



ANALYSIS OF SALES DATA AT WALMART

Farhan Nugraha Pratama






Project Vision

01. Analysis of factors
affecting Weekly Sales

02. Analyzing Weekly Sales
Report using autogluon



DATA PREPROCESSING

01

Collected from
Kaggle

02

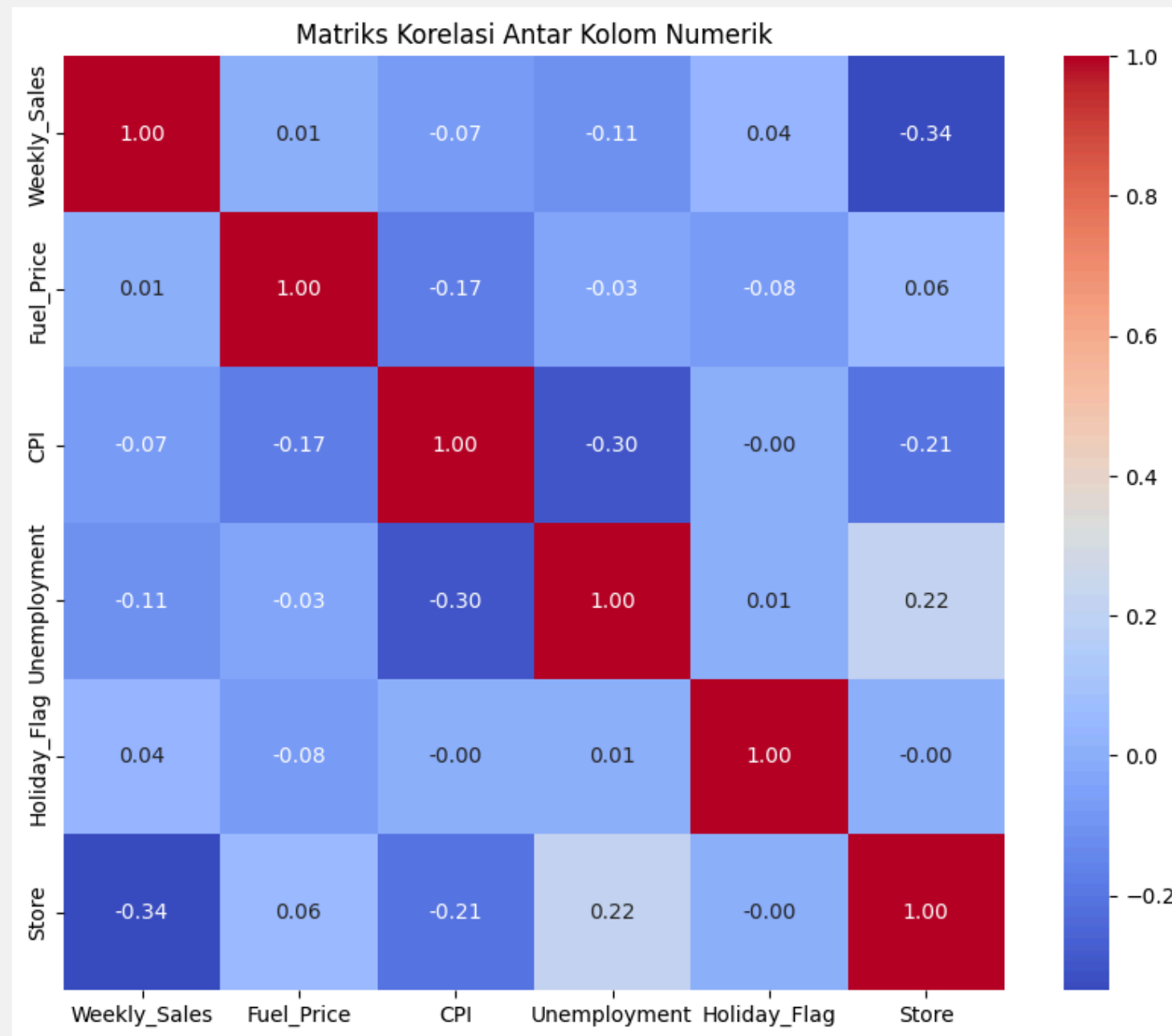
6435 Data with 8
Column

```
# Column Non-Null Count Dtype
---
0 Store 6435 non-null int64
1 Date 6435 non-null datetime64[ns]
2 Weekly_Sales 6435 non-null float64
3 Holiday_Flag 6435 non-null int64
4 Temperature 6435 non-null float64
5 Fuel_Price 6435 non-null float64
6 CPI 6435 non-null float64
7 Unemployment 6435 non-null float64
dtypes: datetime64[ns](1), float64(5), int64(2)
```

03

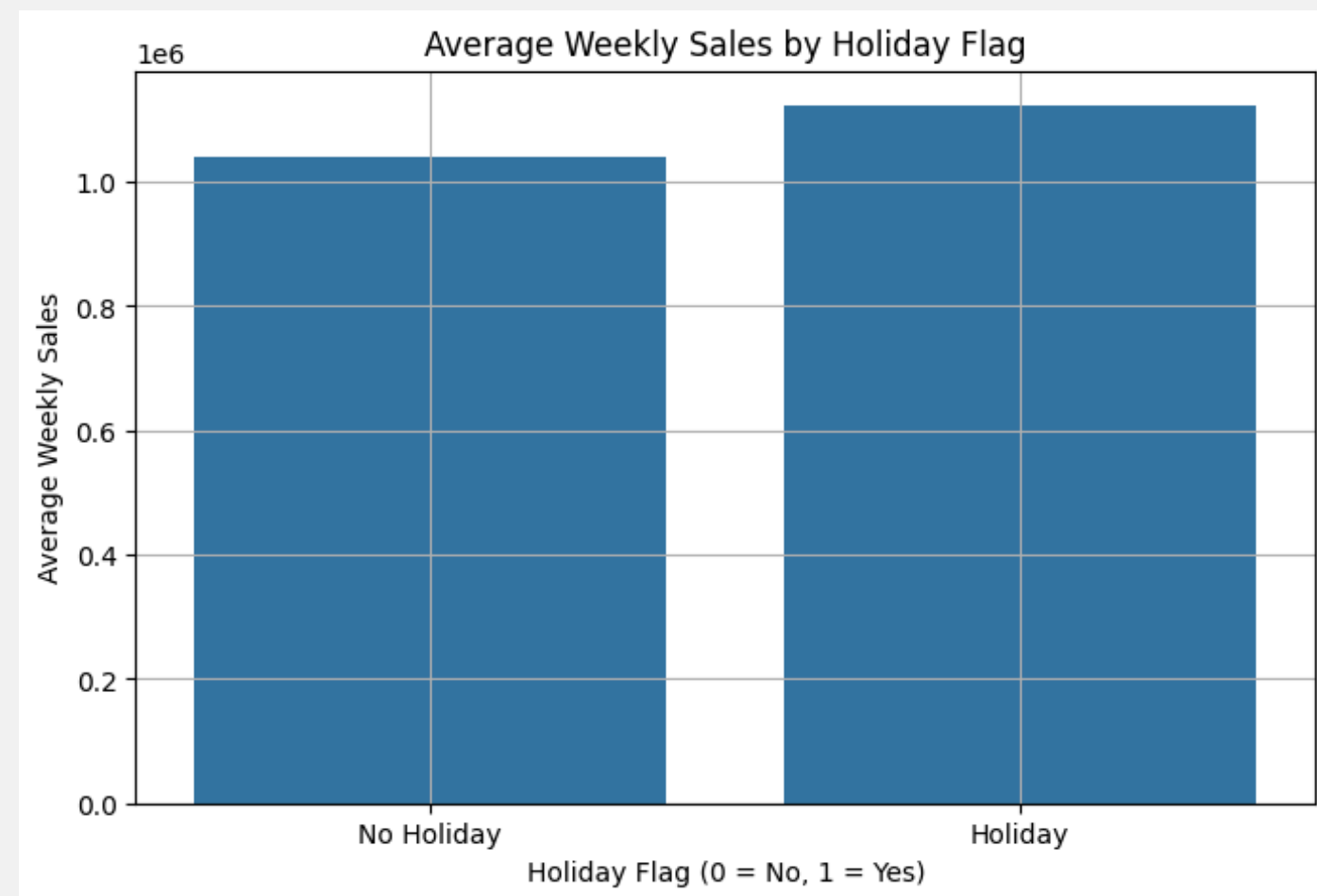
Check Missing
Value and
Duplicates Data
6435 Complete
with 0 Missing
Value and
Duplicates Data

EDA: Matrix Analysis



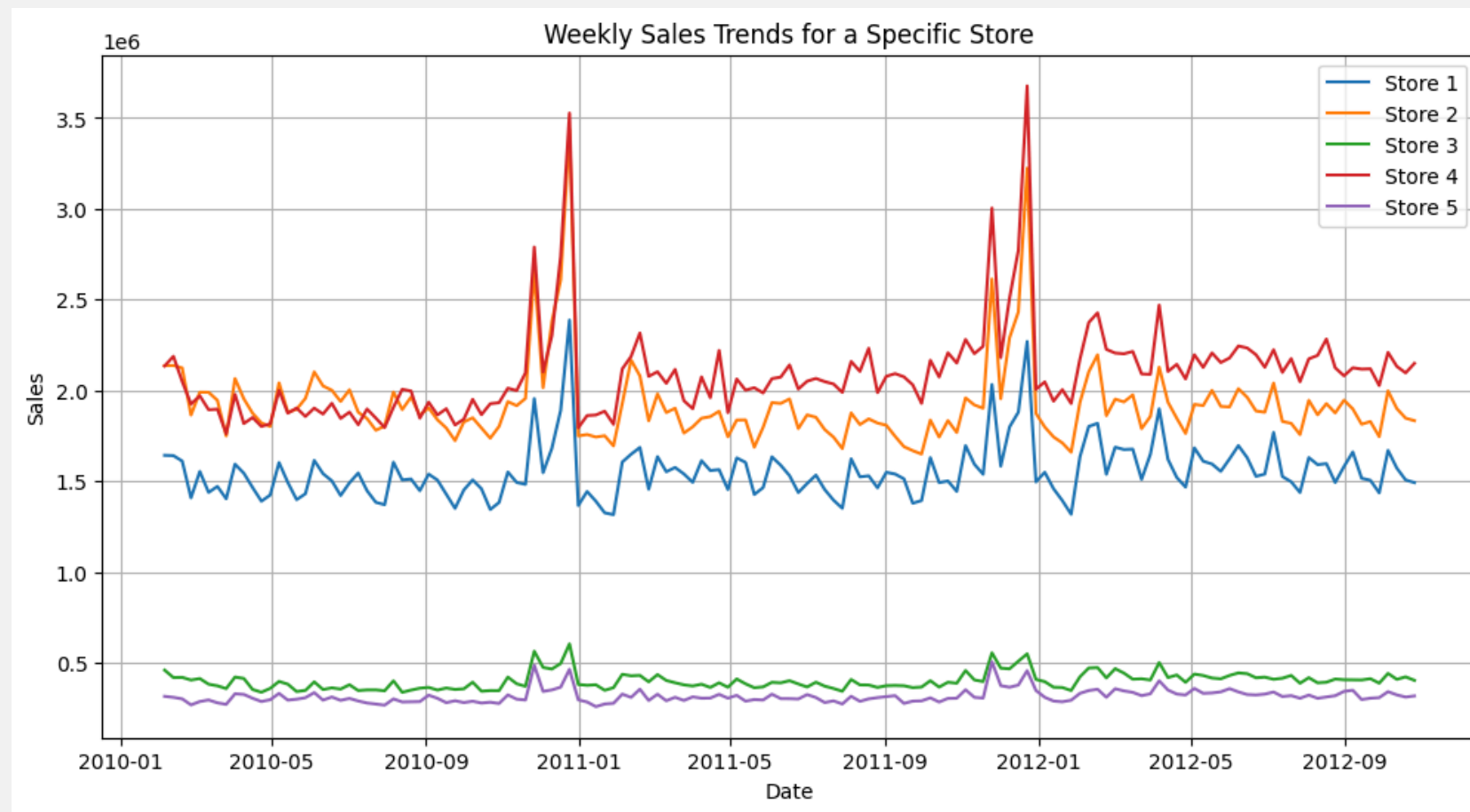
- Store shows a moderate negative correlation with Weekly_Sales, indicating that store identity has a considerable influence on weekly sales levels.
- Variables such as Fuel_Price, CPI, and Holiday_Flag show very weak correlations with Weekly_Sales, so they are unlikely to be the main drivers of variations in weekly sales.

EDA: Weekly Sales Report



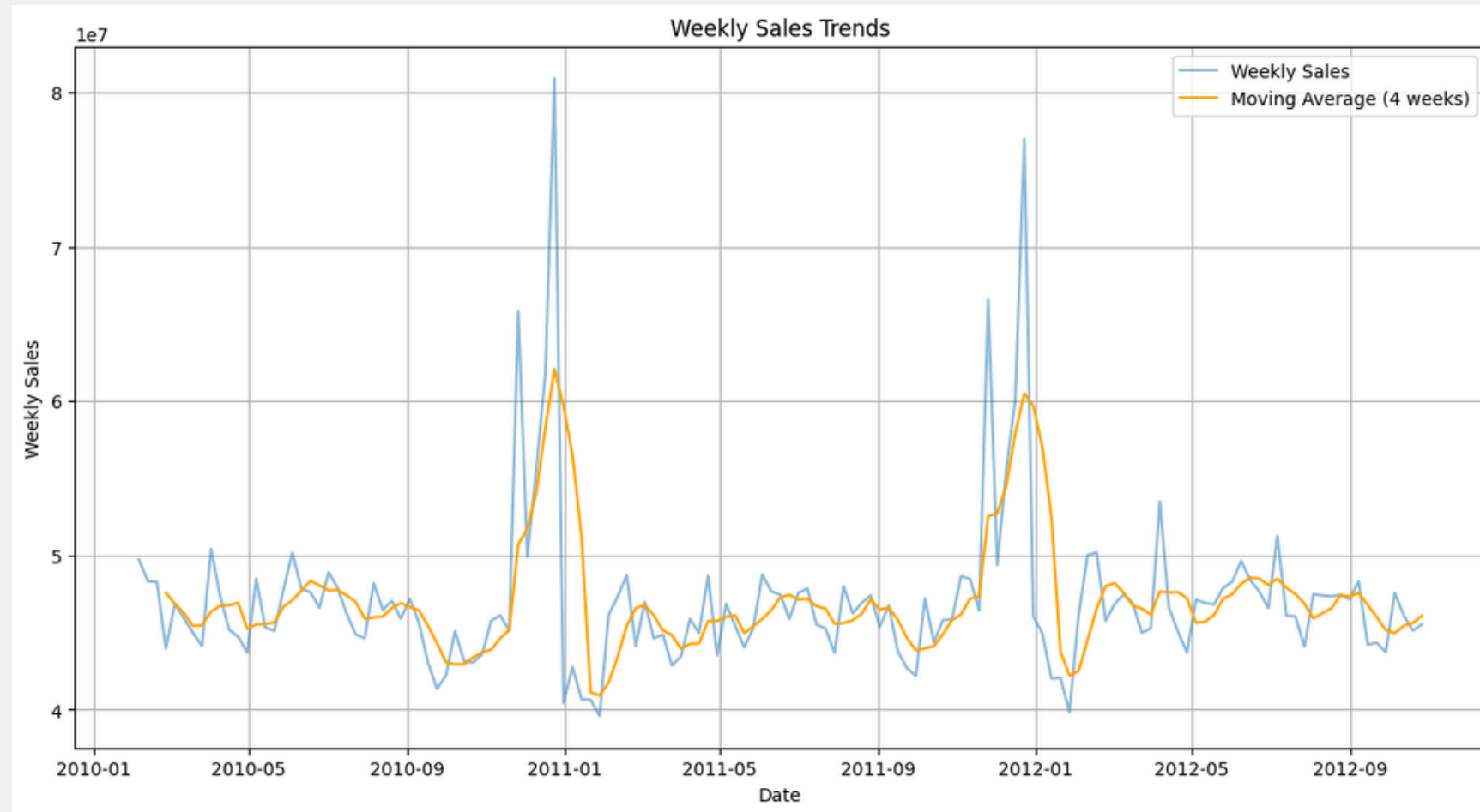
Analysis of average weekly sales based on holiday indicators shows that there is an increase in sales on weeks that include holidays.

EDA: Weekly Sales Trends



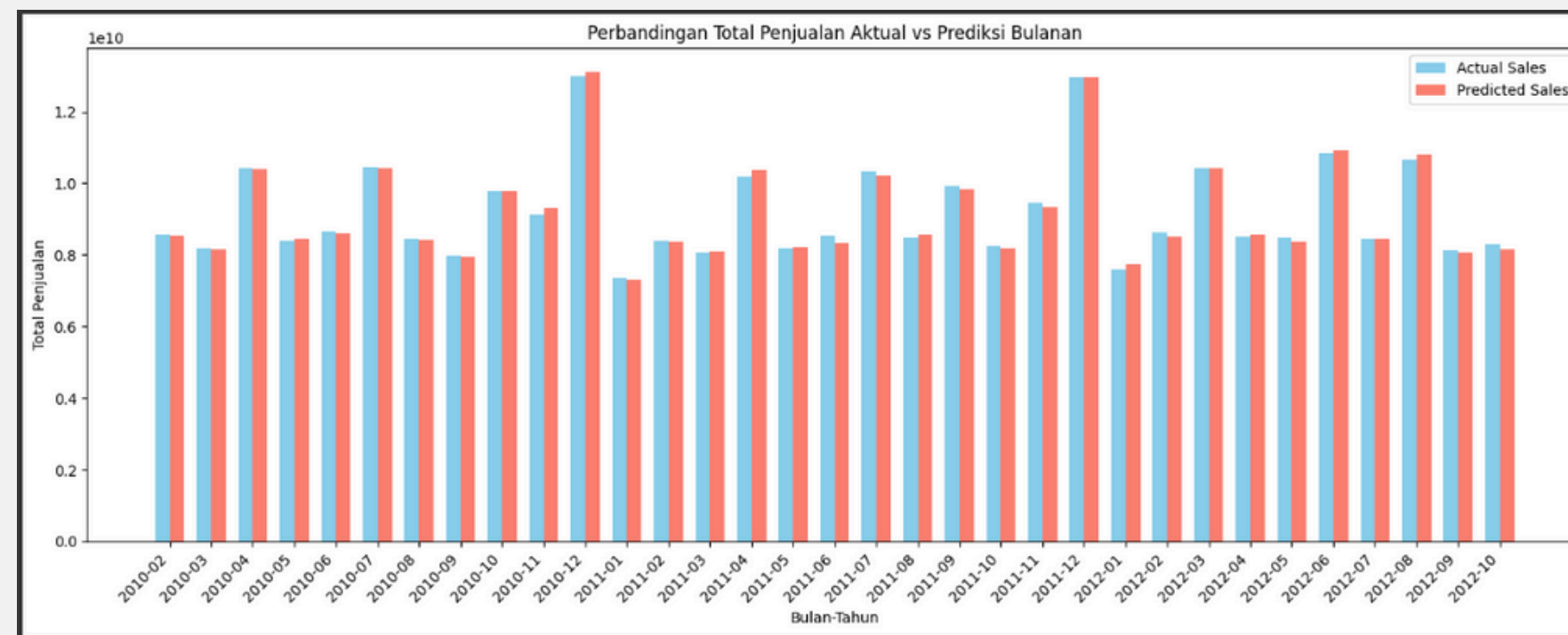
- There is a significant difference in weekly sales levels between the five stores featured.
- Sales at each store are strongly influenced by external factors

EDA: Weekly Sales Trends



There's a clear seasonal pattern in the weekly sales. Notice the significant peaks occurring roughly around the end of the year (late November/December) in both 2010 and 2011. This suggests a strong positive impact of holiday shopping or end-of-year promotions on sales.

EDA: Actual vs Prediction

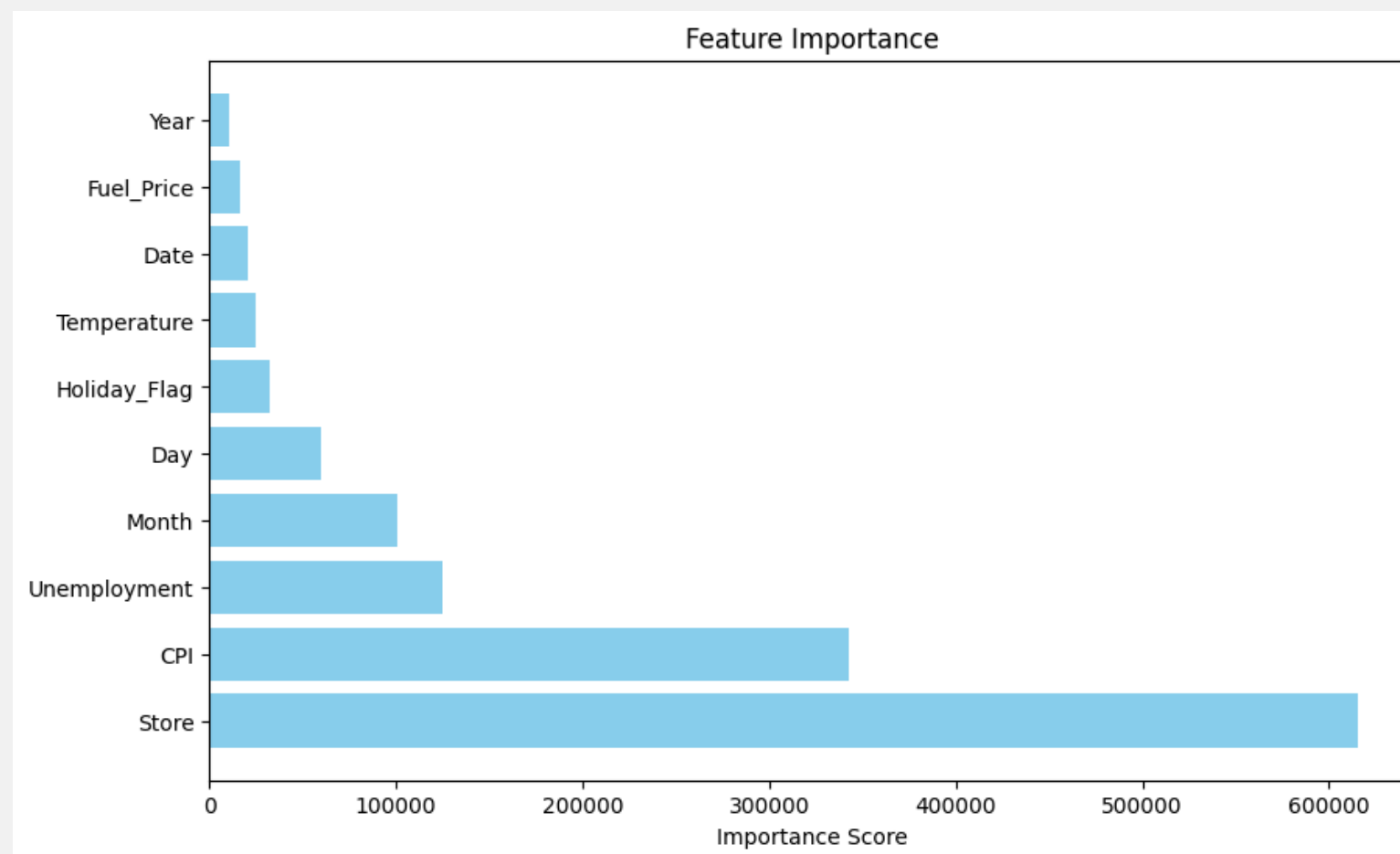


- There is little difference between predicted and actual sales
- In some months, there is a clear difference between actual and predicted total sales. This illustrates the presence of external or internal factors that may not be captured in the prediction model, which could also be changes in consumer behavior or disruptions in supply.

Store Sales Forecasting with AutoGluon

	Feature	Importance
Store	Store	615875.034641
CPI	CPI	342809.511526
Unemployment	Unemployment	124668.682584
Month	Month	101080.562996
Day	Day	59859.111440
Holiday_Flag	Holiday_Flag	32384.516481
Temperature	Temperature	24492.798200
Date	Date	20209.582709
Fuel_Price	Fuel_Price	16125.228833
Year	Year	10335.513395

- Store is dominant factor
- Economic Indicators Matter (Consumer Price Index)
- Time-Based Patterns Exist
- Holiday Impact is Noticeable
- External Factors Play a Smaller Role





Store Sales Forecasting Performance with AutoGluon

- {'root_mean_squared_error': 66420.49509406292,
- 'mean_squared_error': 441682168.540437,
- 'mean_absolute_error': 40742.462632575756,
- 'r2': 0.9864126703804796,
- 'pearsonr': 0.9932250749689345,
- 'median_absolute_error': 26008.0925000000028}



Model demonstrates a very strong ability to predict the target variable, as evidenced by the high R-squared and Pearson r. While there are some errors, as indicated by the RMSE and MAE, they need to be interpreted in the context of data's scale. The median absolute error provides further insight into the distribution of these errors.



Recommendation

- **Prioritize Store-Level Strategies:** Given the overwhelming importance of the "Store" feature, focus your efforts on understanding and optimizing performance at the individual store level.
- **Deep Dive into Store-Level Data:** Conduct more granular analysis at the store level.
- **Monitor and Respond to Economic Conditions:** The significant importance of "CPI" and "Unemployment" suggests a strong sensitivity to the broader economic environment.
- **Leverage Time-Based Patterns:** The moderate importance of "Month" and "Day" indicates predictable sales fluctuations.
- **Capitalize on Holiday Effects:** The importance of "Holiday Flag" confirms the impact of holidays.

The background is a light gray color, decorated with various hand-drawn blue doodles. These include several overlapping circles and loops at the top, a series of concentric arcs at the bottom left, a wavy line at the bottom center, and several small 'v' shapes at the bottom right. On the far right edge, there are some vertical blue strokes. The central text 'Thank you' is written in a bold, black, sans-serif font with a white outline, making it stand out against the gray background.

Thank you