ZEROSTOCK

RETAIL INVENTORY FORECASTING

Presented by -

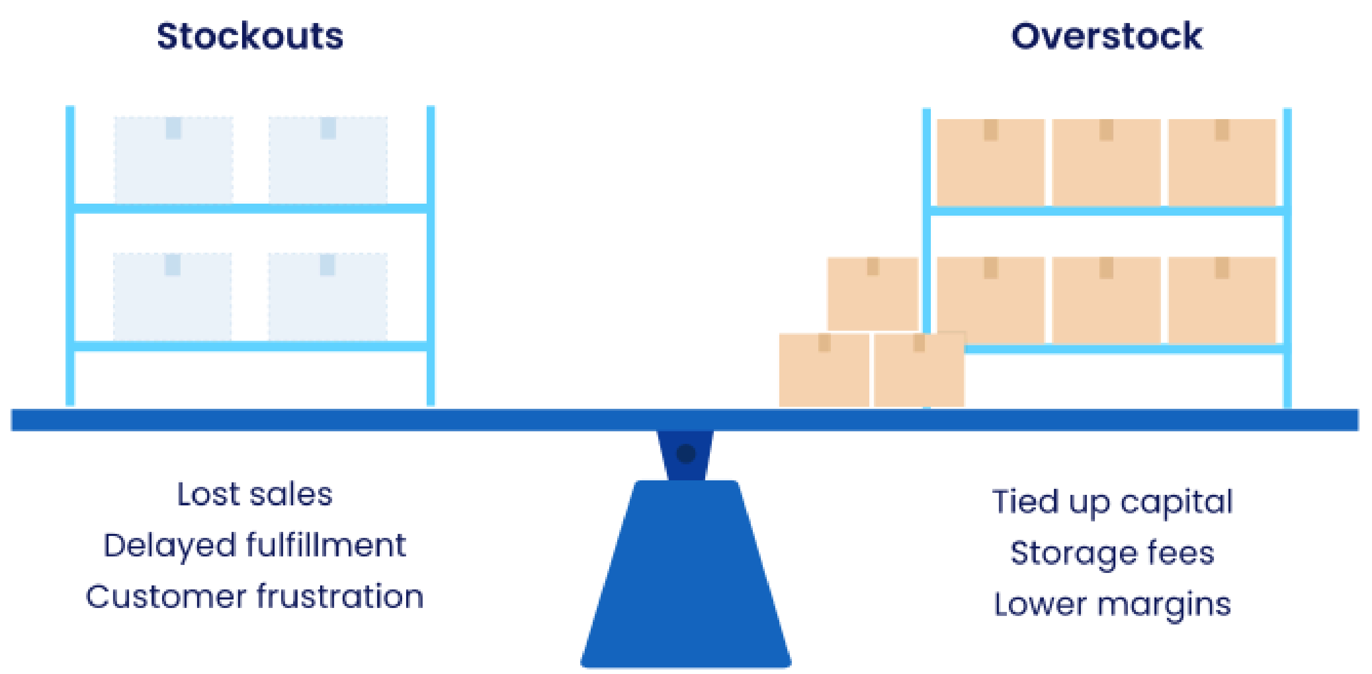
Aneek Saha - 2205355
Akangkshya Goswami - 2205354
Tanay Changoiwala - 2205340
Arnab Mazumdar- 2205277
Sakshi Mishra - 2205320

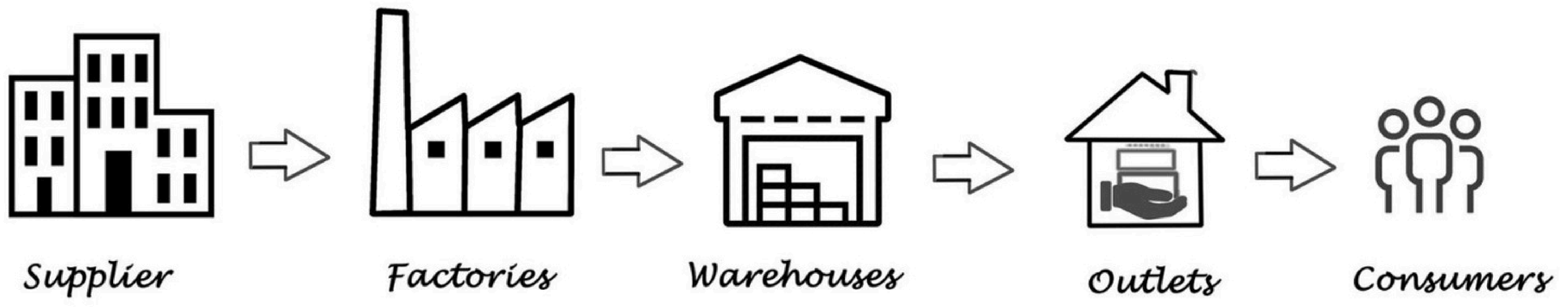
Problem Statement

Problem: Retailers often face overstocking or stockouts due to poor demand forecasting.

Objective:

- Predict daily demand for different products across stores
- Understand impact of factors like holidays, promotions, and weather
- Build dashboards for business insights and decision-making





Dataset Overview

- **Dataset Description:**
  - 73,000+ rows of synthetic but realistic retail data
  - Covers multiple stores and product categories
- **Key Features:**
  - date, store_id, product_id
  - sales, inventory_level
  - price, promotion, holiday, weather
- **Use Cases:** Demand forecasting, inventory optimization, pricing strategies

	Feature Group	Example Columns
🏬	Store/Product Info	Store ID, Product ID, Category, Region
📦	Inventory/Sales Data	Inventory Level, Units Sold, Units Ordered
📈	Forecasting/Price	Demand Forecast, Price, Discount, Competitor Pricing
🌤️	External Factors	Weather Condition, Holiday/Promotion, Seasonality

Date	Store	Product	Category	Region	Sold	Ordered	Forecast	Promo	Season
01-01-22	S001	P0001	Groceries	North	127	55	135.47	0	Autumn
01-01-22	S001	P0002	Toys	South	150	66	144.04	0	Autumn
01-01-22	S001	P0003	Toys	West	65	51	74.02	1	Summer

How the Forecasting System Works

Our forecasting system follows a structured pipeline that transforms raw retail data into accurate demand predictions. It starts with preprocessing and feature engineering, where we extract meaningful features like lagged sales, rolling averages, and calendar events. These engineered features are then used to train the XGBoost model.

Once trained, the model takes recent data as input and generates forecasts for future inventory demand across different stores and products. The final predictions are presented through interactive dashboards for decision-making.

This system helps anticipate stock requirements, optimize inventory levels, and prevent understocking or overstocking.

Process Breakdown

🧹 1. Data Preprocessing

🛠️ 2. Feature Engineering

📊 3. Model Training

🔮 4. Forecast Generation

📈 5. Visualization & Dashboard

Data Preprocessing

Transforming Raw Data into Insightful Signals

To enable effective forecasting, we carefully prepared the data through multiple preprocessing steps. We engineered time-based features like **lag sales** and **rolling averages** to help the model recognize historical demand patterns. Categorical variables like store ID, product category, and promotions were encoded numerically for model compatibility. Scaling was applied to normalize key numeric fields like sales and price.

A time-based split was used to simulate real-world forecasting — training on past data and testing on future dates — ensuring our model learns from realistic business scenarios.

Cleaned & Structured -

We handled missing values in sales and inventory columns and dropped unnecessary or duplicate rows. Columns were renamed and types standardized. This ensured data consistency and model readiness.

Feature Engineering -

We introduced lag features, moving averages, and calendar-based features like month, day of week, and holiday flags. Categorical fields like store ID and product type were label-encoded. These features helped the model capture seasonality, trends, and external influences.

Scaling & Train-Test Split -

Numerical columns like sales and price were scaled using StandardScaler to improve model convergence. A time-based split was used to ensure future dates were not leaked into training, simulating realistic forecasting environments.

EXPLORATORY DATA ANALYSIS

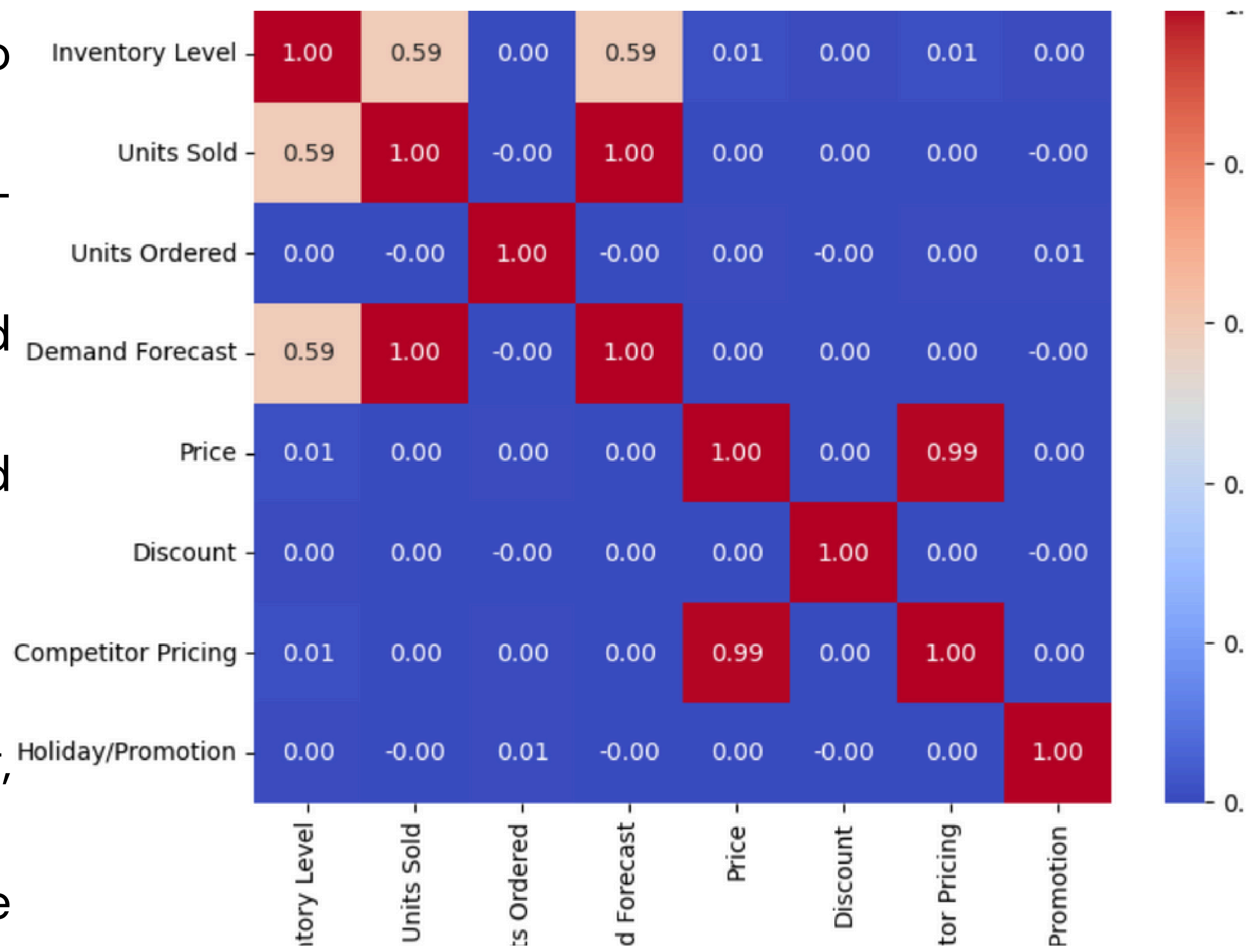
Trend & Seasonality Analysis

· Dataset contains 73,100 records spanning from January 2022 to January 2024

· Analysis covers 5 stores (S001-S005) and 20 products (P0001-P0020) across multiple categories

· Strong correlation (0.589) between Inventory Level and Units Sold suggests proper inventory management

· Extremely strong correlation (0.997) between Units Sold and Demand Forecast indicates highly accurate forecasting[1]



Impact of External Factors

· Seasonality column allows for quarterly sales pattern analysis (Spring, Summer, Autumn, Winter)

· Weather conditions (Sunny, Rainy, Cloudy, Snowy) show variable impact on product performance

· Holiday/Promotion flag (0/1) enables comparison between promotional and non-promotional periods

Forecasting Model

Model Used

- XGBoost Regressor

We used XGBoost due to its strong performance on structured tabular data. It handles non-linear patterns well, works with missing values, and includes regularization to prevent overfitting.

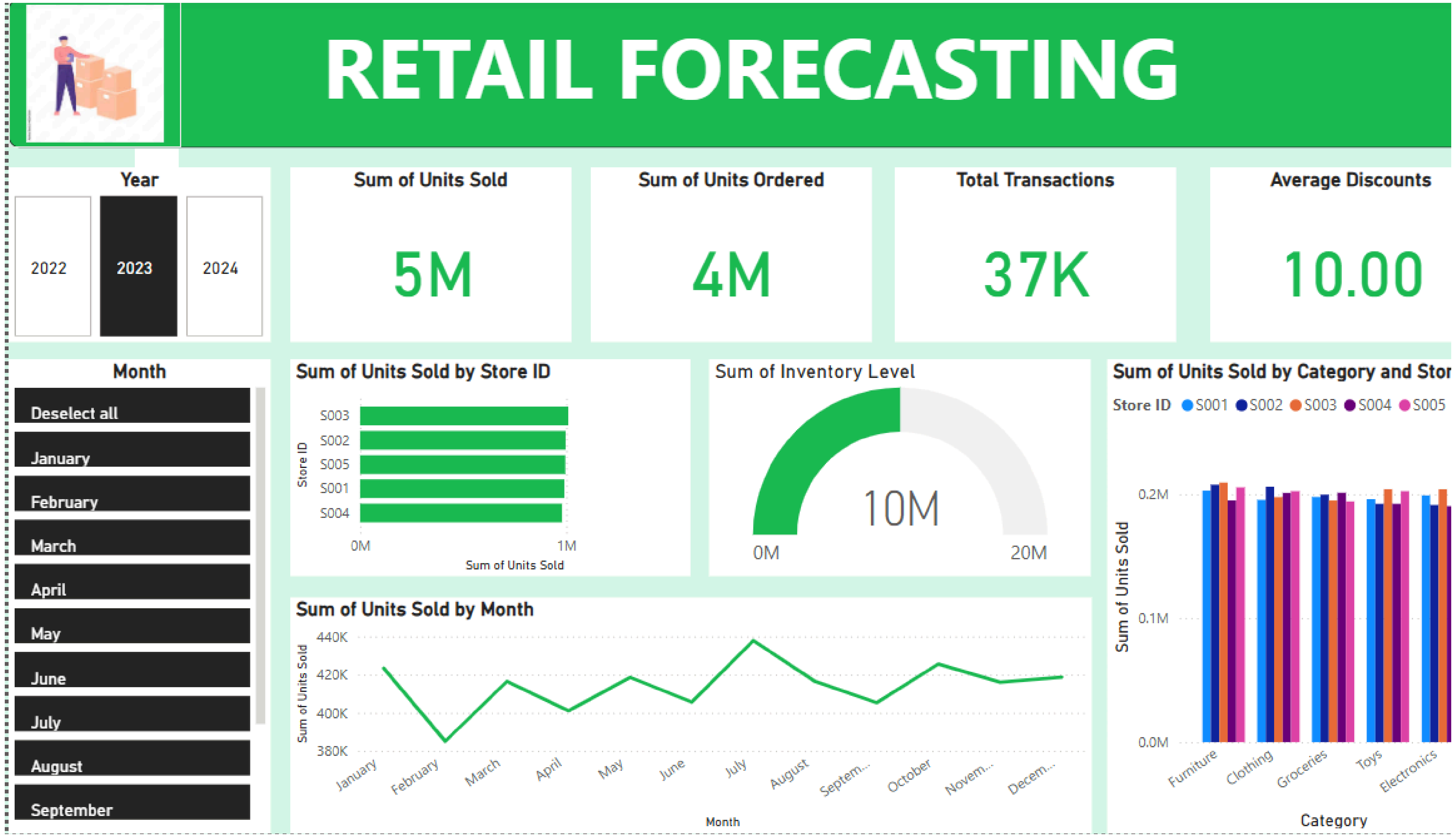
Why XGBoost?

- It efficiently learns from engineered time based variables
- Performs well on large datasets like ours (73k+ rows)
- Robust against overfitting and ideal for time series forecasting when properly configured

Model Performance

- Evaluation Metric: Root Mean Squared Error (RMSE) and R^2 score.
- Time-based train-test split ensured no data leakage and realistic forecasting.
- Model captured seasonal patterns and promotional effects effectively

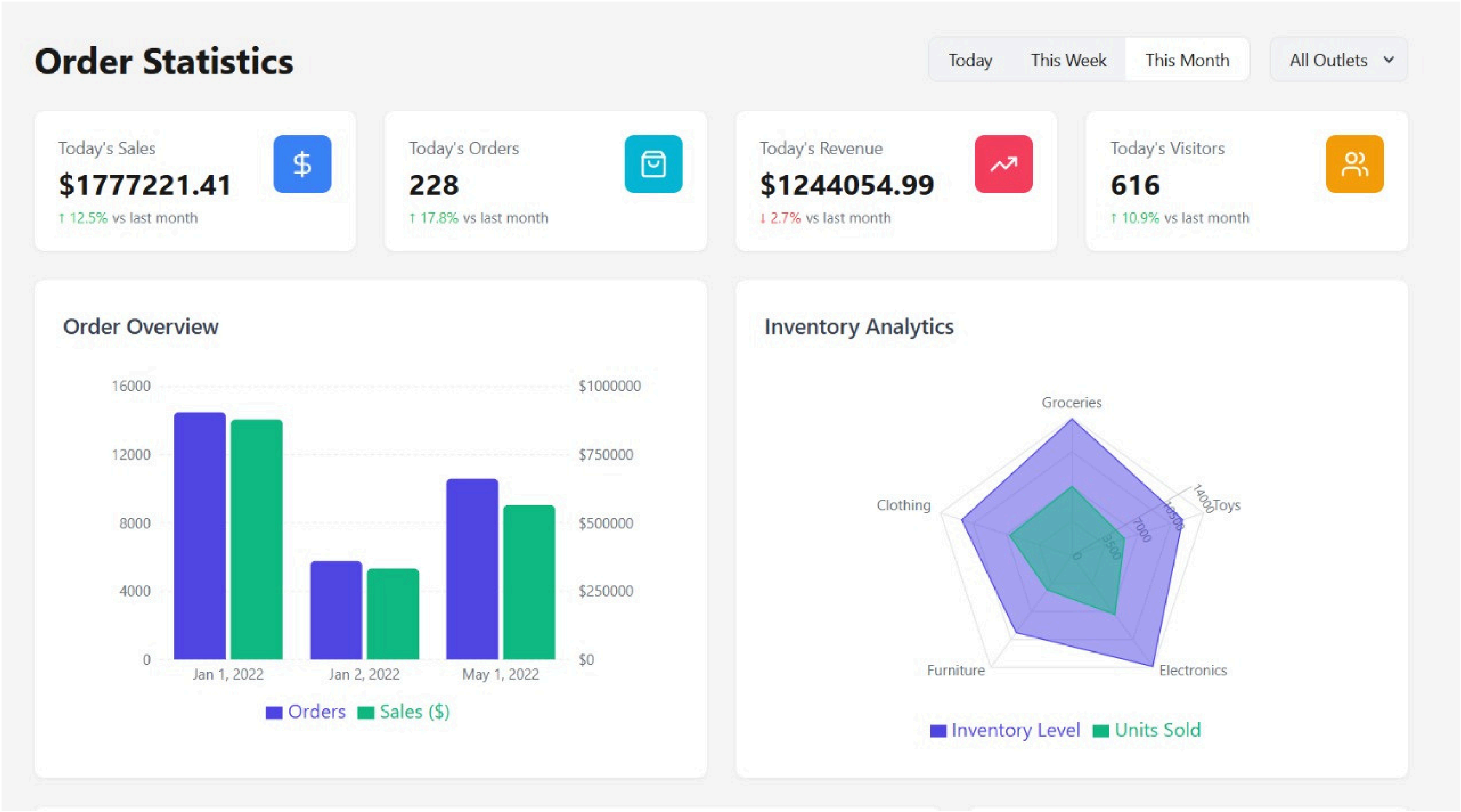
PowerBI Dashboard



Here, we have a Power BI dashboard sample showcasing data from all stores for the year 2023. From this visualization, we can derive the following insights:

- A total of 5 million units were sold in 2023, with Store IDs S003 and S002 emerging as top performers in terms of sales.
- July recorded the highest monthly sales, highlighted by a significant upward trend from June.
- The current inventory stands at 10 million units, reflecting a stable stock status, while the average discount offered remains consistent at 10.00.

Next.js Dashboard



- Real-time Sales Tracking: Displays today's sales amounting to $1.77M with a 12.5% increase from last month.
- Order & Revenue Insights: Shows 228 orders today and $1.24M in revenue; revenue down by 2.7%.
- Order Overview Chart: Visual representation of daily orders and corresponding sales.

Insights & Outcomes

Key Analytical Findings:

- Daily sales rates vary significantly by store location, with Store S002 showing highest average performance
- Weekly sales patterns provide more stable forecasting baseline than daily fluctuations
- "Days of Inventory Remaining" metric reveals potential stock-out risks when below 3 days
- Rolling average sales calculation (implemented in code) improves short-term forecasting accuracy

Impact Analysis:

- Price correlation with competitor pricing (0.994) indicates strong market alignment.
- Minimal correlation between discount level and units sold suggests opportunity for discount strategy optimization.
- Tracking days since last restock (calculated in code) enables more efficient inventory replenishment scheduling

Forecasting Performance:

- · Demand forecasting shows high correlation with actual sales performance
- · Implementation of rolling averages provides adaptive forecasting capability
- · Time-series analysis framework established for continuous improvement

Conclusion & Future Work

Project Achievements:

- Successfully implemented comprehensive retail analytics framework
- Established key inventory management metrics including Days of Inventory Remaining
- Developed rolling average sales forecasting methodology
- Created foundation for advanced demand prediction

Challenges Faced:

- Managing large dataset (73,100 records) with multiple categorical variables
- Integrating diverse factors (weather, promotions, seasonality) into coherent analysis
- Establishing appropriate time windows for rolling averages and forecasting

Future Improvements:

- Real-time forecasting integration with automated alerting system
- Advanced machine learning models to improve prediction accuracy
- Live retail system integration for continuous data updates
- Additional external data sources (economic indicators, social media trends, competitor promotions)
- Supplier lead time optimization based on historical performance