



CONSUMER SENTIMENT

ISA 514 Project Group 3

Our TEAM



LOGAN LOFTUS



AARON BLACK



MAGUIRE SHEPPARD



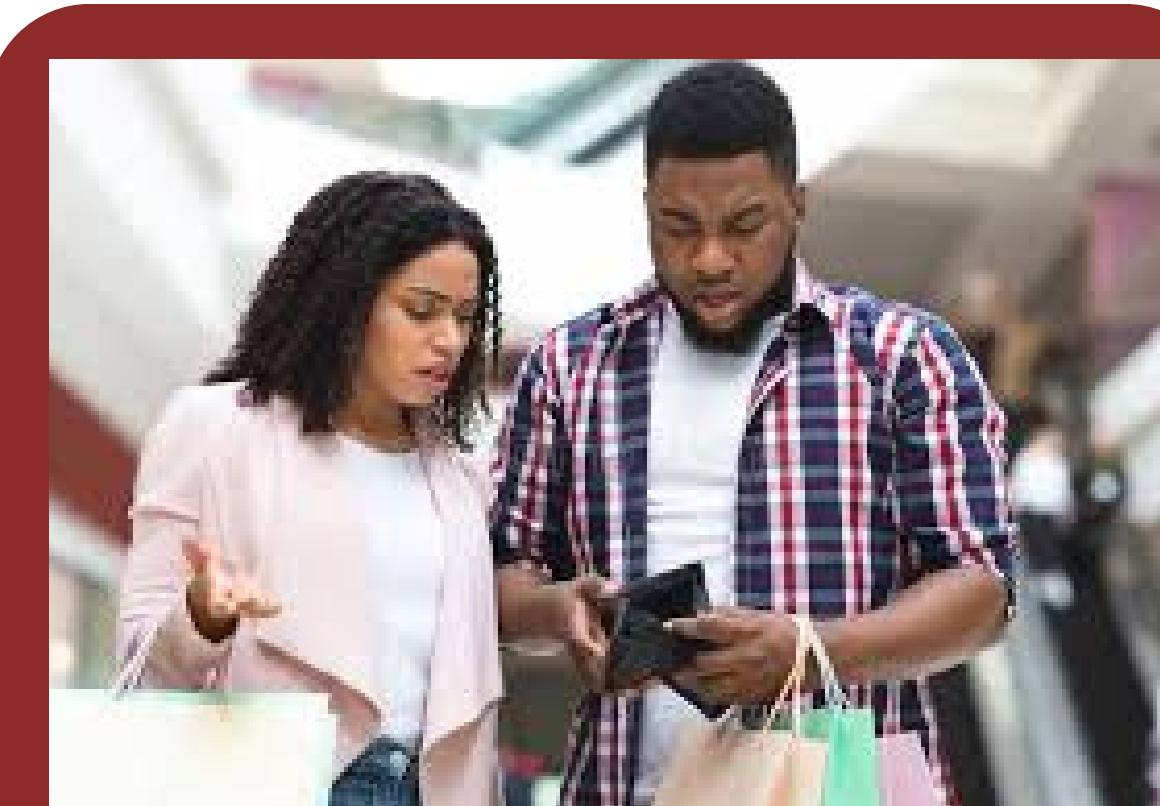
JACKSON CHLEBOWY

Table of CONTENTS

- The Problem
- Data Description
- Predictive Modeling
- Model Insights
- Conclusion

THE PROBLEM

As prices rise, consumers are increasingly worried about their financial stability and the affordability of essential products. Our project explores what economic factors can predict these shifts in sentiment and which are most important to how Midwesterners feel about grocery affordability.



Our FOCUS

- Identify which economic variables influence Midwestern perceptions of affordability
- Build random forest and logistic models to predict and explain future positive/negative Midwestern consumer sentiment based on real economic data and scraped sentiment analysis
- Understand how economic factors shape affordability concerns



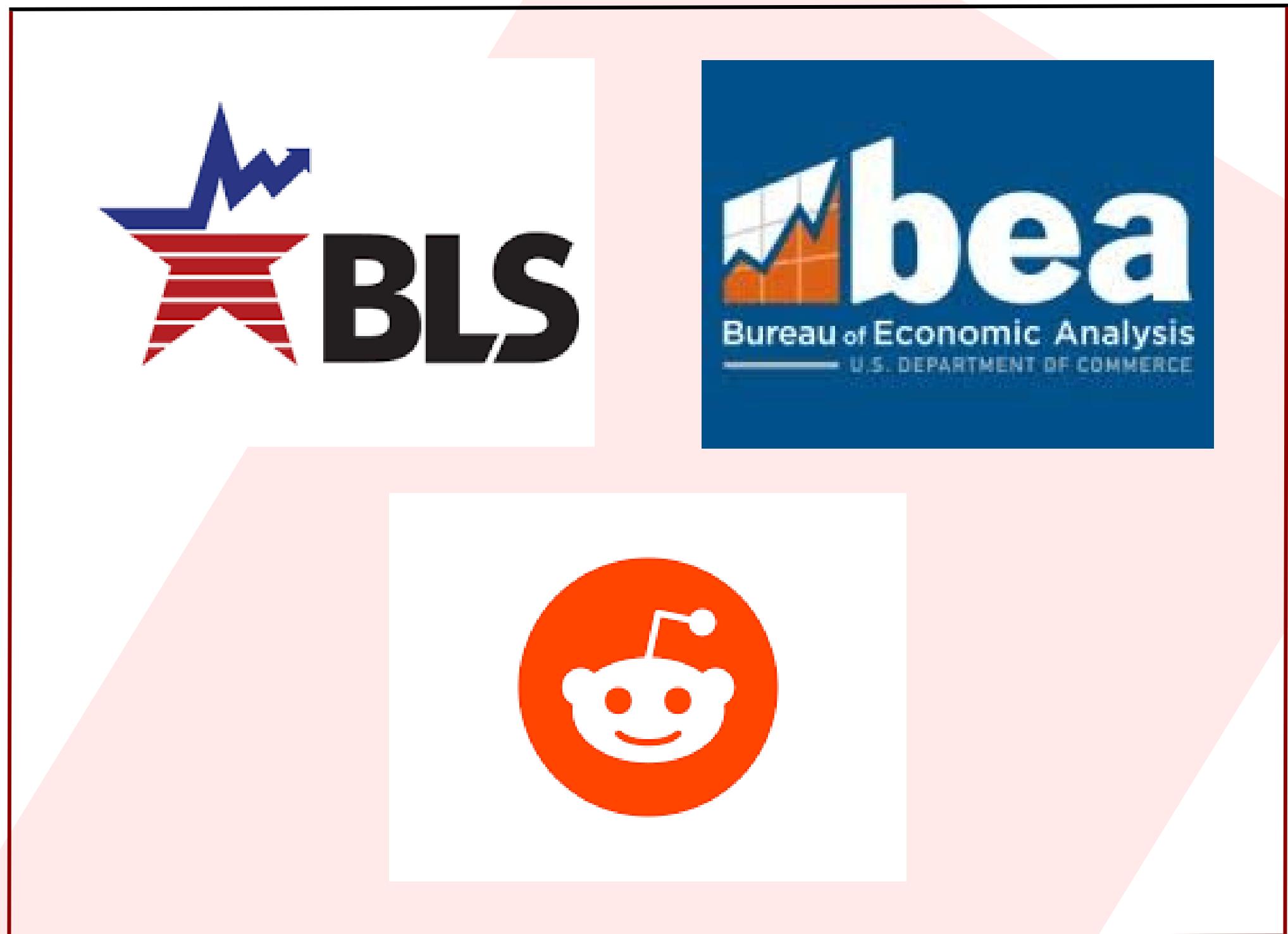
The Process



- 1 Extract the data from BLS and BEA online sources for 5 midwest states
- 2 Scrape Reddit posts for consumer sentiment
- 3 Define Our Target– Positive/Negative Sentiment
- 4 Perform Random Forest and Logistic Regression
- 5 Model Conclusions

DATA DESCRIPTION

- Monthly Economic Data from 2015–2025
 - BLS
 - Unemployment
 - Meat, Dairy, Produce CPI
 - Wholesale, Grocery Store PPI
 - Real Disposable Income
 - BEA
 - Personal Consumption Expenditures (PCE)
- Scrapped Reddit posts by state to determine monthly midwest sentiment



THE TARGET

1103 Reddit posts were API scraped from these 5 midwest states:

Michigan, Ohio, Illinois, Indiana, and Wisconsin

From 2015-2025 and query = "economy {state}"

NLTK vader was used to gather compound sentiment score aggregated by year and month

We aim to predict positive or negative compound sentiment (1/0)



PREDICTIVE MODELS

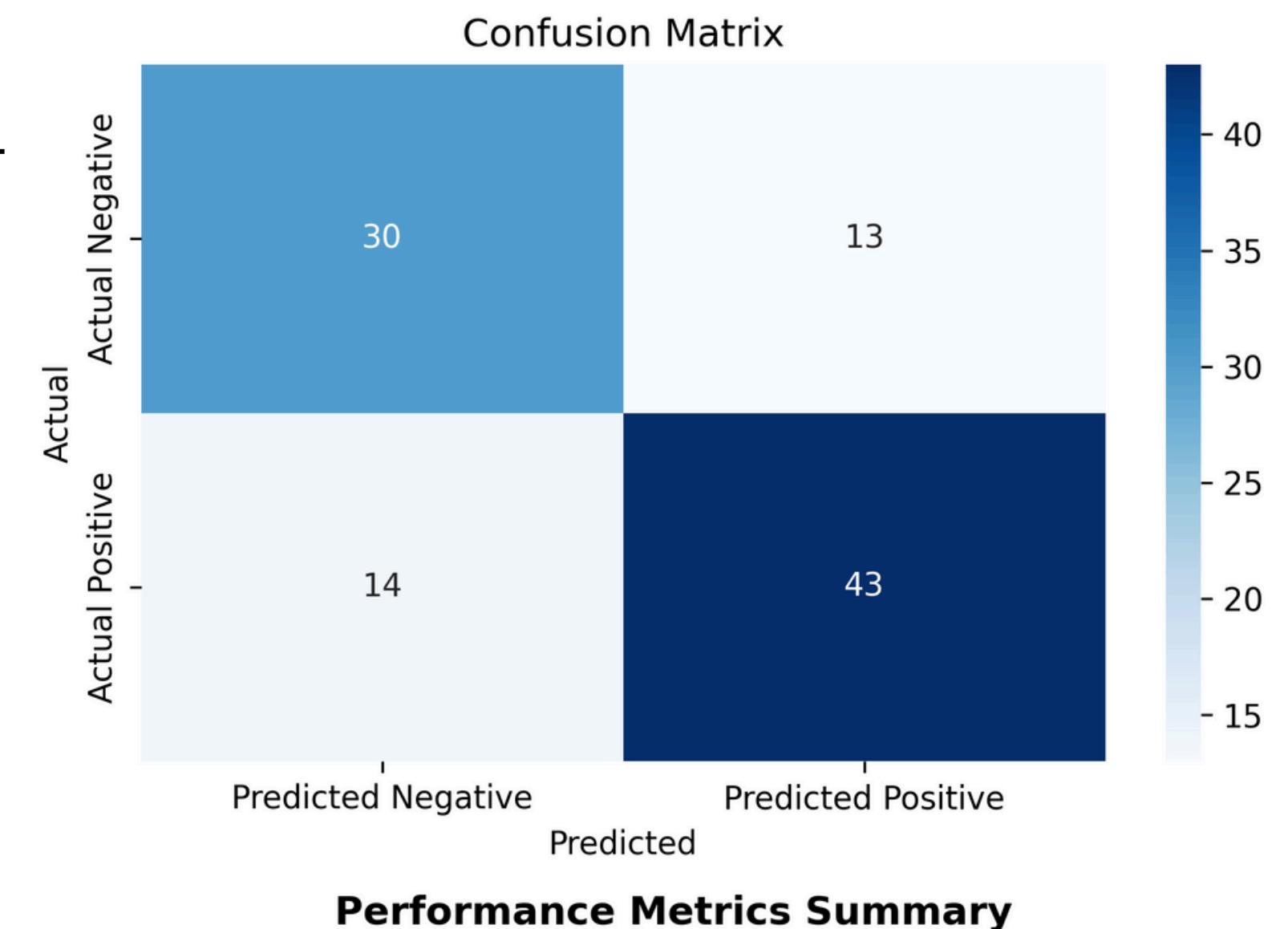
Logistic Regression & Random Forest

- Interpretable model.
- Prediction coefficients show **direction** and **magnitude** of association with the target classification.
- Predictive Power.
- Look at which variables are most **important in predicting** the consumer sentiment.

LOGISTIC REGRESSION

The confusion matrix shows that the logistic regression model performs fairly on predicting overall midwest consumer sentiment.

- 73% of all model predictions are correct.
- Model performs better than random chance.



Metric	Value
Accuracy	0.73
ROC AUC	0.692
Precision (Macro)	0.725
Recall (Macro)	0.726
F1 Score (Macro)	0.725

LOGISTIC REGRESSION

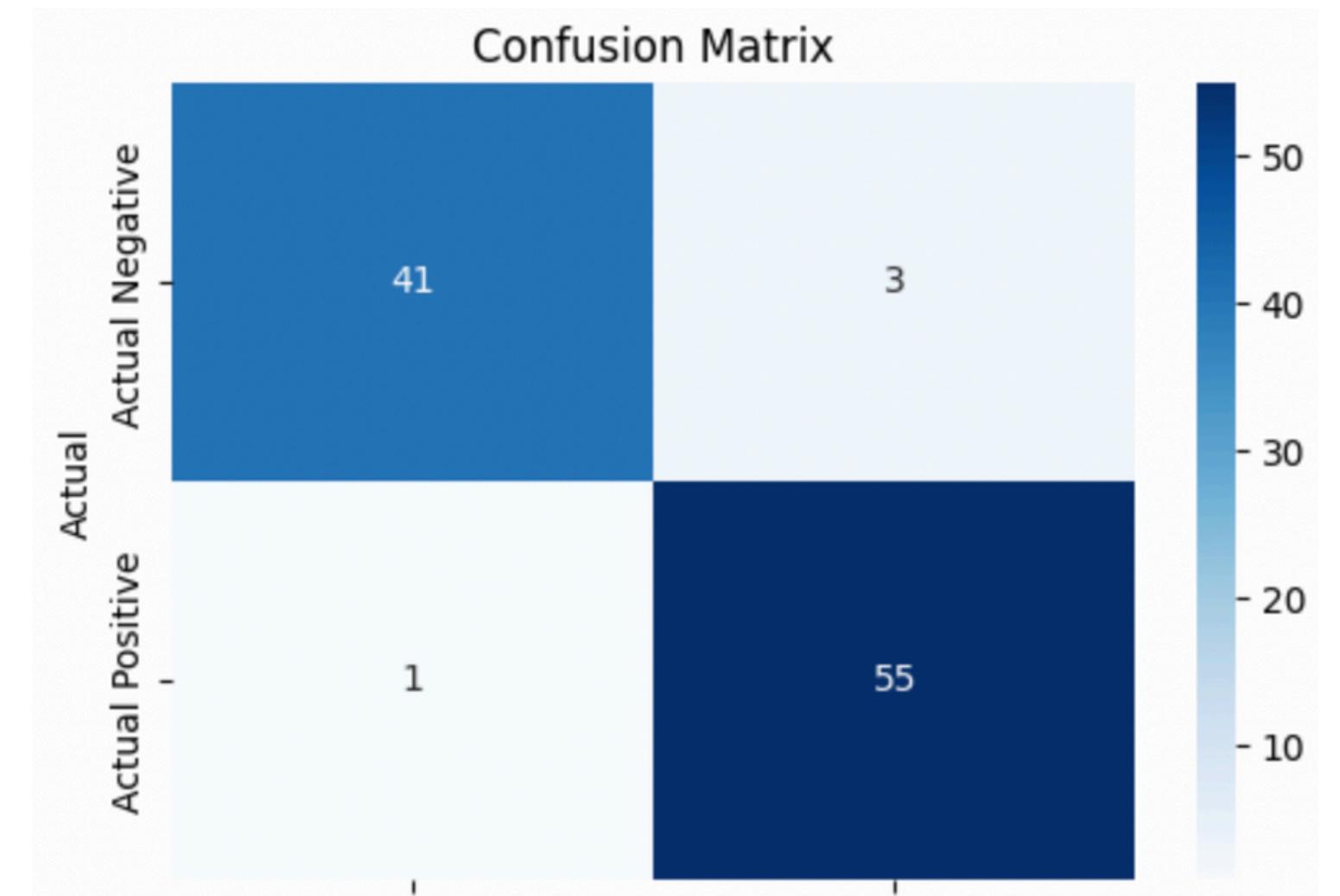
- Counter-Intuitive coefficients.
- Variables scaled before modeling.
→ 1 standard deviation increase in Real Disposable Income **decreases** the odds of positive sentiment by $\exp(-1.741) = 0.175$ times.
- Produce CPI has the highest correlation with positive sentiment.
- Real disposable income has the highest correlation with negative sentiment

Logistic Regression Coefficients

Predictor	Coefficient
mw_produce_CPI	0.755
mw_wholesale_PPI	0.745
mw_supermarket_PPI	0.613
Unemployment	0.573
mw_PCE	-0.064
mw_diary_CPI	-0.353
mw_meat_CPI	-1.004
mw_RealDislIncome	-1.741

RANDOM FOREST OVERALL

The confusion matrix shows that the Midwest economic data can precisely predict both negative and positive overall consumer sentiment.



	Class	Precision	Recall	F1 Score	Support (# Obs)
0	Negative	0.976	0.932	0.953	44
1	Positive	0.948	0.982	0.965	56

VARIABLE IMPORTANCE OVERALL

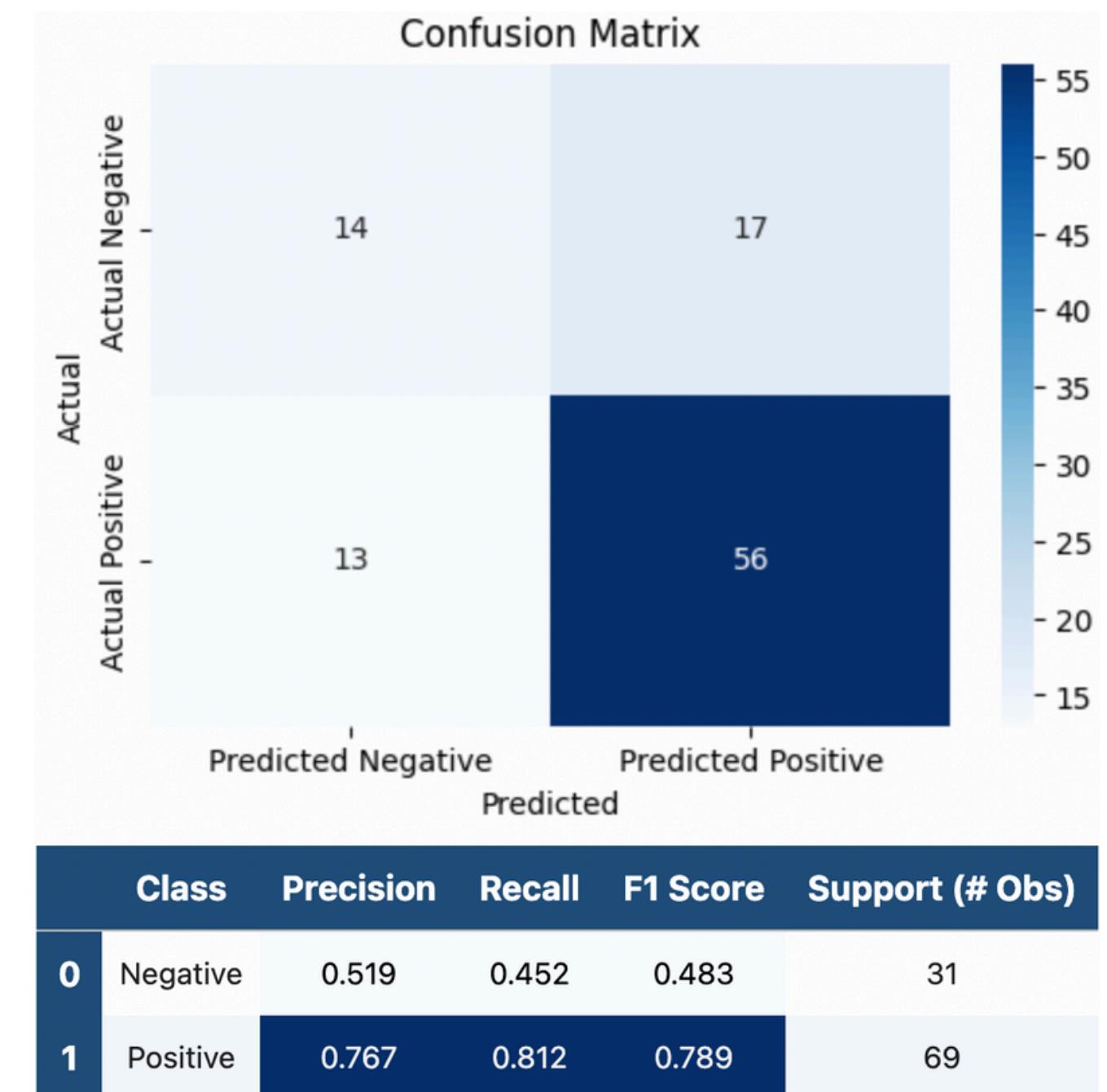
Real disposable income was the most important variable for predicting consumers economic sentiment.

Personal Consumption Expenditures (PCE) was the least important predictor.

Feature	Importance
mw_RealDispIncome	0.1747
mw_produce_CPI	0.1609
mw_meat_CPI	0.1559
mw_wholesale_PPI	0.1522
mw_supermarket_PPI	0.1287
mw_diary_CPI	0.1236
Unemployment	0.0725
mw_PCE	0.0315

RANDOM FOREST STATE

The confusion matrix shows that the Midwest economic data does not as accurately predict the state level sentiment for the 5 chosen states.



VARIABLE IMPORTANCE STATE

Produce CPI was the most important variable for predicting Midwestern state consumers economic sentiment.

Personal Consumption Expenditures (PCE) was the least important predictor.

Feature Importance

Feature	Importance
mw_produce_CPI	0.1681
mw_RealDisplIncome	0.1515
mw_wholesale_PPI	0.1505
mw_meat_CPI	0.1493
mw_supermarket_PPI	0.1427
mw_diary_CPI	0.1263
Unemployment	0.0798
mw_PCE	0.0318

MODELING INSIGHTS

1

CPI is not the only useful predictor of consumer sentiment

2

Real Disposable Income is one of the most influential variables when predicting Positive Sentiment

3

Our Random Forest Model was able to predict sentiment better than logistic regression

4

Random forest predicts overall Midwest sentiment more accurately than at the state level

LIMITATIONS

- We treat each reddit post representing the true sentiment for the Midwest.
 - Sentiment Analyzer limitation: Cannot detect sarcasm in posts
- Overall low number of observations.
 - Models are built with 331 distinct rows
 - Potential overfitting on model
- Missing economic data was imputed.
 - Might not capture full economic conditions at that time



Conclusion

There are factors other than Consumer Price Index (CPI) that influence consumer sentiment on both the midwestern and state levels.

Our models can prove beneficial to business in understanding strategic approaches to consumers based on the overall sentiment.

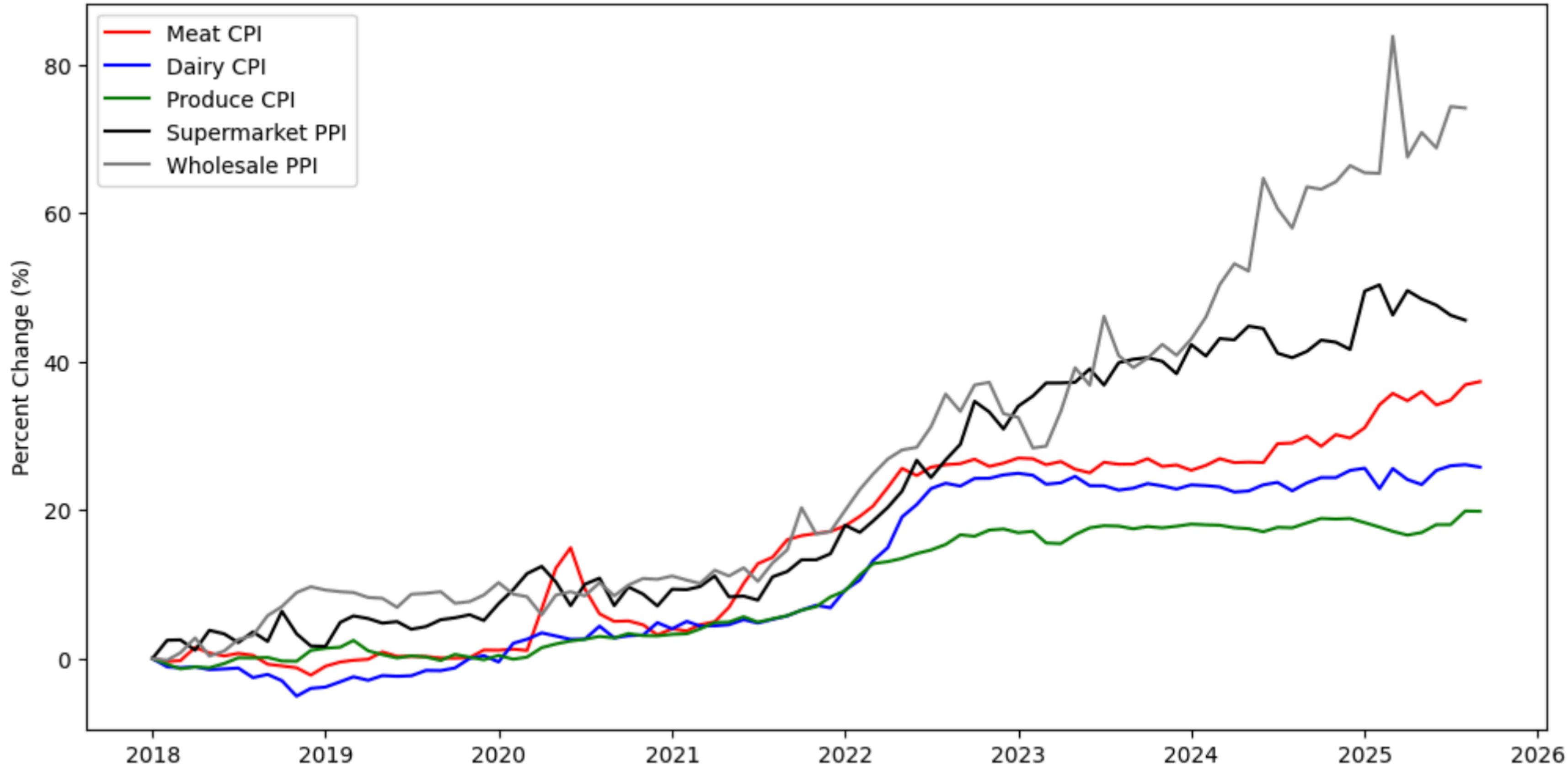


THANK YOU!

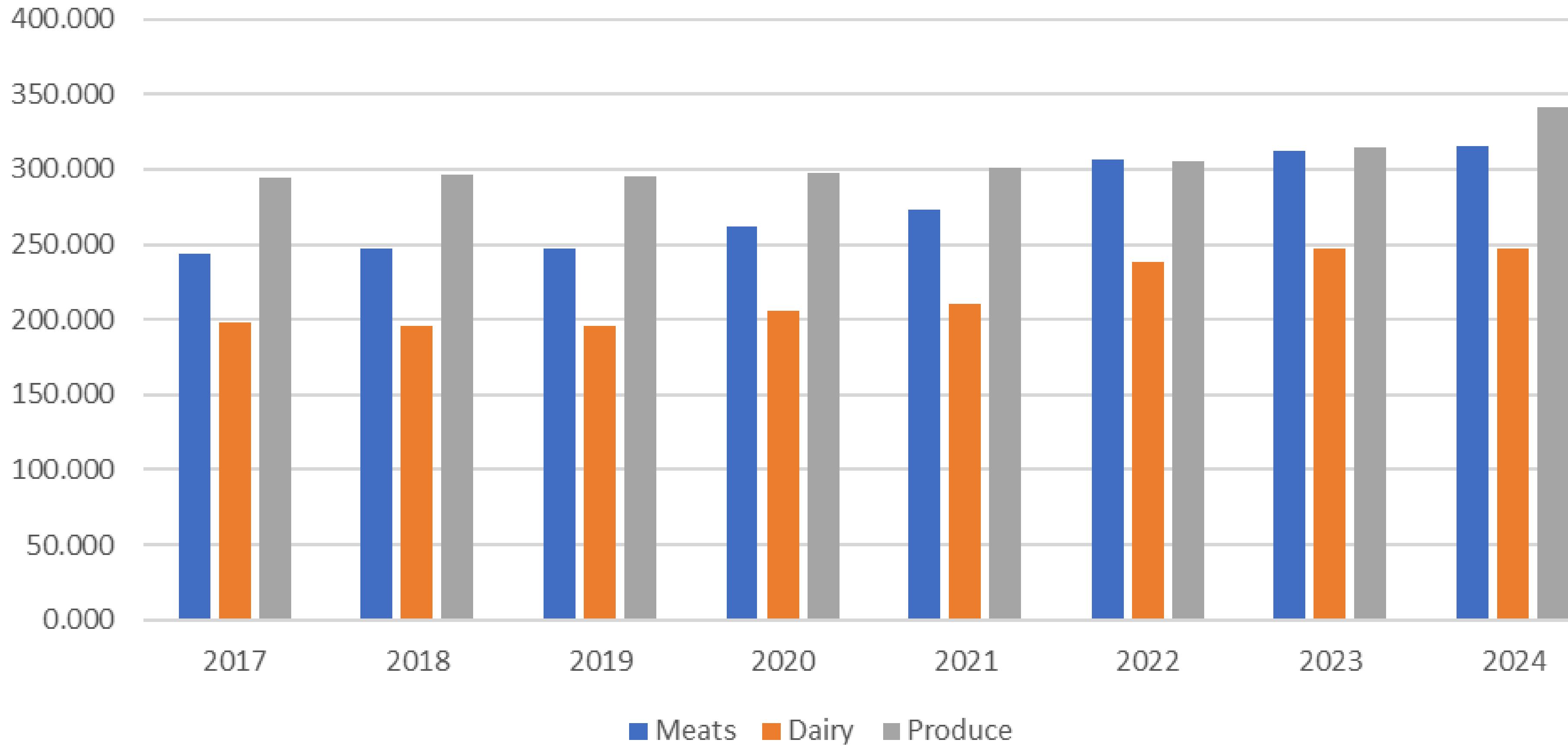
**ANY
QUESTIONS?**

Appendix

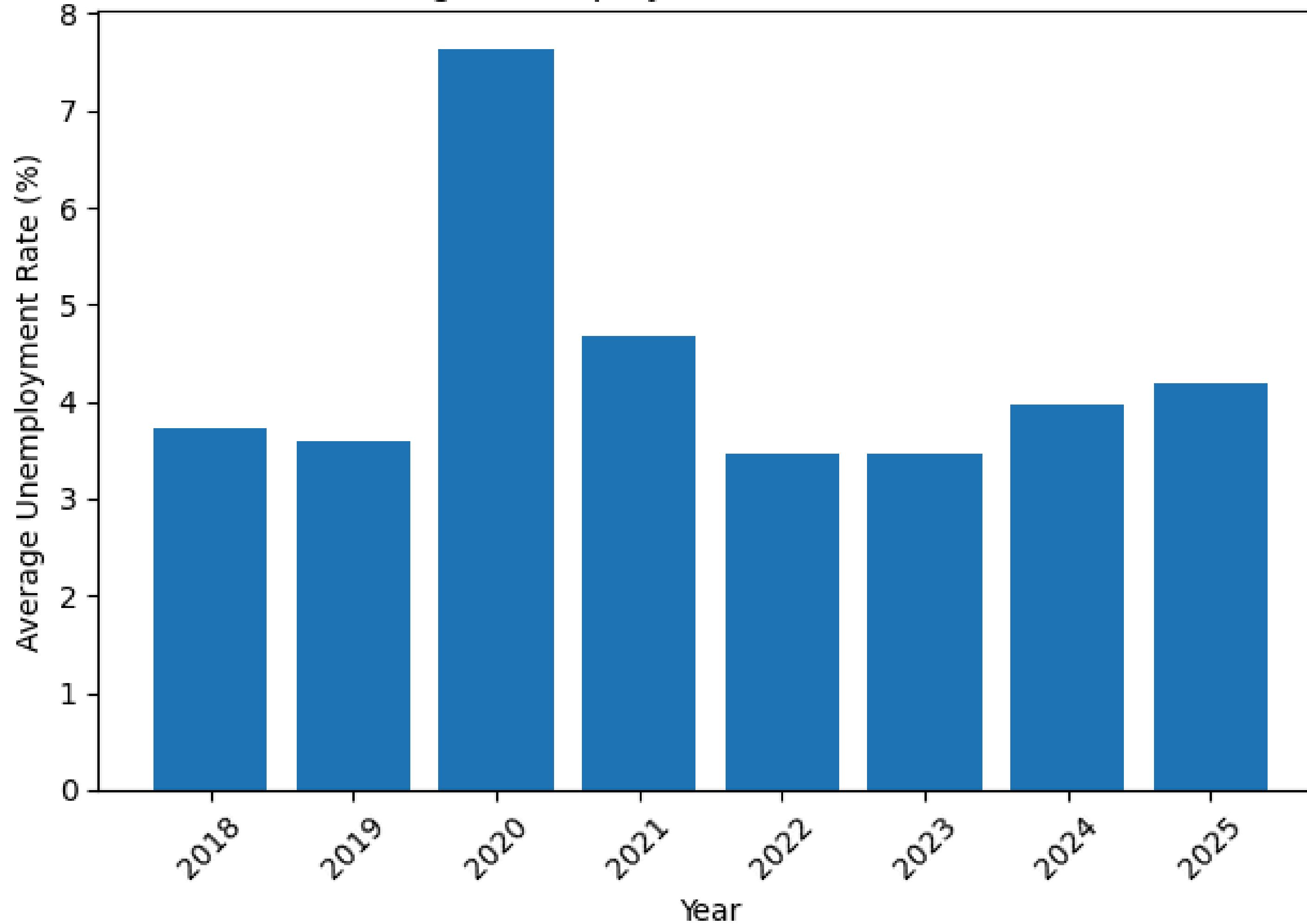
Percent Change Since Start (2018 = 0%)



Average CPI per Year



Average Unemployment Rate (2018-2025)



Real Disposable Income Over Time

