



Artificial Intelligence Bootcamp: Project 2 - Group 3

WALMART ANALYSIS

William Aromando, Dan Becker, Joanne Donohue, Vivin Rajagopalan, and Aaron Swan

UNDERSTANDING OUR BUSINESS VIA STORE SALES DATA & LOCAL ECONOMIC INDICATORS



- 45 Stores, weekly sales
- Feb 5, 2010 to Oct 25, 2012
- Holiday Flag (Super Bowl, Labor Day, Thx Giving and Xmas only)
- Local CPI, Fuel Price, Temp, Unemploy %



Explore and Analyze
data set w/ Pandas



Estimate business impact of
features and predict future
outcomes using ML models



Source:

<https://www.kaggle.com/datasets/yasserh/walmart-dataset/data>

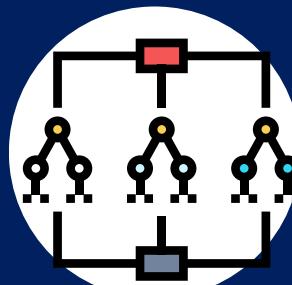
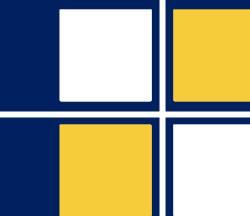
Four diverse ways to analyze and action this data set, varying accuracy rates

Through various machine learning techniques and visualization platforms, we were able to action the Walmart store level sales data to understand our business and make predictions to drive sales



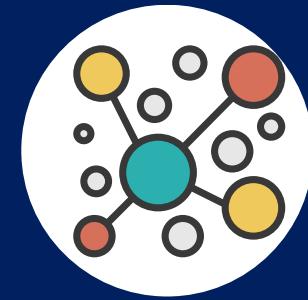
PANDAS,SKLEARN

Analysis



RANDOM FOREST

Supervised



KMEANS CLUSTERING

Unsupervised



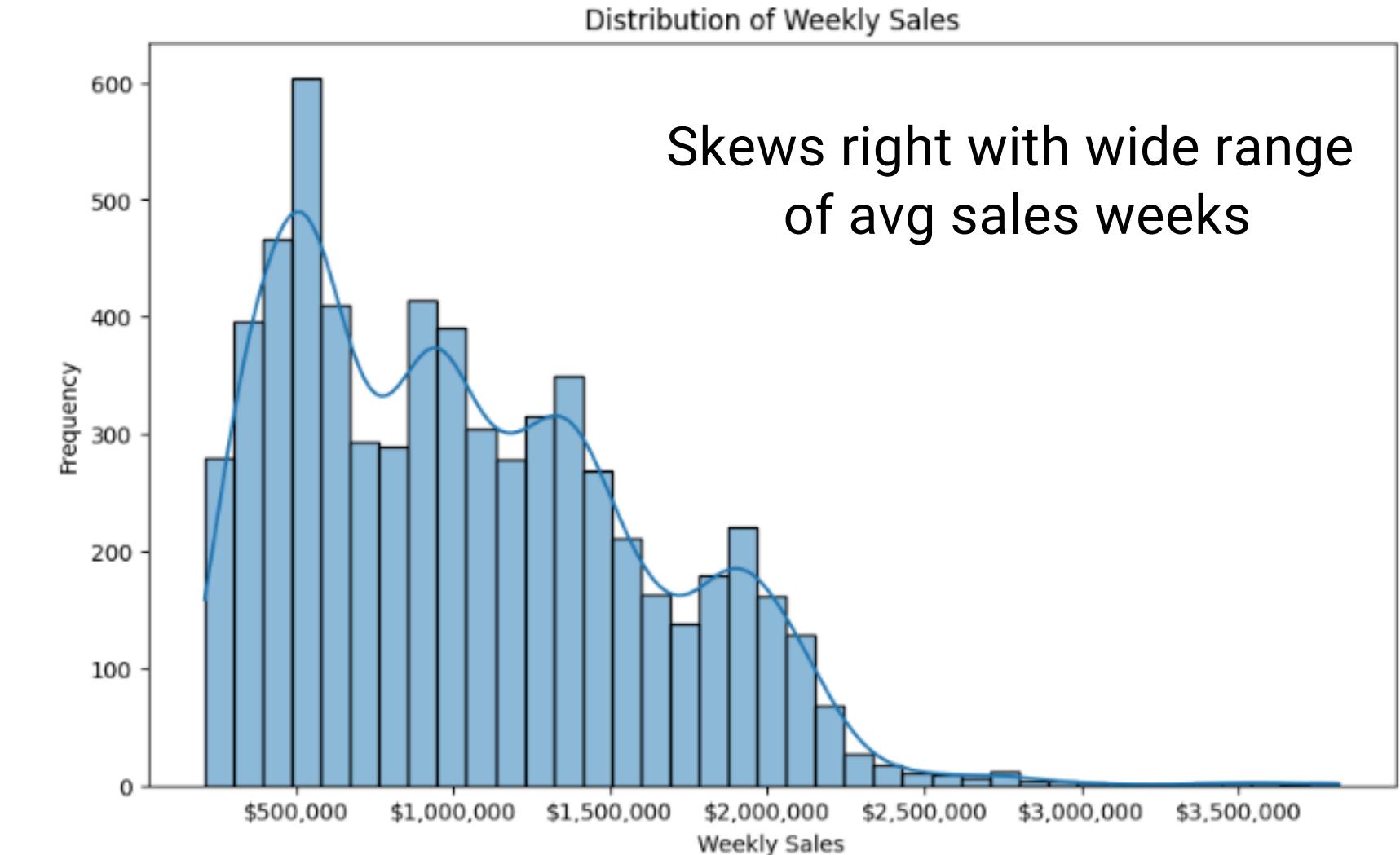
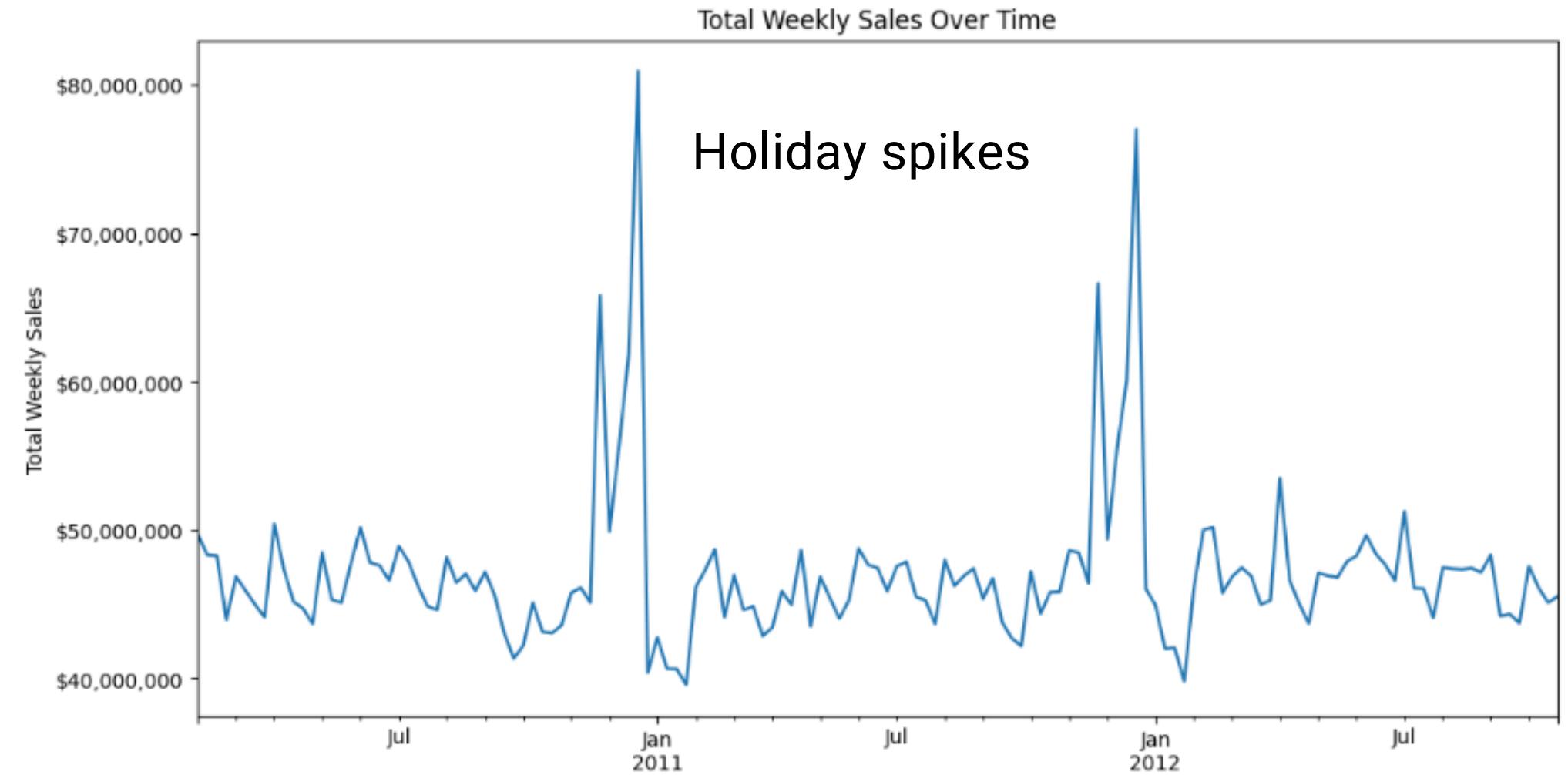
STREAMLIT APP

Application



Sales vary by location

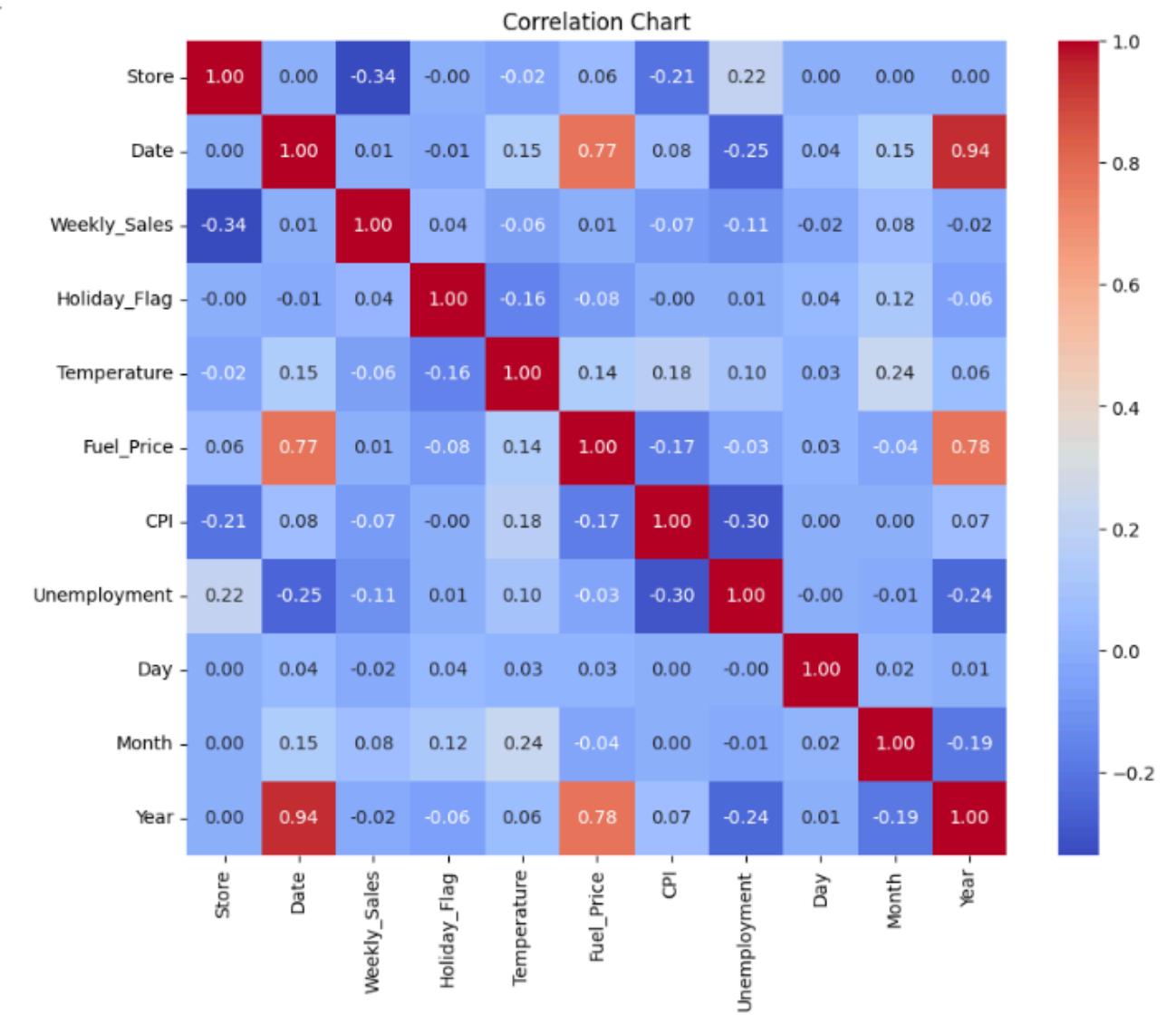
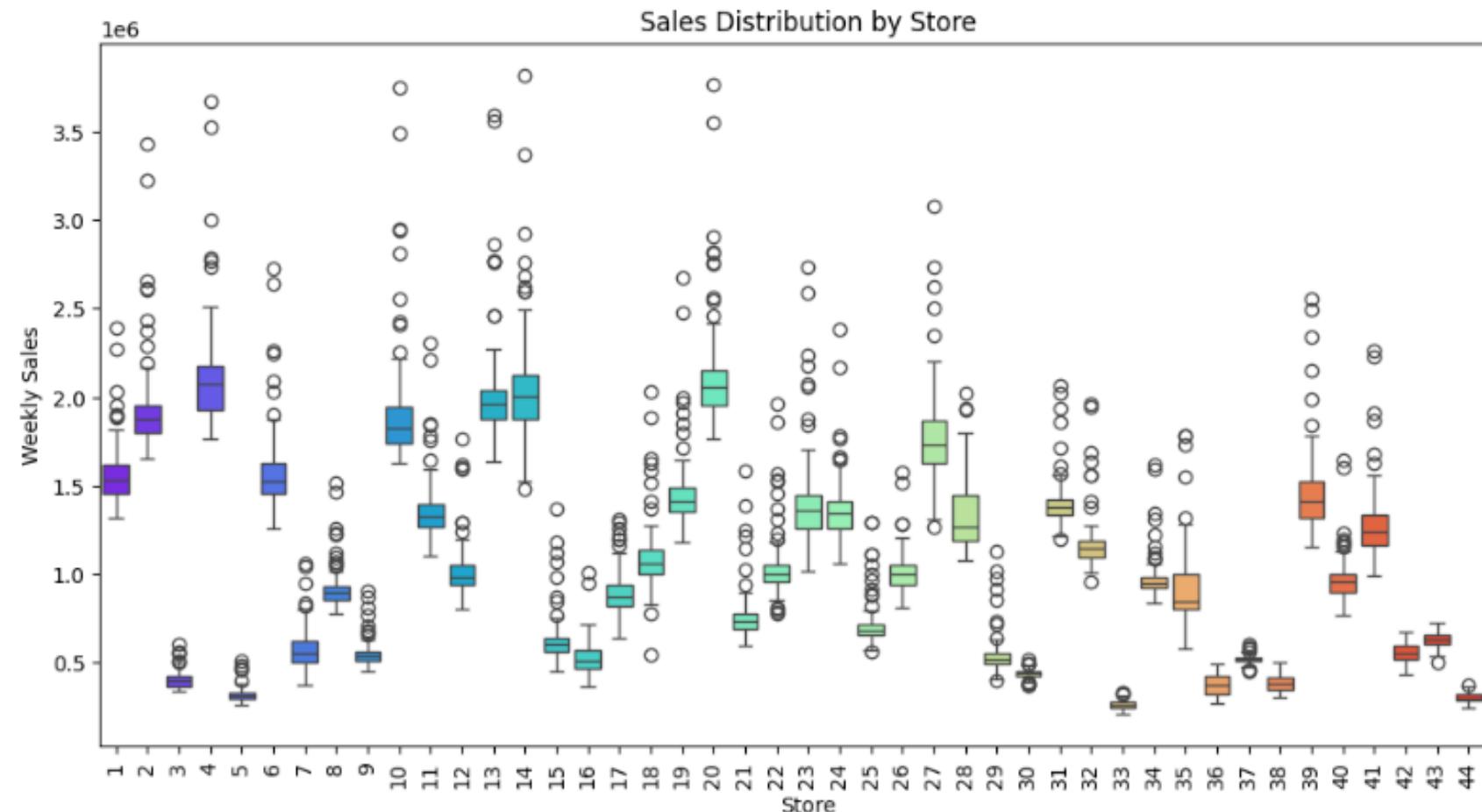
Our 45 stores range in performance, we can leverage economic and weather indicators to better understand sales patterns and manage our business better.





PANDAS,SKLEARN
Analysis

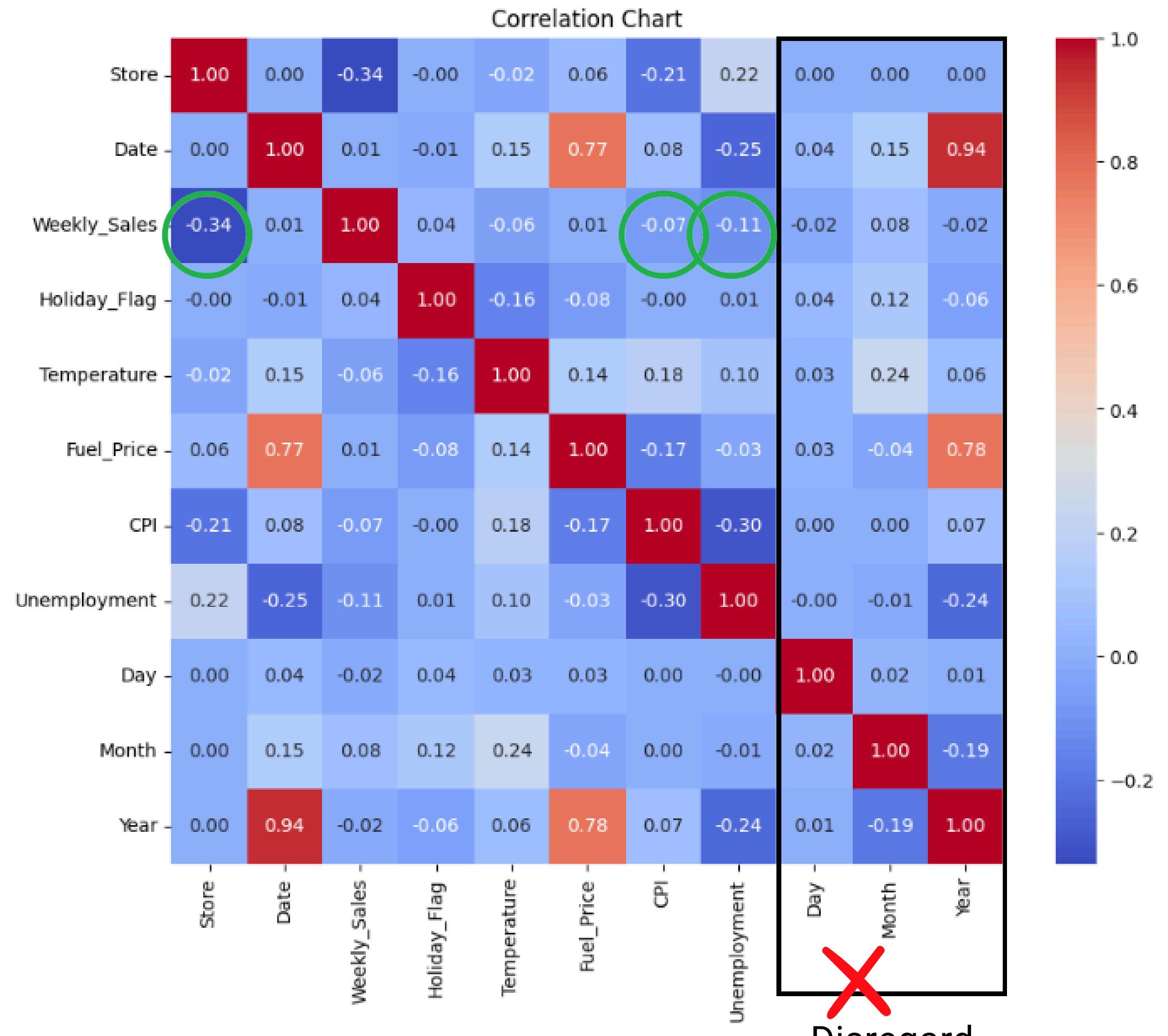
STORE LEVEL STATS DON'T TELL US MUCH, FEATURES TELL A STORY



Local CPI and Unemployment Rate are indicative of weak negative relationship with weekly sales

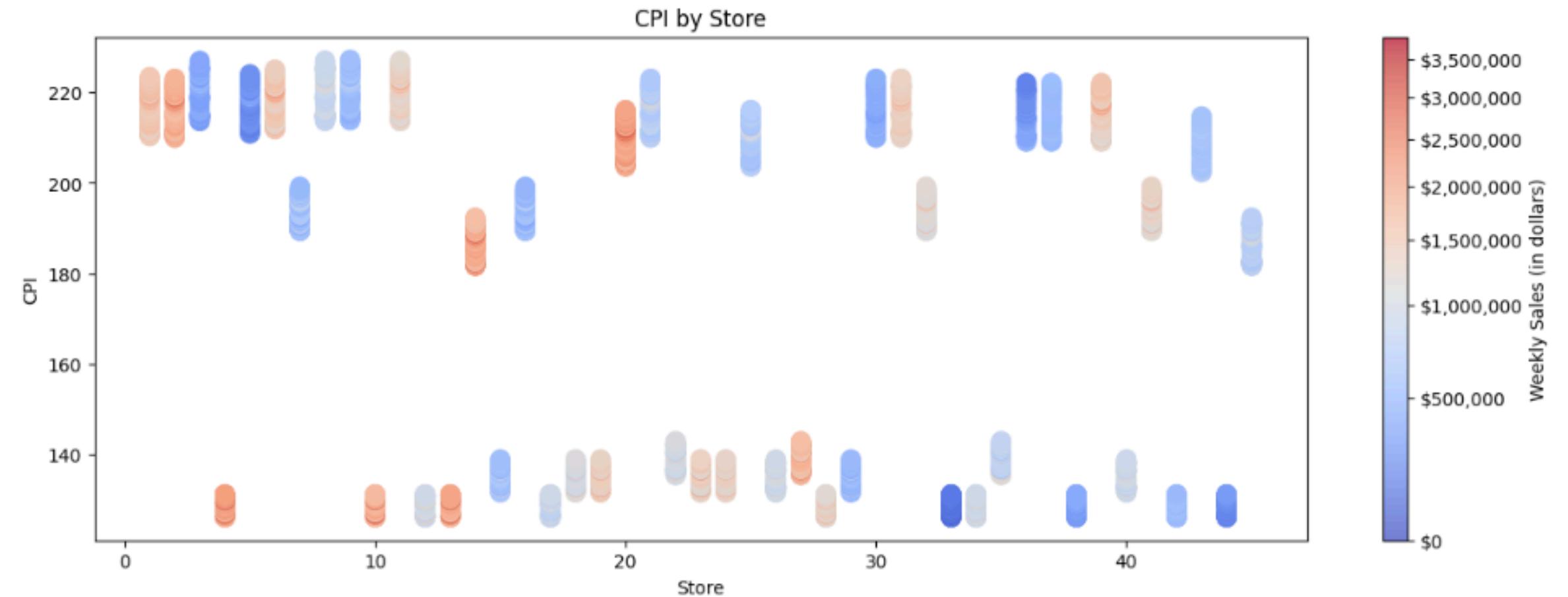
Observations:

- Strongest useable correlation is between **Store # and Weekly Sales at 34%**, which indicates that stores are consistent in their sales performance
- Next best correlation is between **Sales and Unemployment at -11%**, followed by the **CPI at -7%**
- Somewhat surprisingly, **fuel price and holiday had a neutral correlation** with weekly sales.
 - This could be a function of few holiday weeks in the data and store proximity to home, work or public transportation.



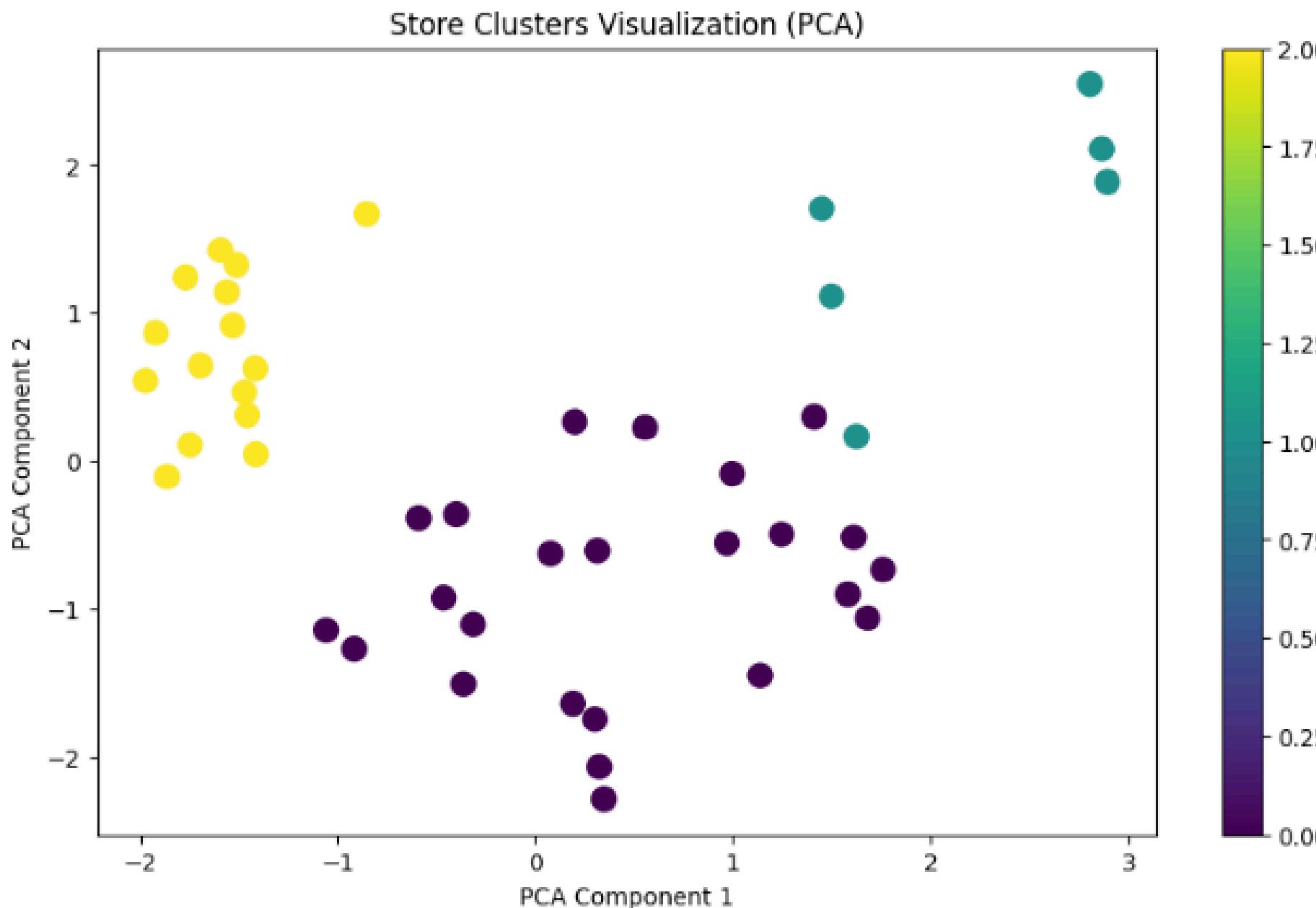
Disregard

Deep dive into important features w/ color emphasis confirms correlations



K MEANS CLUSTERING TO STRATEGICALLY DRIVE SALES INCREASES WHERE MARKET ALLOWS

Store	Weekly_Sales	Cluster
1	\$1,555,264.40	2
2	\$1,925,751.34	2
3	\$402,704.44	2
4	\$2,094,712.96	0
5	\$318,011.81	2

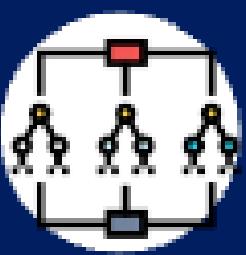


KMEANS CLUSTERING
Unsupervised

Column, bar, and pie charts compare values in a single category, such as the number of products sold by each salesperson. Pie charts show each category's value as a percentage of the whole.

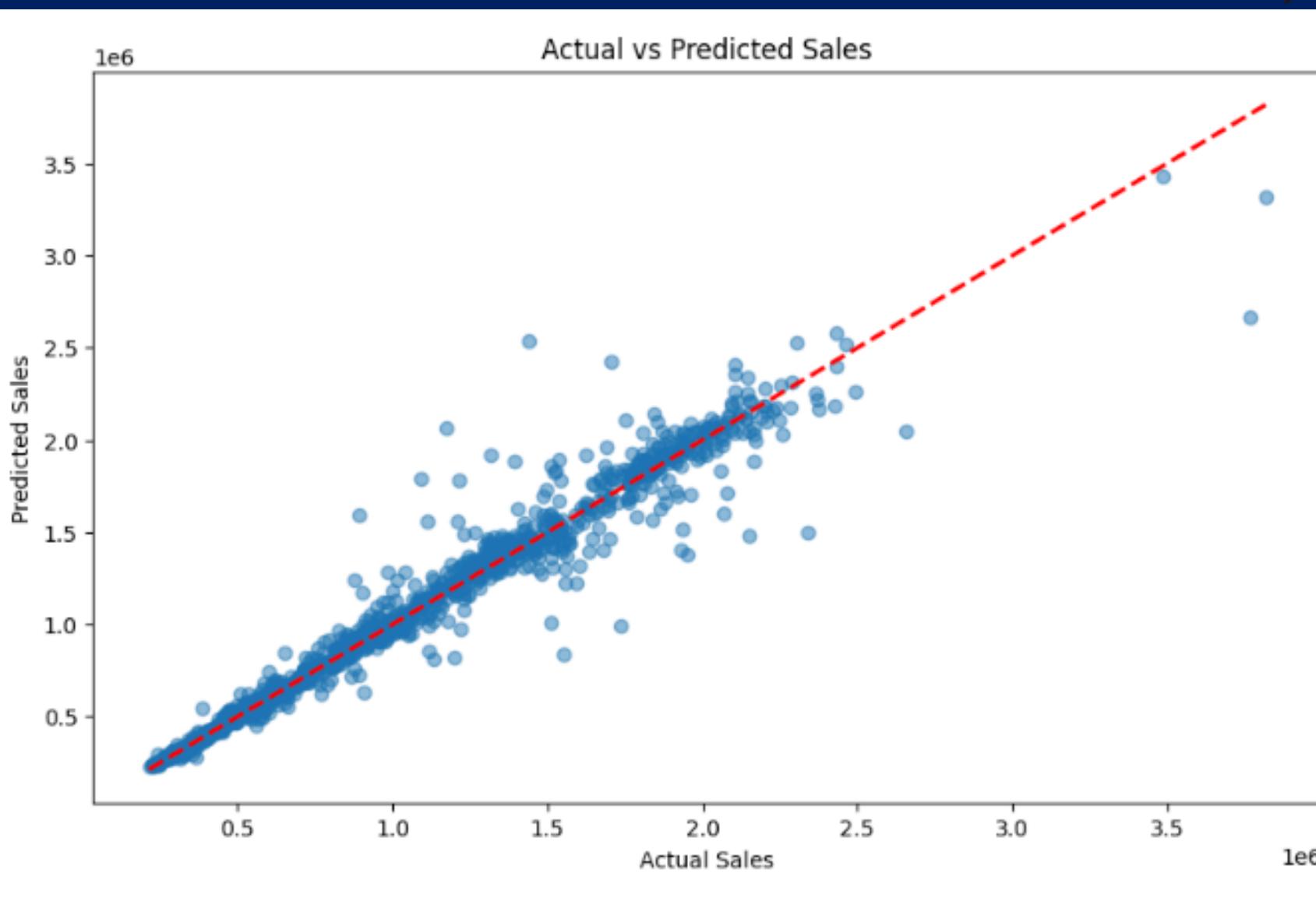
Implement Tiered Pricing Strategy

Use localized economic features to cluster stores into pricing tiers to increase price on items with low price elasticity - example, Every Day Low price on bottom tier, EDLP + 5% on next tier and so on to drive more sales

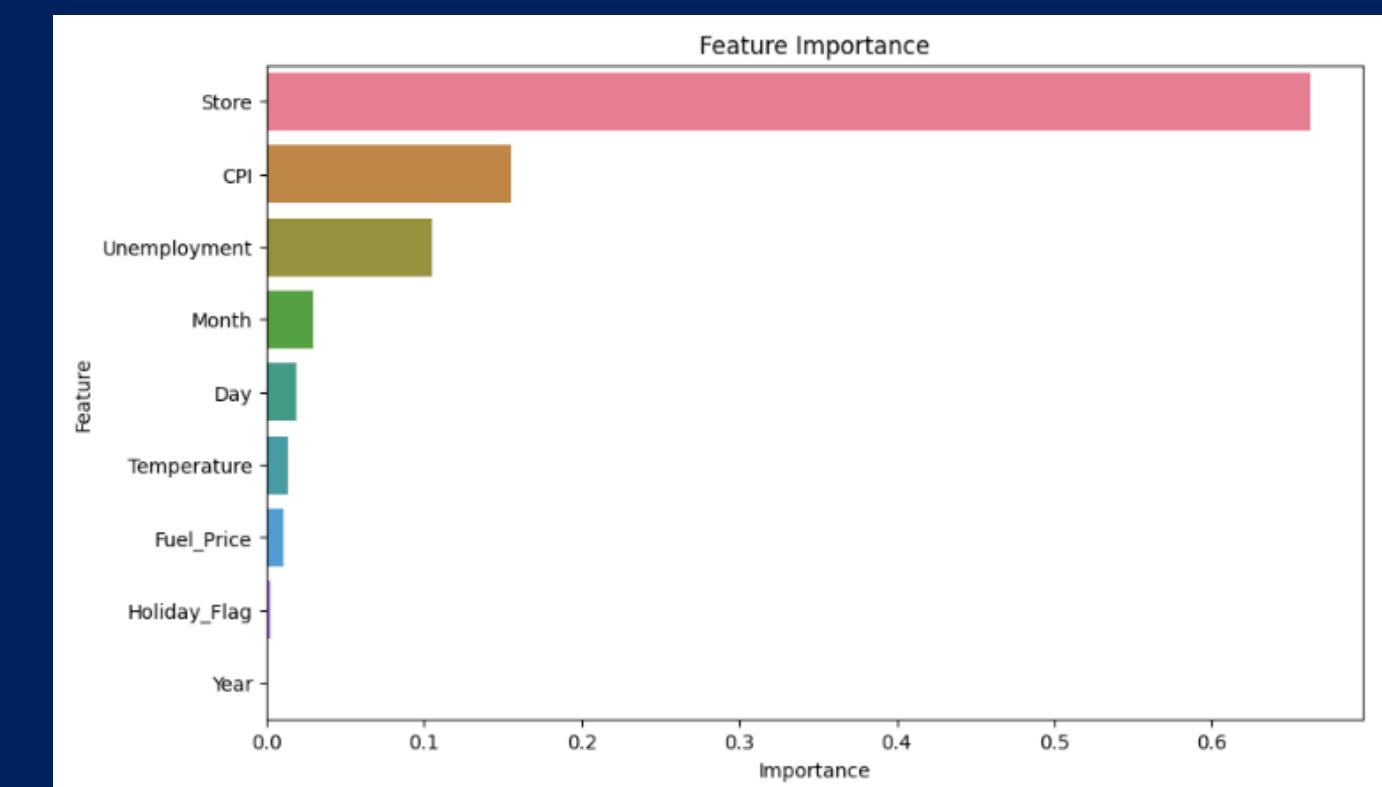


RANDOM FOREST
Supervised

Mean Squared Error: 13504777637.6563
R-squared Score: 0.958079819067563



HIGH ABILITY TO PREDICT SALES USING RANDOM FOREST MODEL





STREAMLIT APP

Application

walmart_app · Streamlit

localhost:8501

Choose a Classification Model

Select a Model

Decision Tree

Train Model

Holiday Week Prediction Model Comparison

st.cache is deprecated. Please use one of Streamlit's new caching commands, st.cache_data or st.cache_resource.

More information [in our docs](#).

Cleaned Data Loaded Successfully!

Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemp
0	2010-02-05 00:00:00	1,643,690.9	0	42.31	2.572	211.0964	
1	2010-02-12 00:00:00	1,641,957.44	1	38.51	2.548	211.2422	
2	2010-02-19 00:00:00	1,611,888.17	0	39.93	2.514	211.2891	
3	2010-02-26 00:00:00	1,469,727.59	0	46.63	2.561	211.3196	
4	2010-03-05 00:00:00	1,554,806.48	0	46.5	2.625	211.3501	

Model Performance: Decision Tree

Accuracy: 0.95

Classification Report:

	precision	recall	f1-score	support
0	0.97	0.97	0.97	1183
1	0.71	0.76	0.71	104
accuracy	0.95	0.95	0.95	1287
macro avg	0.84	0.84	0.84	1287
weighted avg	0.95	0.95	0.95	1287

Screen Recorder is sharing your screen. Stop sharing Hide

1. Predicting given all the data if its weekend
2. Predicting weekly_sales given all the data

FUTURE WORK



Prophet Model

Implement time series analysis techniques for more accurate forecasting



More granular data

Source department-level or item-level sales information by store for indepth analysis



Promotional Modeling

Explore the impact of promotional events on sales and price elasticity modeling



External data sources

Consider incorporating external data sources (e.g., proximity to competitors) to improve predictions



THANK YOU