

# Forecasting for Retail Inventory

Presented by:

Phu Han

Rizvan Ahmed Rafsan

Instructor:

Dr. Diala Ezzeddine

# Contents

|   |                                  |   |                   |
|---|----------------------------------|---|-------------------|
| 1 | Problem Statement                |   |                   |
| 2 | Data Acquisition & Preprocessing | 5 | Modeling          |
| 3 | Exploratory Data Analysis (EDA)  | 6 | Results           |
| 4 | Feature Engineering              | 7 | Challenges        |
|   |                                  | 8 | Future Directions |

# Problem Statement

## The Problem

Inefficient inventory management



Lost Revenue (Sales)  
Wasted Storage  
Higher logistics costs

## Our Goal

Build a predictive model for future demand



Better inventory insights

—  
Better understanding of customer behavior

## Results

- Lower buffer stock
- Lower storage cost
- High-demand items always available
- Improved customer satisfaction



# Data Acquisition



**M5 COMPETITION**

100,000 ENTRIES    100+ COUNTRIES    \$100,000 PRIZE MONEY

A banner for the M5 Competition. The top half features a dark background with a grid of small digital icons. The bottom half has a dark blue bar containing the competition name and three statistics: '100,000 ENTRIES', '100+ COUNTRIES', and '\$100,000 PRIZE MONEY'. Each statistic is accompanied by a small white icon: a document with a pencil, a globe, and a trophy.

**Competition page:** <https://www.unic.ac.cy/iff/research/forecasting/m-competitions/m5/>  
**Dataset Link:** <https://www.kaggle.com/competitions/m5-forecasting-accuracy/overview>

kaggle



# Data Preprocessing

## Datasets

- calendar.csv
- sales\_train\_evaluation.csv
- sell\_prices.csv



| item_id    | dept_id   | store_id | state_id | Day 1 | .... | Day 1940 | Day 1941 |
|------------|-----------|----------|----------|-------|------|----------|----------|
| Item 1     | HOBBIES_1 | CA_1     | CA       | 2     | ...  | 0        | 1        |
| ...        | ...       | ...      | ...      | ...   | ...  | ...      | ...      |
| Item 30490 | FOODS_5   | TX_3     | TX       | 5     | ...  | 4        | 3        |

melted

Merge

Merge

(59181090, 8)

Calendar



Sales Melted



Prices

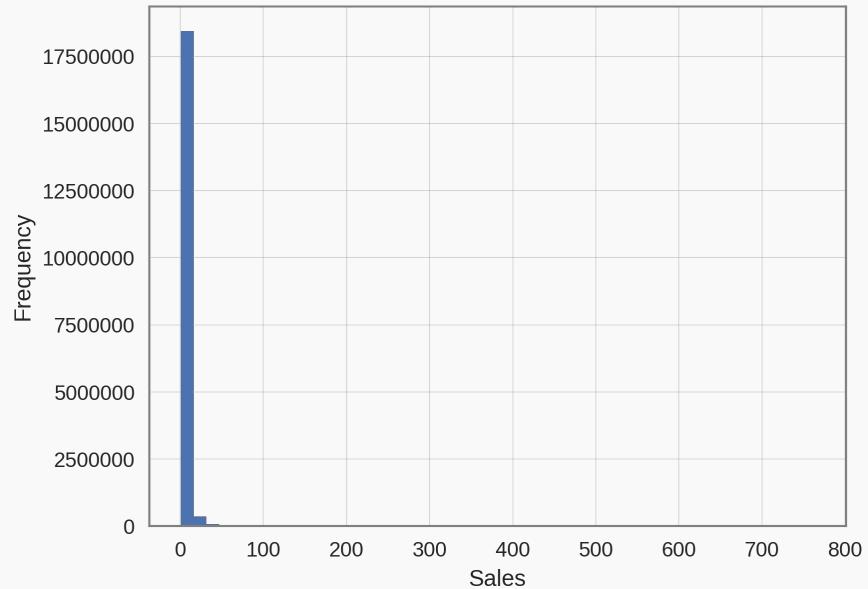
Final  
DataFrame

| item_id    | dept_id     | store_id | state_id | ... | day  | sales |
|------------|-------------|----------|----------|-----|------|-------|
| Item 1     | HOBBIES_1   | CA_1     | CA       |     | 1    | 6     |
| Item 1     | ...         | ...      | ...      | ... | ...  | ...   |
| Item 1     |             |          |          |     | 1947 |       |
| ...        |             |          |          |     |      |       |
| Item 30489 | HOUSEHOLD_5 | WI_2     | WI       |     | 1    | 25    |
| Item 30490 | ...         | ...      | ...      | ... | ...  | ...   |
| Item 30490 | FOODS_5     | TX_3     | TX       |     | 1947 | 3     |

Final DataFrame contains ~60 million rows!

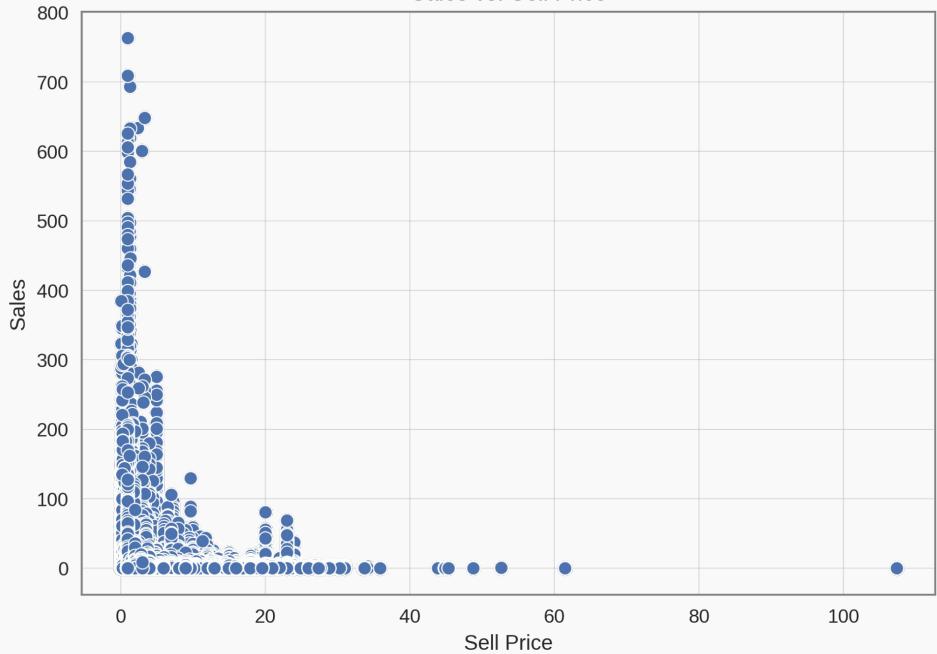
# EDA - Sales Distribution

Distribution of Non-Zero Daily Sales

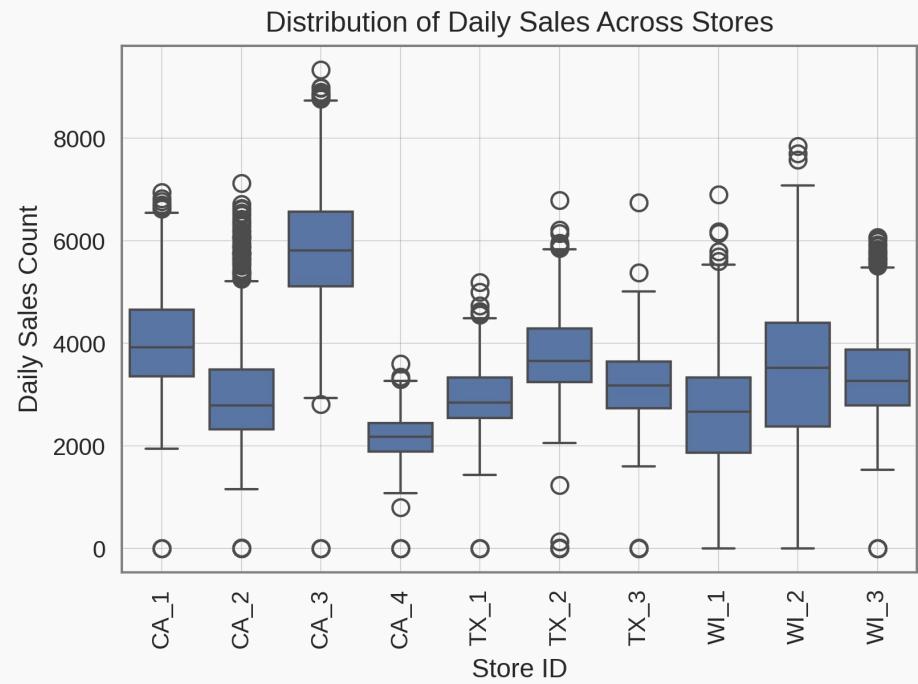
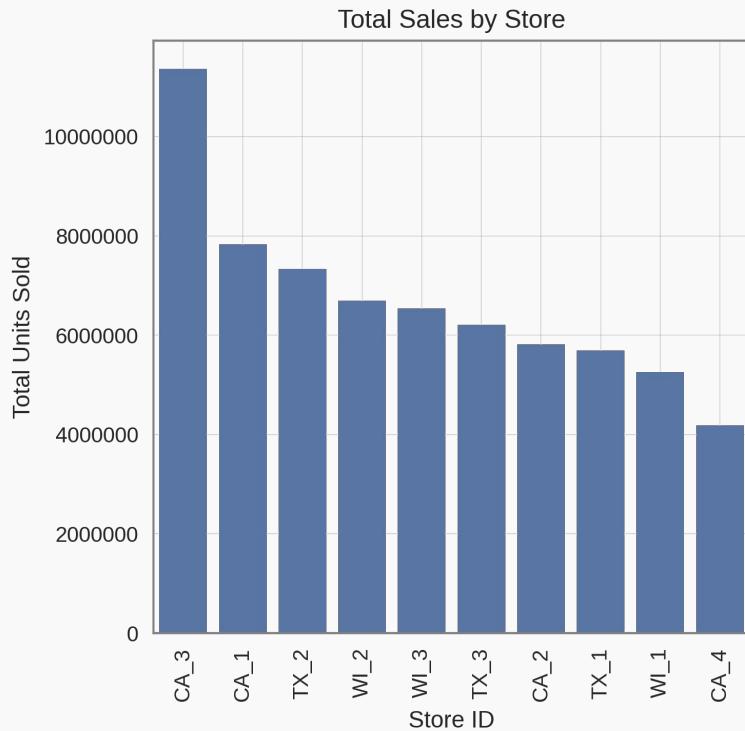


~ 68% of the data instances showed zero sales for items!

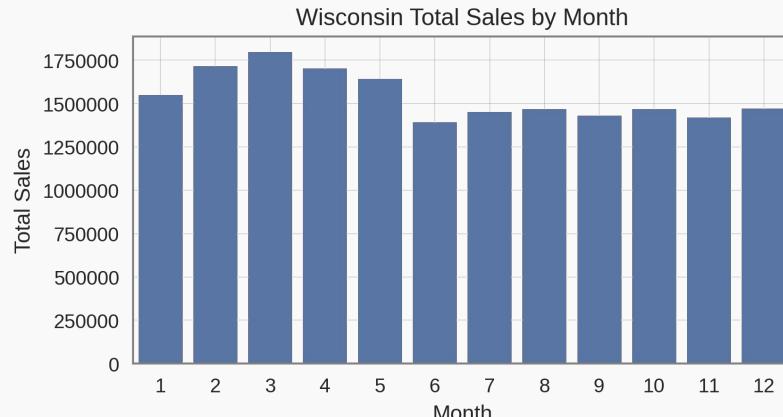
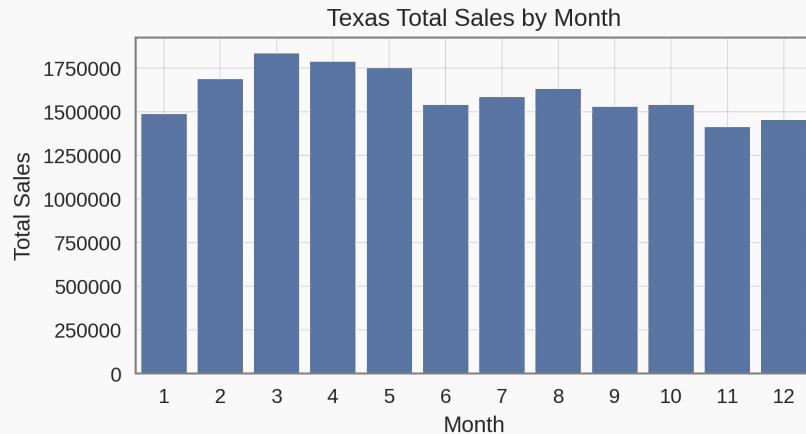
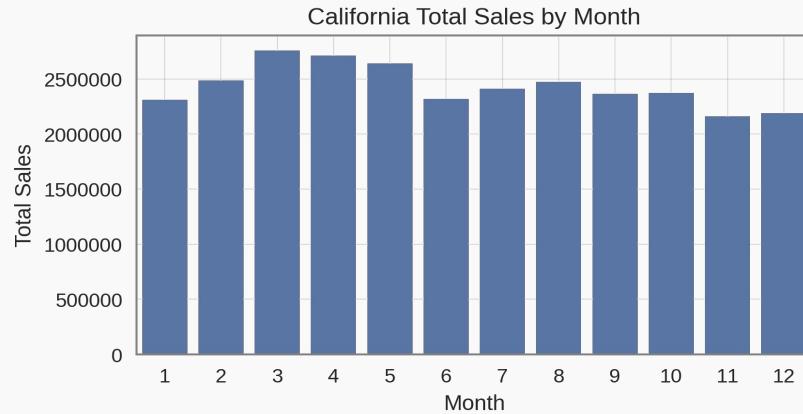
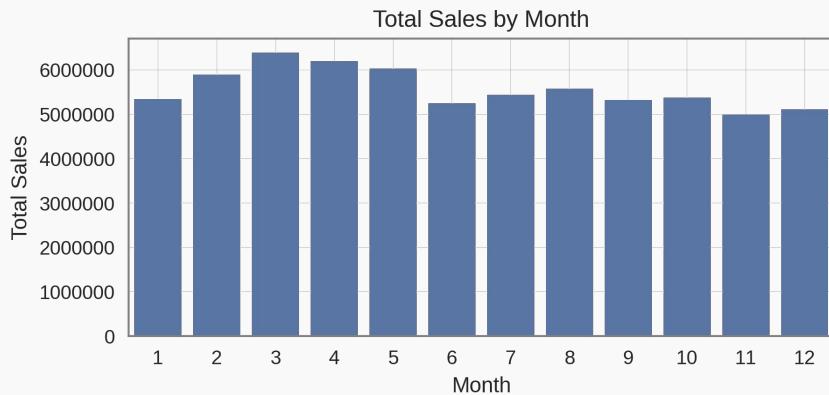
Sales vs. Sell Price



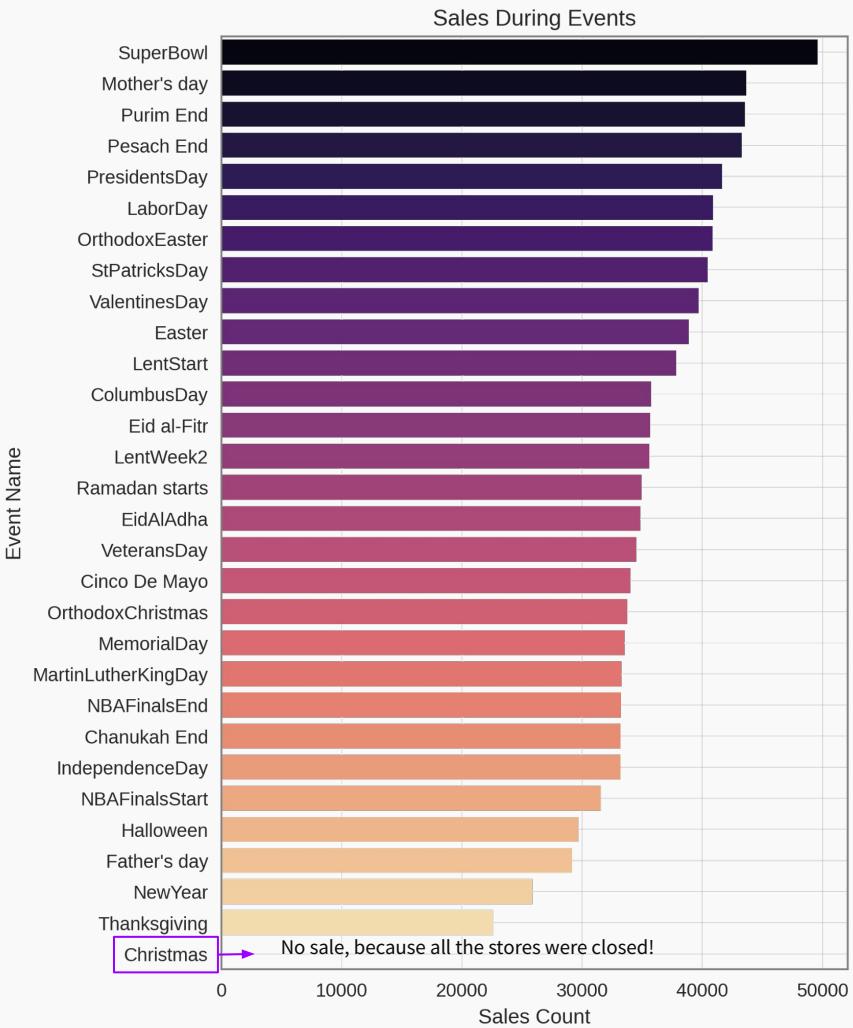
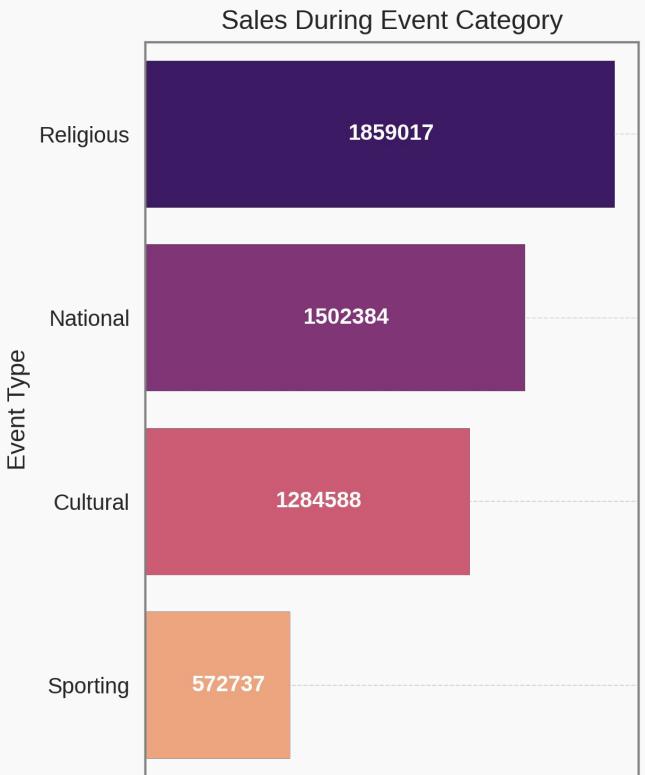
# EDA - Store-wise Distribution



# EDA - Sales by Month



# EDA - Special Events



# Feature Engineering

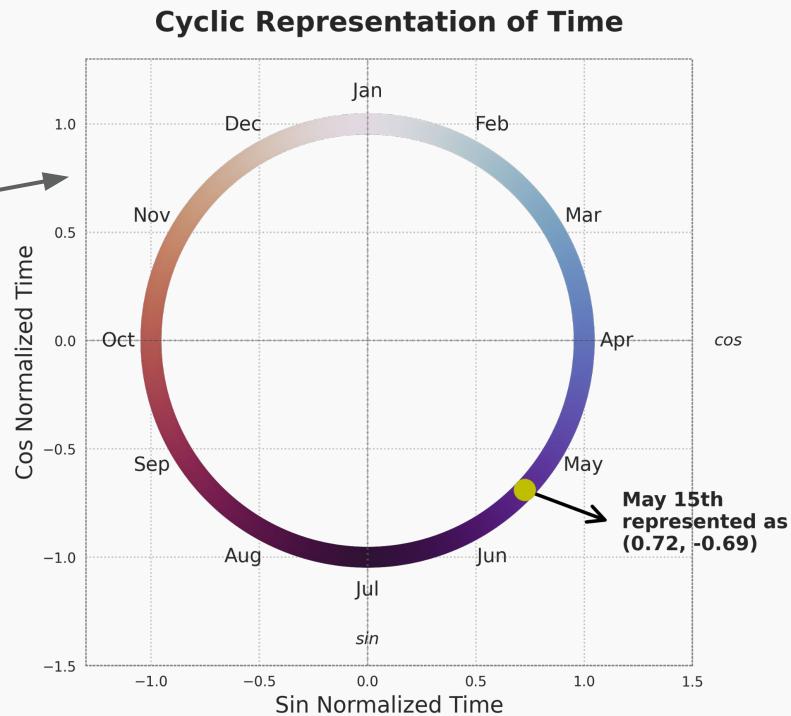
## Time-series dataset

**Temporal Features** (weekend, week of year, payday, ...)

**Cyclic Temporal Features**

**Price Dynamics Features** (lag, differentials, rolling statistics)

**Sales History Features** (lag, differentials, rolling statistics)



---

# Modeling



dmlc  
**XGBoost**

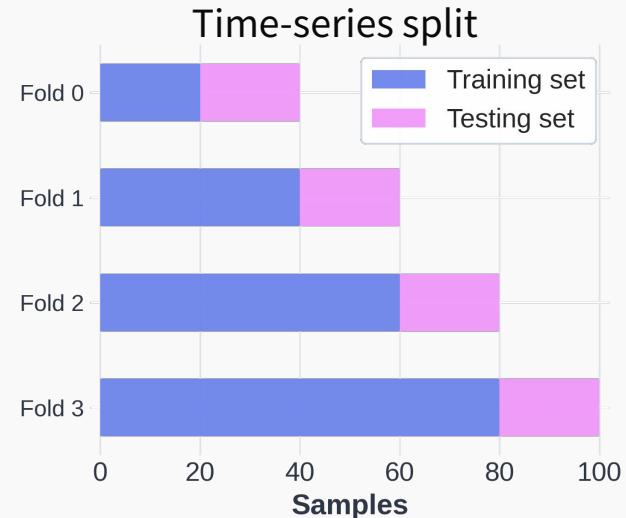


O P T U N A  
Hyperparameter Tuning



# Challenges

| Model           | Tuning Time (minutes) | Training Time (minutes) | Peak RAM Usage (GB) |
|-----------------|-----------------------|-------------------------|---------------------|
| Random Forest   | 321.43                | 35.8                    | 128                 |
| XGBoost         | 296.7                 | 19.2                    | 45.2                |
| <b>LightGBM</b> | <b>49.25</b>          | <b>11.4</b>             | <b>47.5</b>         |

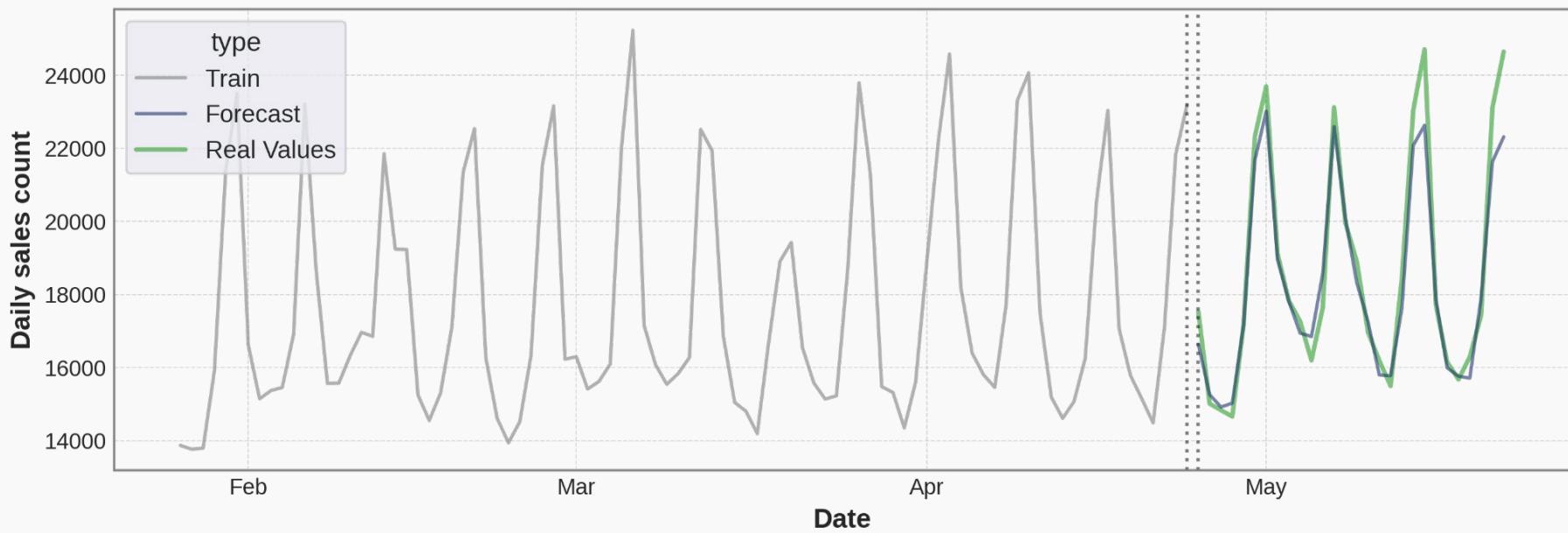


# Model Evaluation

| Model           | RMSE          | MAE           |
|-----------------|---------------|---------------|
| Random Forest   | 1.9392        | 0.9933        |
| XGBoost         | 1.9340        | 0.9872        |
| <b>LightGBM</b> | <b>1.9213</b> | <b>0.9789</b> |

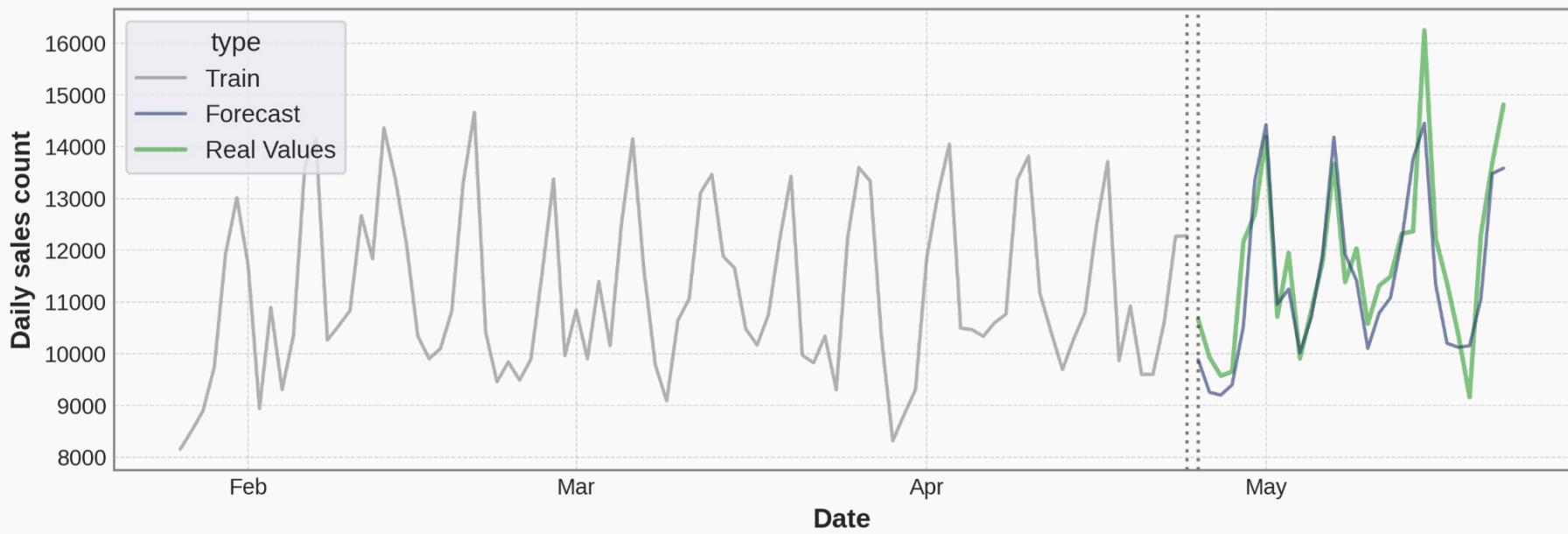
# Final Prediction

Forecast for all stores in CA



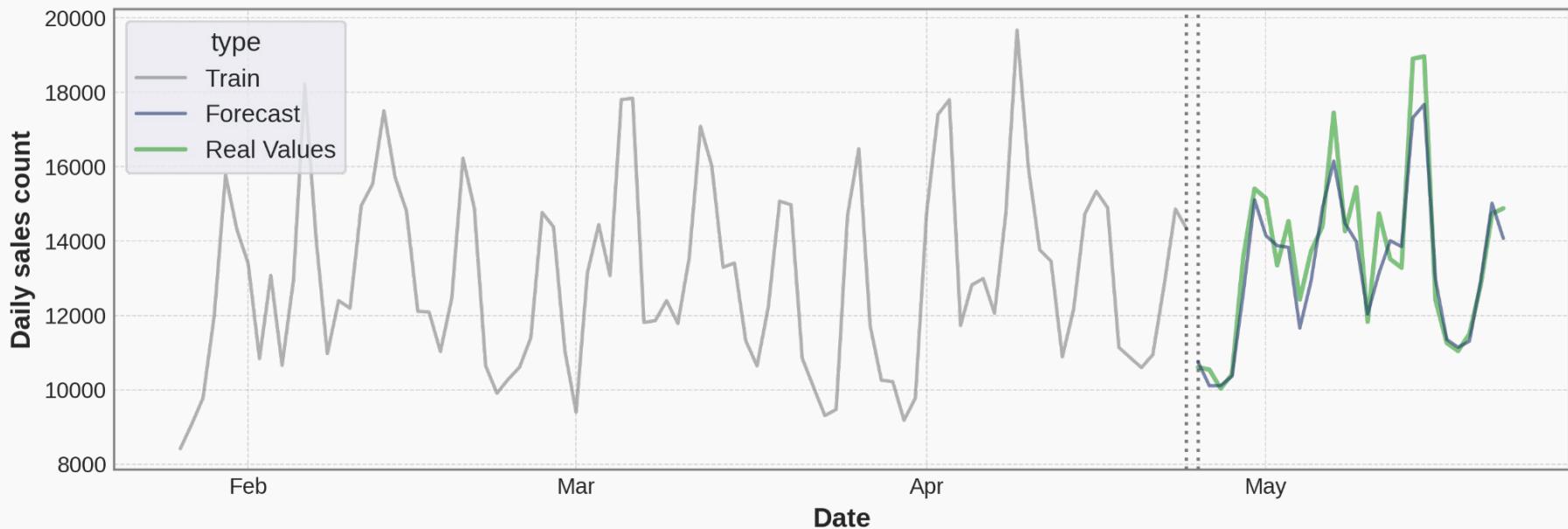
# Final Prediction

Forecast for all stores in TX



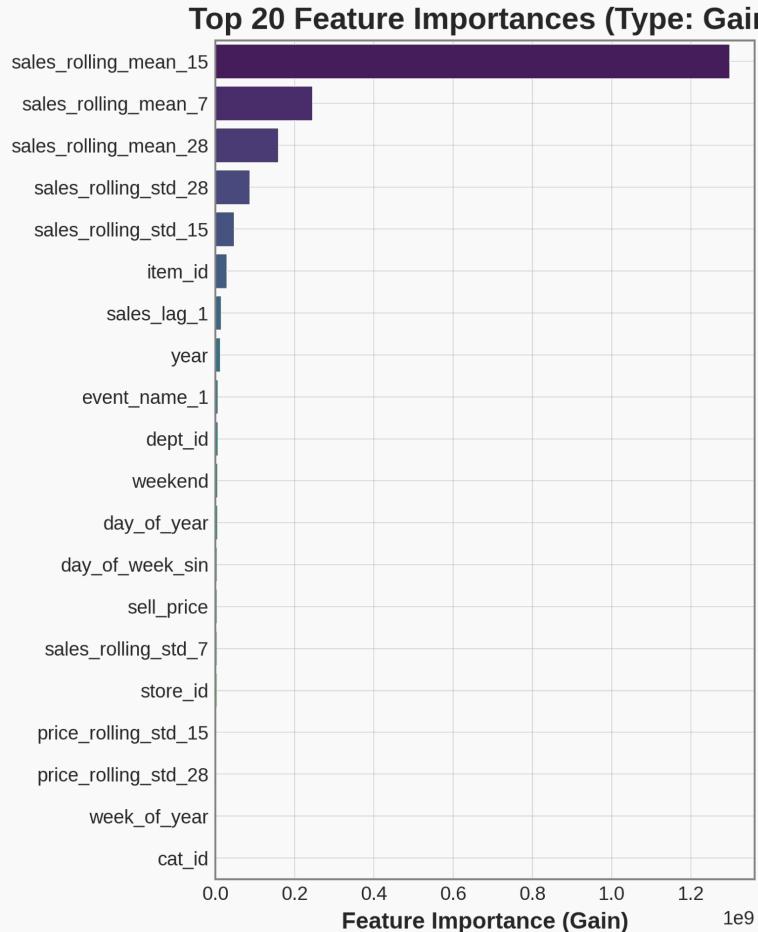
# Final Prediction

Forecast for all stores in WI

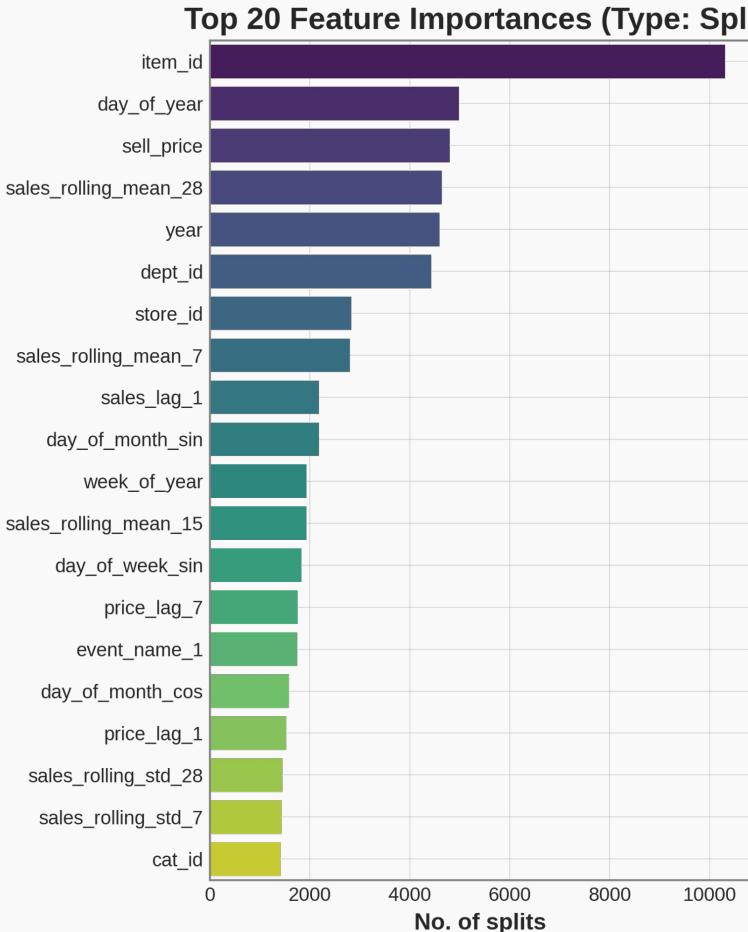


# Features Importance

Feature Name



Feature Name



# Future Directions

- Spark or PySpark
- Deep learning architectures: **Temporal Fusion Transformers (TFT)**
- Using a **classification model** to predict if a product will be sold or not
- Leveraging **hurdle modeling technique** using the classification model to test if can generate better predictions



# Citations

1. Paper: Fildes, R., Ma, S., & Kolassa, S. (2022). Retail forecasting: Research and practice. International Journal of Forecasting, 38(4), 1283-1318.
2. M5 Forecast Accuracy. University of Nicosia · Featured Prediction Competition. Can be accessed at: <https://www.kaggle.com/competitions/m5-forecasting-accuracy>
3. Mastering Time Series Analysis in Python: Tools for Time Series Analysis and Forecasting in Python. Sadrach Pierr. Link: <https://towardsdatascience.com/mastering-time-series-analysis-in-python-8219047a0351/>
4. AdaBoost in Machine Learning. Geeksforgeeks. Link: <https://www.geeksforgeeks.org/machine-learning/AdaBoost-in-Machine-Learning/>
5. Ke, G., Meng, Q., Finley, T., et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.
6. A PySpark Example for Dealing with Larger than Memory Datasets. Georgia Deaconu. Link: <https://towardsdatascience.com/a-pyspark-example-for-dealing-with-larger-than-memory-datasets-70dbc82b0e98/>

# Thank you!

Do you have any questions?

