**Data Preprocessing & Feature Engineering**

* **Data Cleaning**
  + **Convert date columns to datetime format.**
  + **Handle missing values (e.g., interpolate missing oil prices).**
  + **Merge datasets based on date and store numbers.**
* **Feature Engineering**
  + **Extract time-based features (year, month, day, day\_of\_week, is\_weekend).**
  + **Create binary flags for holidays, promotions, and economic events.**
  + **Identify if a day is a government payday.**
  + **Consider earthquake impact (April 16, 2016) as a feature.**
  + **Generate lagged features (sales from previous 7 and 30 days).**
  + **Compute rolling averages for past sales trends.**
* **Exploratory Data Analysis (EDA)**
  + **Visualize sales trends over time.**
  + **Analyze sales impact before and after holidays.**
  + **Check correlations between oil prices and sales trends.**

**Model Training & Forecasting**

* **Train multiple models:**
  1. **Baseline Model (Naïve Forecasting) - Assumes future sales = previous sales.**
  2. **ARIMA (AutoRegressive Integrated Moving Average) - Captures seasonality and trends.**
  3. **Random Forest Regressor - Tree-based model for non-linear relationships.**
  4. **XGBoost Regressor - Gradient boosting model for better accuracy.**
  5. **LSTM (Long Short-Term Memory Network) - Deep learning model for time-series forecasting.**
* **Model Evaluation**
  1. **Compare models using RMSE (Root Mean Squared Error).**
  2. **Visualize actual vs. predicted sales.**
  3. **Identify the best-performing model.**
* **Final Prediction**
  1. **Use the best model (XGBoost) to make final sales predictions.**
  2. **Save predictions for submission.**

**Key Insights & Findings**

**Best Performing Model: XGBoost**

* Achieved the lowest RMSE and MAPE High R² Score, meaning it explains variance well Handles large datasets and missing values efficiently Outperforms ARIMA and LSTM in terms of speed and accuracy.

**External Factors Influencing Sales**

* Holidays & Promotions → Significant spikes in sales Oil Prices → Weak correlation but can impact transportation-related sales Paydays → Higher sales around 15th and last day of the month Store-Specific Trends → Clusters have different buying patterns

**Business Recommendations**

* ✔ Use XGBoost for forecasting with tuned hyperparameters .Consider hybrid models (XGBoost + LSTM) for long-term forecasting ✔ Plan inventory & staffing around holidays using forecast insights Target promotions around payday & peak demand times .

**Final Conclusion**

* 🔹 XGBoost is the best model for short-term sales forecasting LSTM could be explored for long-term forecasting if computational power allows Business decisions should factor in seasonal trends, holidays, and promotions