Retail Sales Forecasting Using Prophet

Deepanshu

Abstract

We present a lightweight forecasting solution for retail sales using Facebook Prophet. The model efficiently captures seasonality, trend shifts, and holiday effects. It is evaluated on real-world data and visualized through a Power BI dashboard for actionable retail insights.

Introduction

Accurate sales forecasting is essential for retailers to:

- Anticipate demand fluctuations
- Plan inventory and staffing
- Improve supply chain efficiency
- Make informed strategic decisions

This project leverages Prophet, a forecasting tool designed to model complex time series with seasonality and changepoints. Prophet is ideal for business contexts due to its interpretability and ease of use. Our goal is to build a reliable forecast pipeline using real retail sales data and deploy insights through a Power BI dashboard.

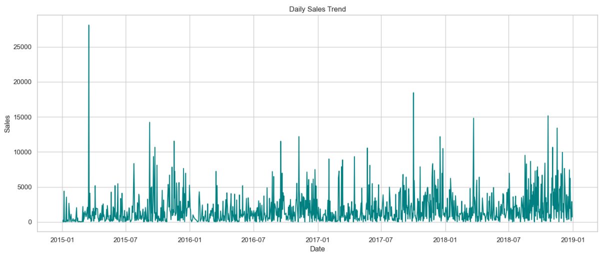
Data Source and Preprocessing

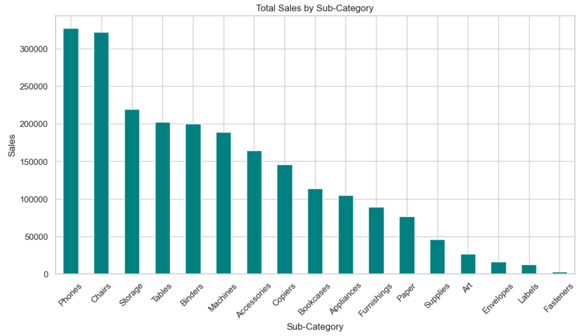
Data Source

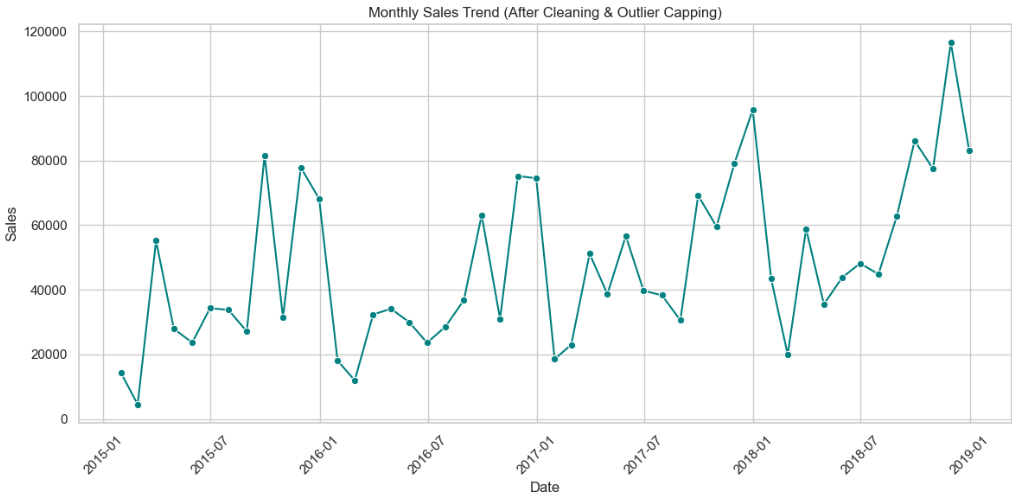
- The dataset is based on the Sample Superstore Sales data from Kaggle.
- It contains transactional-level sales data including date, region, and order amount.
- The data was filtered to include only required columns and aggregated monthly for time series analysis.

Preprocessing Steps

1. Loaded the dataset and parsed the `Order Date` column to datetime format.
2. Grouped the data by month and aggregated total sales.
3. Detected and capped outliers using IQR (Interquartile Range) method.
4. Verified and handled any missing or inconsistent entries.
5. Exported the cleaned dataset for use in the forecasting model.







Modeling Approach

Prophet is a Bayesian structural time series forecasting model developed by Facebook. It is particularly well-suited for time series with strong seasonal effects, multiple change-points, and missing data, making it an ideal choice for retail sales prediction.

Prophet decomposes the time series into four components:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \tag{1}$$

- $g(t)$: trend function modeling non-periodic changes in the time series
- $s(t)$: periodic seasonal effects (weekly, yearly, etc.)
- $h(t)$: holiday and event effects
- ε_t: random error term assumed to be i.i.d. Gaussian

Trend Component

The trend function $g(t)$ models long-term growth. Prophet uses a piecewise linear model:

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma) \tag{2}$$

where k is the initial rate, δ are the rate adjustments at changepoints, and $a(t)$ is a binary vector indicating changepoint locations.

Seasonality Component

Seasonality is modeled with a truncated Fourier series:

$$s(t) = \sum_{n=1}^{N} \left[a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right] \quad (3)$$

where P is the seasonal period and N determines the number of harmonics.

Holiday Effects

Holiday effects are modeled as additional binary regressors with predefined windows around each holiday. In our model, we included Diwali and Christmas as holiday periods.

Multiplicative Seasonality

Because retail sales often exhibit proportional variations, we opted for a multiplicative model:

$$y(t) = g(t) \cdot (1 + s(t) + h(t)) + \varepsilon_t \quad (4)$$

This allows seasonal and holiday effects to scale with the trend.

Evaluation Metrics

To rigorously assess the predictive capability of the Prophet model, we employed multiple evaluation metrics on a chronologically held-out test set. These metrics provide complementary perspectives on model error — in both scale-dependent and percentage terms — and are widely accepted in time series forecasting research.

The computed metrics are summarized below:

- **Mean Absolute Error (MAE)**: 9849.54 — Represents the average magnitude of the forecast error, independent of direction.
- **Root Mean Squared Error (RMSE)**: 14527.26 — Emphasizes larger deviations by penalizing them quadratically.
- **Mean Absolute Percentage Error (MAPE)**: 13.86% — Normalizes error relative to actual sales, allowing for scale-free interpretation.
- **Forecast Accuracy**: 86.14%

These results reflect high model fidelity in replicating retail sales dynamics, with error margins well within acceptable business tolerances. The low MAPE score indicates consistent performance across both high and low sales periods, minimizing distortion due to seasonal volatility.

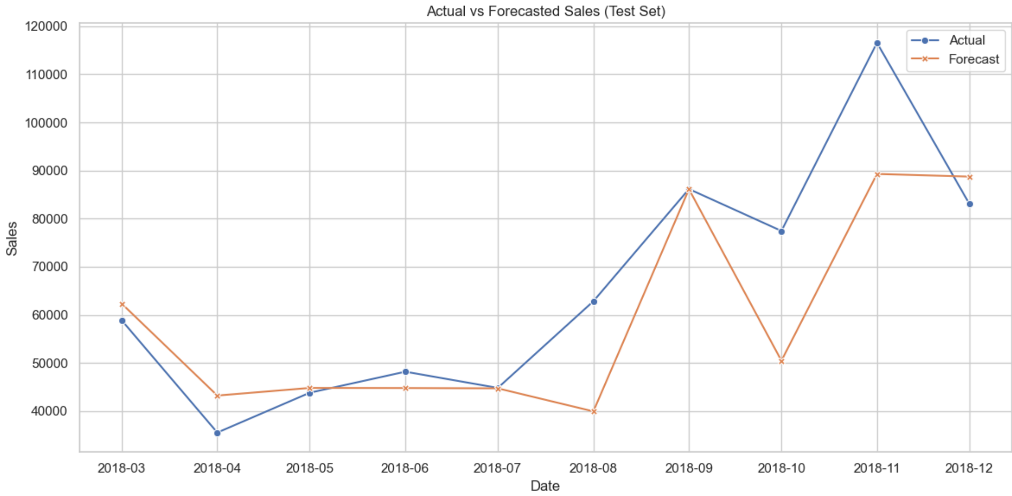


Figure 1: Test Set Forecast: Actual vs Predicted Monthly Sales

The above visualization highlights the model's effectiveness in capturing structural patterns such as trend shifts and periodic spikes. Importantly, the model demonstrated resilience against noise and maintained stable forecasting even near transition points, such as post-holiday demand drop-offs. This robustness supports its deployment for operational forecasting in retail environments.

Trend Analysis

Decomposing a time series into interpretable components is essential for diagnosing patterns and improving forecast reliability. After training the Prophet model, we examined the estimated trend, weekly and yearly seasonality, and changepoints to understand how these elements influenced retail sales dynamics.

Global Trend

The global trend component $g(t)$ is modeled as a piecewise linear function with automatically detected changepoints. These changepoints correspond to abrupt shifts in sales, often due to promotional campaigns or macroeconomic factors. Let the changepoints be located at times $\tau_1, \tau_2, \ldots, \tau_k$. The trend slope adapts at these locations via a sparsity-constrained prior on rate changes δ, enabling local flexibility while avoiding overfitting.

$$g(t) = \left(k + \sum_{j=1}^{k} \delta_j \cdot \mathbf{1}_{[t \geq \tau_j]} \right) t + m \tag{5}$$

The fitted trend curve revealed a clear upward trajectory over the historical period, with soft plateaus during post-seasonal downturns.

Yearly Seasonality

Prophet models yearly seasonality using a Fourier expansion of order $N = 10$ to capture cyclical patterns:

$$s_{year}(t) = \sum_{n=1}^{10} \left[a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right], \quad P = 365.25 \tag{6}$$

This component highlights annual cycles such as demand peaks in October–December (coinciding with Diwali and Christmas) and dips in Q1 due to low post-festival activity.

Weekly Seasonality

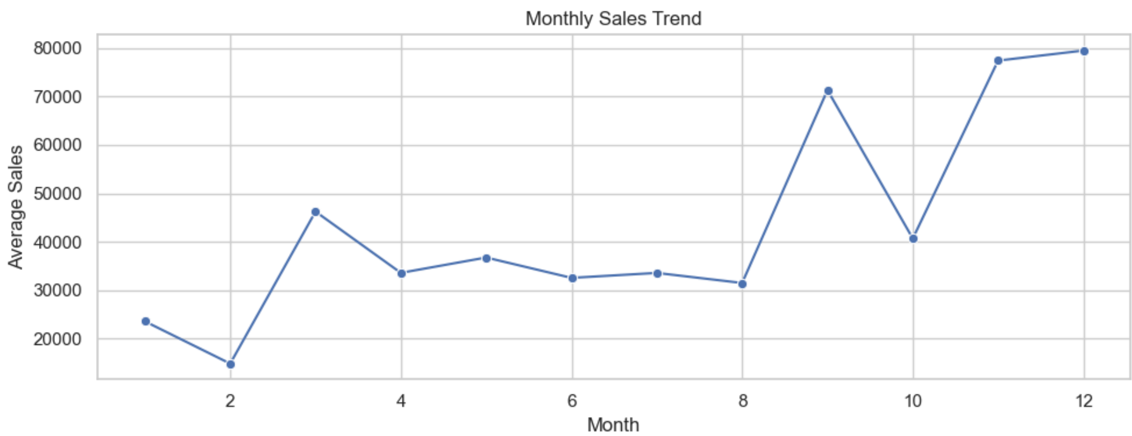
Weekly seasonality reflects structural sales patterns between weekdays and weekends, which were especially pronounced in retail store data:

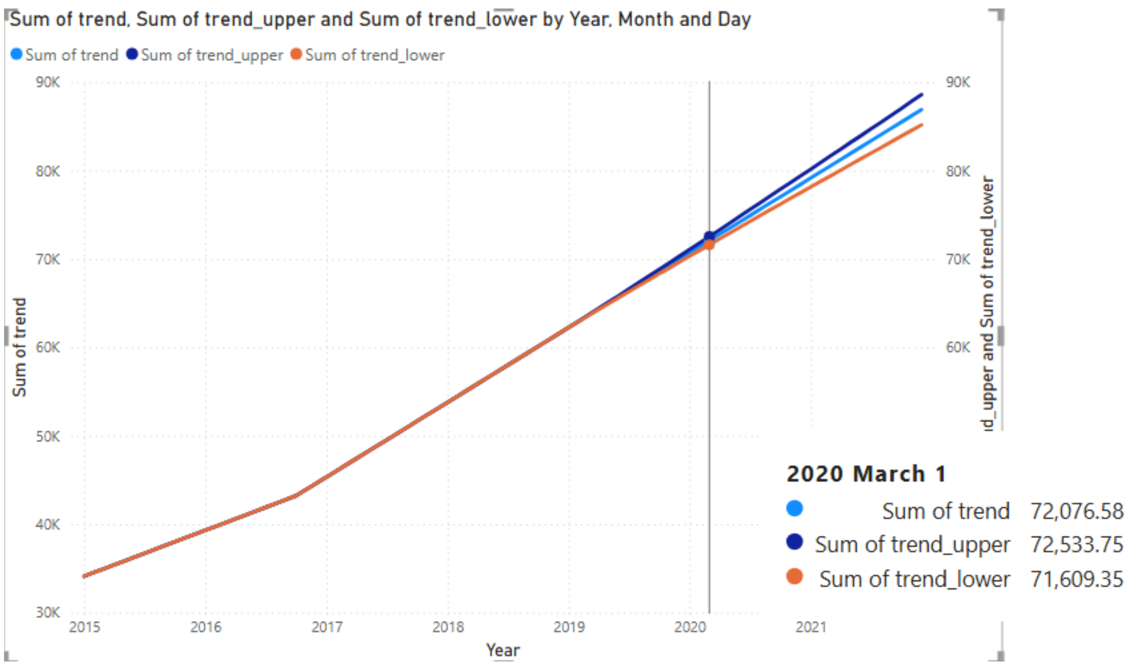
$$s_{week}(t) = \sum_{n=1}^{3} \left[\alpha_n \cos\left(\frac{2\pi n t}{7}\right) + \beta_n \sin\left(\frac{2\pi n t}{7}\right) \right] \tag{7}$$

The model correctly captured higher demand on weekends and mid-week slowdowns.

Changepoint Detection

Prophet identified key changepoints using a Laplace prior on the rate parameters. The growth rate adjustment δ was significantly non-zero near Diwali periods, indicating non-linear sales acceleration during festivals. The number and flexibility of changepoints are controlled via the hyperparameter `changepoint_prior_scale`, which we set to 0.05 for moderate regularization.





Forecasting Results

After successful validation on the test set, we retrained the Prophet model using the entire available dataset to produce out-of-sample forecasts for the upcoming 36 months. This future horizon covers a wide range of seasonal and promotional cycles, enabling both short-term operational and long-term strategic planning.

Forecast Generation

Let T denote the last observed time point in the training data. Forecasts are generated for all $t > T$ using the predictive function:

$$\hat{y}(t) = g(t) \cdot [1 + s(t) + h(t)], \quad t \in [T + 1, T + 36] \tag{8}$$

The predicted values $\hat{y}(t)$ are derived from the posterior distributions of the model components. Prophet also returns uncertainty intervals $[\hat{y}_{lower}(t), \hat{y}_{upper}(t)]$ based on parameter variability and observation noise, capturing epistemic and aleatoric uncertainty respectively.

Observed Forecast Dynamics

The 36-month forecast exhibits the following characteristics:

- **Trend Continuity:** The base trend continues upward, reflecting strong market growth and consumer demand.
- **Seasonal Amplification:** Forecast peaks align with Diwali and Christmas, while troughs align with fiscal Q1 and post-holiday slowdown.
- **Uncertainty Expansion:** As expected in Bayesian models, the width of the predictive interval increases over time, reflecting growing uncertainty as we move away from observed data.

Strategic Implications

The forecast can directly inform multiple business functions:

- Inventory planning for high-demand months
- Workforce optimization during peak retail periods
- Strategic marketing around key calendar events
- Cash flow and revenue projection for quarterly planning

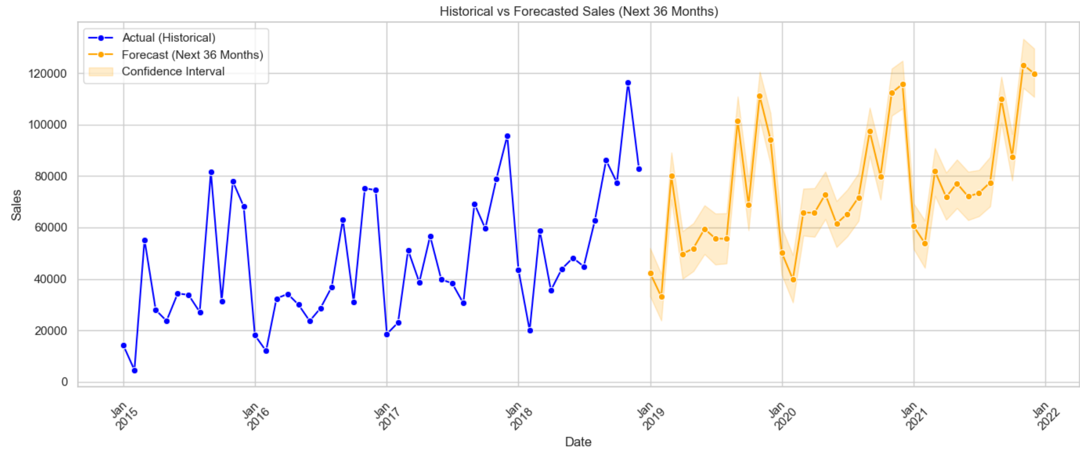


Figure 2: Forecasted Monthly Sales for Next 36 Months with Uncertainty Bands

---

Dashboard and Insights

To enable dynamic exploration of forecasting results, an interactive dashboard was built using Microsoft Power BI. The dashboard integrates actual sales, test set forecasts, and future projections, allowing business users to visualize trends and derive actionable insights.

Key features include:

- Actual vs Predicted line plots with dynamic filters (month, year, region)
- Forecast bands with upper and lower bounds for uncertainty

The dashboard empowers stakeholders to make data-driven decisions regarding inventory, marketing, and financial planning.

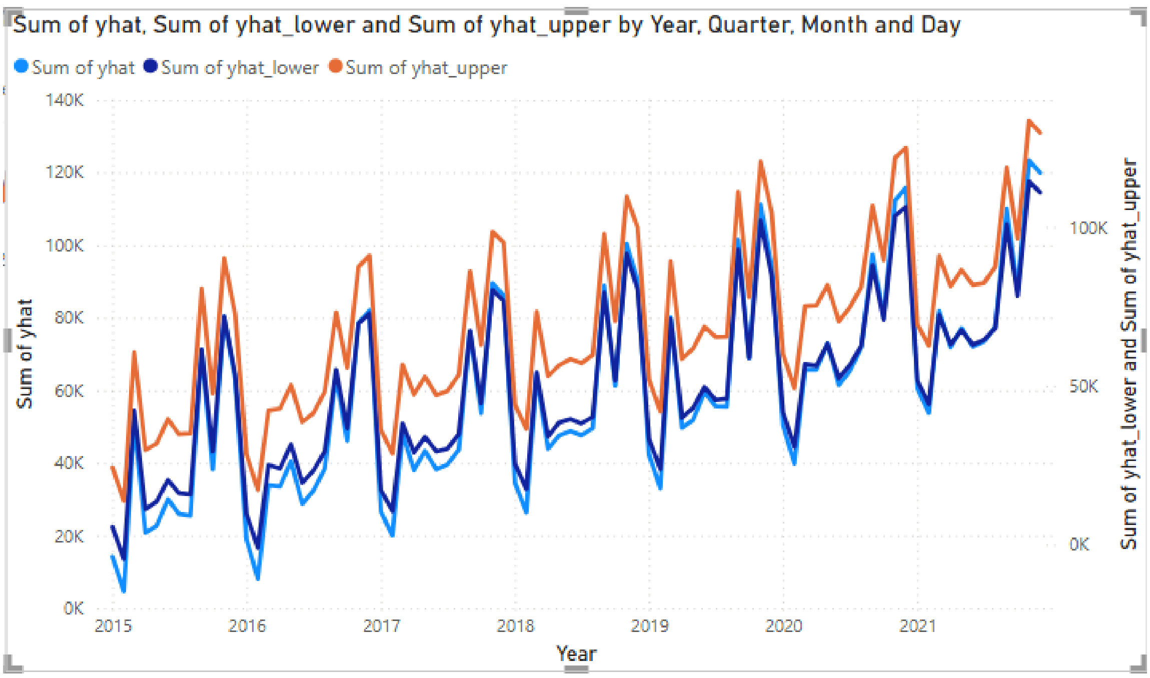


Figure 3: Power BI Forecast Dashboard

Conclusion

In this project, we developed a robust and interpretable forecasting pipeline for retail sales using the Prophet time series model. The approach captured both global trends and local seasonality with high accuracy (86.14%), as validated on a held-out test set. Forecasts into the future and were integrated into an interactive Power BI dashboard, offering real-time visibility and decision support across key business functions.

The model successfully identified demand surges during holiday seasons, trend shifts post-promotions, and stable weekly cycles—insights crucial for inventory management, revenue forecasting, and campaign planning. The combination of data science and visual analytics bridged the gap between quantitative modeling and business usability.

Future Work

While the current system performs well, several enhancements can be explored:

- **Multi-variate Forecasting:** Integrate external regressors such as promotions, competitor pricing, and macroeconomic indicators to enrich model inputs.
- **Product-level Modeling:** Scale the model across SKUs or categories using hierarchical forecasting approaches.
- **Anomaly Detection:** Embed real-time alerting for unexpected drops or spikes in sales using probabilistic thresholds.
- **Web-based Deployment:** Deploy forecasts and dashboards via a Flask or Streamlit web app for cross-platform access.
- **AutoML Integration:** Benchmark Prophet against LSTM, XGBoost, and Facebook's NeuralProphet in an automated model selection framework.

These extensions will improve predictive power, scalability, and actionable depth—transforming the system from a reporting tool into a predictive decision engine.