

RETAIL- DEMAND FORECAST

*Using Machine
Learning & Time Series
Analysis*

COLLABORATORS:

1. ANGELA MWANZIA
2. REAGAN ADAJO
3. DANIEL MUIGAI
4. SAMUEL KAMAU
5. RONALD OGWANG
6. FILDA KIARIE

PROJECT OVERVIEW

Problem statement:

The retail market is ever-changing and marked by overwhelming, poor-quality data, etc

Objectives:

- To predict volumes of items sold across multiple stores
- To identify key factors influencing customer purchasing patterns
- To understand the impact of pricing strategies, promotions, and markdowns on sales volumes

Business Impact(s):

- Improved inventory management
- Optimized pricing strategies
- Enhanced customer satisfaction
- Data-driven decision-making

Data Understanding

Data Source: <https://www.kaggle.com/competitions/ml-zoomcamp-2024-competition/data>

Using 25 months of historical sales data to capture complex patterns in **customer purchasing behavior, price sensitivity, and impact of promotions**

Key datasets:

- **Sales.csv**: daily sales records including quantity sold, pricing, and revenue metrics
- **Online.csv**: store-specific e-commerce transaction data
- **Discounts_history.csv**: detailed promotional pricing history
- **Catalog.csv**: product details including department, class, and specifications

Data Preprocessing

The following steps were done, to prepare the data for modeling:

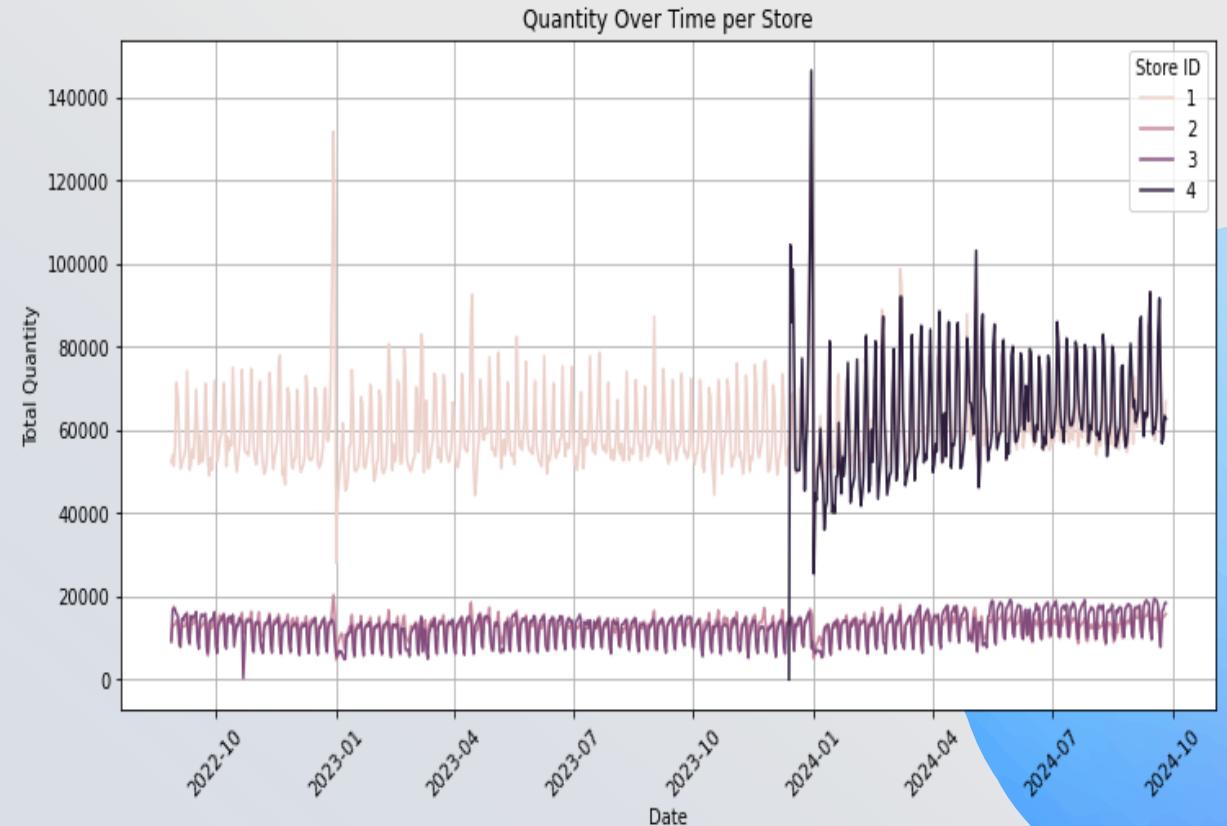
- ✓ Dropped unnecessary columns(e.g. Unnamed:0)
 - ✓ Converted data columns into a uniform format
 - ✓ Filled missing categorical values
 - ✓ Handled numerical data(dropped missing and negative values)
 - ✓ Merged necessary datasets
-
- Identified and transformed skewed columns
 - Handled missing values

Key Insights

Quantity Over Time Per Store:

- **Store 1 and 4** have the highest sales, but **Store 4 only starts in late 2023**, suggesting a new opening.
- **Store 1** shows steady trends with occasional spikes, possibly due to **seasonal demand or promotions**.
- **Stores 2 and 3** have lower but steady sales.

Key Insight: Store 4's rapid growth suggests a strong launch, while spikes in Store 1 and 4 may indicate successful promotions or external demand shifts.

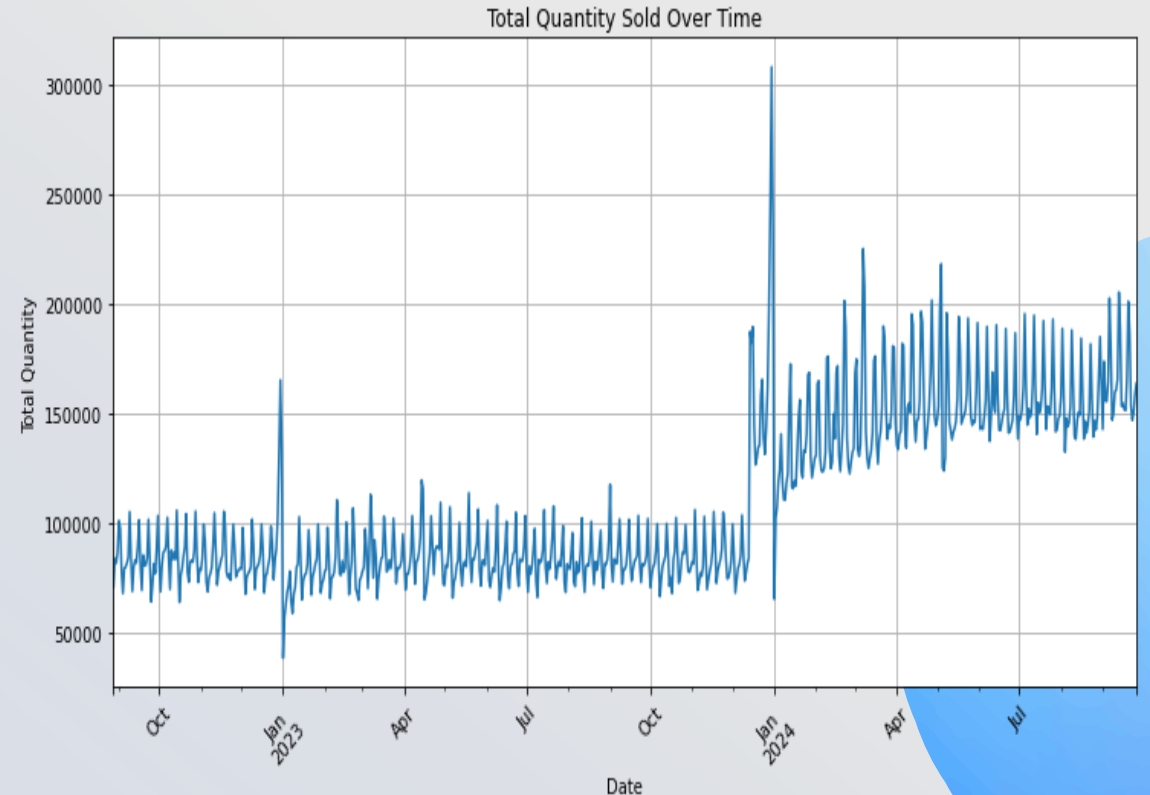


Key Insights

Total Quantity Sold Over Time:

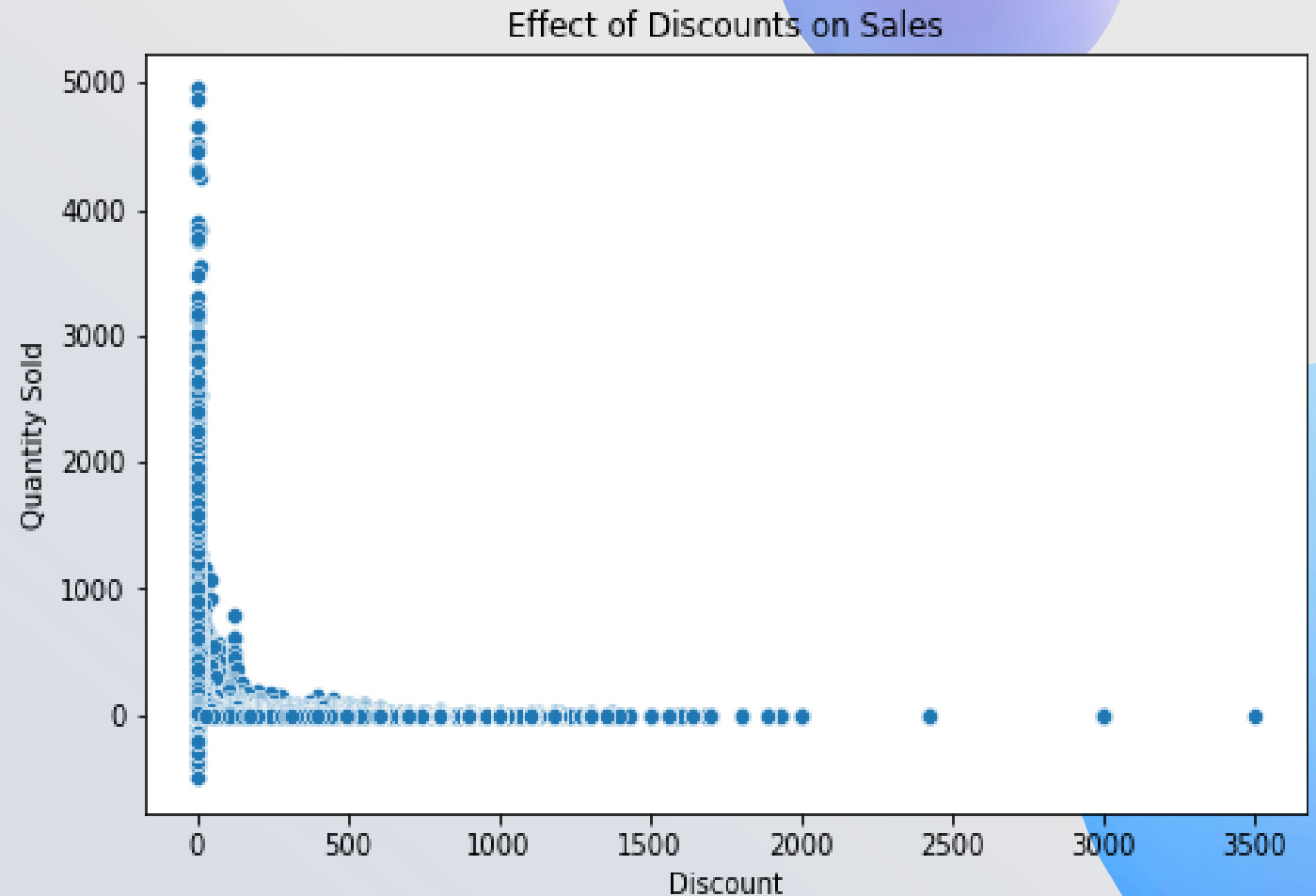
- **Stable sales initially** with minor fluctuations.
- **Sharp spike followed by a shift to higher sales levels**, possibly due to a major event, like store expansion, promotion, or market change.
- **Regular fluctuations after the shift**, indicating a possible seasonal or cyclical pattern.

Key Insight: Something significant **boosted sales permanently** around the midpoint, warranting further investigation into contributing factors.



Impact of Discount on Sales

Discounting as a promotional tool has relatively less impact on quantity sold



Feature Engineering

Create meaningful features to improve model performance:

- ✓ **Created Lag Features** (e.g., lag_1_sales, lag_7_sales) to capture past trends
- ✓ **Generated Moving Averages** (e.g., ewma_7_sales) for smoothing fluctuations
- ✓ **Calculated Average Sales per Area** to normalize store performance

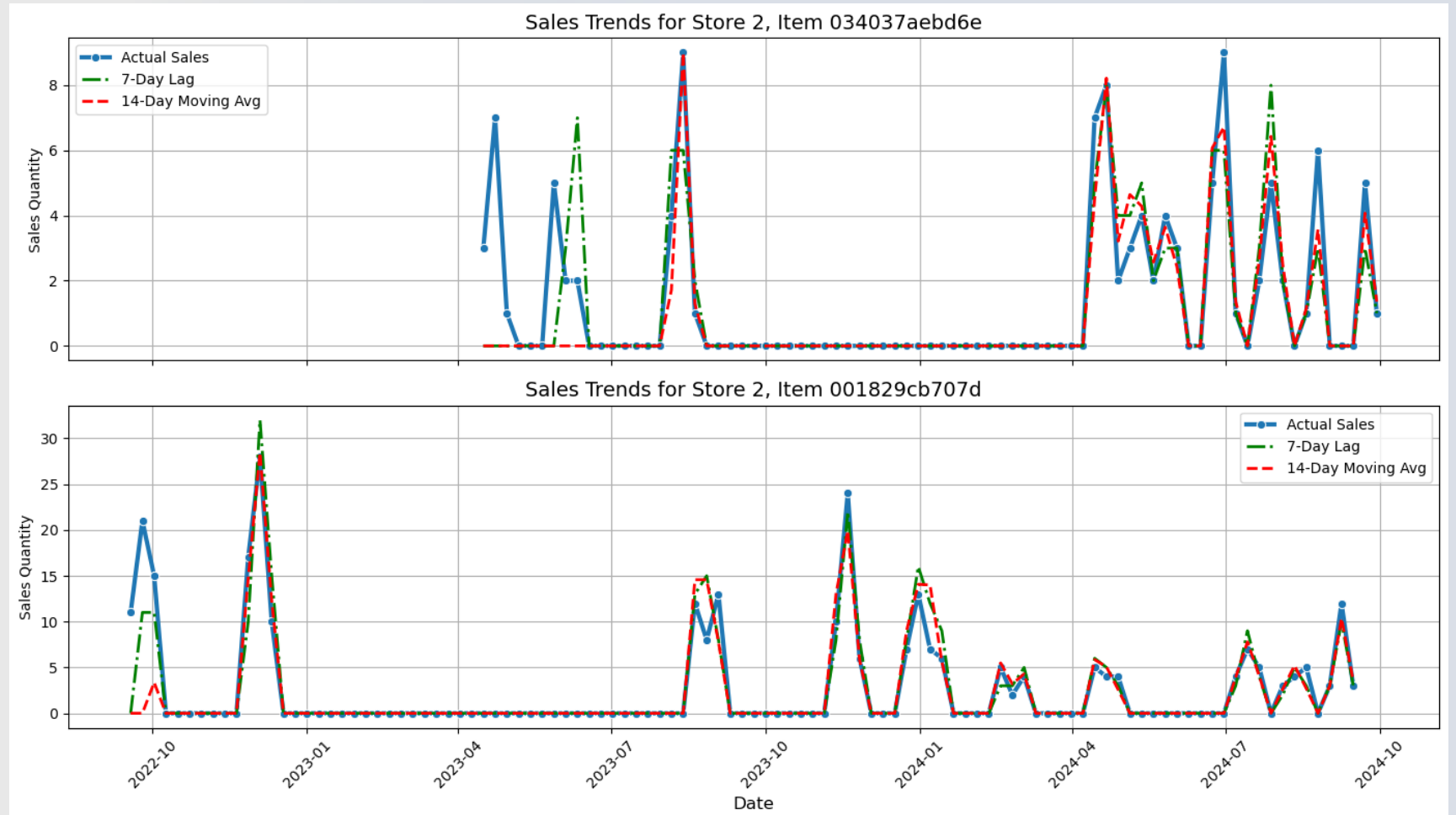
Why This Matters?

- Helps models detect seasonality and trends
- Adds predictive power beyond raw sales numbers

Sales Trend with Lag and Moving Averages

The sales patterns help in identifying;

- Demand cycles
- Seasonal effects



Modeling Approaches

1. LightGBM for Short-Term Demand Forecasting

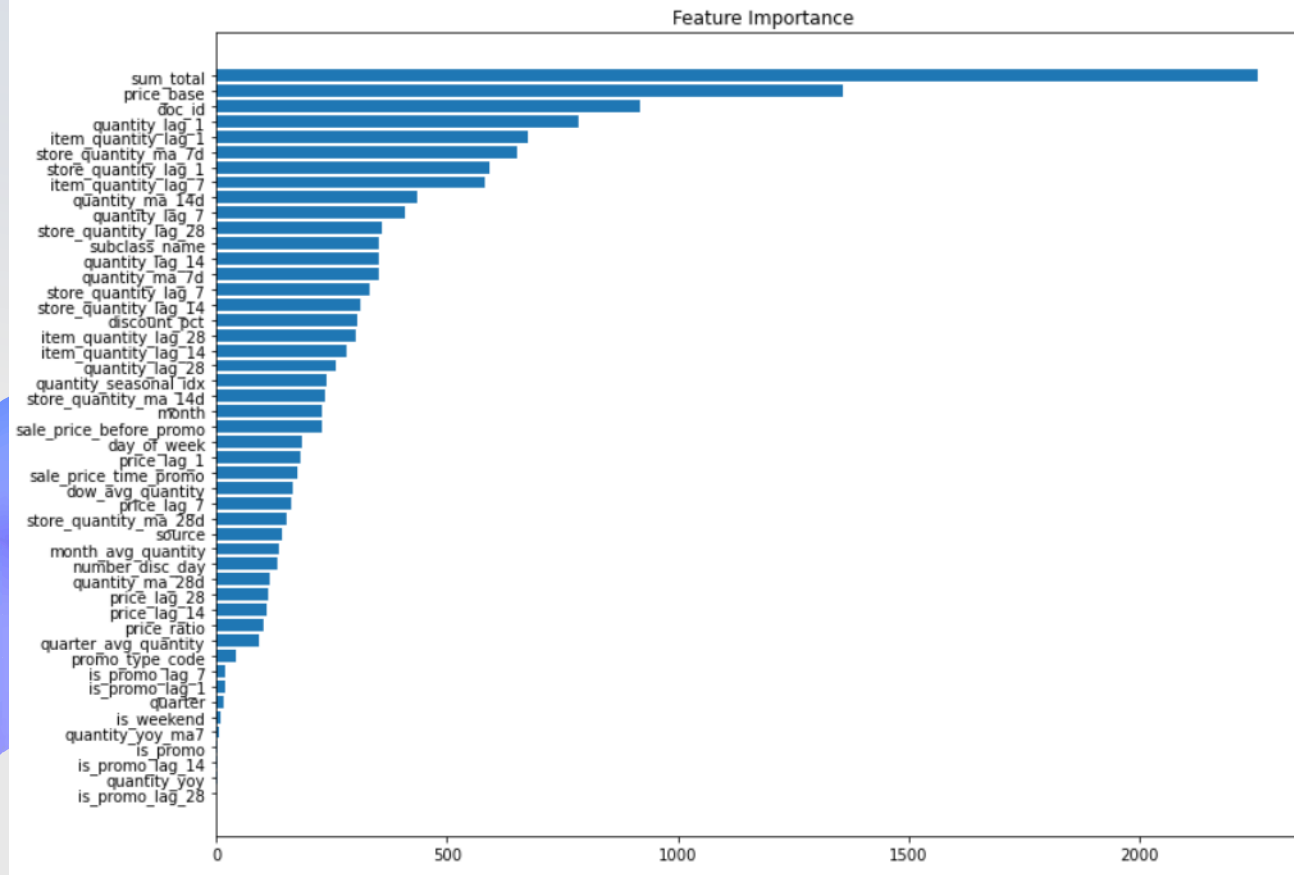
We trained a **gradient boosting model (LightGBM)** to capture key demand drivers, including:

- **Lag Features & Moving Averages** for temporal dependencies
- **Price-related factors** influencing sales
- **Store, Item, and Department-Level Trends**

This model is trained using historical sales data and evaluated using a **time-based validation split** (last 28 days).

LightGBM: Training & Testing

Validation RMSE: 1.2485
Validation MAE: 0.2664



	row_id	quantity
0	0	4
1	1	4
2	2	3
3	3	4
4	4	4
5	5	3
6	6	4
7	7	4
8	8	3
9	9	4

2. Prophet for Time Series Forecasting

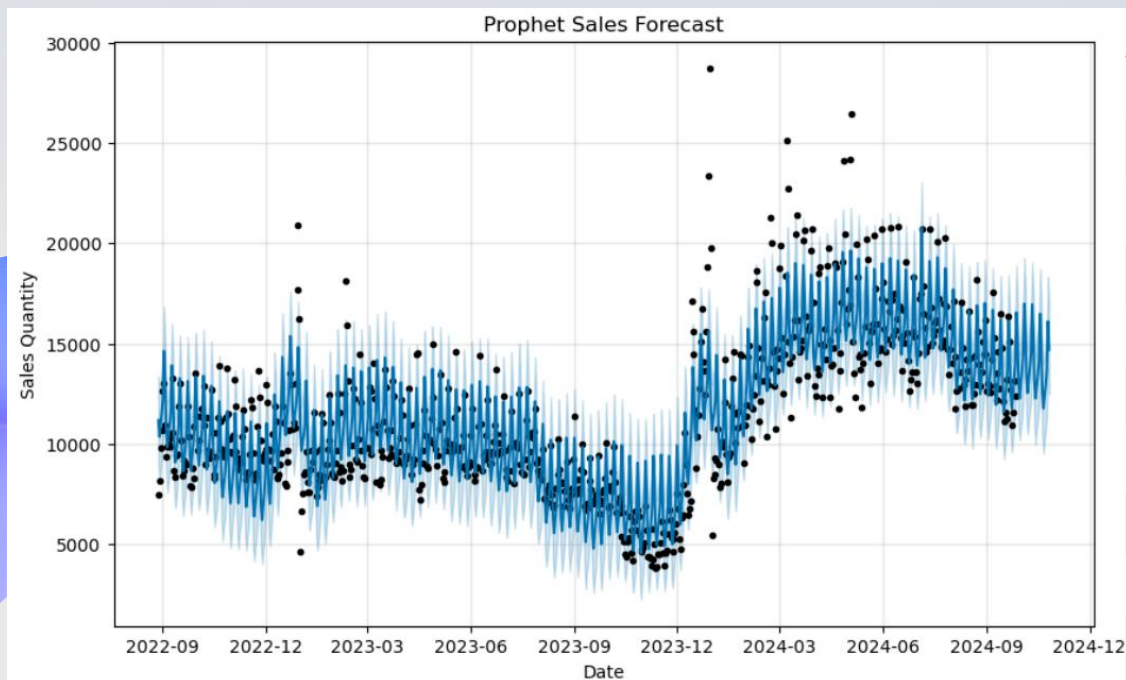
We used **Facebook Prophet** to capture long-term sales trends and seasonality.

Prophet is ideal for:

- **Yearly and Weekly Seasonality** to identify cyclic demand patterns
- **Flexible Trend Adjustments** using additive modeling
- **Generating 30-Day** time inventory planning

LightGBM model provides **high-resolution short-term predictions**, while Prophet enables **long-term trend analysis**.

Prophet: Training & Testing



	ds	yhat	yhat_lower	yhat_upper
781	2024-10-17	14125.568440	11900.644883	16273.660284
782	2024-10-18	16480.597567	14267.700723	18733.434155
783	2024-10-19	15010.743982	12886.298588	17283.426677
784	2024-10-20	12601.172409	10356.494017	14921.659584
785	2024-10-21	11759.823241	9511.178544	14038.691871
786	2024-10-22	12387.236221	10216.844746	14500.487829
787	2024-10-23	12926.214972	10636.184634	15240.721789
788	2024-10-24	13699.693900	11395.626580	15789.386556
789	2024-10-25	16100.243850	13772.444706	18328.657327
790	2024-10-26	14683.890536	12481.067959	17028.819956

Modeling Insights

LightGBM model provided **high-resolution short-term predictions**, while Prophet enabled **long-term trend analysis**.

From LightGBM

- ❖ The validation RMSE suggests that the model performs well, capturing the overall trend of the target variable (quantity) with moderate precision.
- ❖ The relatively low MAE indicates that the model consistently predicts values close to the actual ones, with minimal deviations on average.

Recommendations

True Retail Demand Forecast

1. The stores should leverage the model to determine the true customer demand for a particular product at a given time to minimize overstocking and avoid stockouts

Revenue Growth and Profitability

1. The stores should adapt to the seasonality of demand. For example, all the stores should prepare for the peak sales periods to avoid unexpected market shifts while increasing profitability
2. The stores should not invest heavily on discounts but explore other sales promotions since they do not impact on sales revenues or quantity sold

Adapting to Future Dynamics

1. For inventory optimization, continuous evaluation of the demand forecast is desired based on real-time data to refine the forecasts/ accommodate unseen consumer demands

The slide features a light blue background with a thin white border. On the left side, there is a large, semi-transparent blue sphere. On the right side, there are two smaller blue spheres: one positioned higher and one lower. The text "THANK YOU" is centered in the middle of the slide.

THANK YOU