TEAM 84- EXPLORING TRENDS IN CONSUMER DEBT IN THE UNITED STATES

INTRODUCTION

The collections industry is often overlooked by many, however, with the ballooning growth of U.S. debt, this industry should not be ignored. Like most industries, different actors participate and often share similar challenges.

Objectives

1. "What are the demographic and macroeconomic features that are significant in predicting whether or not a consumer begins repayment of a debt within 5 years of an initial chargeoff date?"

2. "Using those identified features, what is the probability that a current consumer, who has yet to begin repayment, but that is still within that timeframe (charge-off dates on or after 4/1/2019), to begin repaying their debt?"

The answers to these questions are highly sought out by various entities within this industry, from a revenue perspective in business, to that of socioeconomic policy in government, thus this project can reveal significant insights for many that participate within this space.

DATA

The datasets for this project were collected from both:

An active business entity within the collections industry

More than five MySQL tables were joined together and exported to construct two tables consisting of data concerning:
1. account-level information
2. account payment transactions
More than 170,000 accounts were randomly sampled, along with more than 700,000 corresponding payment transactions

A U.S. Government agency

A data table was downloaded from a government website and consisted of macroeconomic data. The macroeconomic data covered 15 different macroeconomic measures and spanned in time from 1998 to 2022.

All three data tables were restricted to 17 designated states, representing 6 regions making up the United States, and subsequently cleaned for further analysis.

OUR APPROACH

Algorithm and Interactive Visualization

- Classification models (Random Forest, Voting Ensembles, etc.) for predicting debt repayment likelihood and satisfactory repayment within 5 years
- Time series analysis for identifying pre and post COVID-19 trends in consumer debt
- Interactive visualization using Tableau and Plotly for exploring regional differences, feature importance and trends

How they work

- Classification models trained on consumer demographic and macroeconomic features to predict repayment probabilities
- Time series analysis uses techniques like seasonal decomposition and SARIMA for forecasting charge-offs and analyzing cost of living trends
- Interactive visualization allows users to filter and explore predicted probabilities based on specific scenarios, and deep dive into details

Why they can effectively solve our problems

- Combining individual and macroeconomic factors provides a comprehensive view for predicting repayment
- Rule-based algorithms like Random Forest have shown strong predictive power in previous studies
- Interactive visualization enables stakeholders to assess specific scenarios for decision-making
- Time series analysis uncovers trends and shifts in debt patterns pre and post COVID-19

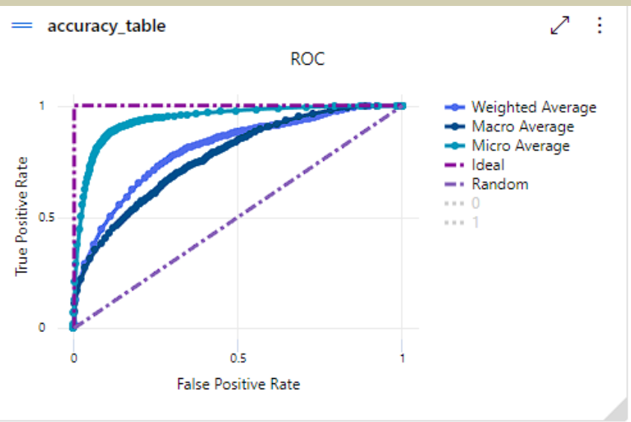
What is new in our approaches

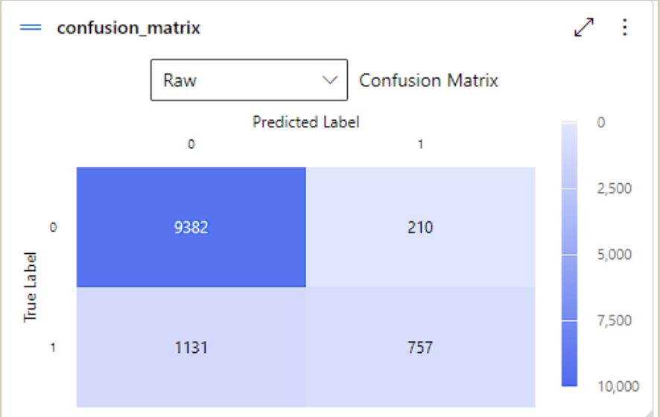
- Integrating both individual and macroeconomic features, unlike previous studies that focus on one or the other
- Employing modern machine learning techniques with Azure AutoML and python libraries for enhanced prediction
- Exploring differences in debt trends pre and post COVID-19, which is previously unseen
- Providing an interactive visualization platform on Tableau for stakeholders to engage with the results backed by actual company data is unprecedented

EXPERIMENTS

Evaluations

- Classification models evaluated using metrics like AUC, accuracy, ROC curves, and confusion matrices
- Time series models assessed for goodness of fit and forecasting performance





COMPARISON TO OTHER METHODS

Feature insights

Novel approach integrating individual and macroeconomic factors generating insightful results on features impacting repayment probabilities and identified trends and regional differences across the US.

ML models

Employs Azure AutoML, enabling >50 models trained; advanced seasonal decomposition and SARIMA, compared to traditional statistical/ manual ML models.

Interactive Viz

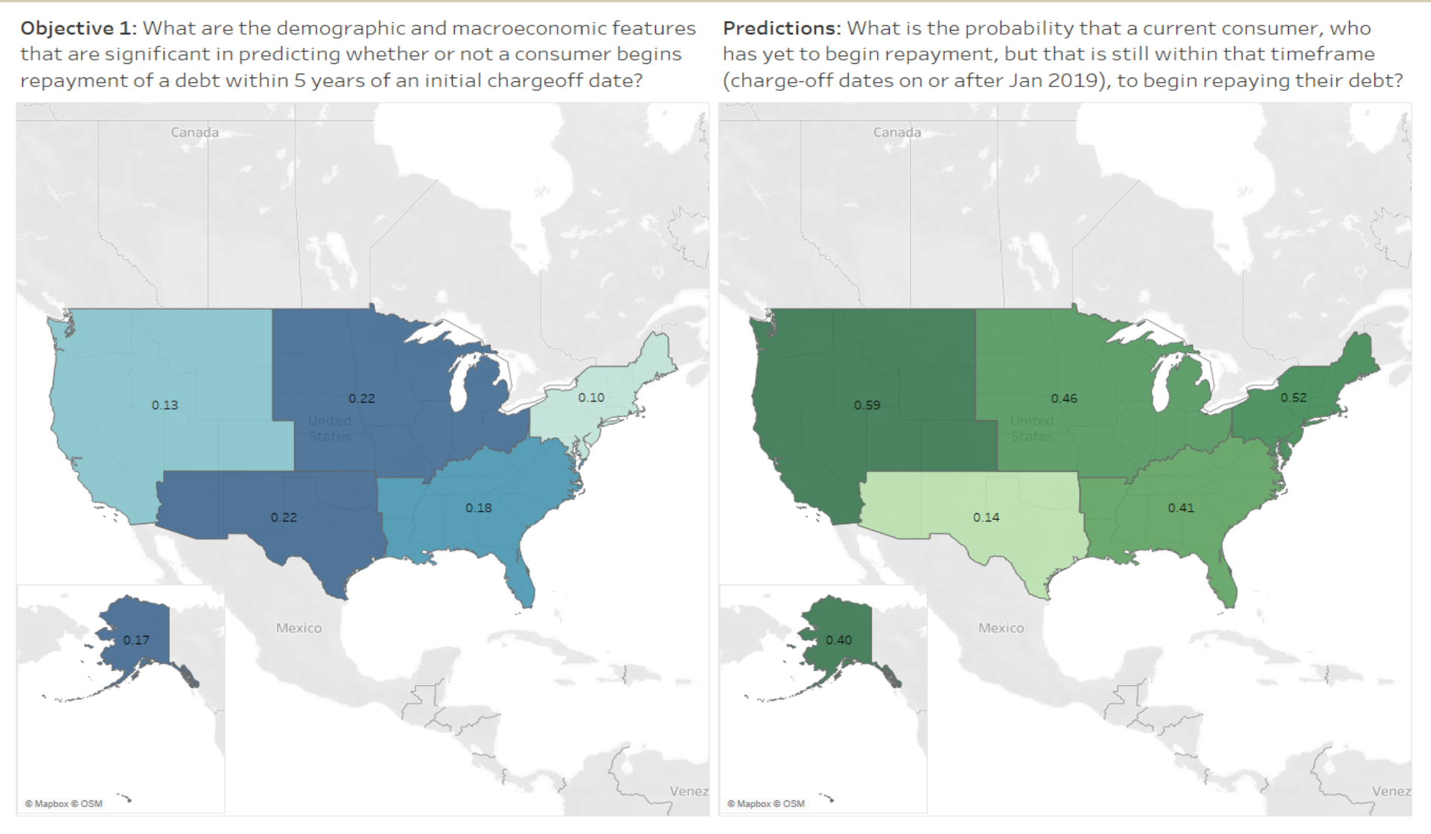
Provides interactive visualization for user engagement and scenario analysis on Tableau Public with potentials of on-going updates

RESULTS

Classification Models

Voting Ensemble achieved best AUC for repayment prediction (0.85) and satisfactory repayment (0.99)

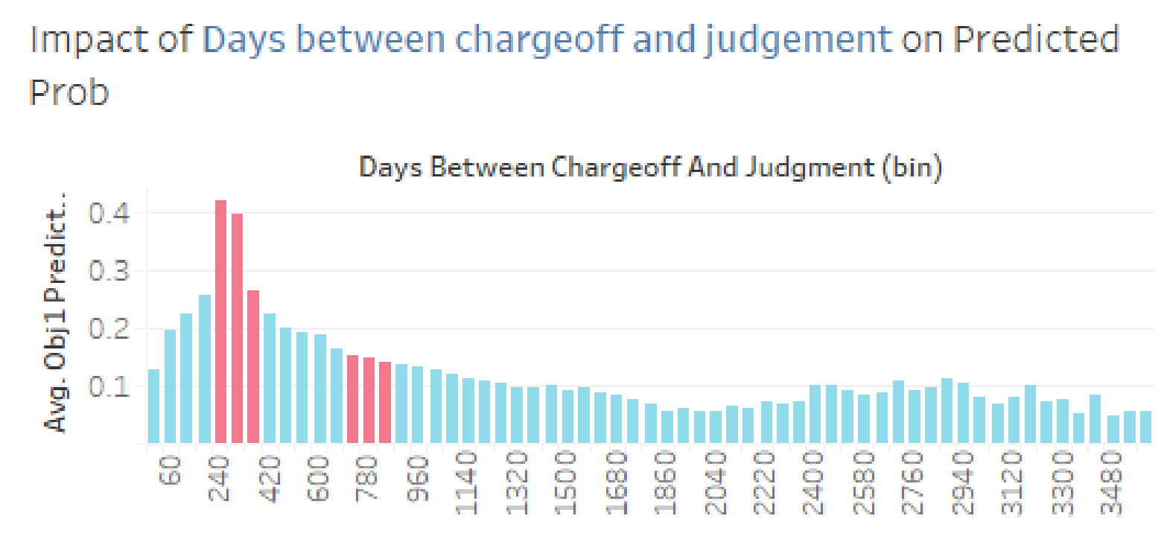
Predicted repayment probability with new data post 2019

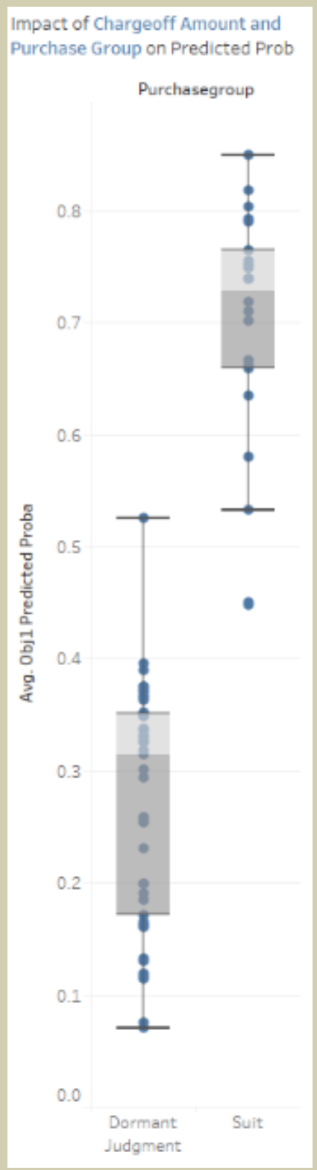


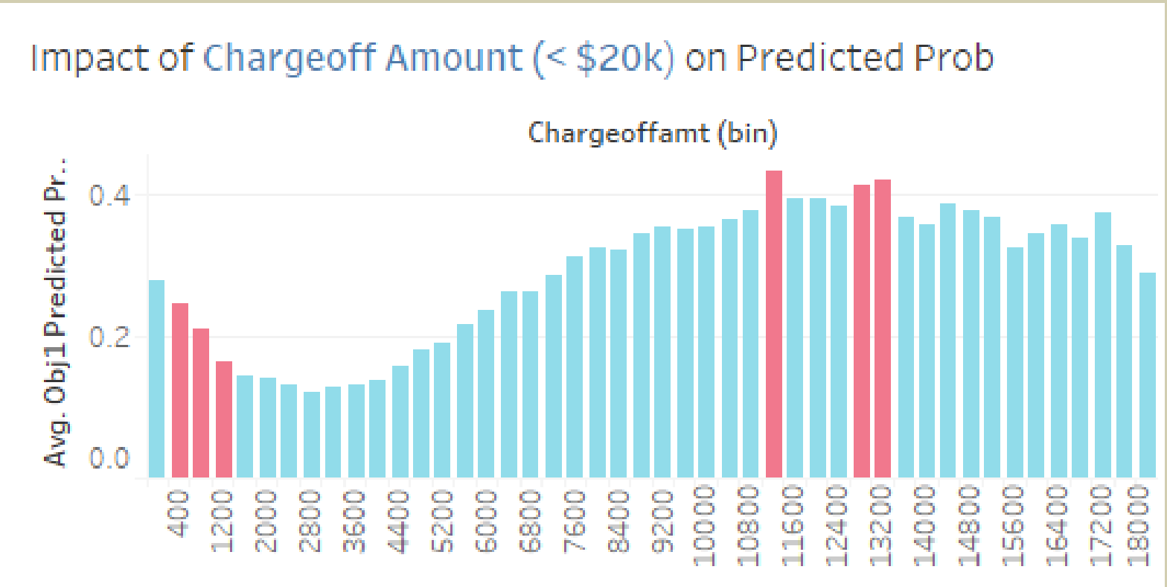
Objective 1: What are the demographic and macroeconomic features that are significant in predicting whether or not a consumer begins repayment of a debt within 5 years of an initial chargeoff date?

Predictions: What is the probability that a current consumer, who has yet to begin repayment, but that is still within that timeframe (charge-off dates on or after Jan 2019), to begin repaying their debt?

Key features identified: time between charge-off and judgment, charge-off amount, purchase group, payment frequency

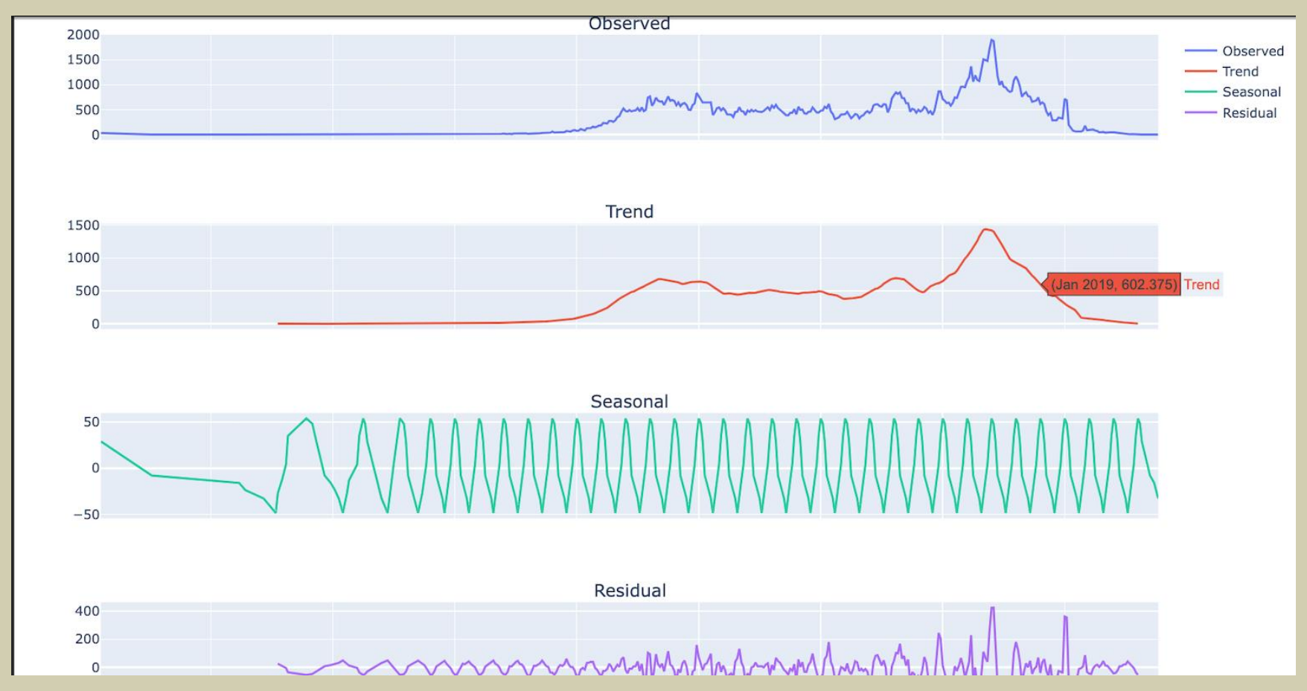






Time Series Analysis

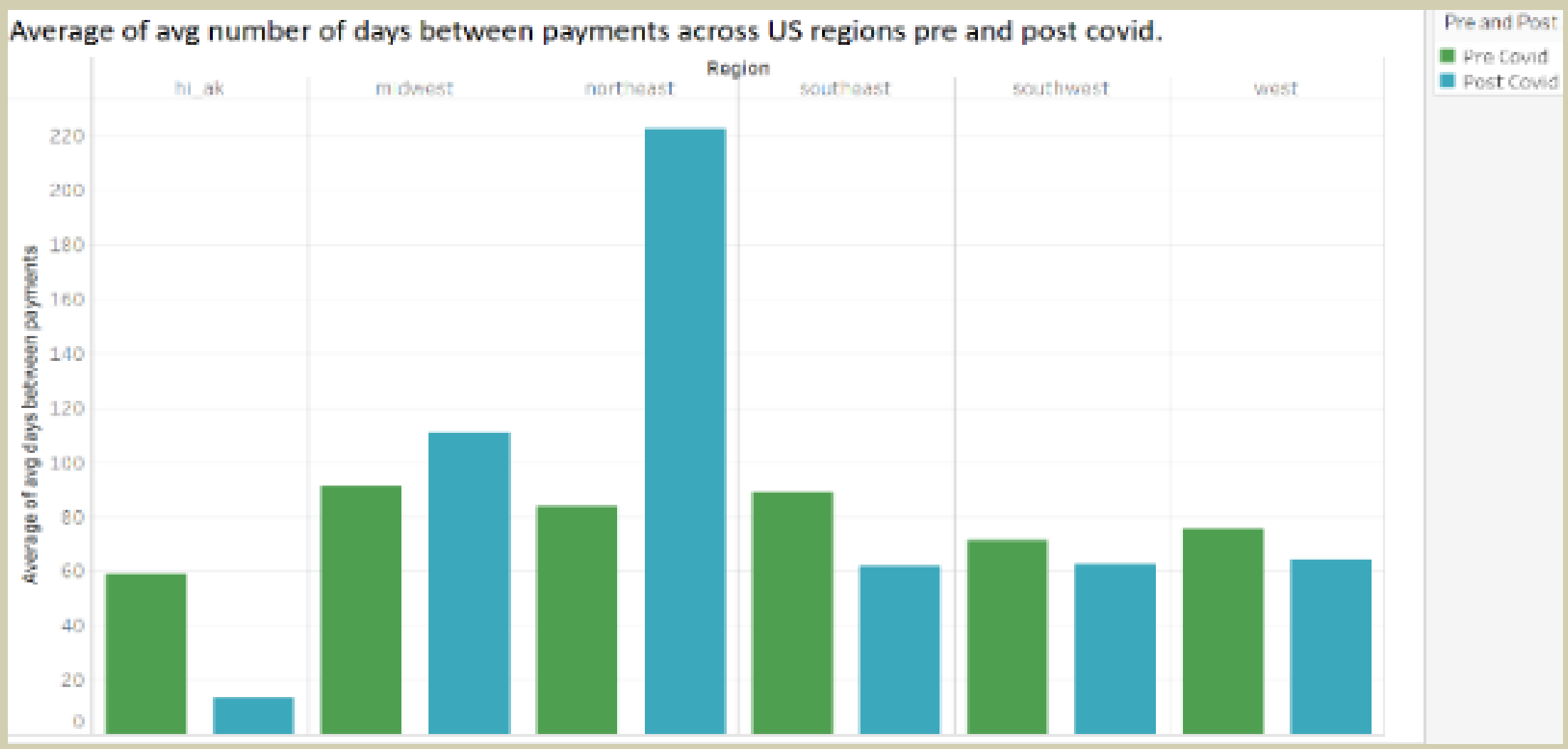
Seasonal Trend identified

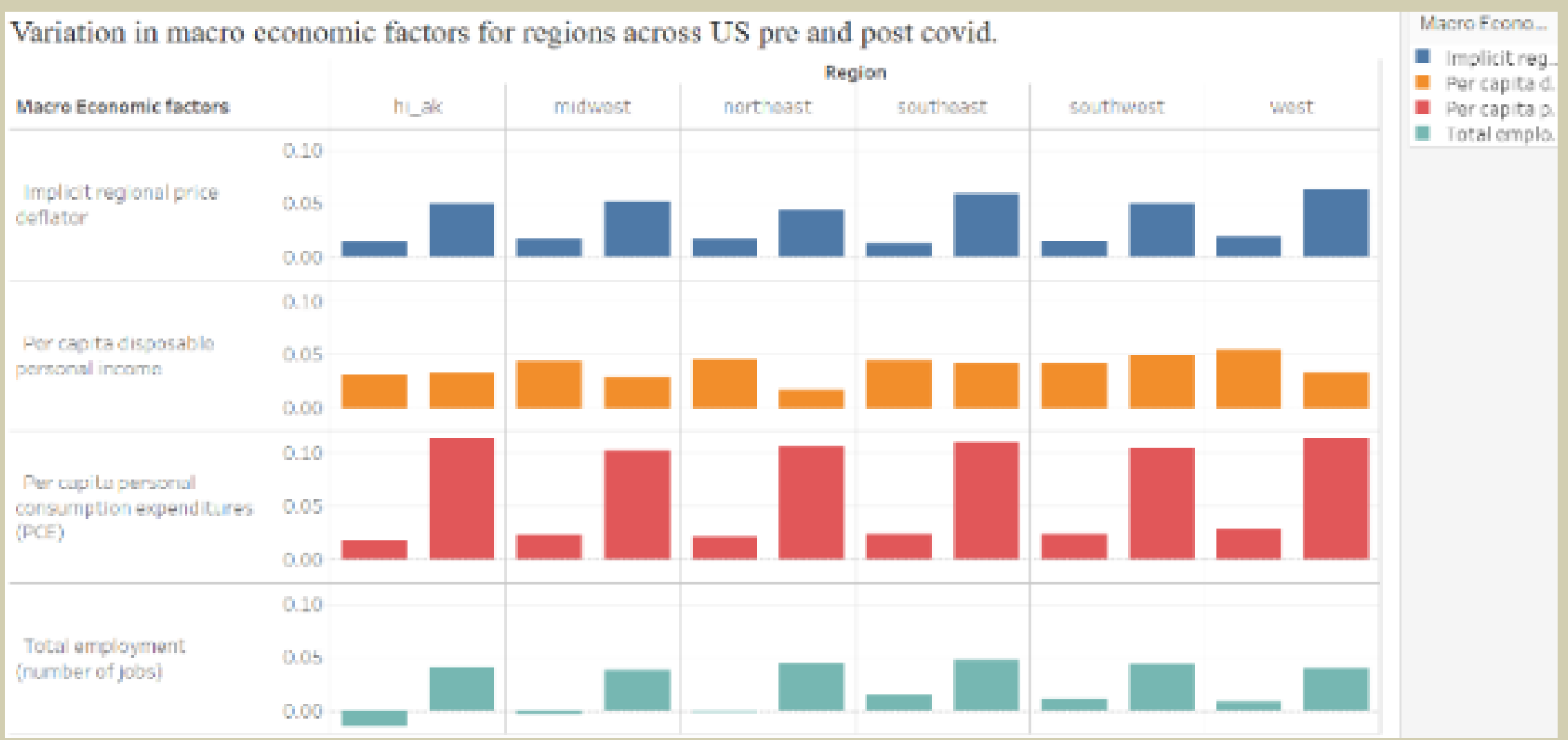


Trend Analysis Pre & Post COVID

Post-COVID: decreased time between charge-off, judgment, and first payment; regional differences in payment frequency

Macroeconomic factors show mixed trends post-COVID, impacting consumer financial health





Mark Yun Kiu Chan
ychan76@gatech.edu

Adam Bishop
abishop41@gatech.edu

Annu Kumari
akumari36@gatech.edu

Megha Gulati
mgulati30@gatech.edu