Retail Sales Insights & Demand Forecasting

Capstone project

Name: Shridevi Kulkarni

PROJECT OBJECTIVE

Objective:

- Analyze Walmart's retail sales data
- Identify key patterns and drivers of weekly sales
- Build a predictive model to estimate sales
- Provide business insights and recommendations

Business Importance:

Better understanding of sales patterns helps optimize:

- Inventory
- Marketing
- Promotions
- Store planning

DATASET OVERVIEW

Files Used:

- train.csv
- features.csv
- stores.csv

Dataset Summary:

- ~421,000 rows
- 45 stores
- 81 departments
- Variables include weekly sales, holidays, store type, size, fuel price, CPI, unemployment.

Final dataset created by merging all three files.

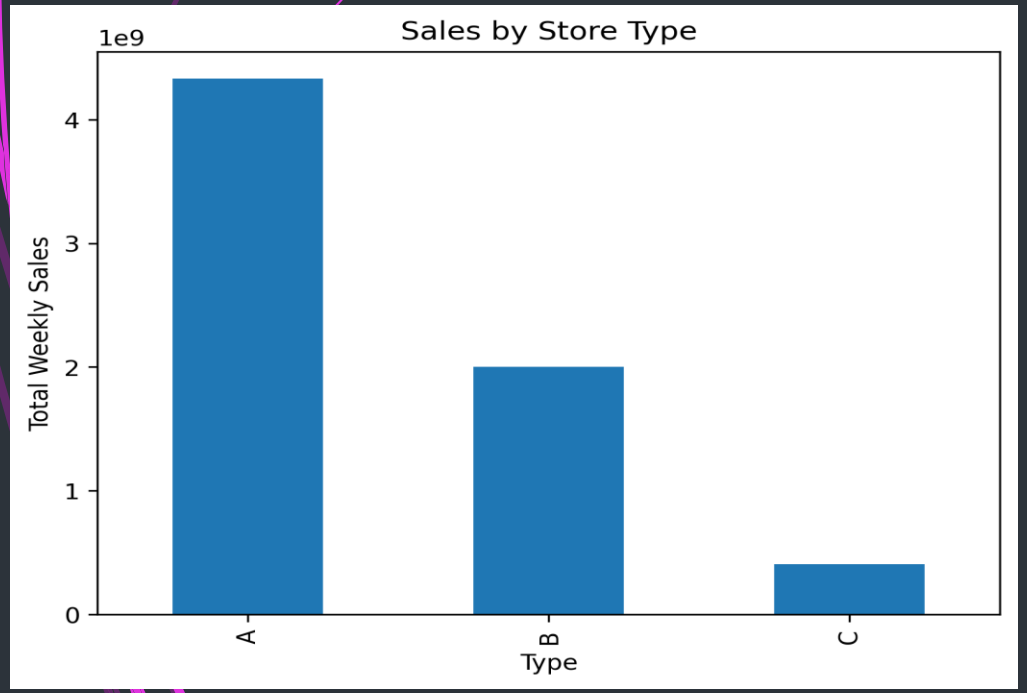
DATA CLEANING PROCESS

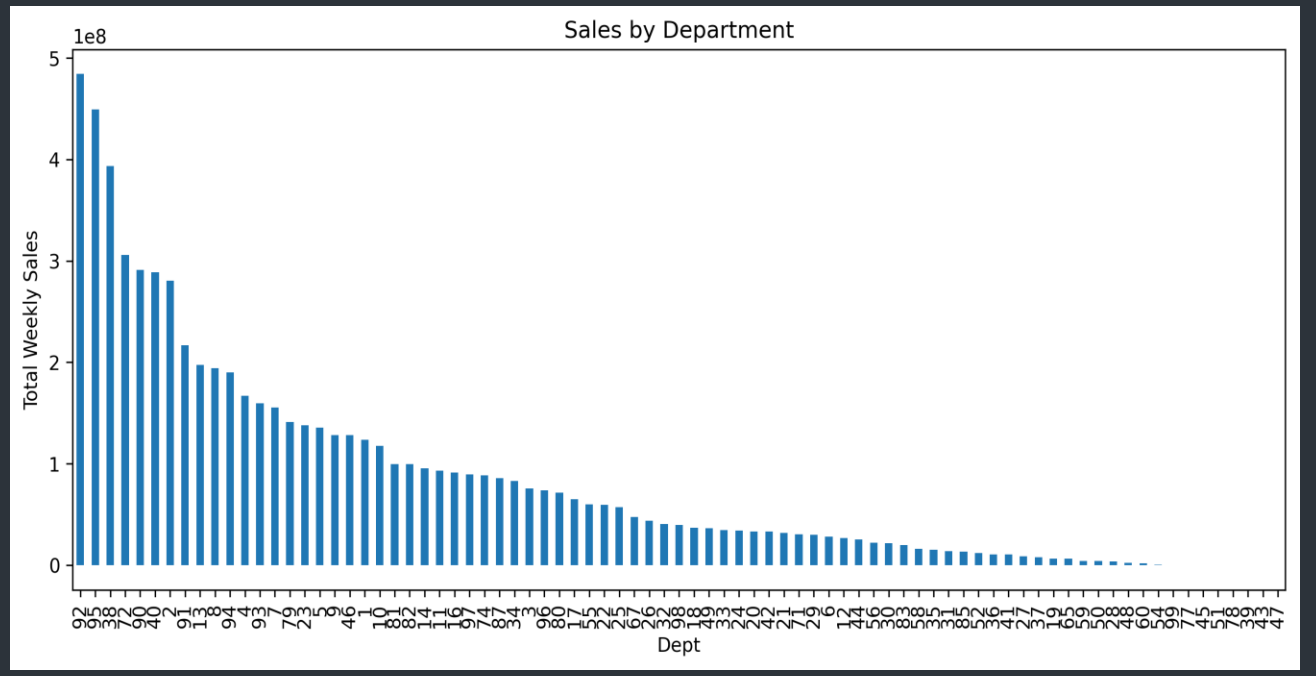
Steps Performed:

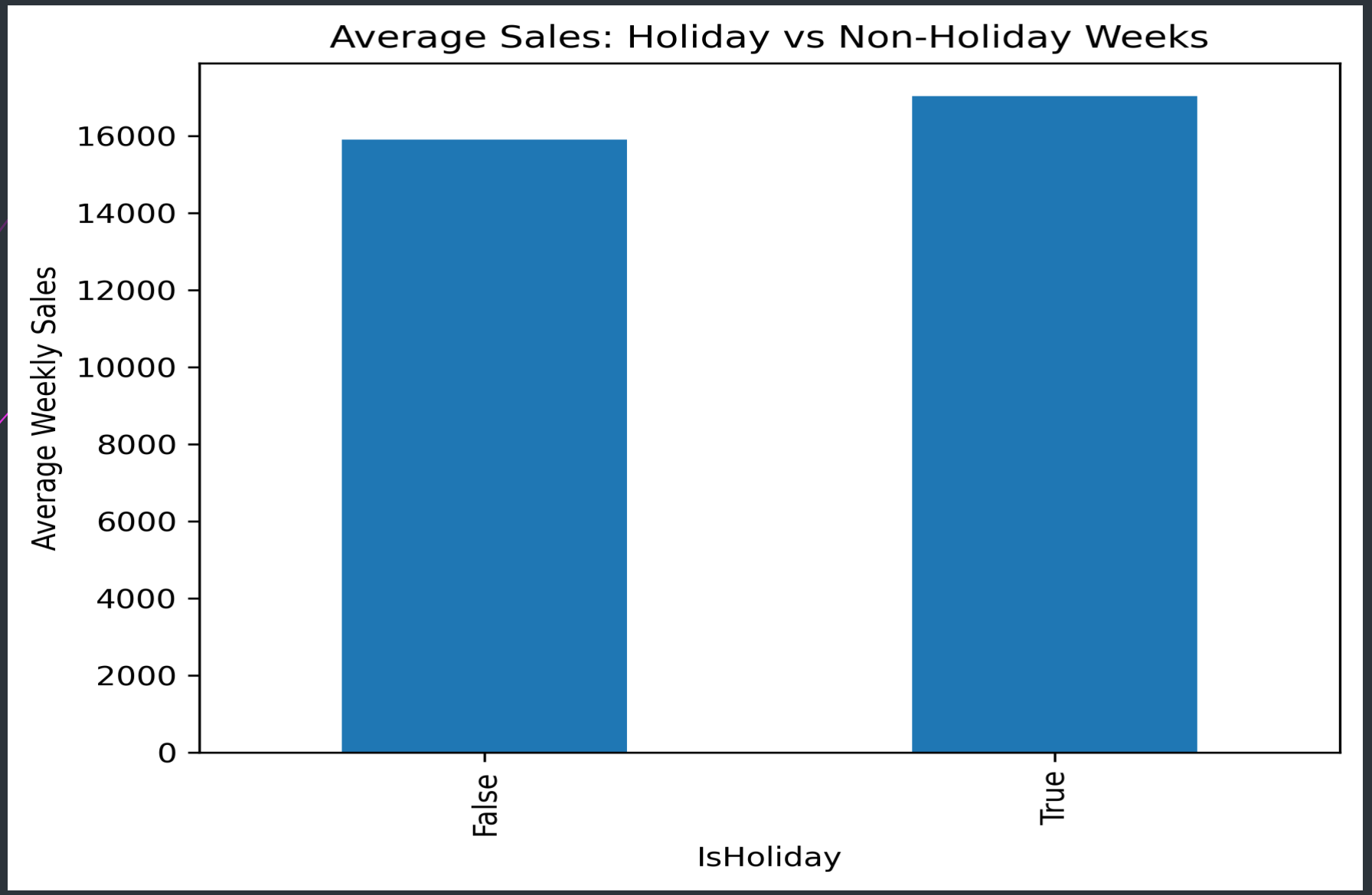
- Merged train, features, and stores datasets
- Converted Date → datetime format
- Created Year, Month, Week features
- Filled missing values using median
- Removed duplicates
- Encoded categorical features
- Final cleaned dataset ready for EDA & modeling

EXPLORATORY DATA ANALSIS

- **Key Insights:**
- Sales show seasonal patterns
- Store Type A has the highest sales
- Holidays increase average sales
- Department contributions vary widely
- Correlations with external factors are weak






Average Sales: Holiday vs Non-Holiday Weeks
Average Weekly Sales
16000
14000
12000
10000
8000
6000
4000
2000
0
False
True
IsHoliday

Statistical Tests (T-Test + ANOVA)

T-Test: Holiday vs Non-Holiday Weeks

- Purpose: Check if holidays impact weekly sales
- t-statistic: 7.00
- p-value: 2.59×10^{-12}
- Conclusion:
- Holidays significantly increase weekly sales
- Reject H_0 → There IS a difference

ANOVA: Sales by Store Type (A, B, C)

- Purpose: Compare sales performance across store types
- F-statistic: 7764.43
- p-value: 0.0
- Conclusion:
- Store Type C > Type A > Type B
- Store type has a strong impact on sales
- Reject H_0 → Store types differ significantly

Key Insight:
Holiday periods and store type heavily influence weekly sales — statistically proven.

MODEL SUMMARY

Model Used: **Linear Regression**
Target: **Weekly Sales**

Input Features: Store, Dept, Holiday, Year, Month, Week, Size, Temperature, Fuel Price, CPI, Unemployment, Store Type

Model Performance:

- **R^2 Score: 0.09**
- **MAE: 14,570**
- **RMSE: 21,767**

Interpretation:
Retail sales are highly variable due to promotions/events not in the dataset → so low R^2 is expected.
But the model is still useful for identifying key sales drivers.

KEY DRIVERS & FINAL INSIGHTS

Top Positive Influencers

- Type C Stores
- Holiday Weeks
- Certain Months
- Fuel Price
- Dept Number

Top Negative Influencers

- Type B Stores
- Unemployment
- CPI
- Year (downward trend)

CONCLUSION

- • Holidays significantly boost sales
- • Store Type C performs best
- • Economic factors affect demand
- • Sales show seasonal trends
- • Insights can guide inventory, marketing & pricing decisions