INVENTORY STRATEGY REDESIGN USING MACHINE LEARNING FORECASTING FOR A GROCERY CHAIN

1. Background and Problem Statement:

A regional grocery chain with over 30 stores struggled with inventory inefficiencies, particularly in fast-moving and perishable categories like dairy, snacks, and beverages. Manual forecasting and vendor-driven restocking led to frequent stockouts, excess expiry-related waste, and poor shelf availability. The company required a **machine learning–based forecasting system** to build a more responsive, data-driven inventory strategy.

2. Objectives:

- To forecast weekly product-level sales using advanced machine learning and time series models
- To compare model performance between traditional ARIMA and Random Forest
- To design reorder points and stock levels tailored to product perishability and demand variability
- To reduce inventory holding costs while maintaining high product availability

3. Methodology:

Data Source:

- Two years of POS data across 150 SKUs in dairy, snacks, and beverages
- Weekly sales volume, shelf stock levels, expiry records, promotions, and vendor lead times

Tools and Techniques:

- **Python**: pandas, scikit-learn, ARIMA from statsmodels, xgboost
- Visualization and evaluation: matplotlib, seaborn, sklearn.metrics
- Model evaluation metrics: MAE, RMSE, and MAPE

Steps:

1. **Data Preprocessing:**
    - Missing value imputation

- Encoding promotions and holidays as binary features
- Feature engineering: rolling averages, lagged variables, freshness index

2. **Modeling:**
    - **Model 1**: ARIMA for baseline time series forecasting
    - **Model 2**: Random Forest Regressor using engineered features
    - **Model 3** (Bonus): XGBoost for highly volatile SKUs
3. **Inventory Planning Logic:**
    - Reorder point based on forecasted weekly sales + safety stock (calculated from sales std. deviation during lead time)
    - Product shelf-life used to cap maximum stockholding

4. Results:

Forecast Accuracy:

Model	MAE	RMSE	MAPE
ARIMA	18.5	24.2	21.4%
Random Forest	12.1	17.6	13.6%
XGBoost	10.8	15.9	11.9%

- Random Forest and XGBoost outperformed ARIMA, especially for high-variance SKUs and promotion-heavy weeks
- XGBoost chosen as the model for SKUs with low shelf life (e.g., fresh milk)

Operational Improvements:

- Forecast-driven reorder points reduced stockouts by **38%**
- Waste from expired stock dropped by **22%** over three months
- Forecast model updated weekly and retrained monthly for ongoing accuracy

5. Interpretation and Insights:

- Machine learning models capture non-linear trends and respond better to promotions and holidays than traditional models

- Demand forecasting granularity at SKU level allows for vendor-specific restocking strategies
- Inventory decisions must incorporate perishability constraints, not just forecasted volume
- Forecast dashboards helped store managers gain confidence and act early

6. Recommendations:

- Integrate forecast models into the store's ERP for auto-generating order quantities
- Use different forecast models per category (e.g., ARIMA for beverages, ML for dairy)
- Conduct quarterly model evaluations to handle concept drift
- Build an alert system for SKUs with unusual demand spikes

7. Future Work:

- Expand model scope to include pricing, weather, and competitor promotions
- Create dynamic safety stock levels based on real-time sales volatility
- Use clustering to group similar SKUs for bulk forecasting

8. Stakeholder Relevance:

Academic:

- A strong case for integrating ML in operations analytics, showcasing model comparison and inventory impact
- Teaches end-to-end workflow: data prep → modeling → inventory logic

Corporate:

- Applicable to retail chains and supermarkets looking to cut waste and improve on-shelf availability
- Demonstrates real ROI from predictive analytics in inventory management