Adhil Payingalil

Forecasting Weekly Demand for a Soft Drink Product – A Time Series Modeling Challenge

Abstract— **This report analyzes high-frequency (15-minute interval) demand data for a flagship soft drink product from March 1st to March 21st, 2005, to understand demand patterns and provide a 2-week forecast. Exploratory data analysis revealed dominant daily (evening peaks) and weekly (weekday highs, weekend lows) seasonality, alongside high volatility. Statistical tests confirmed short-term stationarity. Three models (SARIMA, Holt-Winters, Prophet) were evaluated, with Prophet showing the best performance (RMSE: 12.43). Prophet forecasts an average demand of ~21.65 units per 15-minute interval for the next two weeks. The core recommendation is to align production with recent weekday averages while maintaining operational flexibility and monitoring real-time sales, given moderate forecast confidence due primarily to the short historical data window.**

I. EXECUTIVE SUMMARY

Context & Objective: This report analyzes high-frequency (15-minute interval) demand data for a flagship soft drink product, recorded from March 1st to March 21st, 2005. The primary goal was to understand the underlying demand drivers within this period and generate a reliable 2-week demand forecast to improve production planning, minimize stockouts, and reduce spoilage costs.

Key Findings: Analysis revealed demand is overwhelmingly driven by predictable time-based patterns rather than a consistent trend over this short period. The dominant factors are:

- **Strong Daily Seasonality**: A clear cycle exists within each day, characterized by near-zero demand overnight and significant peaks in the evening (18:00-19:00).
- **Strong Weekly Seasonality**: A stark difference exists between weekdays (higher, more variable demand) and weekends (consistently very low demand).
- **High Volatility**: Demand fluctuates significantly between consecutive 15-minute intervals.
- **Stationarity**: Statistical tests indicated the series was stationary over the observed ~3 weeks, meaning its basic statistical properties didn't systematically drift.

Modeling & Forecast: Three time series models (SARIMA, Holt-Winters, Prophet) were rigorously evaluated. The Prophet model significantly outperformed the others on unseen test data, yielding the lowest Root Mean Squared Error (RMSE) of 12.43 units per 15-minute interval. Based on Prophet, the forecast for the next two weeks anticipates an average demand of approximately 21.65 units per 15-minute interval, translating to a total of ~29,093 units over the period. This level aligns closely with the average demand observed during recent weekdays.

Core Recommendation: The primary recommendation is to align initial production planning for the next two weeks with recent average weekday demand levels. However, confidence in the precise forecast level is moderate due to the inherent high volatility of the data and the significant limitation of having only ~3 weeks of historical data. Therefore, close monitoring of real-time sales and maintaining operational flexibility are critical. External factors not captured by this model (e.g., weather, promotions) will influence actual demand.

II. VISUAL INSIGHTS & DATA CHARACTERISTICS (EDA)

Exploratory Data Analysis (EDA) provided crucial insights into the demand dynamics:

Overall Time Series: The plot of demand over the ~3 weeks immediately highlighted intense activity during weekdays and near-complete lulls during weekends. While visually noisy, the 1-day rolling average helped emphasize the dominant weekly cycle. No clear long-term upward or downward trend was discernible in this short timeframe.

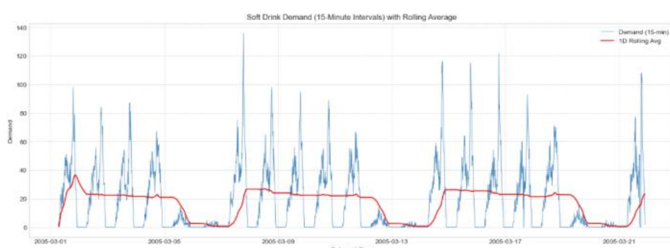


Figure 1. Full Time Series Plot

Intra-Day Pattern: Average demand analysis by hour starkly illustrated the daily rhythm. Demand is virtually non-existent from 22:00 to 05:00, ramps up gradually through the morning and early afternoon, experiences a significant peak between 18:00 and 19:00, and drops off rapidly afterward. This indicates critical evening hours for potential sales.

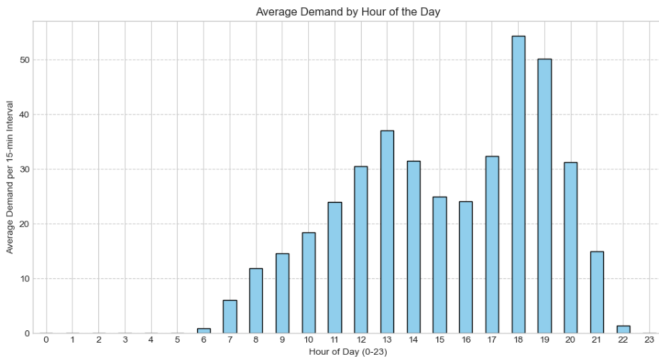


Figure 2. Average Hourly Demand Plot

Weekly Pattern: Comparing average demand across days of the week showed substantially higher levels from Monday to Friday compared to Saturday and Sunday. Weekends consistently averaged near-zero demand per interval. Box plots further confirmed this, showing not only higher median demand on weekdays but also much greater variance (spread) and the presence of numerous high-demand outliers, which were absent on weekends. (Reference: Average Daily, Weekday/Weekend, and Box Plots)

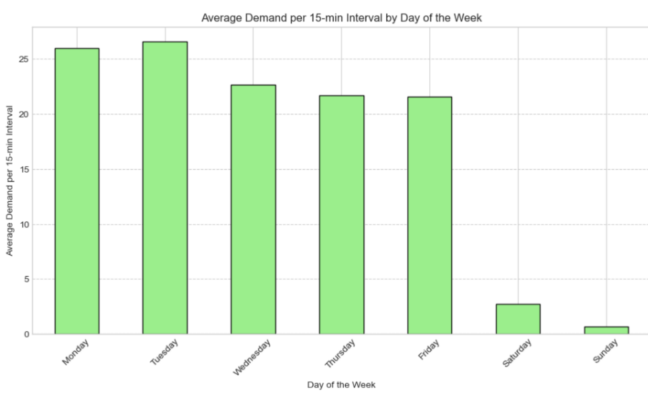


Figure 3. Average Daily Plot

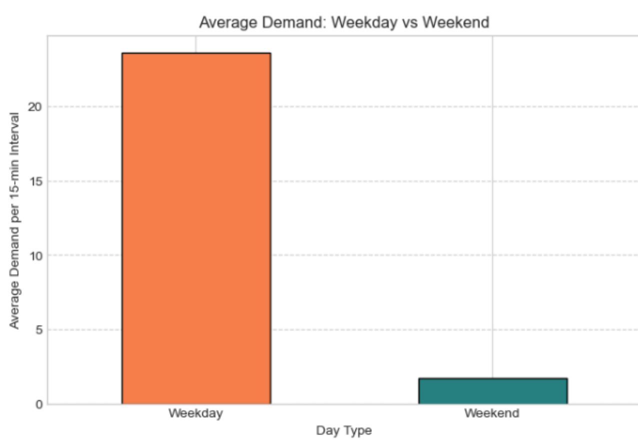


Figure 4. Average Weekday/Weekend Plot

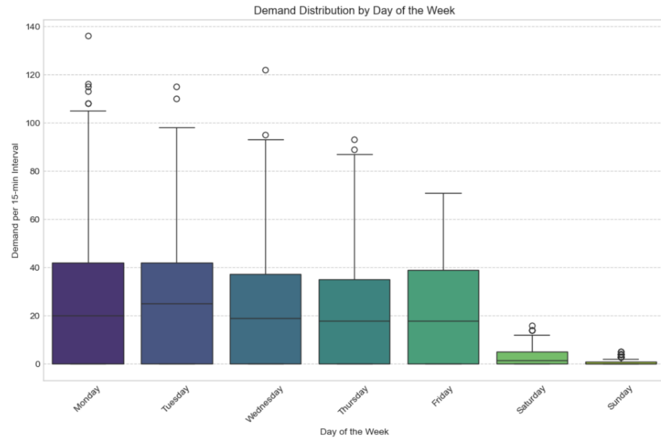


Figure 5. Daily Distribution Box Plot

Noise & Volatility: The raw 15-minute data exhibit considerable noise, meaning frequent and sometimes large jumps between consecutive intervals. This suggests factors influencing demand can change rapidly even within short periods, making precise forecasting inherently challenging.

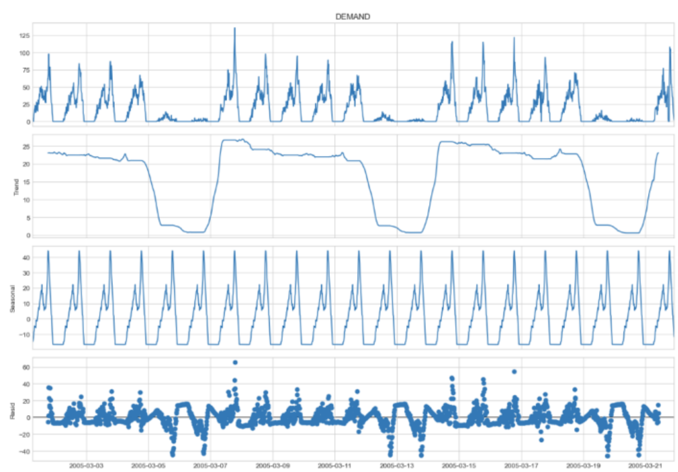


Figure 6. Demand Trend & Seasonality

III. METHODOLOGY

A. *Data Preperation*

Raw 15-minute data was loaded and preprocessing to ensured data integrity by:

- Parsing date/time and setting a chronological Date time Index.
- Imputing any missing 'DEMAND' values with 0.
- Removing duplicate timestamps (keeping first occurrence).
- Ensuring time series continuity by reindexing to a full 15-minute frequency range and filling any created gaps in 'DEMAND' with 0.

- Extracting 'HOUR', 'DAY_OF_WEEK_NAME', 'DAY_TYPE' for EDA.

B. *Exploratory Analysis*

Utilized visualizations (time series plots, rolling averages, bar charts of hourly/daily averages, box plots) to identify key patterns, seasonality, and noise characteristics.

C. *Time Series Analysis*

Stationarity Assessment: The Augmented Dickey-Fuller (ADF) test was performed on the original demand series. The result (p-value < 0.05) indicated stationarity, meaning differencing was not required for modeling.

Seasonal Decomposition: An additive seasonal decomposition (with a period of 96, representing the daily cycle in 15-min intervals) was used to visually separate the short-term trend, the strong daily seasonal component, and the remaining residual noise.

D. *Modeling Strategy*

The dataset was split chronologically, reserving the last 5 days (480 intervals) for testing model performance. Three distinct time series forecasting models were implemented:

- **SARIMA**: A Seasonal Autoregressive Integrated Moving Average model. Orders were selected based on stationarity (d=0, D=0) and assumed simple AR/MA structures for seasonal (P=1) and non-seasonal (p=1, q=1) components, optimized for speed given the high frequency.
- **Holt-Winters Exponential Smoothing (TES)**: Utilized additive trend and additive seasonality (period=96), allowing the model to learn smoothing parameters automatically. Additive seasonality was chosen for its ability to handle zero-demand values.
- **Prophet**: Facebook's forecasting tool, configured with daily and weekly seasonality detection and a demand floor of 0 to prevent unrealistic negative forecasts.

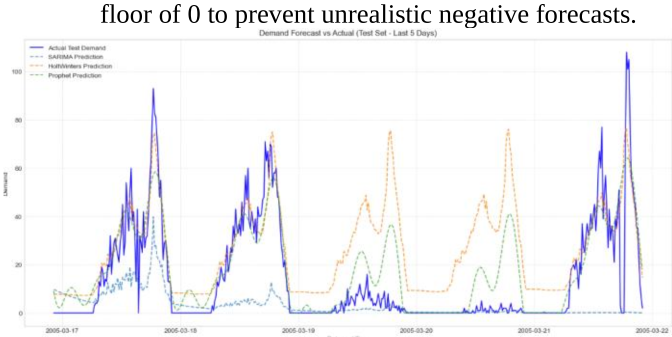


Figure 7. Actual vs Demand Forecasted

E. *Evaluation & Selection*

Model performance was compared on the test set using MAE, MSE, and RMSE. The model achieving the lowest RMSE was selected as the 'best' for forecasting. Prophet significantly outperformed the others.

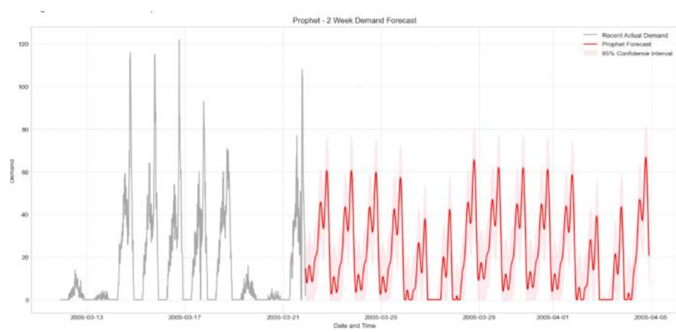
Evaluation Metrics Summary:

	MAE	MSE	RMSE
SARIMA	12.488510	484.579734	22.013172
HoltWinters	15.023013	440.803944	20.995331
Prophet	7.681310	154.382746	12.425085

Figure 8. Evaluation Metrics Summary

F. *Forecasting*

A 2-week forecast was generated using the best model (Prophet), including 95% confidence intervals.



IV. Recommendations

Best Performing Model: The Prophet model demonstrated superior predictive accuracy on the test data, with an RMSE of 12.43 units per 15-minute interval, significantly lower than Holt-Winters (21.00) and SARIMA (22.01).

Two-Week Forecast: Prophet predicts an average demand of approximately 21 units per 15-minute interval over the next two weeks (totaling 29,093 units approximately). It also provides a 95% confidence interval, indicating the likely range of demand (e.g., typically between ~7 and ~36 units for a mid-forecast point, although this varies).

Production & Inventory Implications: The forecast average aligns well with recent weekday demand levels. Recommendation: Use recent weekday average demand as the primary baseline for production planning for the next two weeks. Avoid over-reacting to the model's interval-level fluctuations due to inherent noise. Ensure sufficient stock, particularly for peak evening hours (18:00-19:00) on weekdays. Weekend requirements appear minimal based on this historical snapshot.

Confidence Assessment: Confidence in the forecast is moderate. While Prophet was the best model, an RMSE of ~12.4 units indicates that predictions for any specific 15-minute slot can still be off considerably. The primary value lies in capturing the overall daily and weekly patterns and average levels. The very short historical data window (only 3 weeks in March) is a major limitation – seasonal patterns observed might not represent other times of the year.

Key Actions:

Implement Baseline: Set initial production/stocking targets based on recent weekday averages, informed by the Prophet forecast level (~21.65 units/interval average).

Monitor Actively: Establish a process to track real-time sales against the forecast daily. Identify significant deviations quickly.

Maintain Flexibility: Ensure production schedules and logistics can be adjusted within the two-week period if actual demand trends higher or lower than expected.

Account for Externalities: Factor in known upcoming events, planned promotions, significant weather changes, or competitor activities that could override historical patterns.

V. CONCLUSION

This analysis provides valuable insights into the short-term demand dynamics, highlighting strong daily and weekly cycles. The Prophet model offers the most reliable forecast based on the available data. However, due to the data's short duration and high volatility, the forecast should be used as informed guidance rather than a precise prediction. Proactive monitoring and operational adaptability are paramount for successful inventory management over the next two weeks. A future analysis incorporating a longer time series (ideally >1 year) is strongly recommended for more robust forecasting