

Project Plan

Warehouse and Retail Sales Forecasting

Business Questions

➤ **Primary:**

How can we accurately forecast monthly warehouse sales to improve inventory management, supplier relations, and operational planning?

➤ **Secondary:**

- What are the key drivers of warehouse sales (supplier, item type, seasonality)?
- How can predictive models help identify trends and prevent overstock/understock situations?
- Can we generate actionable insights to support strategic decision-making in procurement and sales?

Dataset Used

Source: Warehouse and retail sales transaction records (307,645 rows, 9 columns: YEAR, MONTH, SUPPLIER, ITEM CODE, ITEM DESCRIPTION, ITEM TYPE, RETAIL SALES, RETAIL TRANSFERS, WAREHOUSE SALES) is a dataset from Montgomery County, Department of Liquor Control in the state of Alabama in the United of America. The dataset is updated monthly and is available at [dataMontgomery](#) public repository. It contains a list of sales and movement data by item and departments.

Content:

- Transactional sales data from warehouse and retail channels updated monthly.
- Supplier and item metadata for analysis and feature engineering.

Ethical Considerations:

- ✓ Dataset contains no personally identifiable information (PII) or customer data.
- ✓ All data usage is in compliance with Montgomery County, MD privacy policies and ethical standards for public use.

Methodology and Tools

Methodology:

1. **Exploratory Data Analysis (EDA):** Identify trends, outliers, missing data, and drivers of sales.
2. **Data Cleaning:** Remove or impute missing values, filter negative/outlier sales, prepare dataset for modeling.

3. **Feature Engineering:** Encode categorical variables, aggregate at relevant time intervals, and prepare for machine learning.

Model Development:

- Compare linear regression and tree-based regression (Random Forest) for sales prediction.
- Tune model hyperparameters and evaluate on hold-out/test data.

Validation & Interpretation: Visualize results, analyze feature importances, check for overfitting.

Deployment/Reporting: Save model and define recommendations for business integration.

Tools:

- Python (pandas, numpy, matplotlib, seaborn, scikit-learn, joblib)
- Jupyter Notebook for reproducible analysis and code sharing.
- Github for public sharing

Timeline and Milestone

Week	Milestone	Deliverables
1	Project kickoff, EDA, Data cleaning and feature engineering.	EDA summary, initial plots, cleaned dataset, report.
2	Model development & baseline evaluation, Model tuning, validation, and interpretation, and Final reporting.	Regression results, plots, Tuned model, diagnostics, Project report/notebook

Summary

This project will provide data-driven forecasts of warehouse sales, supporting better business planning and operational decisions. The approach combines rigorous data science with practical business application, and delivers clear, actionable recommendations.