Predicting Retail Sales

Presented by: Jie Liang

Problem Statement

Can we forecast sales based on **internal** and **external factors**?

Data from Major Retailer

Dataset

- 2010-2012
- 420k records
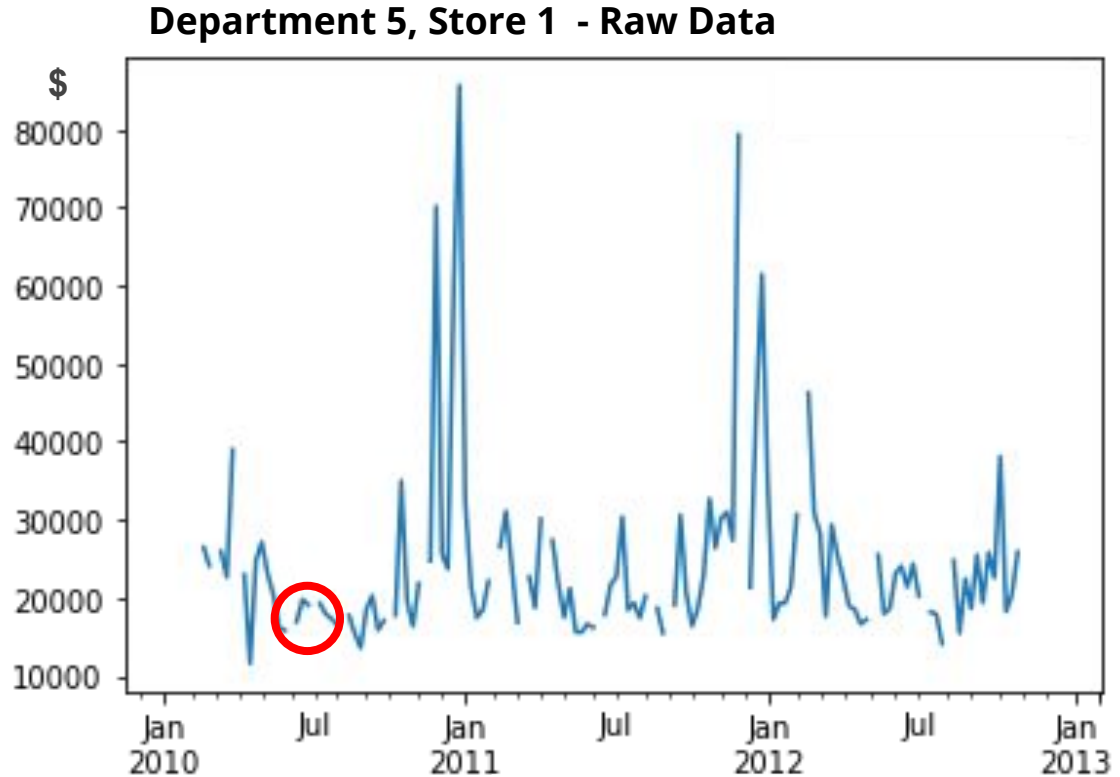- 3k departments
- 45 stores

Features

- Weekly Sales
- Markdowns
- Holidays
- Temperature
- Gas Price
- Unemployment Rate
- Consumer Price Index (CPI)

Data Cleaning

Issues identified in raw data:

- Missing data points
- Multiple entries per week

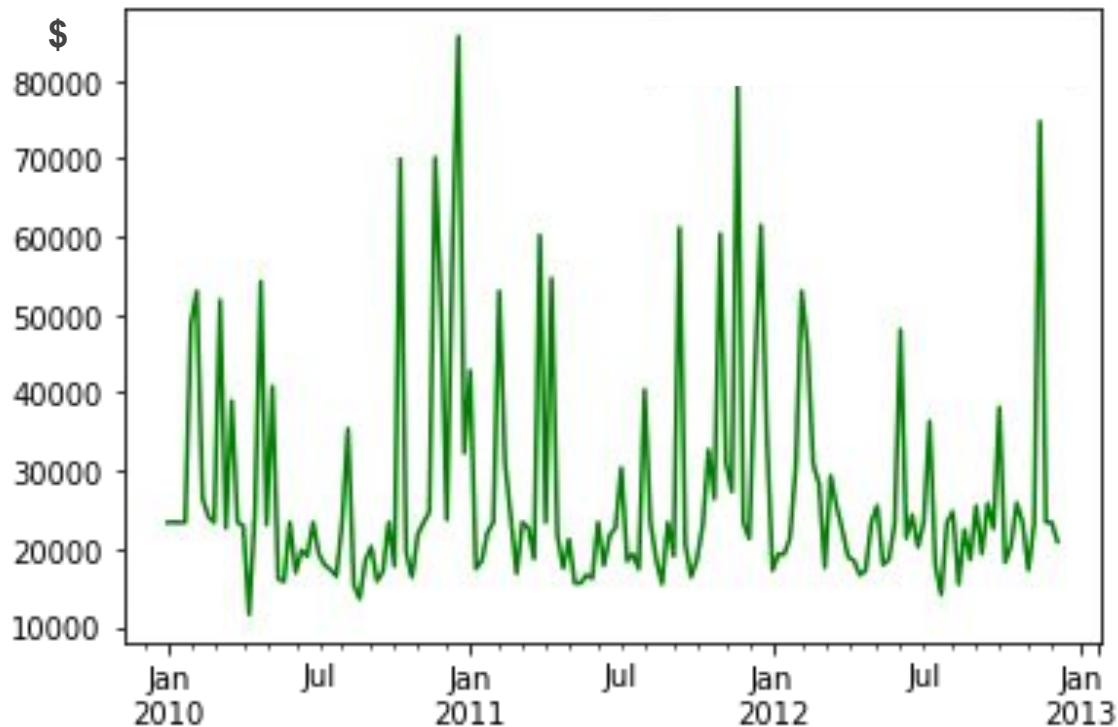**Department 5, Store 1 - Raw Data**
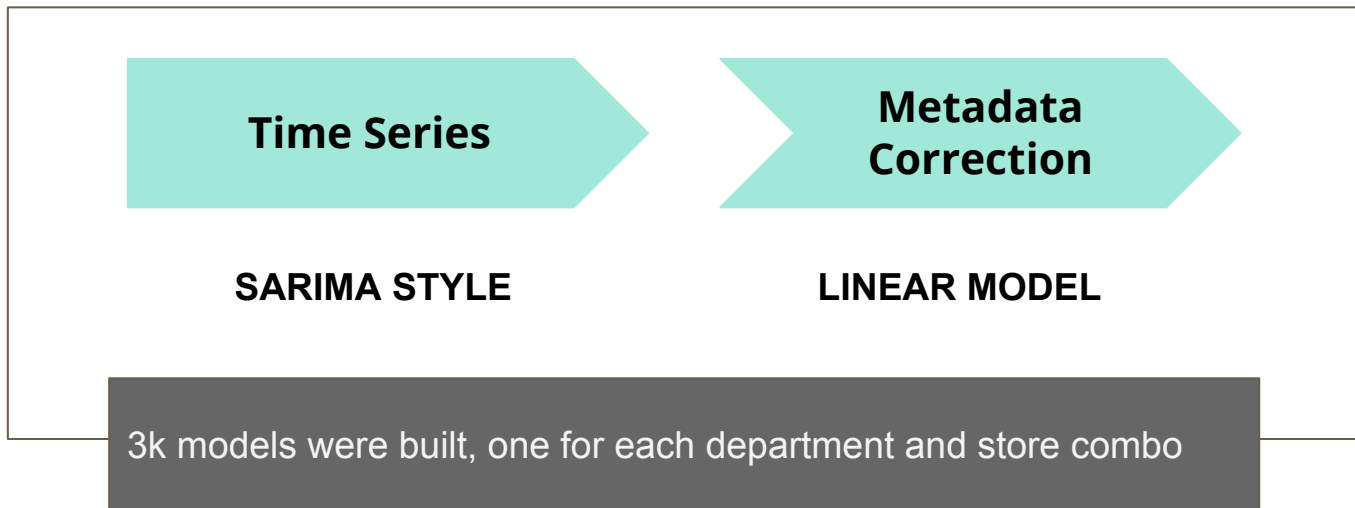
Modeling Time Series

Fixes made for final modeling series:

- Missing values imputed
- Multiple entries in a week aggregated

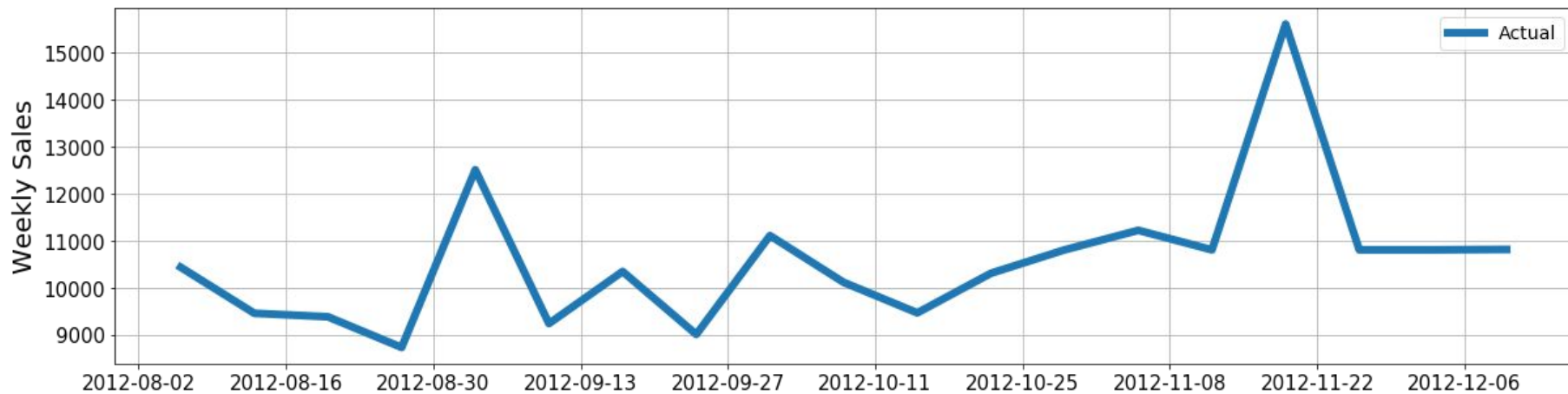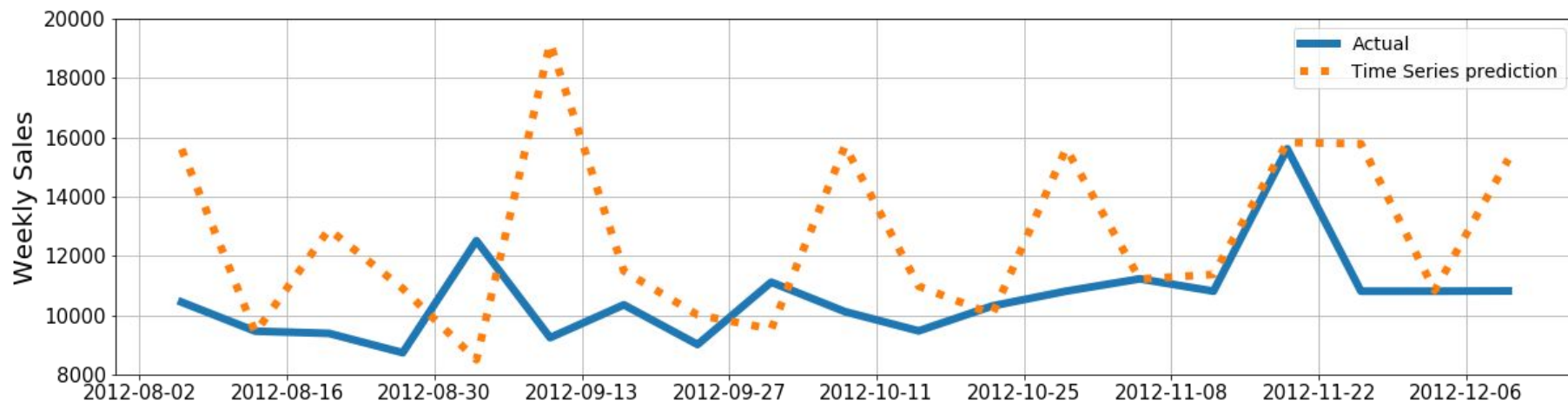**Department 5, Store 1 - Clean Data**

Two-Step Modeling

Time Series

Metadata Correction

SARIMA STYLE

LINEAR MODEL

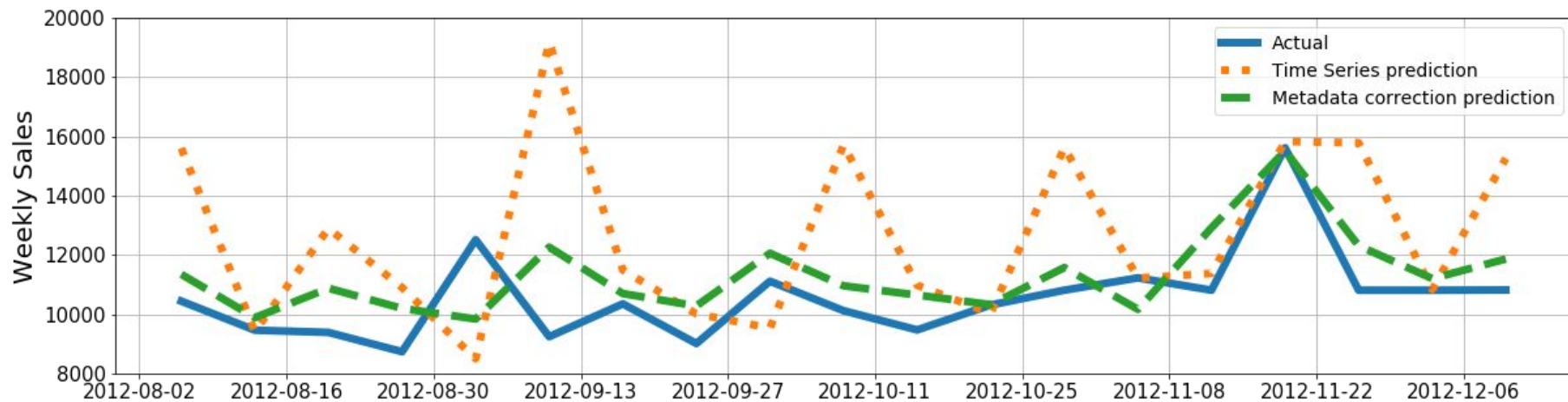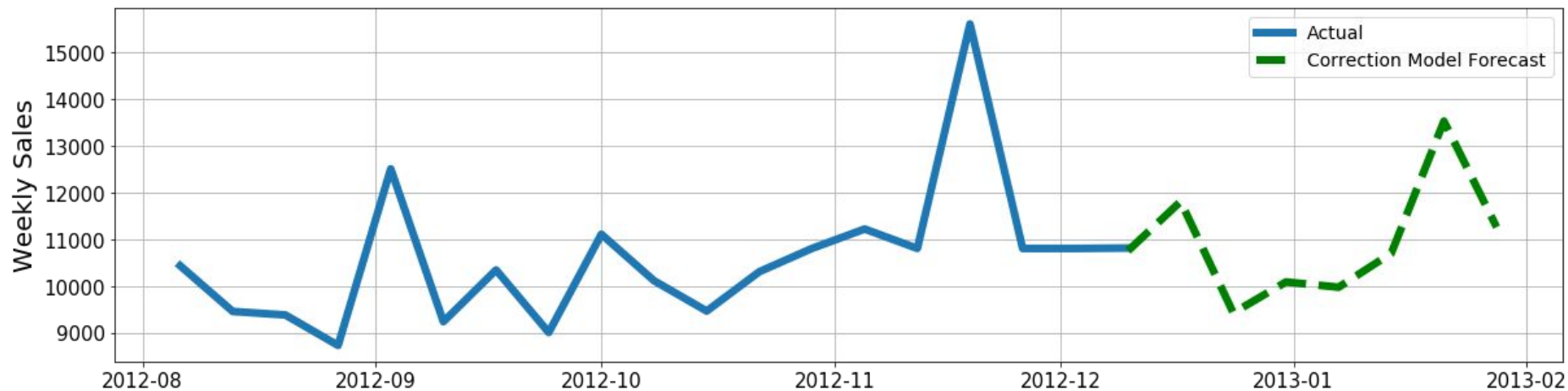3k models were built, one for each department and store combo
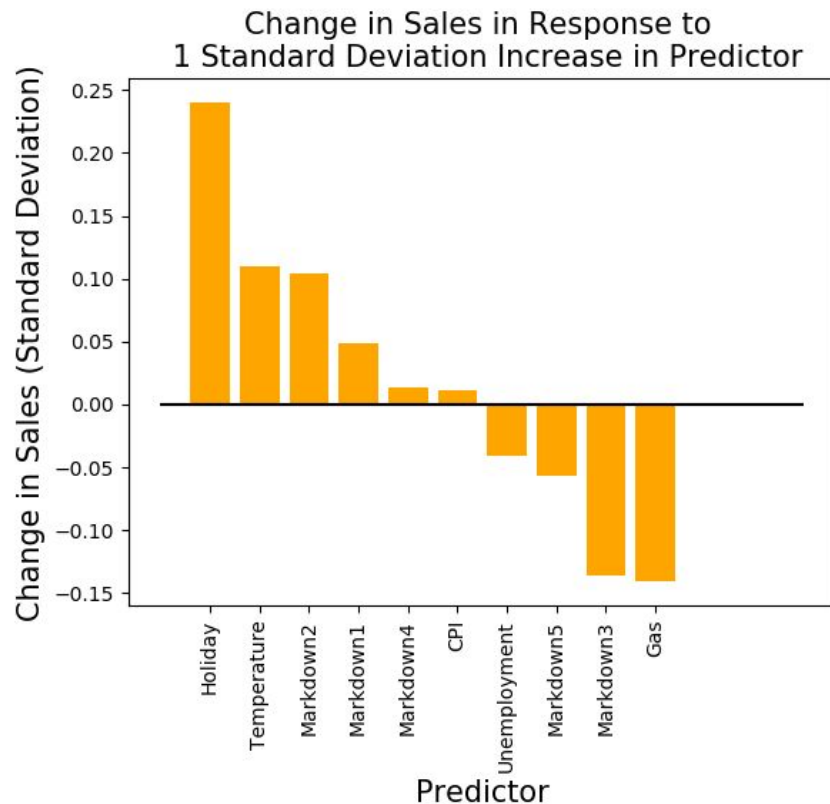
Modeling Time Series

In-Sample Prediction

Correction Model Improves In-Sample Accuracy

Out-of-Sample Prediction

Effects of Internal and External Factors

Change in Sales in Response to
1 Standard Deviation Increase in Predictor

Business Takeaways

	Recommendations
Internal factors	• Promote markdowns positively correlated to sales • Prioritize departments expected to see biggest jumps during holidays
External factors	• Sales **negatively correlated with gas price**: promote discount on **shipping/delivery** • Sales **negatively correlated with unemployment**: reduce high end products and **increase low end products**

Applications of Forecasting

INVENTORY

STAFFING

Thank You

Jie Liang

https://www.linkedin.com/in/jieliangusa/
https://github.com/jieliang
jieliang_usa@yahoo.com