

# Retail Demand Forecasting & Driver Analysis (Synthetic Data)

## Project Overview

This project demonstrates how a retail analytics team approaches **demand forecasting and sales driver analysis** in a large-scale omnichannel retail environment. Rather than relying on externally sourced datasets, the project intentionally uses **synthetically generated but realistic retail data** to simulate real-world analyst workflows while avoiding data privacy and NDA constraints.

The focus of this work is not data collection, but **business framing, analytical logic, and decision support**—the core responsibilities of a retail data analyst.

## Business Objective

Retail organizations must continuously forecast demand and understand the key drivers of sales to support:

- inventory planning,
- pricing strategy,
- promotion effectiveness,
- and operational efficiency.

The objective of this project is to simulate how an analyst would:

1. design a realistic retail dataset,
2. identify factors influencing weekly sales,
3. build an explainable demand forecasting model,
4. and translate analytical results into business-relevant insights.

## Key Questions Addressed

- How can weekly demand be forecasted at a store–SKU level?
- How do pricing changes impact unit sales?
- What is the effect of promotions on demand?
- Which variables most strongly influence weekly sales performance?

## Dataset Description

The dataset is **generated programmatically within the notebook** using Python (numpy and pandas).

No external files are used.

Each record represents weekly sales for a product in a retail store and includes:

- Store and product identifiers
- Weekly unit sales
- Product price
- Promotion indicator
- Seasonal and environmental proxies (e.g., temperature)

The synthetic data mirrors realistic retail behavior, such as price sensitivity and promotional lift, enabling meaningful analysis without relying on proprietary datasets.

## Analytical Approach

### 1. Data Generation

Retail-like data is simulated using controlled random distributions to reflect plausible business patterns.

### 2. Exploratory Analysis

Relationships between sales, pricing, and promotions are examined to understand demand drivers.

### 3. Feature Engineering

Relevant variables are structured to reflect how retail data is modeled in production analytics environments.

### 4. Demand Forecasting

A linear regression model is used to forecast weekly demand.

This model was chosen intentionally for its **interpretability**, allowing business stakeholders to clearly understand how each variable impacts sales.

## Key Insights

- Promotions consistently increase weekly unit sales.
- Higher prices are associated with reduced demand, reflecting expected price elasticity behavior.
- Simple, explainable models can provide strong baseline forecasts suitable for operational planning.

## Why Synthetic Data?

In many real-world retail environments, analysts prototype models using simulated or anonymized data due to confidentiality constraints. This project reflects that reality. The emphasis is on **how decisions are supported**, not on the source of the data.

## Tools & Technologies

- Python
- Pandas & NumPy
- Matplotlib
- Scikit-learn (Linear Regression)