

Walmart Sales Forecasting

June 10, 2024

Agenda

- Walmart Sales Forecasting
- Analyzing and Predicting Weekly Sales
- Data Sources and Description
- Data Preprocessing
- Store Types Distribution
- Sales Distribution
- Holiday Impact on Sales
- Correlation Matrix
- Feature Engineering
- Modeling
- Evaluation
- Visualization
- Conclusion
- Q&A

Analyzing and Predicting Weekly Sales for Walmart Stores

This presentation delves into the methods and processes used to forecast sales for Walmart stores based on historical data. It highlights the importance of sales forecasting for determining seasonal demands, preventing financial losses, positively influencing stock prices, managing inventories, and planning effective marketing campaigns.

Data Sources and Description

Source

Walmart's historical sales records.

Key Columns

Store: The ID of the store. Date: The week of sales. Weekly_Sales:
The sales amount for the week. IsHoliday: Indicates if the week
includes a holiday.

Additional Features

Data includes features such as store number, department, date, weekly
sales, and other variables.

Data Preprocessing

Merging Datasets

Using the `pd.merge()` function in pandas to combine datasets on the 'Store' column.

Date Conversion

Converting date strings to datetime objects using `pd.to_datetime()`.

Handling Missing Values

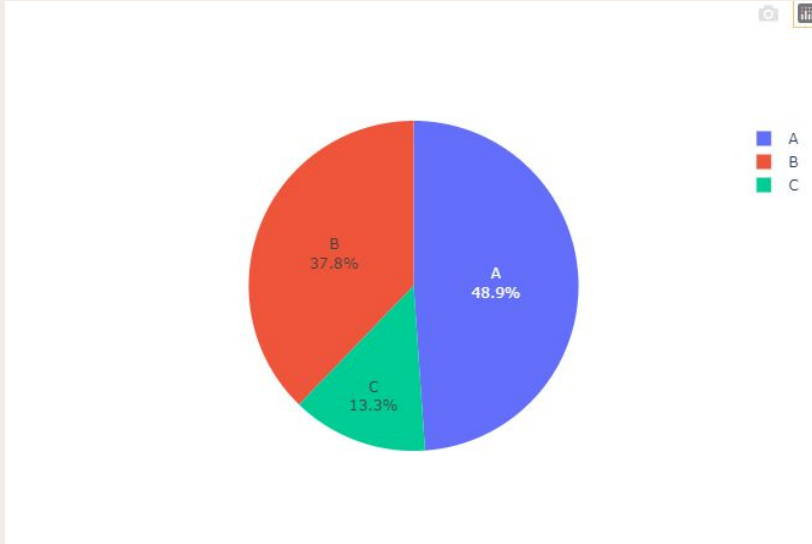
Techniques: Filling missing values with mean/median values using `fillna()`; Dropping rows with missing values using `dropna()` if the proportion is small; Using interpolation methods where appropriate.

Handling Outliers

Methods: Identifying outliers using the Interquartile Range (IQR) method; Using Z-score to detect outliers; Treating outliers by capping or using transformation techniques like log transformation.

Store Types Distribution

Distribution of Store Types



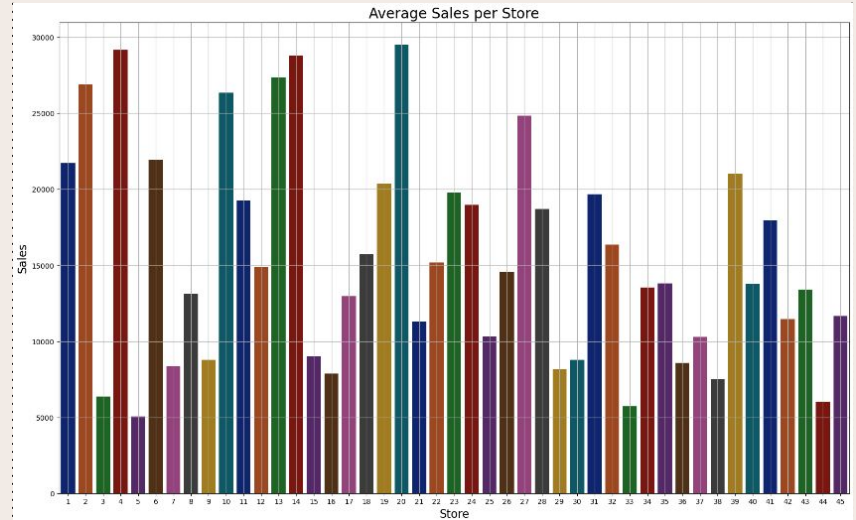
Type A stores are the most numerous among all store types, which helps tailor forecasting models.

The trend that larger stores are more prevalent helps maintain market demand and supply in check.

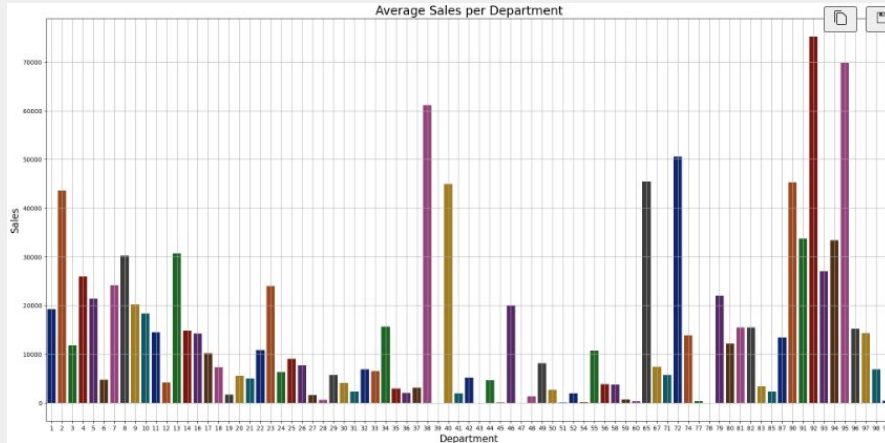
Sales Distribution

Overall Sales

- Distribution of weekly sales across different stores and departments.
- Provides insight into sales patterns and variability over time.
- Key for understanding the performance of stores and departments over time.



Sales Distribution Cont.



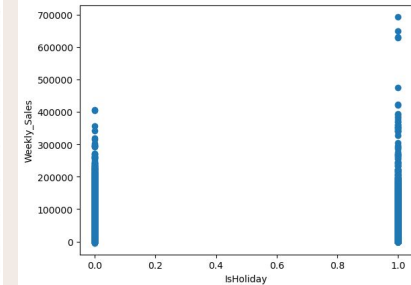
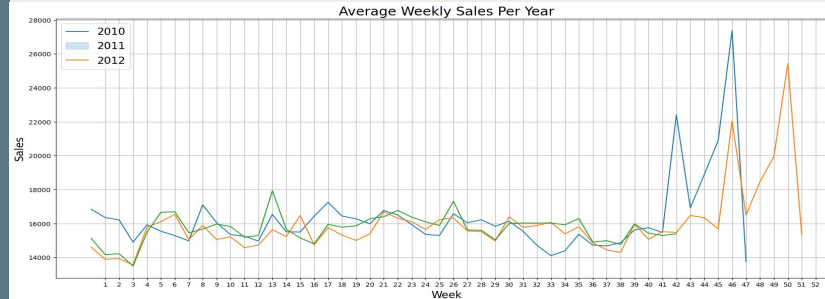
Insight

- Helps in identifying high-performing stores and departments.
- Useful for tailoring forecasting models.
- Supports inventory and marketing strategy adjustments.
- Demonstrates the necessity of supply storage during times of High Demand

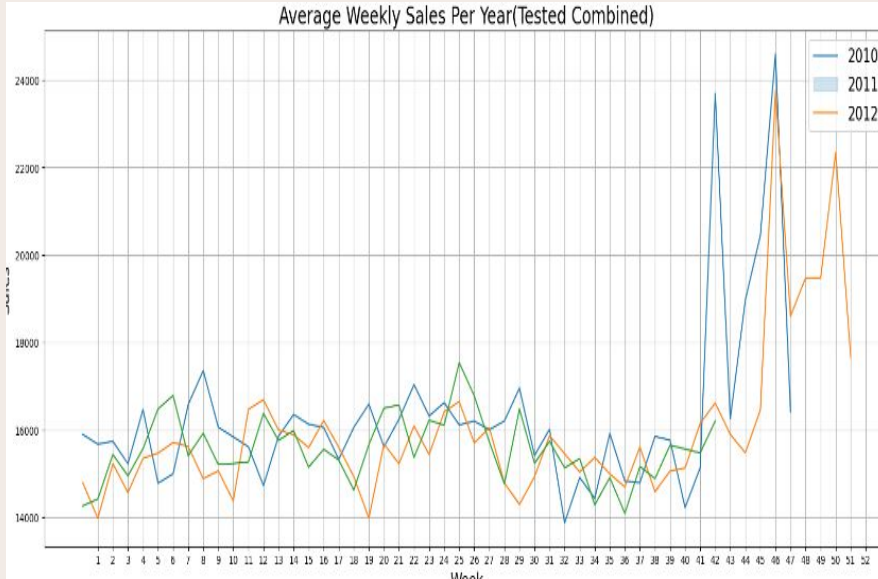
Holiday Impact on Sales

Impact of Holidays on Sales

- Historical sales data shows significant spikes during holiday weeks, such as Thanksgiving and Christmas.
- Holidays influence consumer behavior, leading to increased spending on gifts, decorations, and holiday-specific items.
- Accurately forecasting holiday sales can help Walmart manage inventory and staffing more effectively.



Correlation Matrix



Correlation Analysis

Analyzed relationships between features and weekly sales using a correlation matrix. Key predictors identified are the holidays and far more erratic pathing of election years.

Insights

Key relationships: Strong correlations between store types, holidays, and weekly sales.

Feature Engineering

Creating New Features

Generated new features such as multiple graphs/charts in one cell, multiple line graphs on one chart, as well as learning multiple ways to create different models.

Aggregate Features

Calculated rolling averages, cumulative sums, etc. These aggregates help in capturing trends over time and improve model accuracy.

Temporal Features

Extracted day of the week, month, and holidays. These features help in understanding the temporal patterns in sales data.

Feature Selection

Selected important features based on correlation analysis and domain knowledge.

Modeling

Model Selection

Chose Ridge and Lasso regression models for forecasting.

Hyperparameter Tuning

Used GridSearchCV and RandomizedSearchCV for hyperparameter tuning. Optimized parameters for the best model performance.

Training and Validation

Split data into training and testing sets. Trained models and validated their performance using cross-validation.

Evaluation



Model Performance Metrics

Evaluated models using metrics such as Mean Squared Error (MSE) and R-squared (R2) score. For example, Ridge MSE: ``ridge_mse``, Lasso MSE: ``lasso_mse``, Ridge R2: ``ridge_r2``, Lasso R2: ``lasso_r2``.



Visualizations of Results

Visualized actual vs predicted sales. For instance, plotting the actual sales against Ridge and Lasso predictions to observe performance.



Comparison of Models

Compared the performance of Ridge and Lasso regression models.



Future Improvements

Build different models for specific stores or departments. Conduct market basket analysis to identify high-demand items in departments.

Conclusion

- 1 Summarized key insights from the analysis and modeling.
- 2 Key factors influencing sale volume: holidays, store types, and department performance. Other factors may include sales, shortages, stormy weather, etc.
- 3 Best-performing model: Exponential Smoothing with the lowest error.
- 4 Future work: additional feature engineering, experimenting with other models, and improving markdown effect modeling.

View The Code:

We will now go to present our code!

Q&A

Thank you for your attention. Now, we would like to open the floor for any questions or comments you may have regarding our Walmart Sales Forecasting project. Feel free to ask about the data sources, preprocessing steps, modeling techniques, or our findings and future improvements. We are here to discuss and clarify any points.