ANALYSIS FINDINGS

HOW THE DATA ANSWERS CORPORATE QUESTIONS

OVERVIEW

Reasons for Analysis

Answer business questions – both intra and inter department assistance

Discover potential cause & effect relationships that drive our metrics

Diagnose potential areas of inefficiency

General Topics

The dataset

Unemployment

Effect of CPI on weekly sales

Holidays

The Dataset

Walmart Sales is a dataset which records data about 45 different Walmart locations.
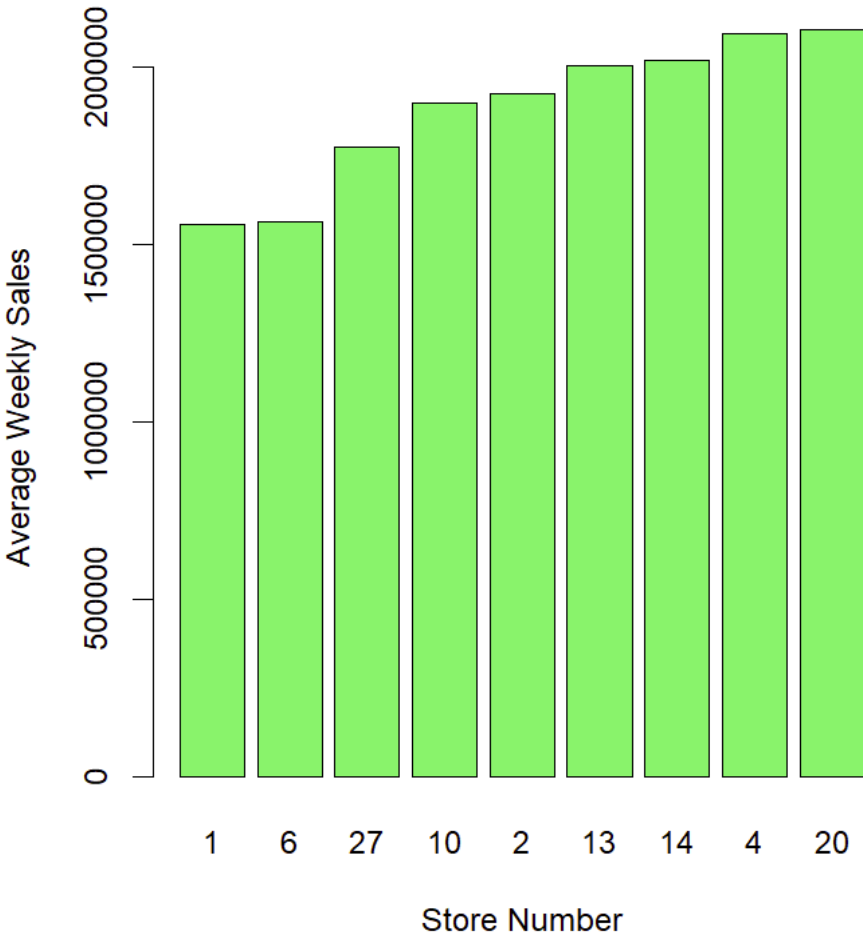
Data includes both retailer data & economic data

KEY VARIABLES

- **Store (location number, identifier)**
- **Unemployment**
- **Weekly Sales**
- **CPI**
- **Holiday Flag**

SOLUTIONS VIA ANALYSIS

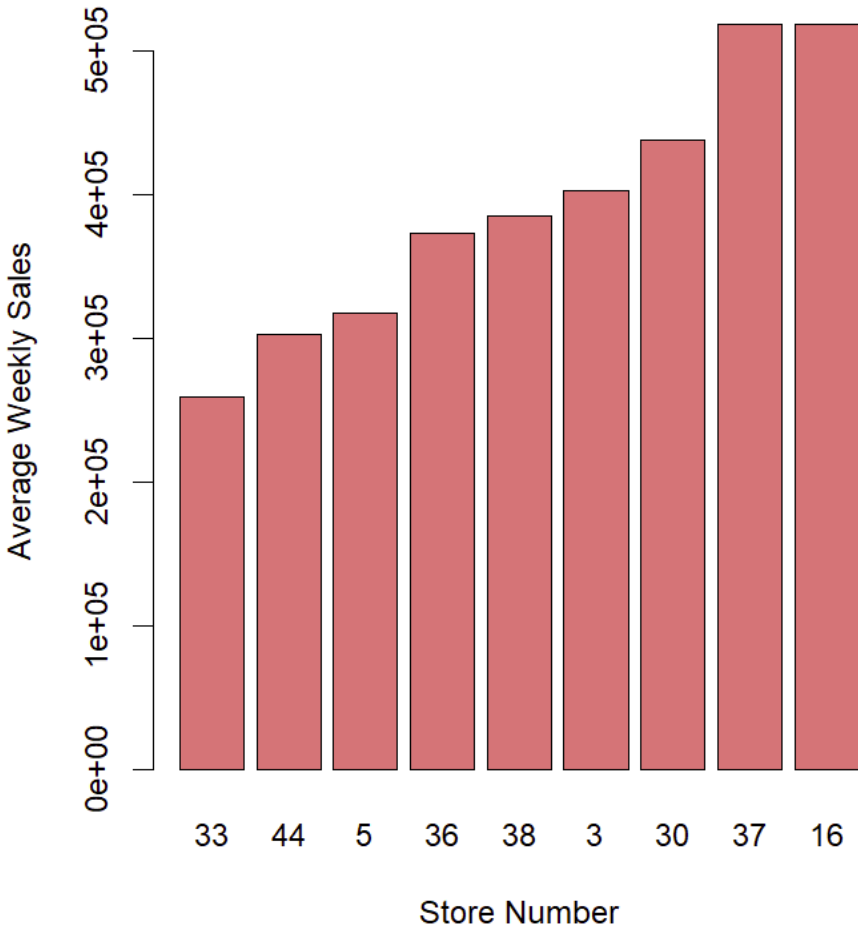Answering business questions with statistical methods

Top 20% of Stores

Which stores in the district have the lowest and highest average weekly sales?

Locations with the highest sales include 13, 14, 4, & 20, all having averages of more than $2,000,000/week

Bottom 20% of Stores

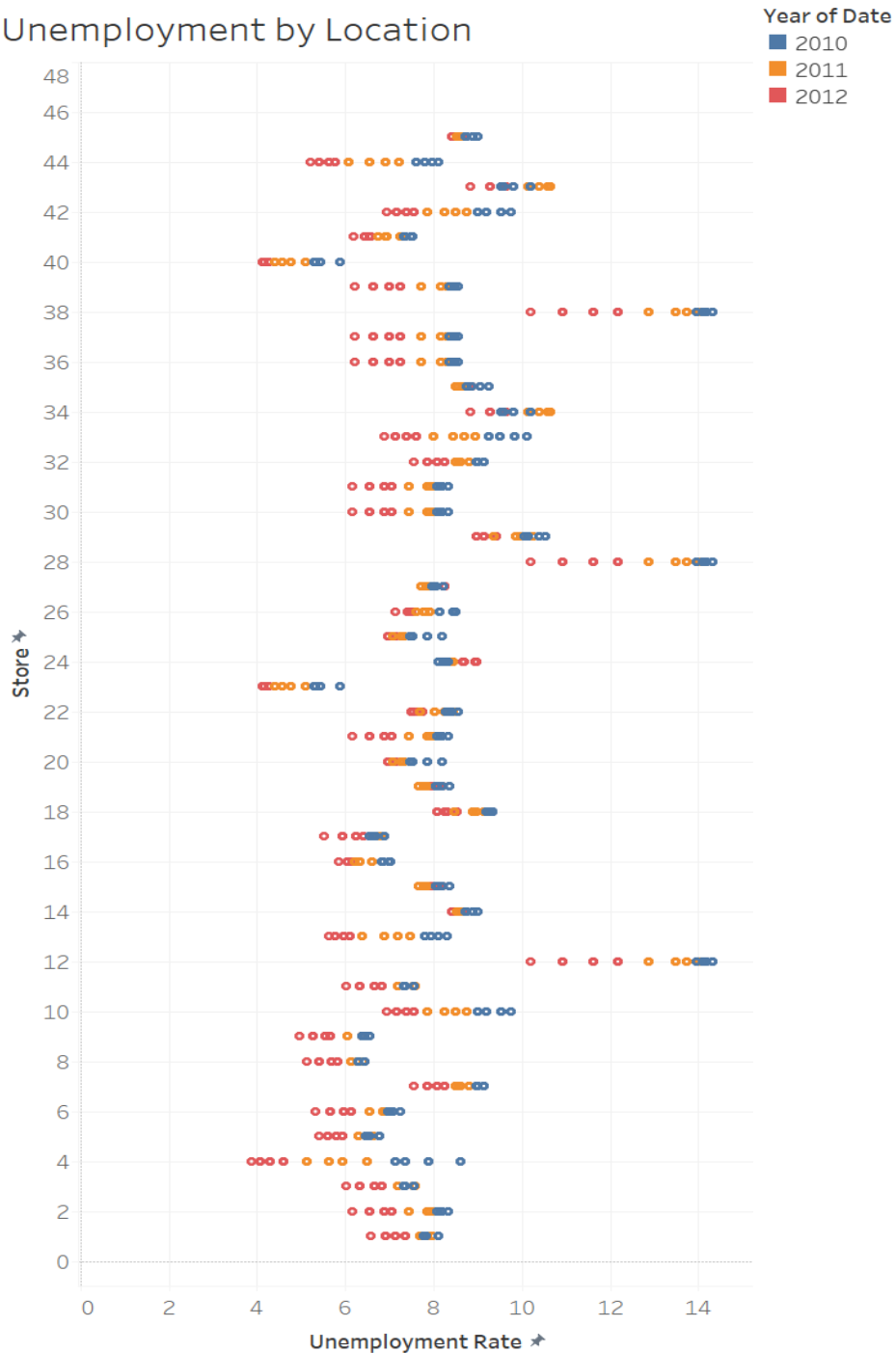**Which stores in the district have the lowest and highest average weekly sales?**

Locations with the lowest sales include 33, 44, 5, & 36, all having averages of less than $400,000/week

The Model

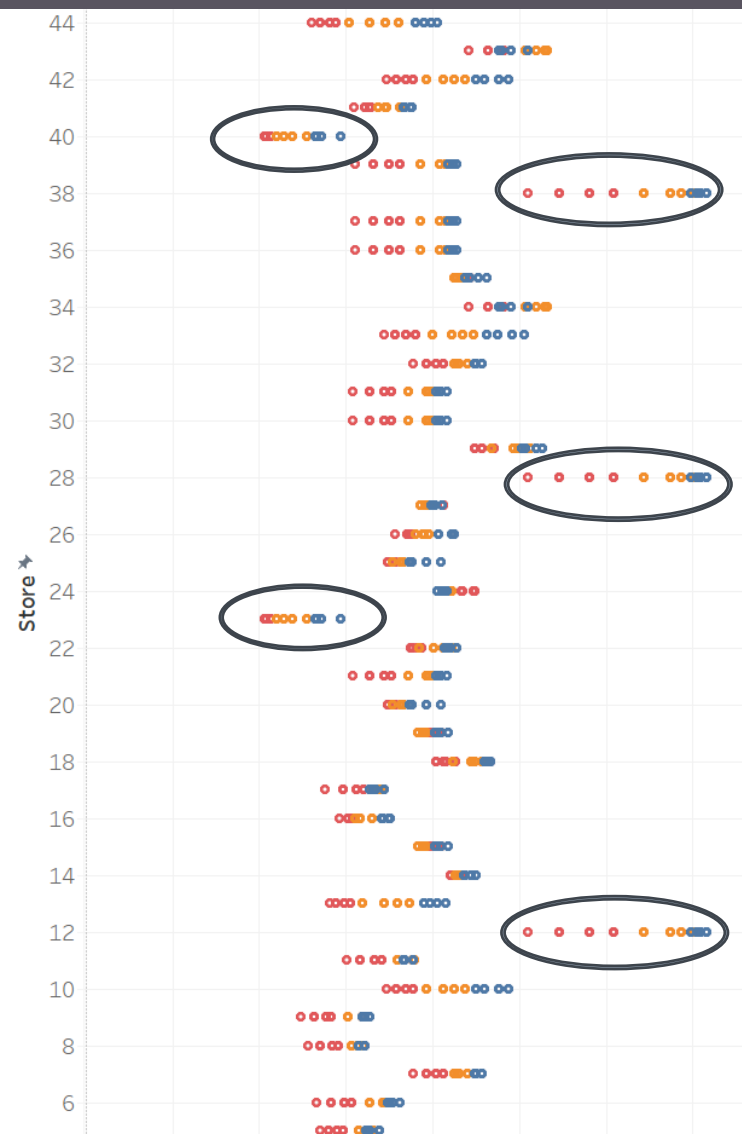We can predict weekly sales with the following model from our data.

Weekly Sales($) = 1664939.1 + 84509.6(If it's a holiday week) - 1652.8(CPI) - 42542.2(Unemployment Rate (%) * 100)

Which stores in the district have the lowest and highest unemployment rates?

Locations with the highest unemployment include 12, 28, & 38

Locations with the lowest unemployment include 4, 23, & 40

Big Red Flag –
repeated observations

It appears that there are groups of stores reporting the same values for unemployment. This is likely due to these stores being in the same COUNTY, and the unemployment statistic collected is county unemployment. HOWEVER, as this is a public dataset that I did not have a hand in creating, I am unable to confirm this theory with the publisher, and this is simply an educated guess at the reasons behind the repetitions.

# What factors do you think are impacting the unemployment rate?

- Location

  - Unemployment varies drastically by store

  - We lack information about whether these stores are urban or not – lurking variable

- CPI

  - Our data supports the generally accepted inverse relationship between inflation (inferred via CPI) and Unemployment

  - In our model, when CPI increases by 1 point, we predict a fall in Unemployment by 0.0144 points

Correlation matrix
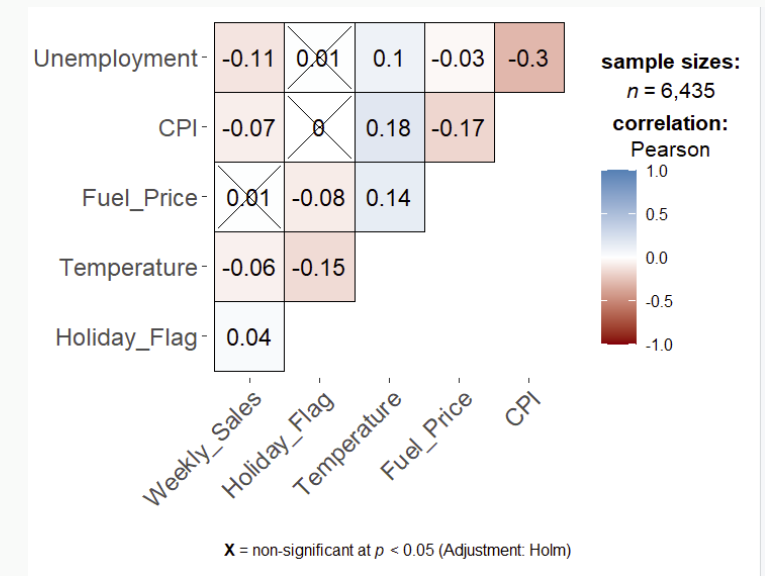
Is there any correlation between CPI and Weekly Sales?

- Negative

- Very weak correlation

- The intuition holds that a rise in prices would correlate with a fall in consumption
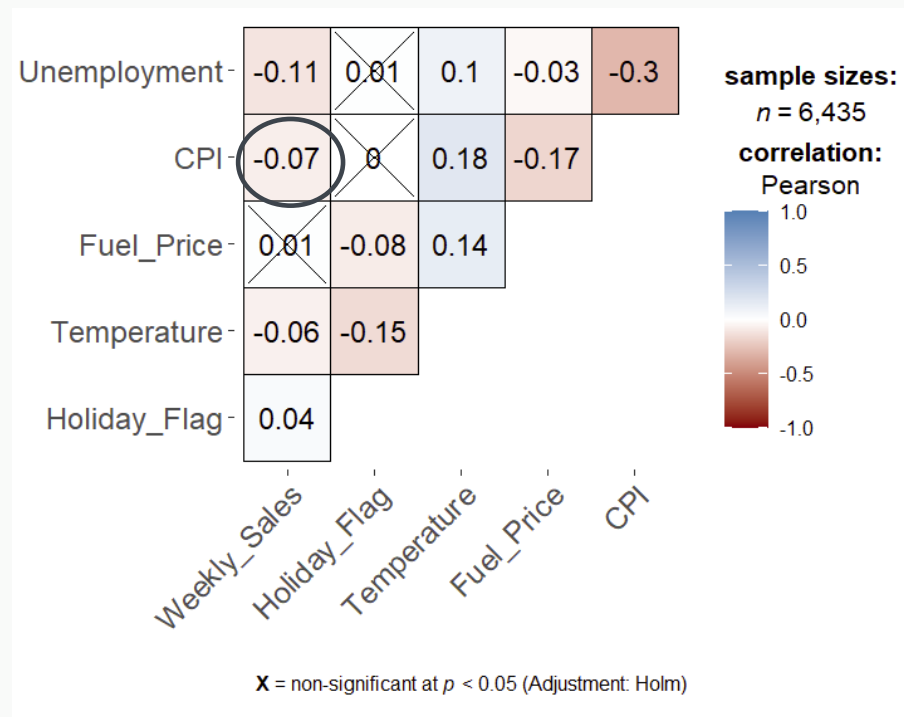
Do holidays impact the correlation between CPI and Weekly Sales?

- Correlation is slightly weaker on holidays

  - It stands to reason that consumption around holidays may be slightly more inelastic

- Potentially concerning small number of observations

  - Our sample of holiday observations is a bit small for inference to be conclusive

Holiday	Regular Day
-0.29962	-0.3021949

Why is Fuel Price is included in this dataset?

- Fuel price, while volatile in price, has relatively inelastic demand in the short term

- The creator of the dataset likely believed it to be a useful *loose* benchmark of short term, localized CPI

- The creator may also have used this dataset to examine relationships between localized fuel price and other variables in the dataset.

THANK YOU

Will Au

willau000@gmail.com

https://github.com/willau00

SPECIAL THANKS TO

Mikhail1681 on Kaggle for the public dataset

Mary the Analyst on TikTok who inspired the project and suggested some interesting questions to include