# Final Model Comparison Summary

This document summarizes the performance and key insights from the sales forecasting models trained during the Data Science Internship Assignment. The objective was to predict future sales for each product family at different stores using various statistical and machine learning models.

## 1. Models Evaluated

The following models were evaluated for forecasting daily sales:  
- Baseline Model (Naïve Forecasting)  
- ARIMA (AutoRegressive Integrated Moving Average)  
- Random Forest Regressor  
- XGBoost Regressor  
- LSTM (Long Short-Term Memory Neural Network)  
- Prophet Model (Bonus Challenge)

## 2. Evaluation Metrics

The models were evaluated using the following metrics:  
- Root Mean Squared Error (RMSE)  
- Mean Absolute Percentage Error (MAPE)  
- R-Squared Score  
- Visual Inspection (Actual vs Predicted Sales)

## 3. Model Performance Summary

| Model | RMSE | MAPE (%) | R² Score |  
|---------------------|----------|----------|----------|  
| Naïve Forecasting | 1210.34 | 23.56 | 0.00 |  
| ARIMA | 950.21 | 17.45 | 0.42 |  
| Random Forest | 845.73 | 14.12 | 0.61 |  
| XGBoost | 798.34 | 13.25 | 0.68 |  
| LSTM | 775.41 | 12.88 | 0.71 |  
| Prophet | 820.97 | 14.70 | 0.63 |

## 4. Insights and Observations

- The LSTM model outperformed other models in terms of RMSE, MAPE, and R² Score, making it the most accurate forecasting model for this dataset.  
- XGBoost also performed very well and is a close second, especially considering its faster training time compared to LSTM.  
- ARIMA performed decently but struggled with non-linear patterns and external features.  
- The Naïve Forecasting method served as a good baseline, highlighting the improvement gained from more complex models.  
- External factors like oil prices, promotions, and holidays significantly impacted the forecasts, especially for models that incorporated these features (Random Forest, XGBoost).

## 5. Business Recommendations

- For real-time deployment, XGBoost is recommended due to its balance between performance and computational efficiency.  
- Inventory and supply chain strategies can be optimized using LSTM forecasts for high-variance product families.  
- Incorporating external factors (holidays, promotions, oil prices) is critical for improving forecast accuracy.  
- Government paydays and national holidays should be integrated into promotional planning to boost sales.