Weekly Summary - Group 5

Emily Ye zy367@drexel.edu
Junkai Ge jg3944@drexel.edu
Jerry Li jl4533@drexel.edu
Shengyang Dong sd3666@drexel.edu

Summary

This week, we will focus on continuing and completing the EDA analysis and report. Our goal is to understand the data better, uncover patterns, and identify key insights that will inform the modeling phase.

Team Contributions

Emily Ye

- Time Series Analysis: Looked at sales from January 2013 to May 2015 across different stores to spot patterns related to seasons or time.
- Closed Store Analysis: Made charts to see when and why stores closed, with a focus on holidays.
- Impact Analysis: Studied how sales changed on days stores were closed because of promotions

Junkai Ge

- Customer Spending Analysis: Looked at how much money customers spent on average and figured out which stores did best.
- Sales Distribution Visualization: Made charts to show

how sales numbers spread out, comparing days stores were open to days they were closed
- Weekday Sales Analysis: Investigated how sales vary by day of the week, especially on Sundays and holidays.

Jerry Li

- Correlation Study: Checked how different data points, like sales and promotions, relate to each other using a heat map.
- Skewness Check: Looked at how the sales numbers lean towards high or low values to help decide how to process the data for better predictions.
- Promotion and Holiday Sales Analysis: Explored how promotions and holidays affect sales, helping to plan better sales strategies.

Shengyang Dong

- Sales and Customers Relationship Analysis: Made plots to see how the number of customers relates to how much they buy.
- Binary Data Analysis: Examined how yes/no data, like if a store was open or had a promotion, impacts sales.

Future Work

In future work, we will continue refining our sales forecasting

models by experimenting with various time series techniques such as ARIMA, SARIMA, Prophet, and LSTM, and incorporating exogenous variables like holidays and promotions to improve accuracy. We will conduct thorough cross-validation and evaluate models using metrics like RMSE and MAE, ensuring their reliability across different scenarios.