Walmart Sales Intelligence System

Model Details
- Hybrid Global Regression Ensemble (RF + XGB)
- Store–Department Time-Series Ensemble (Prophet, RF, XGB)
- High-Sales Classifier (Weighted RF)
- Store Clustering (K-Means, 3 clusters)
- Built using open-source libraries

Intended Use
- Inventory planning, promotion planning, workforce scheduling, chain-wide and store-level forecasting
- Not for automated decisions or individual behavior prediction.
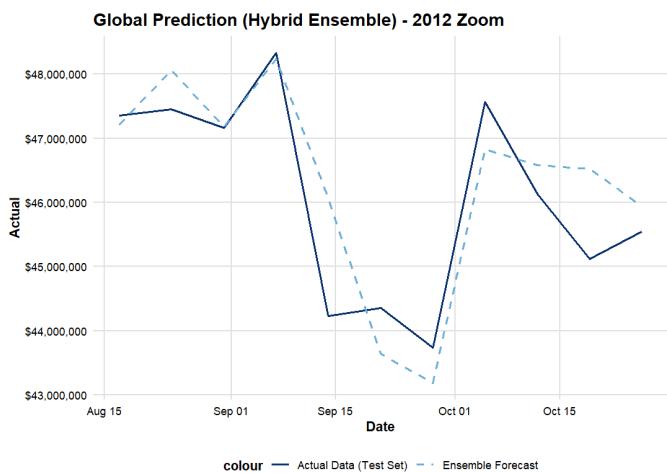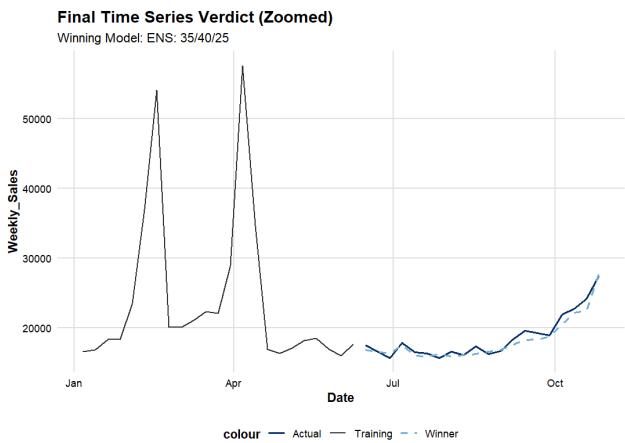
Factors
- Temporal: holidays, seasonality, Lag 52
- Economic: CPI, unemployment
- Store characteristics: size, type, cluster
- Promotional: MarkDown1–5

Metrics
- Global Model: RMSE ≈ 2,563, $R^2 \approx 0.986$
- Time-Series Ensemble: RMSE ≈ 765, $R^2 \approx 0.94$
- Classification: AUC (Weighted RF > Logistic)

Quantitative Analysis:

Training Data
- Walmart Kaggle dataset (stores, depts, weekly sales)
- Economic indicators, promotions, holidays
- Engineered features: Lag1, Lag52, pre-event windows, clusters

Evaluation Data
- Held-out 2012 weeks (10-20-week horizon)
- Time-series tested on specific Store–Dept
- Global model tested on all stores/departments
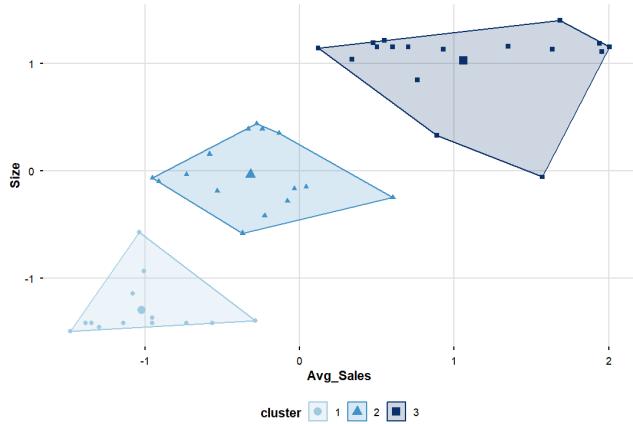
Caveats
- Lag 1 removed in global forecasts to avoid leakage
- Economic variables not forecasted externally
- Cluster features based only on sales + size
- Model assumes regular holiday behavior

Ethical Considerations
- No personal data usage
- The model suggests a price reduction in times of economic hardships, aiming to benefit the customers.
- Project aims to provide an advanced classification of stores that isn't decided by its location, to make sure that any decision regarding the inventory is made, not on the bases of any individual characteristics, but purely to encourage sale of items that already have high sales,

Final Time Series Verdict (Zoomed)
Winning Model: ENS: 35/40/25

Store Segmentation (Sales vs. Size)

Global Prediction (Hybrid Ensemble) - 2012 Zoom

XGBoost SHAP Importance