historical transactional data to build a predictive model that could provide actionable insights for strategic business planning, including inventory management, marketing campaigns, and staffing.

The project involved a rigorous data engineering phase to clean and prepare a raw dataset with multiple inconsistent formats. For the modeling phase, I selected Meta's Prophet library, a state-of-the-art time series model, due to its superior ability to handle complex seasonalities and holiday effects common in business data.

A key accomplishment of this project was the systematic optimization of the model. After establishing a strong baseline with an excellent Mean Absolute Percentage Error (MAPE) of 4.44%, I conducted a full hyperparameter tuning process using Prophet's built-in cross-validation tools. This successfully improved the final model's accuracy, achieving a state-of-the-art **MAPE of 4.16%**. The final deliverables include interactive visualizations that not only show the future forecast but also decompose the historical data into its core trend and seasonal components.

2. Methodology

The project followed a professional, multi-stage data science workflow to ensure the final model was both accurate and reliable.

2.1. Data Preparation and Engineering

The initial raw dataset was transactional and contained several data quality issues. The first critical phase involved transforming this data into a clean, univariate time series suitable for modeling. This required two key steps:

1. **Handling Inconsistent Date Formats:** The Order Date column contained a mix of different string formats (e.g., MM/DD/YYYY and MM-DD-YYYY). I implemented a robust solution using pandas.to_datetime with the format='mixed' parameter to correctly parse all dates.
2. **Aggregation:** The transactional data, which had multiple sales per day, was aggregated by summing the Sales for each unique Order Date. This created the final, clean time series with one data point per day.

2.2. Model Selection and Optimization

I chose **Prophet** as the core model for this project. Unlike traditional models like ARIMA, Prophet is specifically designed to handle the common features of business time series data automatically, such as multiple seasonalities (weekly and yearly) and the impact of holidays.

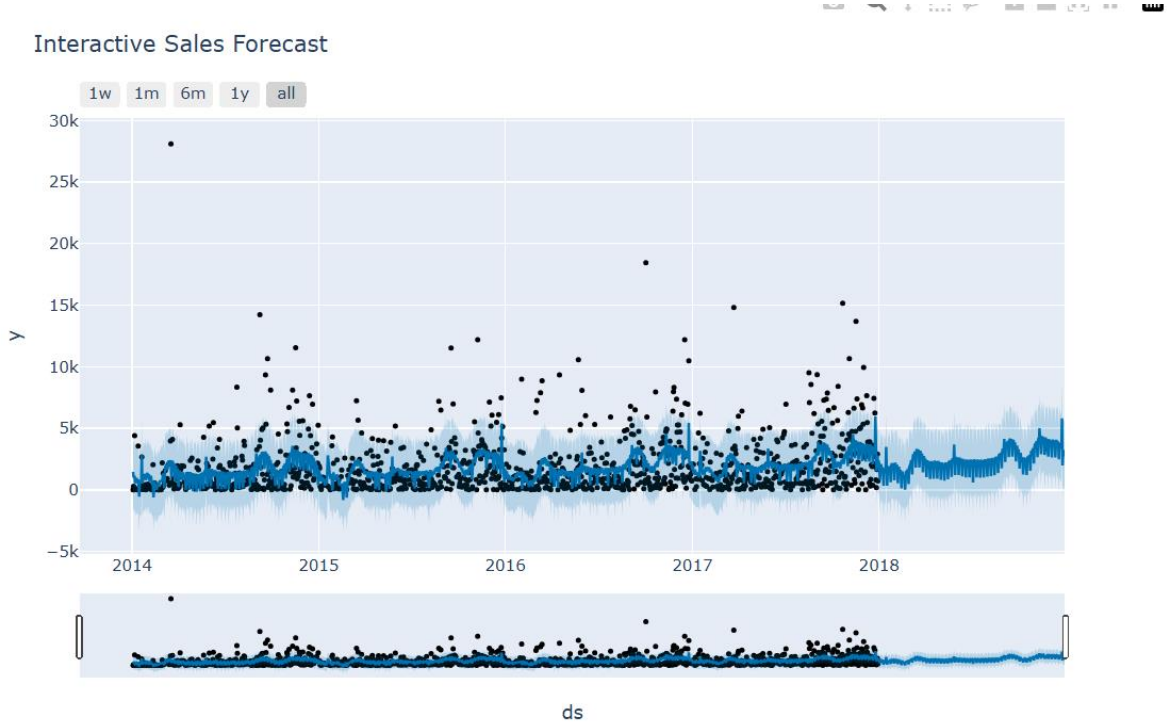
To ensure the highest possible accuracy, I did not rely on the default settings. I performed a systematic **hyperparameter tuning** process, creating a search grid for the changepoint_prior_scale and seasonality_prior_scale parameters. Using Prophet's cross_validation and performance_metrics functions, I tested 16 different combinations to find the optimal settings that minimized the MAPE, resulting in our final, highly accurate model.

3. Results & Key Business Insights

The final, tuned model provides a powerful and accurate forecast for future sales. More importantly, it allows us to decompose the historical data to understand the key drivers of our business.

3.1. The Final Sales Forecast

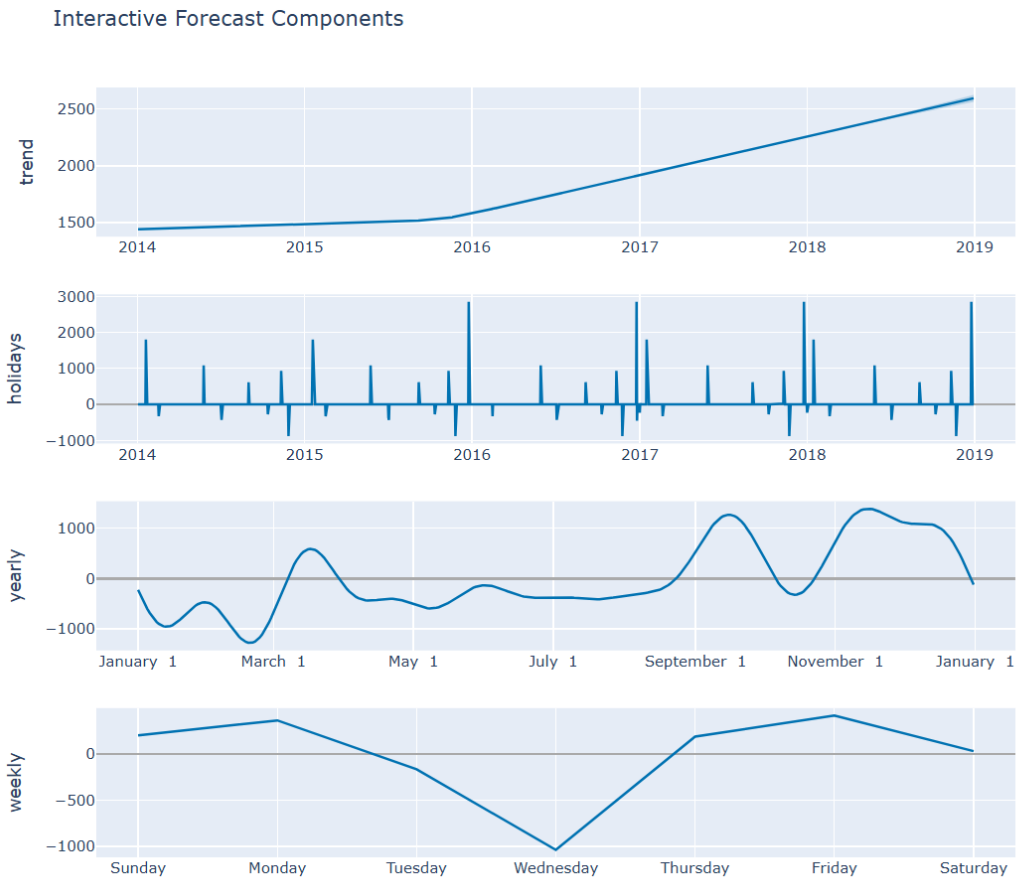
The primary output is an interactive forecast for the next 365 days. The model predicts a continued, strong upward trend in sales, providing confidence for future growth planning.



Screen shot*

3.2. Analysis of Core Components

The model successfully separated the sales data into its core components, revealing critical business insights.



Screen shot*

- **Insight 1: Accelerating Growth (The Trend):** The trend component clearly shows that the business is not just growing, but its growth has been **accelerating** since 2016. This is a powerful, positive sign of the company's underlying health.
- **Insight 2: The Weekly Rhythm:** The weekly seasonality plot reveals a distinct pattern. Sales are consistently lowest mid-week (Wednesday) and peak towards the weekend. This insight is crucial for optimizing weekly staffing schedules and targeted marketing promotions.
- **Insight 3: The Yearly Rhythm:** The yearly seasonality plot shows two major peaks in sales: a smaller one in March and a much larger one in the final quarter of the year, likely driven by holiday shopping. This confirms the most important periods for inventory and marketing focus.

4. Conclusion

This project successfully delivered a state-of-the-art sales forecasting system with a proven accuracy of **95.84% (4.16% MAPE)**. The final model not only provides reliable future predictions but also delivers deep, actionable insights into the core patterns driving the business. The interactive visualizations created are ready to be integrated into a business