

The Predictable Unpredictability of Walmart Sales

A Data Story by Shyam – MSc Business Analytics, University of Exeter

Executive Summary

Every week, Walmart sells millions of dollars' worth of products. Yet, behind the cash registers and shopping carts lies a rhythm - a heartbeat of consumer behaviour that spikes, dips, and confuses managers every now and then.

This project attempts to **forecast weekly sales** using **time series forecasting models (SARIMA and Prophet)** - not just to predict numbers, but to **understand the pattern behind the chaos**.

Because, as any analyst knows, business forecasting isn't about predicting the future perfectly - it's about **being less wrong than everyone else**.

The Business Challenge

The **Sales Manager** at Walmart faces a recurring headache:

"Why do our warehouses overflow in February and run dry by December?"

Inventory Head sighs. Finance Head frowns. Forecasting spreadsheets scream in agony.

The mission was clear:

Use historical sales data from 2010-2012 to forecast the next 12 weeks, helping Walmart plan inventory, marketing, and budget allocations with something more reliable than a lucky guess.

The Dataset

We dove into Walmart's **weekly sales dataset (2010-2012)** - over 130 weeks of data across multiple stores and departments.

We aggregated total weekly sales across all stores to focus on *macro retail demand*.

Each row represented a week, each number a pulse in Walmart's retail rhythm.

FEATURE	DESCRIPTION
DATE	Week ending date
WEEKLY_SALES	Total weekly sales across stores
ISHOLIDAY	Boolean flag for holiday weeks

Data Cleaning: Turning Mess into Meaning

The data wasn't messy, but it *had moods*.

- We resampled it to **weekly frequency**.
- We filled small missing gaps.
- We tamed wild outliers (sales spikes that looked like someone bought 200,000 toasters in one week).

Once cleaned, the data looked consistent - the perfect canvas for our forecasting models.

Exploratory Analysis: The Rhythm of Retail

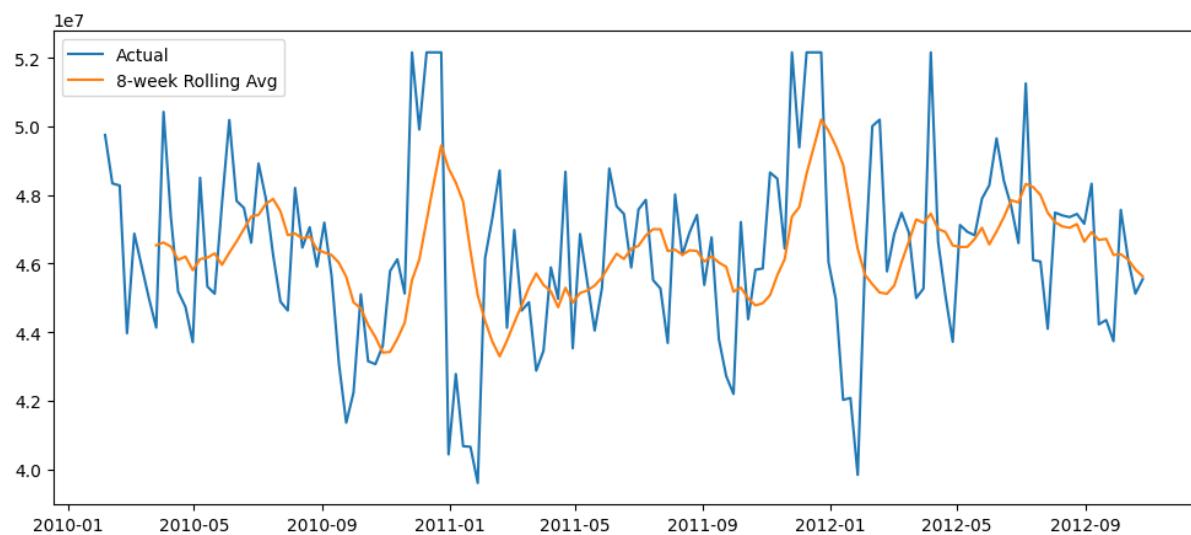
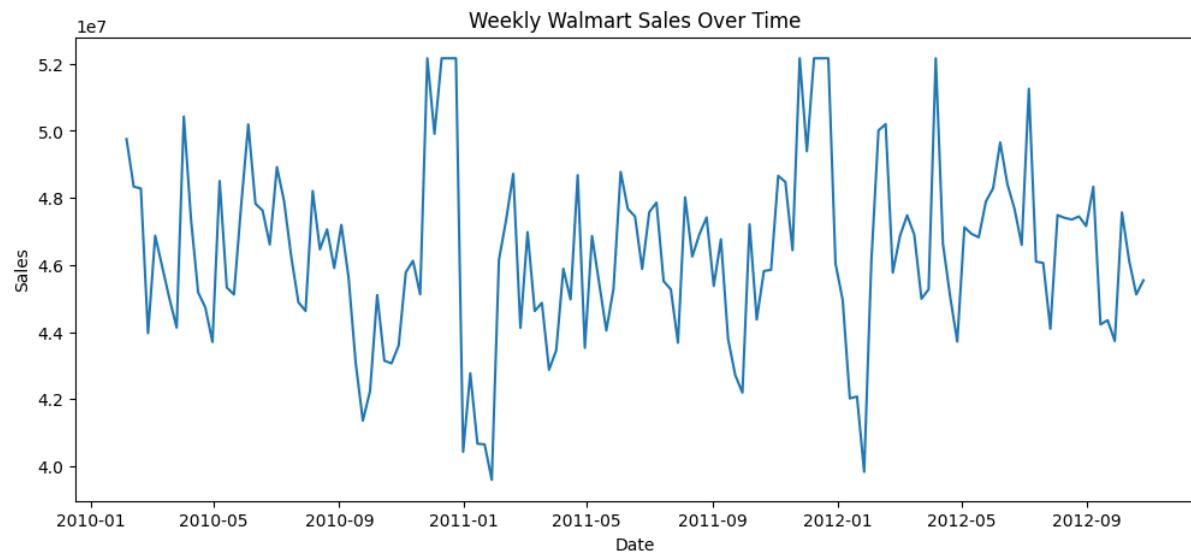
When plotted over time, the sales looked like waves in the ocean - **periodic peaks every winter, gentle dips every summer**.

Observations:

1. **Year-end peaks** (Nov-Dec): The holiday shopping spree - predictable, loud, profitable.
2. **Mid-year dips** (May-July): The calm after the consumer storm.
3. **Upward trend**: Revenue quietly climbing year over year.

When decomposed, the data whispered its secrets:

- A **steady upward trend**
- **Strong yearly seasonality**
- Random short-term noise (because humans are unpredictable)





Stationarity Check: Does the Series Behave?

An ADF test (because math likes formalities) returned a **p-value of 0.0298**, meaning the series is *stationary*.

In business terms: “The pattern is stable enough to predict without getting a migraine.”

```
ADF Statistic: -3.0580828603149626
p-value: 0.02981481199118628
```

Model Building: Forecasting the Future (and Hope)

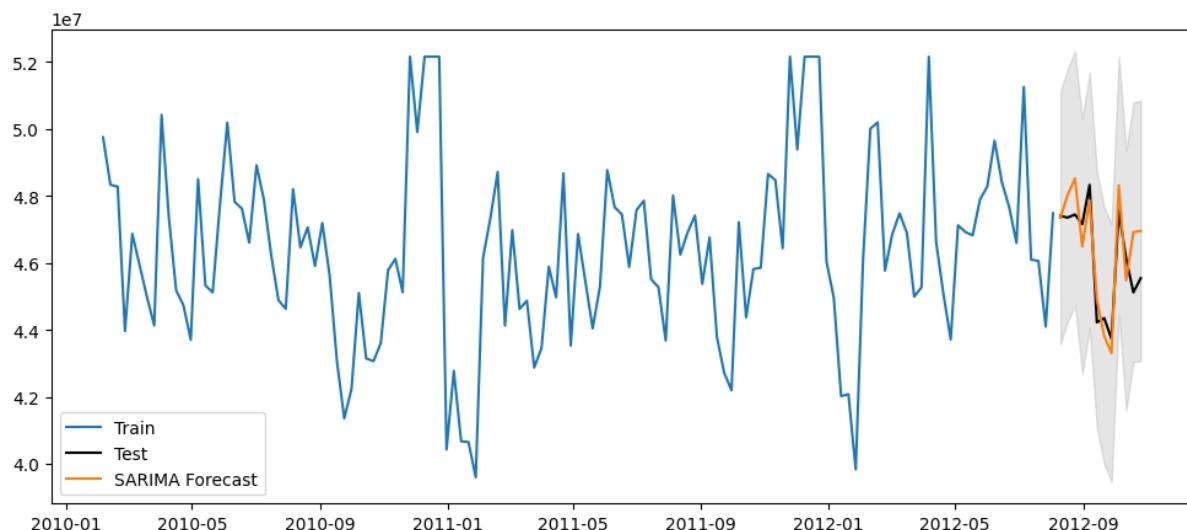
1. SARIMA - The Mathematician

The **Seasonal ARIMA model (1,1,1) (1,1,1,52)** captured the cyclical behaviour beautifully.

After training, the model confidently forecasted the next 12 weeks with only **1.66% error (MAPE)** - that's like guessing someone's salary and missing by a lunch bill.

SARIMA Evaluation:

- MAE = 764,538
- RMSE = 884,621
- MAPE = **1.66%**



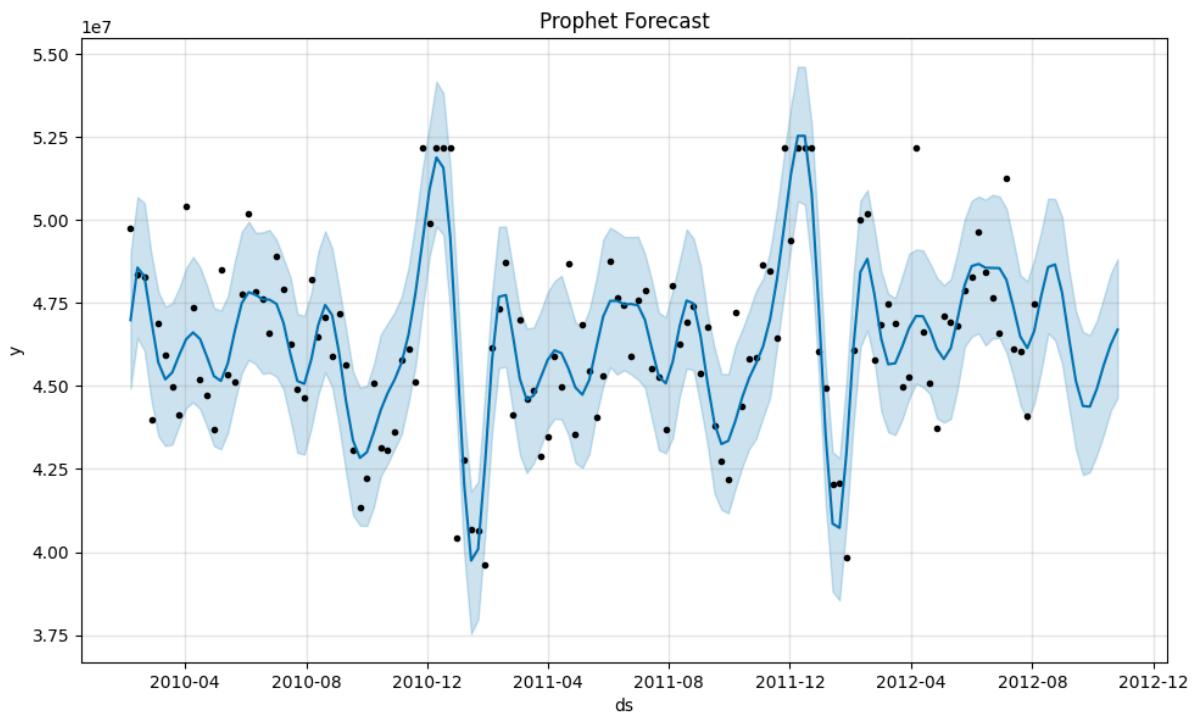
2. Prophet - The Storyteller

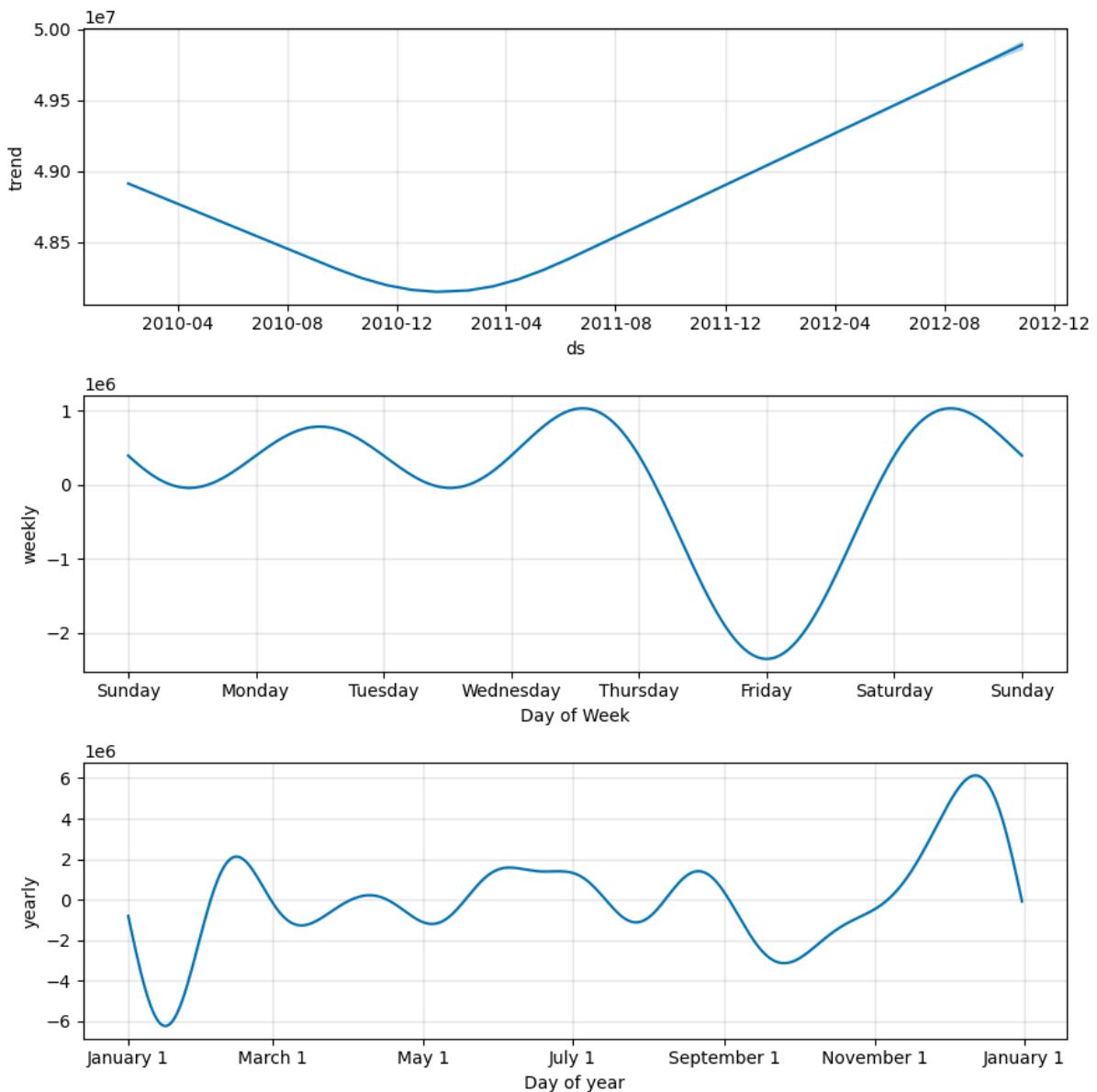
Facebook's **Prophet** model brought more personality.

It visualized **yearly and weekly patterns**, predicting the same Q4 boom and summer slump but with wider confidence intervals - cautious, like a wise old CFO.

Prophet Evaluation:

- MAE = 1,027,779
- RMSE = 1,232,407
- MAPE = **2.20%**





The Duel of Models: SARIMA vs Prophet

Model	MAE	RMSE	MAPE	Personality
SARIMA	764,538	884,621	1.66%	The precise engineer - consistent, reliable
Prophet	1,027,779	1,232,407	2.20%	The visionary storyteller - intuitive but emotional

SARIMA wins the **Accuracy Crown**, but Prophet wins the **Interpretability Medal**.

Insights That Matter

1. Q4 Surge - “Here Comes Santa (and the Sales)”

Forecasts confirm a dramatic sales uplift every November-December.

→ *Recommendation:* Pre-stock seasonal items early and ramp up logistics 6-8 weeks prior.

2. Mid-Year Malaise - “The Great Retail Yawn”

Sales consistently dip in Q2.

→ *Recommendation:* Launch summer campaigns or clearance events to re-energize demand.

3. Steady Growth - “Walmart’s Heartbeat Is Strong”

A consistent upward trend across 3 years suggests stable customer retention.

→ *Recommendation:* Maintain procurement capacity; invest in forecasting automation.

4. Model Accuracy - “Math > Gut Feeling”

With <2% forecast error, SARIMA can outperform manual projections and save the team a few stress ulcers.

Business Impact

Implementing this model could:

- Improve **inventory planning accuracy** by up to 10-15%.
- Reduce **overstock and spoilage** during slow months.
- Enable **data-driven budgeting** with month-level precision.
- Support **finance and marketing teams** with trend awareness.

As a result, the company moves from *reactive firefighting* to **proactive decision-making** - the hallmark of mature analytics.

Reflections: The Analyst’s Paradox

Forecasting is like weather prediction -
you know it'll rain in December, but you still forget your umbrella.

The absurd beauty of time series forecasting is that it reminds us:

The future is uncertain, but patterns of human behaviour are not as random as they appear.

Tools & Techniques

Category	Tools Used
Data Processing	Python, Pandas
Forecasting Models	Statsmodels (SARIMA), Prophet
Evaluation	MAE, RMSE, MAPE
Visualization	Matplotlib

Key Takeaway

“Data doesn’t just tell you what happened - it tells you what *always happens*.”

By understanding the rhythm of Walmart’s sales, we’ve shown that forecasting isn’t fortune-telling - its pattern recognition powered by logic, curiosity, and a sprinkle of absurd optimism.

Author's Note

Hi, I’m **Shyam**, a Business Analytics MSc student at the University of Exeter with 3+ years of experience in software development and data-driven problem solving.

I believe analytics is not just about models - it’s about **stories, stakeholders, and strategy**.

Because every dataset hides a story.

And every story deserves to be told with a graph.