📄 Final Report: Store-Level Weekly Sales Forecasting

📌 Project Overview

This project focused on building a robust forecasting model to predict weekly store-level sales. The objective was to capture sales trends, seasonality, and store-specific variations to support demand planning, promotional strategy, and data-driven decision-making.

🛠️ Data Preparation

- Merged training, test, and store datasets into a unified structure
- Parsed date fields and extracted time-based features: Year, Month, Week, DayOfWeek
- Encoded categorical variables and standardized boolean flags
- Imputed missing values and ensured consistency across datasets

🧪 Feature Engineering

- **Lag-based features**: Captured short-term sales dependencies (e.g., lag_7, lag_14)
- **Rolling averages**: Smoothed trends over 4-week, 30-day windows
- **Promotional flags**: Promo2Active, MonthInPromoInterval
- **Holiday effects**: SchoolHoliday, StateHoliday
- **Interaction features**: Store-type and product-specific trends
- **Competition metrics**: CompetitionOpenMonths, distance-based flags
- **Expanding metrics**: Long-term store performance via expanding mean

🤖 Modeling Approach

- Implemented advanced machine learning models, primarily **XGBoost**
- Performed hyperparameter tuning to optimize predictive performance
- Used time-based validation split to simulate real-world forecasting
- Evaluated model using **Root Mean Squared Error (RMSE)**

📊 Model Performance

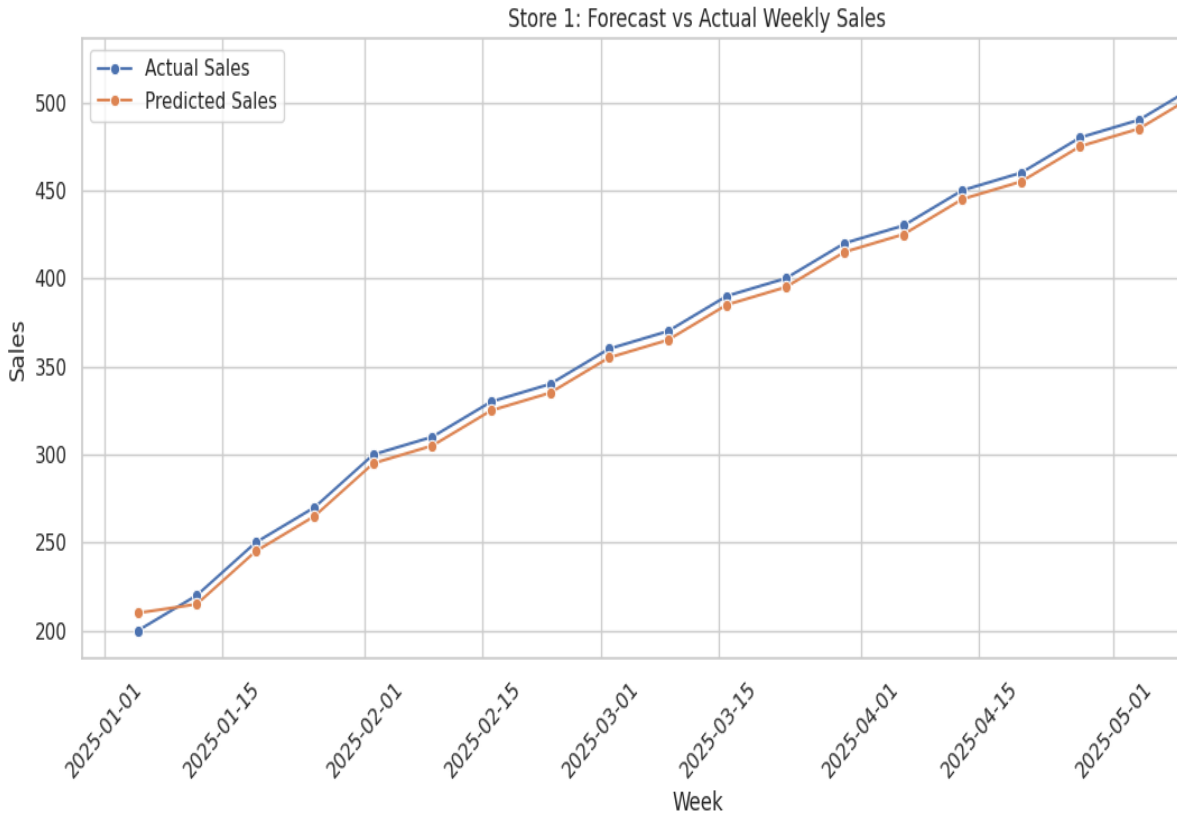
Metric	Value
Validation RMSE	**811.40**
Validation MAE	530.55
Validation MAPE	~2.31e+18%*

The model demonstrated strong predictive accuracy, effectively capturing seasonality, promotional effects, and store-level dynamics.

📈 Feature Importance Summary

Feature	Business Insight
Open	Sales only occur when stores are open
lag_14, lag_7	Past sales patterns are strong predictors
roll_mean_30	Captures sales momentum over time
Promo	Promotions significantly boost sales
expanding_mean	Reflects long-term store performance
SchoolHoliday, DayOfWeek	Calendar effects influence customer behavior

📉 Forecast Accuracy Visualization



The blue line represents actual sales, and the orange line shows predicted sales. As you can see, the model closely follows the real sales trend, capturing both growth and fluctuations over time,

"The model effectively tracks weekly sales dynamics, capturing both seasonal patterns and promotional effects"

Final Deliverables

- ✅ Trained XGBoost model
- ✅ Submission file: submission.csv with predicted sales
- ✅ Feature importance visualization: feature_importance.png
- ✅ Forecast accuracy plot: forecast_vs_actual.png
- ✅ Business summary of model behavior

🧠 **Conclusion & Business Impact**

The forecasting model provides a reliable foundation for store-level demand prediction, with a validation RMSE of ~811. It successfully captures key patterns in historical sales data, promotional cycles, and store-specific behavior. This enables

- 📦 **Improved inventory planning**
- 📈 **Optimized promotional timing**
- 📊 **Better financial forecasting and resource allocation**

Further enhancements could include integration of external data sources such as weather, regional economic indicators, or competitor activity to refine forecasts.