
Walmart Sales Forecast

BY ALICIA QU



PROBLEM

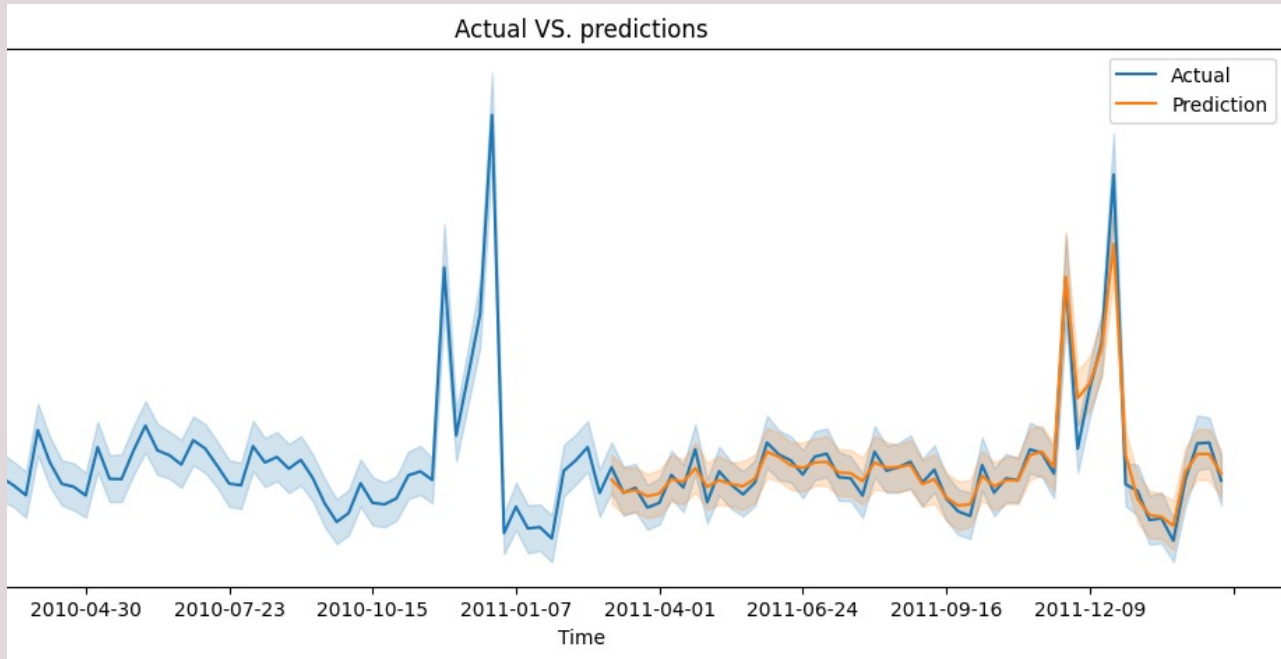
Current prediction on future sales is based on business intuition which is not accurate.



GOAL

Improve the forecast accuracy to optimize Walmart's inventory management, logistics management and promotion strategy.

OUTCOME



- Generated a model with a 96.9% accuracy rate.
- Detected seasonal patterns and trends in sales.
- Identified key features that significantly influence sales.

Approach

INPUTS:

1. Historical sales data
2. Store information
3. Environmental and economic factors

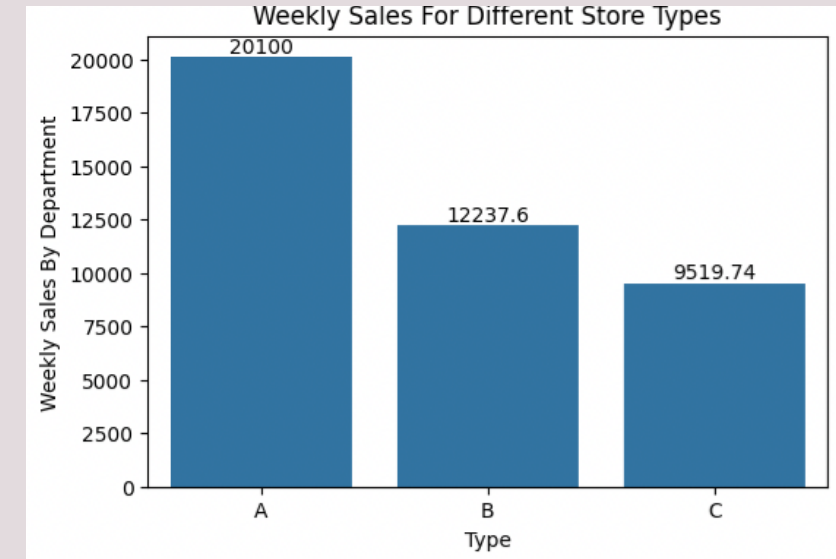
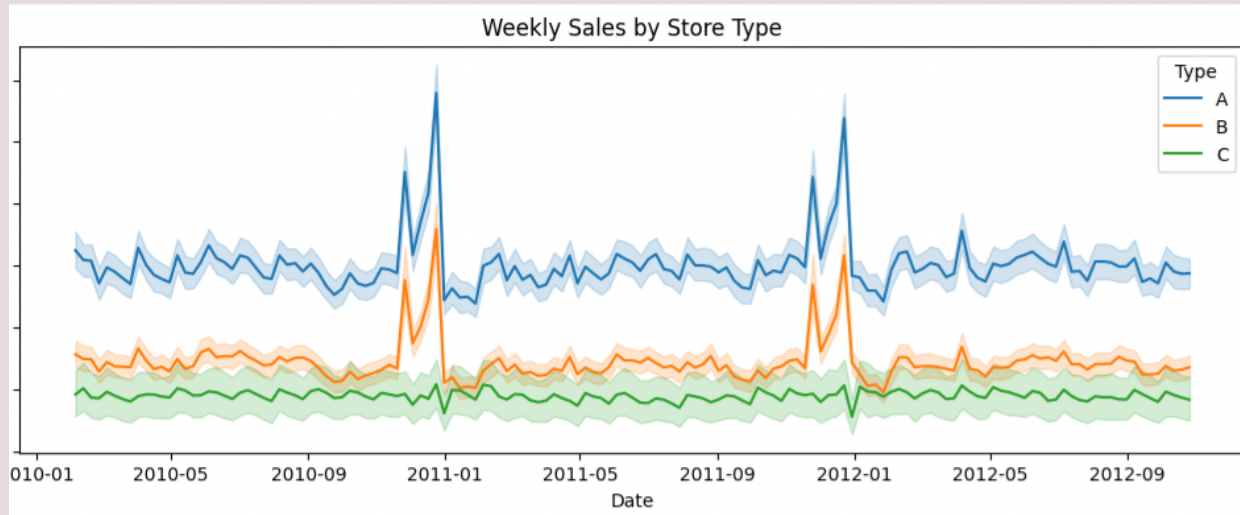
MODELING

1. ML Models
2. Statistic Time Series Model
3. Deep Learning Model

OUTPUTS:

Weekly Sales by
Department Forecasting

EXPLORATORY DATA ANALYSIS

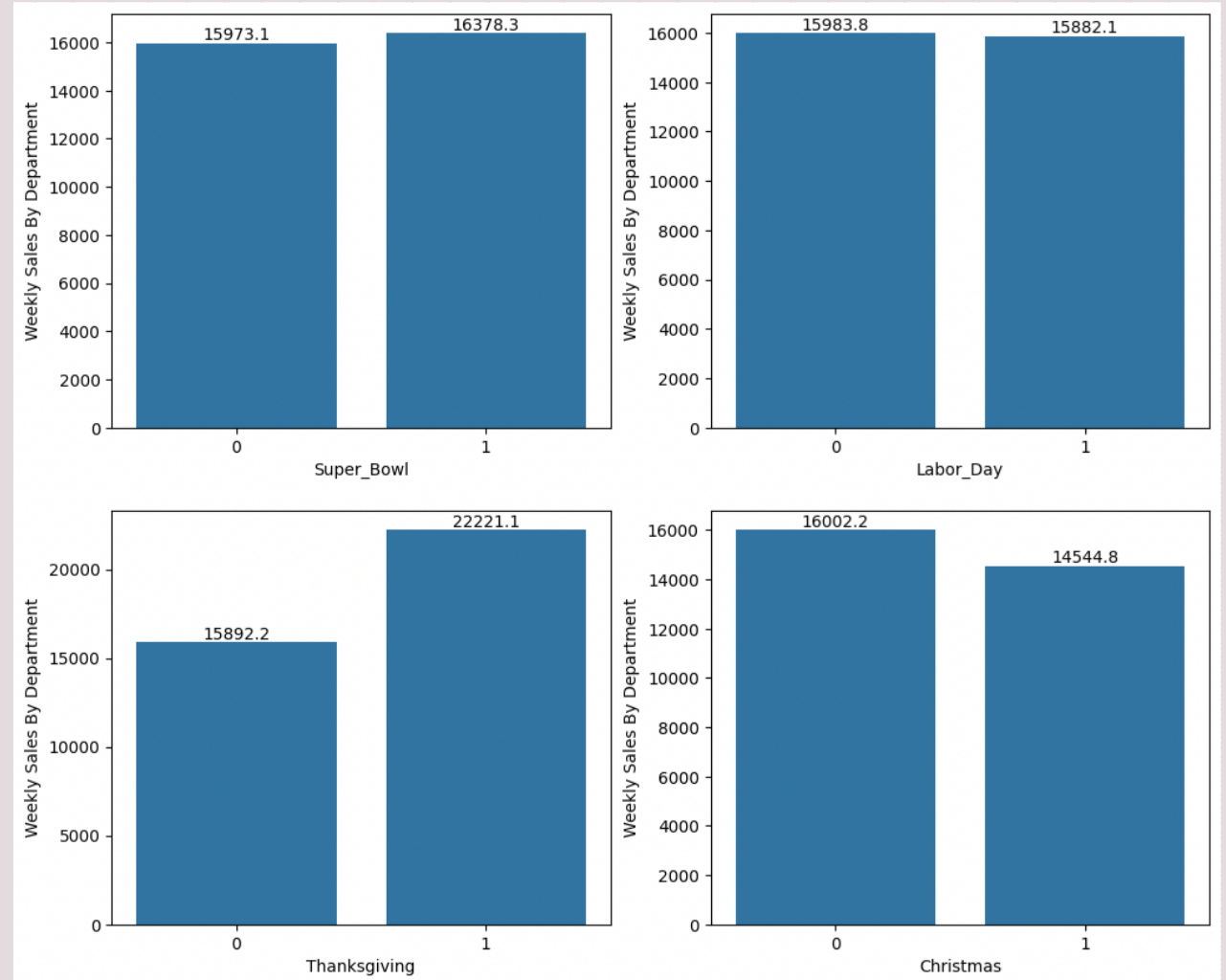


Larger stores consistently have higher sales.

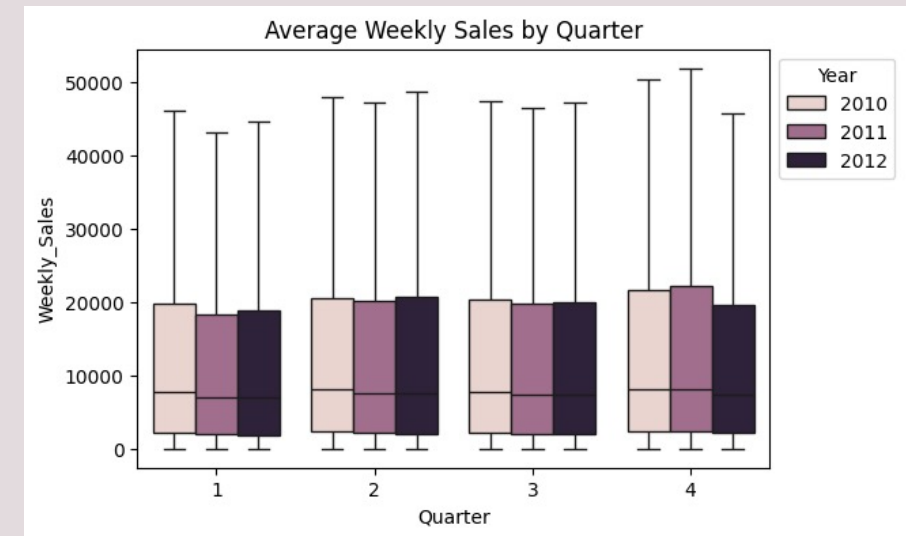
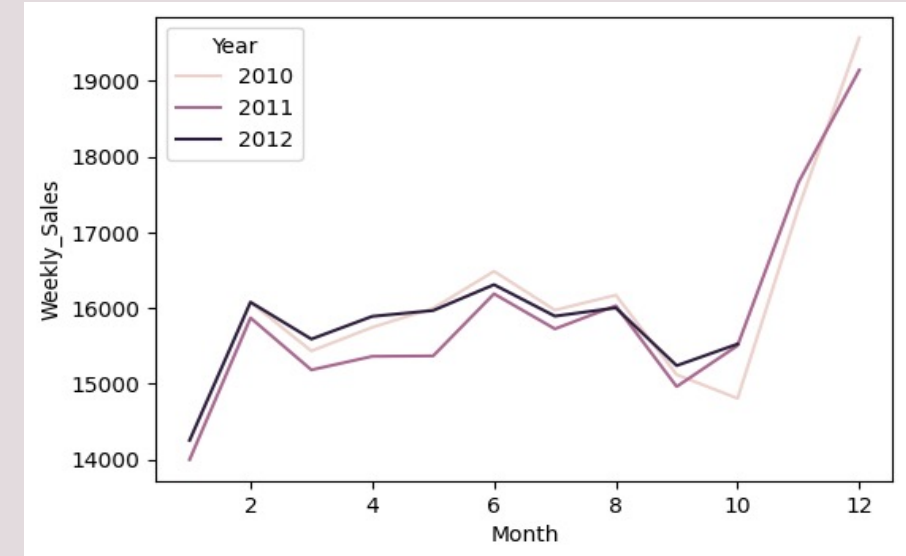
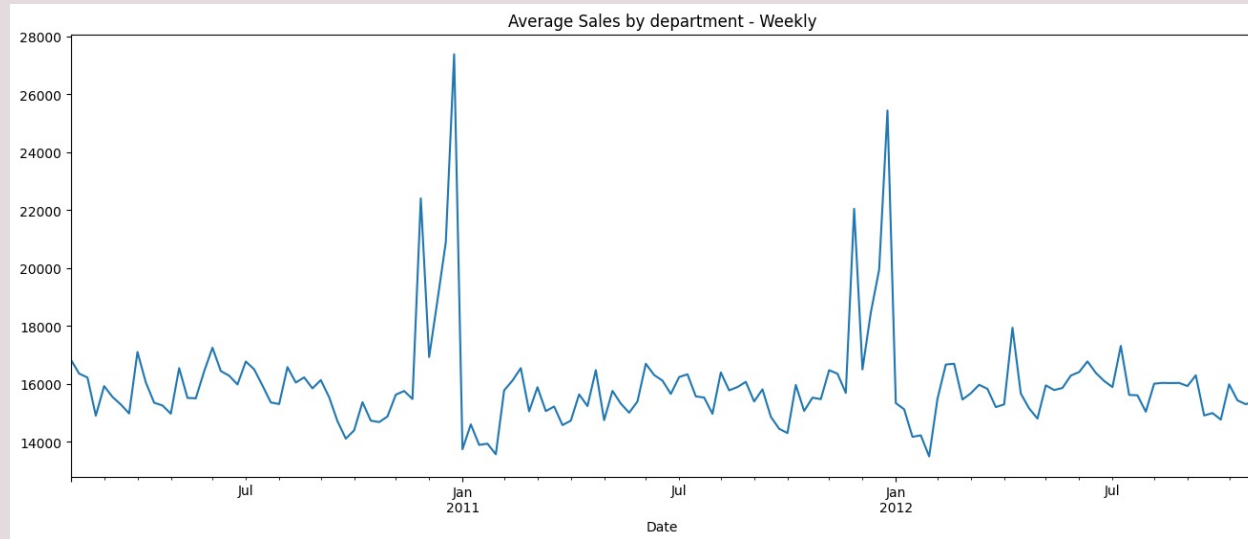
Type C stores exhibit a distinct pattern.

EXPLORATORY DATA ANALYSIS

Thanksgiving and Christmas,
significantly impact sales



EXPLORATORY DATA ANALYSIS



MODELING

ML Models

- Linear Regression
- Random Forest Regressor
- XGBoost Regressor

Statistic Time Series Model

- SARIMA Model

Deep Learning Model

- RNN/LSMT Model

MODEL SELECTION

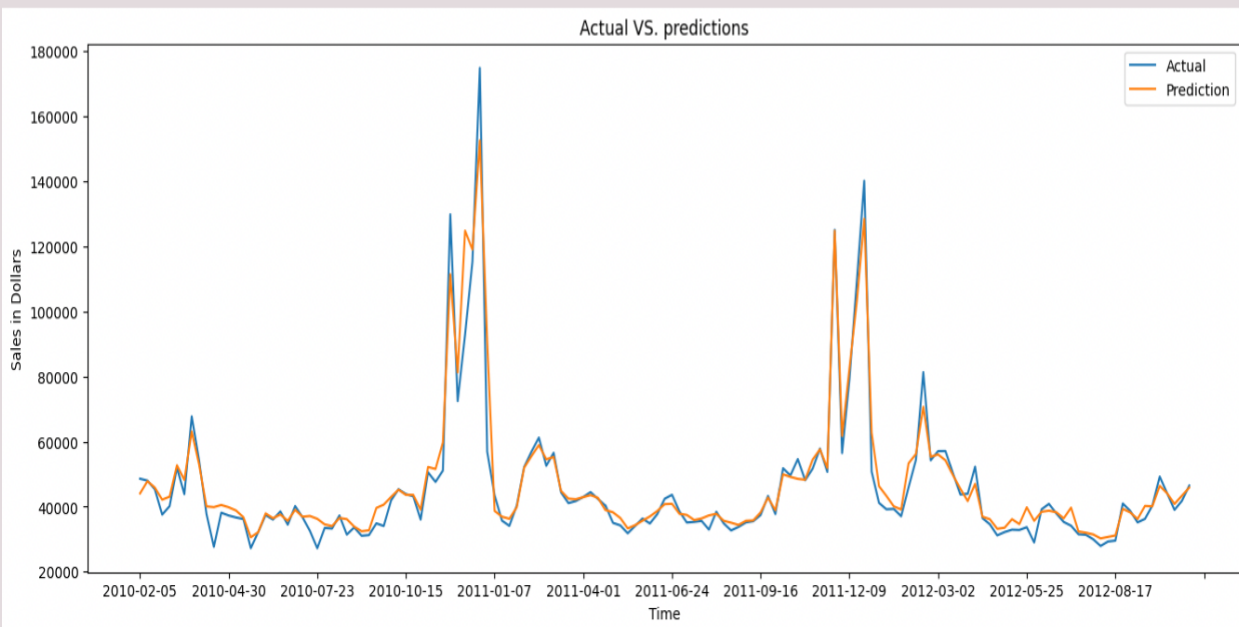
Performance Metrics				
Metrics	Linear Regression	Random forest Regressor	XGBoost Regressor	SARIMA
Train RMSE:	21737.601	1517.267	5316.638	
Test RMSE:	21851.049	4033.612	5594.730	896.360
Train MAE:	14582.799	578.454	3026.045	
Test MAE:	14589.474	1575.786	3093.823	677.566
Train R2:	0.081	0.996	0.945	
Test R2:	0.084	0.969	0.940	0.597

Table 1 Performance Metrics Table

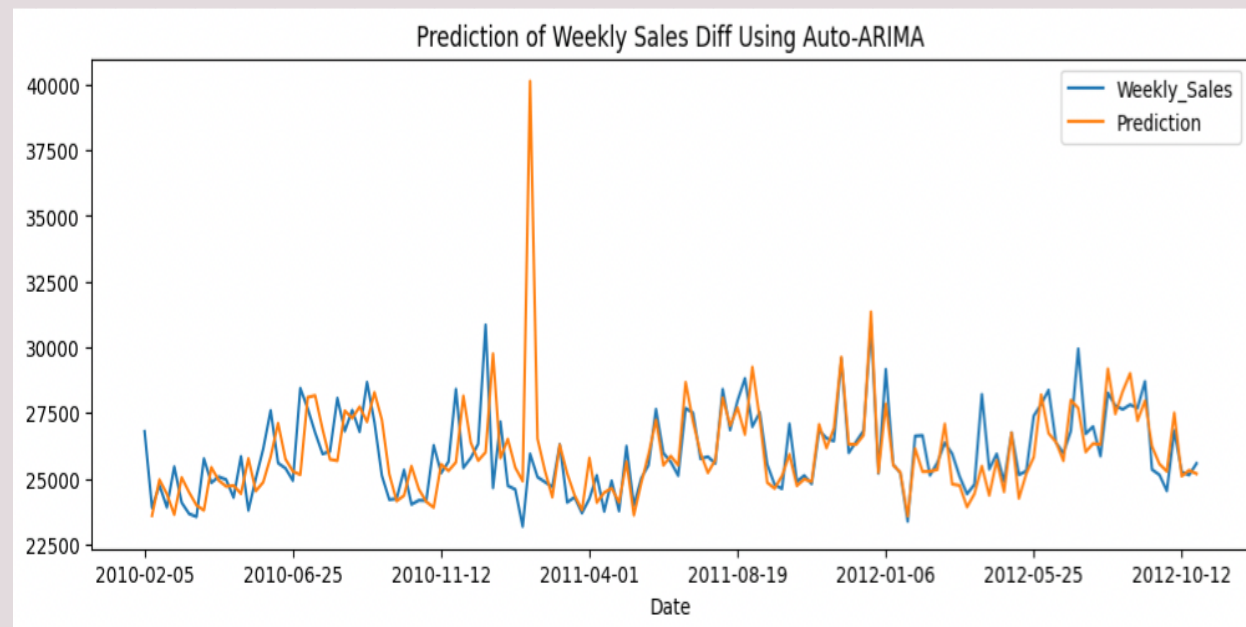
MODEL SELECTION

Random Forest VS. SARIMA

Random Forest



SARIMA



TAKEAWAYS

- ❖ My final model has a high test R^2 score of 0.969, and the prediction plot aligned with the true data perfectly.
- ❖ Department is the key predictor for the weekly sales prediction.
- ❖ Clients can optimize their inventory management and logistics management, target the right time for promotions by utilizing this model.

Scenario:

Based on the passed two weeks sales, manager of store X predict that for department Y, the sales would like the past two weeks, but by using my model, its show's a drop in the next few weeks.

FUTURE RESEARCH

I aim to gather additional data in three aspects:

- ❖ Additional information about the store's location and surrounding demographics.
- ❖ I will include the complete markdown to improve model predictive capabilities .
- ❖ More historical data beyond current timeframe.

**THANK YOU
&
QUESTIONS?**

