SALES FORECASTING PROJECT SUMMARY

1. Project Overview

This project is aimed at building robust forecasting models to predict future sales using historical data.
Multiple machine learning and deep learning approaches were used including **Random Forest**, **XGBoost**, and **LSTM (Long Short-Term Memory)** to identify the most accurate model.

2. Data Preprocessing

The dataset was cleaned and transformed to fit the needs of time series modeling. Key steps included:

- Handling missing values
- Encoding categorical variables
- Creating date-time features (month, day, etc.)
- Feature scaling using MinMaxScaler for LSTM model
- Converting the time series data into supervised format for LSTM

3. Model Development

Three different models were implemented:

- Random Forest Regressor
- XGBoost Regressor
- LSTM Neural Network

Each model was trained on 80% of the data and validated on the remaining 20% using appropriate time-series validation techniques.

4. Final Model Comparison

Model	RMSE	MAPE (%)	R² Score
Random Forest	327.34	8.76	0.8898
XGBoost	328.04	9.15	0.8893
LSTM	310.67	7.93	0.9021

---

5. Key Insights

1. ☑ **LSTM outperformed** both ensemble models in all metrics — lower RMSE and MAPE, and higher R².
2. 📈 **LSTM benefits** from modeling sequential dependencies, making it more suitable for time series forecasting tasks like sales prediction.
3. ⚠ **Random Forest and XGBoost**, while powerful, treat each row independently and miss time-based trends unless engineered manually.
4. ⚡ **If simplicity and speed are priorities**, Random Forest is a good baseline. But for production-grade accuracy, LSTM (with tuning and training time) is superior.
5. ▦ MAPE below 10% across models indicates overall **good forecasting performance**.