Job Preservation Dynamics: A Comprehensive Study of Payment Protection Program Loan Efficacy

Ruhani Walia

University of Toronto

April 2024

Abstract

This paper investigates the effectiveness of the Paycheck Protection Program (PPP) in minimizing unemployment during the economic downturn caused by the COVID-19 pandemic. Using data from over 600,000 small businesses, we conduct a multifaceted analysis incorporating exploratory data analysis, geographic insights, and advanced statistical techniques including OLS regression, regression trees, and random forest models. Our findings reveal significant variability in the program's impact, influenced by business type, owner demographics, and regional economic conditions. Notably, more than half of the PPP loans were concentrated in the 150,000 to 350,000 USD range, with corporations being the most prevalent recipients. Geographic analysis indicated a North-South divide in terms of loan amounts per capita and jobs retained, highlighting the role of regional economic policies in shaping outcomes. Statistical models further suggest that factors such as the number of employed individuals and state-specific party control significantly predict job retention, with our preferred model (Model 9) exhibiting the highest explanatory power. The study concludes that while the PPP was pivotal in mitigating job losses for many, its effectiveness varied widely, suggesting the need for more tailored approaches in future economic relief efforts.

1 Introduction

As the COVID-19 pandemic made its way across the United States, businesses faced an unprecedented storm of economic hardship, grappling with closures, layoffs, and financial uncertainty. In response, the U.S. government introduced the Paycheck Protection Program (PPP) to provide timely financial relief. The PPP was established through the Coronavirus Aid, Relief, and Economic Security (CARES) Act enacted in March 2020 and implemented by the U.S. Small Business Administration (SBA). Since then, the SBA has reported more than seven million approved loans through the program, distributing nearly 700 billion dollars. This makes the PPP one of the farthest-reaching business relief programs in U.S. history.

The PPP allowed entities to apply for low-interest private loans, aiming to help businesses cover payroll costs, rent, utilities, mortgage interest, and other costs. In terms of the program's design, the loans would be partially or fully forgiven if businesses met certain criteria, such as maintaining employee counts or stable wages. Loans provided by the PPP were 2.5 times the applicant's average monthly payroll costs.

The success of the policy has been a topic ripe with differing opinions. For instance, Autor et al. (2022) estimate that only 23 to 34 percent of PPP dollars went directly to workers who would have otherwise lost their jobs. Instead, they find that most of the money went to shareholders and their creditors. As such, the authors argue that the program was highly regressive; a policy that ended up favouring the rich.

The authors use loan-level data from the PPP to determine the size of each firm that received a PPP loan. They also use data from the Census Bureau's Statistics of U.S. Businesses (SUSB) to calculate the "takeup rate", which they define as the employment-weighted share of firms that received PPP loans. They use this measure to evaluate the program's timeliness and targeting effectiveness.

Others argue that the program did well in terms of preventing business closures. Dalton (2021) found that businesses that had received a PPP loan were 5.8% less likely to be closed one-month post-receipt and 3.5% less likely to be shut down after seven months. The study used administrative wage records and a doubly robust dynamic difference-in-difference routine.

The literature has considered several measures to elucidate the PPP's effectiveness in minimizing unemployment. The primary research question guiding this investigation is whether the PPP was an effective policy in reducing unemployment during times of economic hardship caused by the COVID-19 pandemic. Notably, in this study, the effectiveness of the policy will be measured by the number of jobs retained due to PPP loan approval. The PPP loan dataset used for this research is sourced from the U.S. Department of the Treasury. This analysis aims to provide valuable insights into the nuanced dynamics of the PPP, offering a holistic understanding of the factors contributing to program efficacy in supporting small businesses during challenging times.

This research embarks on a comprehensive analysis of over 600,000 small businesses that participated in the PPP, focusing on critical variables as they relate to economic recovery. Thus, variables of interest to this study include the business type, business owner ethnicity, business owner gender, the lender involved, loan amount, and the number of jobs retained by the business due to receiving approval for a PPP loan. Additionally, data on unemployment rates and layoffs are investigated to create a measure of how strongly impacted particular geographies were by the pandemic.

As briefly described, the motivation behind this study stems from a multifaceted exploration of the impact of government business relief measures like the PPP on job retention. For example, Solovyeva et al. (2023) find that job-retention schemes, alongside other fiscal support measures like the PPP, helped mitigate the rise in the unemployment rate by approximately three percentage points during the pandemic. The paper used a microsimulation approach (EUROMOD) and household data to assess the effectiveness of those schemes in stabilizing household income during the pandemic across EU countries.

However, specific nuances emerge in the literature. Selley (2023) finds that payroll-incentive policies like the PPP are ineffective in the short term for health-related economic downturns despite playing a critical role in keeping businesses open, which later positively contributed to long-term employment. This study uses data from the SBA on approved PPP loan applications to create a lagged dependent variable model, and growth model.

More specifically, Selley finds that loans had a significant and positive effect on employment but only peaked approximately six months post-receipt.

In terms of the recipients, Dalton (2021) also reveals that the PPP had a stronger positive effect on smaller firms rather than larger firms. Dalton finds that the smallest firms rebounded much faster than others, indicating that the PPP had a significant and immediate effect, with a 14% increase in employment within a month of PPP loan approval. The study finds that this is much larger in comparison to the smaller effects on wages for larger firms.

Another dimension that the literature has considered is the industry of recipients. Mumford et al. (2023) note negligible impacts on employment for nonprofits through the PPP. Using literature on nonprofit resilience, propensity score matching with survey data, and publicly available PPP data, the study finds that the PPP supported nonprofits in terms of reserve liquidity but minimally in supporting staff retention.

Overall, the literature considers measures such as how many dollars of the loan made it directly to workers, whether the PPP prevented business closures as opposed to job loss, how the program performed in the short term, and the traits of businesses that tended to benefit the most from the distribution methods set out by the federal government.

This paper aims to contribute to the literature in terms of identifying key traits that influenced the effectiveness of the PPP on job retention and expand upon the analysis of which businesses tended to benefit the most. By investigating patterns in geographic distribution and business owner demographics, the study aims to provide guidelines for future business relief programs to improve the efficacy of aid distribution.

Beyond examining the conventional factors such as business type and location, this research aims to delve into a nuanced geographic analysis and demographic scrutiny of business owners. The demographic analysis is anticipated to unveil potential areas of opportunity for business owners and highlight saturated markets. This might provide insight into how PPP loans impact owners differently. The ultimate goal is to define a predictive set of criteria that maximizes job retention and business sustainability after a business has received a loan. The study also seeks to contribute to the ongoing discourse on policy implications by exploring patterns of success and standards in economic relief programs.

Findings of this study's preliminary exploratory data analysis found that the majority, approximately 59%, of approved PPP loans fall within the range of 150,000 to 350,000 USD. Corporations emerge as the dominant recipient business type, constituting 48% of all businesses in the dataset. White owners account for 84% of the total, with males representing nearly 82% of all owners. The most frequent lender in the data is The Huntington National Bank, though there is great diversity in participating lenders. The average number of jobs retained is notably influenced by a few businesses with exceptionally high job retention. Businesses with higher loan amounts tended to retain more jobs due to loan approval, and the most common business types securing PPP loans are corporations, LLCs, and nonprofit organizations. Nonprofits stand out with the highest average number of jobs retained, while self-employed individuals, independent contractors, and sole proprietorships tend to have the lowest job retention. Finally, the number of loans provided significantly decreases over the three months covered by the dataset, from April 2020 to June 2020.

The geographic analysis conducted yields insightful comparisons between the distribution of per capita loan amounts, jobs retained, and diversity scores. Particularly intriguing is the observation that states situated in the central North consistently exhibit the lowest loan amounts but boast the highest job retention rates and diversity scores among business owners. As such, a discernable divide in jobs retained becomes apparent between the Northern and Southern states, with the latter lagging. This finding is interesting in the context of literature that argues COVID-19 accelerated a pre-2020 trend of population and jobs shifting from the northern industrial states to southern states as per CityJournal. Finally, the upper east states tended to receive slightly higher loan amounts than the central North, though they did experience nearly as much job retention.

Merging web-scraped data on CARES Act funding yields insights regarding employment and worker class. The PPP program received funding from three tranches. The first tranche came from the CARES Act when the program was established. However, the PPP then received funding from the Health Care Enhancement Act and the Consolidated Appropriations Act. The data indicate a noticeable correlation between higher CARES Act funding per capita and higher jobs retained per capita, especially evident in states located in the central and upper North regions. Interestingly, states receiving higher per capita PPP support tended to exhibit lower per capita CARES Act support, indicating potential trade-offs in the implementation of federal relief initiatives.

Humphries et al. (2020) found that the smallest businesses were the least aware of government assistance programs, including both the CARES Act and PPP. Studying the distribution of CARES Act funding thus highlights areas where small businesses were unable to catch up with larger firms. These findings were informed from survey data collected by the authors from over 8,000 U.S. small business owners. This data was collected starting one day after the CARES Act was passed and continued until mid-April 2020.

Furthermore, Cortes et al. (2020), using Current Population Survey (CPS) panel data, found that individuals who were low-earning were disproportionately more likely to lose their jobs during the pandemic. In connection to the CARES Act, they find that the absence of this policy response would have caused an exacerbation of earnings inequality.

Therefore, to similarly study the connection between the PPP and employment data, a dataset is merged from the United States Census Bureau for the year 2020 to include data on employment, income and benefits, and worker type at the state level. This study finds that states receiving higher per capita loan amounts, such as California or Texas, tended to have relatively moderate per capita employment levels. It is also found that there is a large cluster of states that tend to receive lower per capita loan amounts but exhibit a range of low to high per capita employment levels. This makes it difficult to ascertain whether there is a strong relationship between employment and loan amounts per capita by state.

The merged data also provide insights in terms of allocation based on worker type. This study finds that, among states with the highest per capita jobs retained, the dominant category of worker type is private wage and salary workers. This indicates the PPP's effectiveness in preserving jobs in the private sector and underscores the importance of tailoring support mechanisms to address the diverse needs of different sectors of the economy, while also recognizing potential limitations in supporting public sector employment through such programs.

Overall, the findings emphasize the complex interplay between funding allocation, sectoral dynamics, and regional economic resilience in shaping the effectiveness of federal relief initiatives in mitigating unemployment and supporting economic recovery efforts across states. Understanding these dynamics is crucial for policymakers to design targeted interventions that address the specific challenges faced by different regions and promote equitable economic recovery nationwide.

As we navigate through the complexities of the PPP's implementation and its varied outcomes, the subsequent sections of this paper will delve into the quantitative and qualitative analysis of the data. We will explore the specific variables that influenced job retention rates across different states and business types, providing a detailed breakdown of the factors that contributed to the effectiveness or shortcomings of the program. By examining these elements in depth, we aim to shed light on how future relief measures can be optimized to better serve the diverse needs of American businesses and their employees during times of crisis. The following section will start by detailing our methodological approach, setting the stage for a comprehensive examination of our findings.

1.1 Variable Selection

For this study, the Y variable or outcome of interest is Jobs Retained. This variable is numeric and provides the number of jobs retained or preserved within each business as a result of receiving approval for a PPP loan. In other words, it provides valuable information about the impact of PPP loans on employment within each business. It provides a measure of the success of the program in preserving jobs during challenging economic times.

To answer the research question, several covariates of interest may be meaningful in explaining Jobs Retained. Below are the five X variables chosen for this study alongside definitions and a rationale for their inclusion in this research.

Business Type This variable categorizes each of the businesses in the dataset by its legal structure. Some examples of values that this variable takes include Corporation, Sole Proprietorship, and Limited Liability Company (LLC). This variable was chosen so that it may be studied whether there is a connection between the Loan Amount granted to a business, its legal structure and the number of jobs retained. This would help answer the research question by clarifying whether the effectiveness of a PPP loan, measured by Jobs Retained, is improved for businesses of a particular structure. Should there appear patterns, further study into the differences in policies and legal requirements across business types could be interesting to better elucidate the conditions under which PPP loans best preserve jobs.

Ethnicity This variable indicates the ethnicity of the owner of a business for each of the businesses in the dataset. Put simply, it is the ethnicity of the company representative receiving the PPP loan. This variable includes categories of values such as White, Hispanic, and Asian. The motivation behind including this variable in the study is to identify whether there is a connection between those businesses preserving the most jobs or receiving the largest amounts of support from the PPP and the ethnicity of the owner of the business. A study of this variable may delve into disparities or inequities in the distribution of PPP support across ethnicities and highlight areas of opportunity (i.e. does investing more in diverse owners increase job retention?) or saturation.

Gender The dataset provided two values for this variable; male-owned and female-owned. The rationale behind including this variable is in the same vein as the Ethnicity variable. By examining the distribution of loans across businesses owned by males and females, insights into potential disparities or variations in job retention can be gained. Understanding the impact of PPP loans on job preservation within the context of gender-owned businesses provides a lens through which policymakers and researchers can evaluate the inclusivity and equitable outcomes of the loan program. By comparing job retention rates between businesses owned by males and females, the analysis can reveal whether PPP loans had a differential impact on preserving jobs based on gender. This insight is crucial for understanding the program's effectiveness in supporting diverse business owners.

Loan Amount This variable provides the range of the loan approved. Analyzing it as a covariate is vital for identifying patterns, assessing the impact of loan sizes on job retention, and informing support strategies created for specific industries (NAICS), business types, or regions. The findings contribute to strategic decision-making, policy optimization, and the overall improvement of support programs for small businesses. By examining the distribution and range of approved loans, analysts gain insights into the scale of financial assistance offered to businesses. This variable is crucial for understanding the economic impact and capacity of the loan program to meet diverse business needs. The variation in loan amounts can signal differences in the financial requirements of businesses and shed light on the program's ability to address a spectrum of economic challenges faced across entities. Evaluating the distribution of loan amounts helps policymakers and researchers gauge the program's flexibility and effectiveness in meeting the financial demands of businesses across different sectors and sizes.

Lender Studying the variable "Lender" as a covariate on loan amount is essential to understanding the influence of different lenders in the PPP loan program. This analysis provides insights into whether certain lenders were more effective in facilitating loans for small businesses. By exploring patterns and variations, the study may help optimize lender selection strategies and inform potential enhancements in the lender participation process. The findings contribute to refining the PPP program and ensuring efficient collaboration with lenders for the benefit of small businesses. It may also help to identify criteria under which job retention is improved by comparing lender size, loan amount, and jobs retained. This variable may justify merging additional data to conduct a more thorough lender analysis.

2 Exploratory Data Analysis

In any data analysis project, especially in the field of economics, the quality of the conclusions of a study is influenced directly by the quality of the underlying data. This section describes the dataset used.

2.1 The Dataset

Raw datasets, such as the one loaded into this study, often include imperfections, inconsistencies, and missing values that must be addressed before meaningful EDA can be undertaken.

This phase, known as data cleaning, involves a series of steps that aim to ensure complete, consistent and accurate data. An examination of the presence of missing values in the dataset is completed to better understand the baseline number of observations.

The data used for this research is a PPP loan dataset sourced from the U.S. Department of the Treasury. The original dataset provides 661,218 observations.

The PPP loan dataset comprises data collected from small businesses and non-profit organizations that received financial support through the PPP. This program, administered not directly by the Small Business Administration (SBA) but through delegated lenders, was designed to help businesses keep their workforce employed during the COVID-19 pandemic. The dataset is released publicly to ensure transparency while protecting sensitive business information.

The level at which data is recorded is quite granular. It includes information at the Zip code level for businesses participating in the PPP in the United States in 2020. Each record in the dataset represents a single loan issued to a small business or non-profit organization. The level of observation is thus the individual loan as received by a specific entity. Data fields included in the dataset are:

- **Loan Range**: Indicates the approved loan amount range.
- **Business Name**: Name of the business that received the loan.
- **Address**: Physical address of the business.

- **City**: City where the business is located.
- **State**: State where the business is located.
- **Zip**: Zip code of the business location.
- **NAICS Code**: North American Industry Classification System code that indicates the industry type of the business.
- **Business Type**: Type of business (e.g., Corporation, Sole Proprietorship).
- **Race Ethnicity**: Race/ethnicity of the business owner, as voluntarily provided.
- **Gender**: Gender of the business owner, as voluntarily provided.

While aiming to be transparent, the dataset is structured to avoid revealing too much about the financial specifics of the businesses involved, focusing instead on general loan information and demographic data.

The inclusion of a business in this dataset indicates only that the loan was approved by a lender and guaranteed by the SBA. It does not imply eligibility for loan forgiveness or compliance with PPP rules, as all loans are subject to further SBA review.

Demographic details such as race and gender are included based on voluntary submission by borrowers. Approximately 75% of records do not contain this demographic information.

This dataset is critical for evaluating the effectiveness of the PPP in supporting the economic stability of small businesses during the pandemic. It also assists in assessing the equitable distribution of funds across diverse business demographics.

By understanding these details, users of the PPP loan dataset can more effectively analyze the reach and impact of the program while being aware of the limitations and scope of the data provided.

2.2 Data Cleaning

The first step taken to clean the data involved dropping drop missing data to ensure a complete dataframe is available for modelling and analysis later on. To maintain as much of the integrity of the original dataset as possible and keep as many observations as possible, missing values are dropped only in columns of interest (i.e. for variables chosen in Variable Selection.) After this step is complete, the dataframe contains 83,422 observations

Next, the data types of each column are inspected. Variables that are non-numeric and contain only two unique values, which include Gender, Veteran, and Non-Profit, are converted to binary categorical variables, or dummy variables. For example, Male Owned is converted to 0 and Not Male Owned (Female Owned) is converted to 1, Veteran converted to 0 and Non-Veteran converted to 1, and Non-Profit converted to 0 and Not Non-Profit converted to 1.

Date Approved is converted into datetime format for ease of sorting and grouping in subsequent analysis. Columns with numbered data such as Jobs Retained, NAICS Code, Zip, and Congressional District, are made numeric. Data that is written in all-caps or inconsistent casing is converted to sentence-case for readability. Loan Amount is converted to a categorical variable with bins numbered 0-4 representing different ranges of PPP loan amounts. All other columns are checked for consistency in their formatting. Finally, the dataframe is sorted by Loan Amount from high to low, Jobs Retained from high to low, and Date Approved from most to least recent.

Finally, to better capture the interaction between 'Ethnicity' and 'Gender', a new variable is created called Ethnicity-Gender.

2.3 Summary Statistics

In this section, summary statistics are shown for all variables of interest (i.e. those declared and described under Variable Selection.) The categorical summary table includes key insights into the distribution and frequency of various attributes, while the numerical summary table provides a statistical snapshot of the quantitative variables, which in this dataset includes just one variable, Jobs Retained. Note that all dollar amounts are in USD.

Accompanying each table is an interpretation of the implications of these statistics for a clearer understanding of each variable. For 'Jobs Retained', which is the only numerical variable, additional statistics, such as the mean and standard deviation, are included.

Variable	Value	Count	Percent Obs
Loan Amount Range	150k - 350k	49341	59.146
	350k - 1M	25333	30.367
	1M - 2M	5988	7.178
	2M - 5M	2340	2.805
	5M - 10M	420	0.503
Business Type	Corporation	40222	48.215
	Limited Liability Company	21775	26.102
	Subchapter S Corporation	16496	19.774
	Non-Profit Organization	1766	2.117
	Partnership	1052	1.261
	Limited Liability Partnership	849	1.018
	Sole Proprietorship	576	0.690
	Professional Association	300	0.360
	Cooperative	208	0.249
	Employee Stock Ownership Plan	52	0.062
	Non-Profit Childcare Center	47	0.056
	Self-Employed Individuals	36	0.043
	Trust	17	0.020
	Joint Venture	11	0.013
	Independent Contractors	10	0.012
	Tenant in Common	4	0.005
Ethnicity	White	70263	84.226
	Asian	5953	7.136
	Hispanic	5295	6.347
	Black	1484	1.779
	American Indian or Alaska Native	426	0.511
	Puerto Rican	1	0.001
Gender	Male	68178	81.727
	Female	15244	18.273
Lender	Huntington N.B	5041	6.043
	City N.B	2089	2.504
	KeyBank	1404	1.683
	East West	1319	1.581
	Truist	1037	1.243
	Other	72532	86.946

Table 1: Summary statistics of categorical variables of interest.

Loan Amount The count statistic correctly verifies that there are 83,422 observations of this variable. It is shown that there are 5 unique values, which also pertain to how the data were cleaned. Specifically, Loan Amount was turned into a categorical with 5 bins, each representing a loan amount range as shown in the table below. The provided data presents insights into the distribution of loans obtained through the PPP across various loan amount categories. One notable observation is that the majority of loans fall within the lower loan amount categories, with approximately 59% of loans categorized in the range of 150,000 to 350,000 USD.

This indicates that a significant portion of PPP loans are relatively smaller in size, which could suggest that smaller businesses, or those with fewer employees or lower operating expenses, are the primary recipients of PPP funds. As the loan amount categories increase, the number of loans decreases, with only a small percentage of loans falling within higher loan amount ranges such as 2-5 million USD and 5-10 million USD. This pattern might imply that larger businesses or those with greater financial needs are less represented among PPP loan recipients.

The distribution of loan amounts highlights the program's effectiveness in providing support to a broad spectrum of businesses, ranging from small enterprises to larger organizations, thereby aiding in maintaining employment and economic stability during times of uncertainty, such as the COVID-19 pandemic. The distribution of these bins can be found in the next section.

Business Type It is shown that there are 17 different business types found within the data. The distribution of business types among recipients of PPP loan approvals provides valuable insights into the composition of businesses that sought and received assistance during the COVID-19 pandemic. One striking observation is the prevalence of traditional corporate structures, with corporations representing nearly half of the approved loans. This suggests that larger, more established companies were active participants in the PPP program, possibly seeking support to sustain their operations and retain employees amidst economic uncertainties. Limited Liability Companies (LLCs) and Subchapter S Corporations also constitute a significant portion of the approved loans, indicating the importance of small and medium-sized enterprises (SMEs) in the program.

The presence of Non-Profit Organizations among loan recipients highlights the diverse range of entities that relied on PPP funding, including charitable organizations and community services providers. Additionally, the relatively small representation of sole proprietorships, partnerships, and self-employed individuals suggests potential challenges faced by smaller businesses in accessing PPP loans or a shift towards alternative forms of financial assistance. Overall, the distribution of business types among PPP loan recipients underscores the program's broad reach and its role in supporting businesses of various sizes and structures, contributing to economic resilience and job retention efforts during times of crisis

Ethnicity There are 6 unique values taken by Ethnicity in the data. The distribution of ethnicity among business owners participating in the PPP program offers insights into the economic landscape and highlights certain demographic trends. One notable observation is the predominance of white business owners among PPP loan recipients, comprising over 84% of the observed cases. This suggests that businesses owned by individuals of white ethnicity were disproportionately represented in the program, possibly reflecting pre-existing disparities in access to financial resources or institutional biases in lending practices. Conversely, the relatively smaller shares of Asian, Hispanic, Black or African American, American Indian or Alaska Native, and Puerto Rican business owners indicate underrepresentation within the program.

This discrepancy may underscore systemic challenges faced by minority-owned businesses in accessing government assistance programs, potentially exacerbating existing disparities in economic opportunity and wealth accumulation. Addressing these disparities and ensuring equitable access to financial support mechanisms like the PPP is essential for promoting economic inclusion and fostering a more resilient and diverse entrepreneurial ecosystem. Moreover, understanding the demographic composition of PPP loan recipients can inform targeted policy interventions aimed at supporting underrepresented communities and facilitating inclusive economic recovery efforts in the wake of crises such as the COVID-19 pandemic.

Gender The original dataset labelled the values of this variable as "Male Owned" and "Female Owned." For future analysis and modelling, this variable was converted to a dummy variable. The distribution of gender among business owners participating in the PPP program provides important insights into gender disparities in entrepreneurship and access to financial support during times of economic crisis. In this dataset, male-owned businesses significantly outnumber female-owned businesses among PPP loan recipients, with males representing over 81% of observed cases compared to females at around 18%. This observation likely reflects broader gender disparities in business ownership and access to capital, where male entrepreneurs historically have had greater access to resources and opportunities compared to their female counterparts.

Understanding these gender disparities is crucial in assessing the effectiveness of the PPP in minimizing unemployment. Research suggests that female-owned businesses have been disproportionately affected by the economic impacts of the COVID-19 pandemic. Female entrepreneurs often operate in sectors such as retail, hospitality, and services, which have been severely impacted by lockdowns and social distancing measures. Additionally, women-owned businesses tend to be smaller in scale and have less access to capital, making them more vulnerable to economic shocks.

Given the significant presence of male-owned businesses among PPP loan recipients, there is a need to ensure that the program effectively addresses the needs of female entrepreneurs and supports their businesses during times of crisis. This includes targeted outreach efforts to increase awareness and accessibility of PPP loans among female-owned businesses, as well as measures to address underlying barriers such as gender bias in lending practices and access to networks and mentorship opportunities.

By addressing gender disparities in access to financial support and fostering an inclusive entrepreneurial ecosystem, the PPP can play a more effective role in minimizing unemployment by supporting the sustainability and resilience of businesses owned by women, ultimately contributing to broader economic recovery efforts.

Lender The statistics provide insight into the great diversity of lenders in this dataset. With over 3000 unique lenders, banks and other financial providers participating in the

PPP were numerous. It was found that the most frequent lender was The Huntington National Bank, lending to approximately 6% of all businesses. The distribution of loans among various lenders in the PPP program sheds light on how different banks and financial institutions participated during the COVID-19 pandemic. The table reveals that while there are over 3,000 lenders involved in disbursing PPP loans, a significant concentration of loans is observed among the top five lenders. Notably, The Huntington National Bank, City National Bank, KeyBank National Association, East West Bank, and Truist Bank collectively account for a substantial portion of the loans, with their shares ranging from around 1% to over 6% of total loan disbursements.

This concentration of loans among a few major banks and financial institutions suggests that larger and more established lenders played a crucial role in facilitating access to PPP funds for businesses. These lenders likely had the resources, infrastructure, and expertise to efficiently process and distribute loans to a wide range of businesses. Additionally, their prominence may reflect existing relationships with business clients, as well as their ability to quickly adapt to the requirements and guidelines set forth by the PPP program.

The "Other" category, which includes loans disbursed by lenders beyond the top five, constitutes the majority of loan disbursements, highlighting the diverse landscape of participating lenders. While individual lenders in this category may have disbursed fewer loans compared to the top five, collectively, they have played a significant role in providing crucial financial assistance to businesses during the pandemic.

Analyzing the distribution of loans among different lenders in the PPP program provides insights into the program's effectiveness in minimizing unemployment. Understanding which banks and financial institutions played a significant role in disbursing loans can indicate how effectively the program reached businesses in need of financial support to retain their employees. By examining the distribution of loans among lenders, policymakers and researchers can assess how effectively the PPP reached businesses of different sizes and sectors, providing valuable insights into its overall impact on job retention and unemployment mitigation efforts.

The next table includes summary statistics for Jobs Retained, which is a continuous variable.

Count	Mean	Std Dev	Min	25%	Median	75%	Max
83422	49.477	62.906	0	18	30	54	500

Table 2: Summary statistics for Jobs Retained.

This variable can be explained as the number of jobs that were preserved as a result of obtaining approval for a PPP loan. It is used as a measure of PPP loan efficacy; if the jobs retained by a business improved due to the loan, this can help identify criteria surrounding the success that prove the effectiveness of the policy.

By scrutinizing the distribution and statistical measures associated with this variable, we gain valuable insights into the effectiveness of the PPP policy. The average number of retained jobs across all businesses in the dataset stands at 49.48, with a notable standard deviation of 62.91. This high standard deviation underscores the substantial variability in job retention levels among businesses. The wide-ranging values, stretching from 0 to 500 jobs, highlight the heterogeneous nature of job retention outcomes. Further analysis reveals that half of the businesses retained 30 jobs or fewer, while 75% retained 54 jobs or fewer. This distribution underscores the diverse impact of PPP loans on job preservation, emphasizing the necessity for a nuanced understanding of its effectiveness. Such insights into the distribution of job retention metrics provide crucial context for evaluating the PPP's role in mitigating unemployment and fostering economic resilience.

2.4 Plots and Histograms

In this section, a systematic exploration of the dataset is detailed through a variety of visual representations. Using a distinct dataframe grouped by date, some of the plots below aim to elucidate correlations between the outcome, loan amount, and jobs retained. A new variable, 'Ethnicity-Gender,' is introduced to provide a nuanced perspective by amalgamating categorical attributes. Subsequent bivariate analyses delve into relationships between critical variables, encompassing examinations of jobs retained and loan amount, loan amount and business type, jobs retained and business type, gender and ethnicity, and loan amount and ethnicity. The economic intuition behind each plot is provided along with a rationale for the selection of variables in certain plots. These visualizations contribute substantially to enhancing comprehension of the dataset and, as will be explained below, are directly aligned with the overarching research question.

2.4.1 Numerical Data Distribution

As mentioned briefly in the previous section, a histogram is used to better interpret the variable Jobs Retained and visualize the distribution of the numerical data.

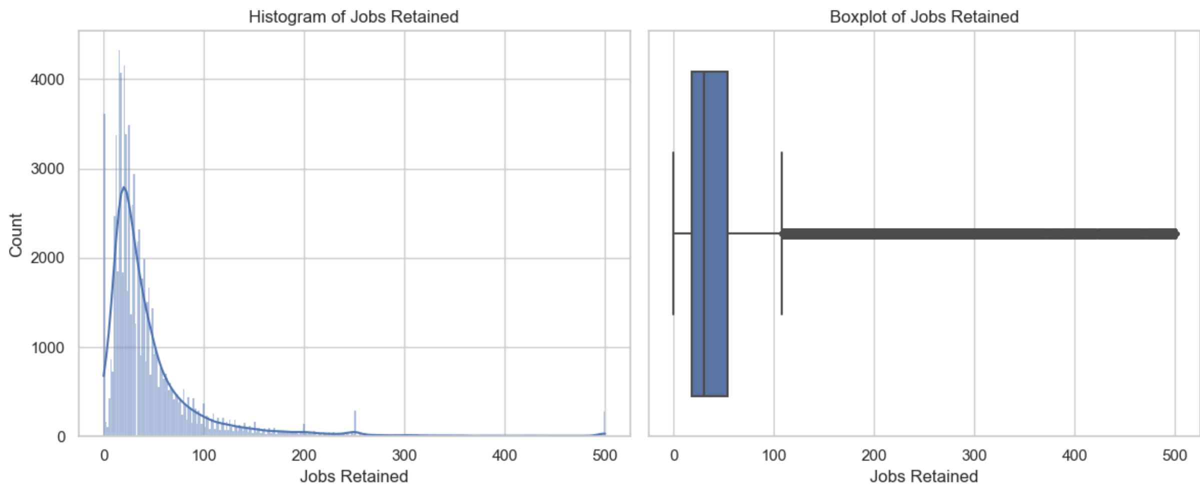

Figure 1: Distribution of Jobs Retained.

The histogram on the left illustrates the distribution of Jobs Retained and showcases a distinct right-skewness in the data. Statistically, this means that the mean is being pulled to the right by the presence of a few high values on the right side of the distribution. Practically, this means that the average number of jobs retained is being influenced by a few businesses with very high job retention, influencing the standard deviation.

This fact is seen more clearly in the boxplot on the right. The box indicates the interquartile range (range between Q1 and Q3) while its length represents the spread of the middle 50% of the data. The whiskers extend from the box to the minimum and maximum values within a certain range, which is typically calculated as 1.5 times the IQR. Any data points that lie beyond the whiskers are considered potential outliers. As is shown, most outliers according to this measure do indeed lie to the right (e.g. a few businesses with higher job retention). Given the volume of outliers, they are not removed at this stage of the project to prevent loss of data in subsequent stages.

The variable, 'Jobs Retained,' was chosen for this specific plot to determine whether further exploration would be necessary of industries, business types, or regions where job preservation was more pronounced. The right-skewness indicates that a few businesses significantly impact average job retention, potentially reflecting unique characteristics in certain sectors or business types. Thus, these visualizations directly contribute to understanding PPP loan impact by guiding further analysis of other covariates.

2.4.2 Bivariate Analysis

This section explores the relationships and interactions between pairs of variables identified under Variable Selection. By examining the joint distribution of key variables, patterns and potential insights are found. The analyses include visualizations such as cross-tabulations and stacked bar charts to share a nuanced understanding of how variables may influence each other within the dataset.

Jobs Retained vs. Loan Amount First, a variation of the boxplot allows visualization of the distribution of Jobs Retained for different categories by Loan Amount.

There are two important aspects of the boxplot in Figure 2. Firstly, the differences in height of each of the boxes indicate differences in the spread of the number of jobs retained for each loan amount range. Secondly, the differences in the medians indicate differences in the central tendency, or median, of the number of jobs retained across loan amounts. The higher position of the right-most box suggests that a large portion of the jobs retained are within the highest loan amount category. Importantly, this implies that businesses with higher loan amounts tend to retain more jobs, suggesting a positive correlation between the two variables. This may indicate that as the loan amount increases, there is a tendency for a higher number of jobs to be retained.

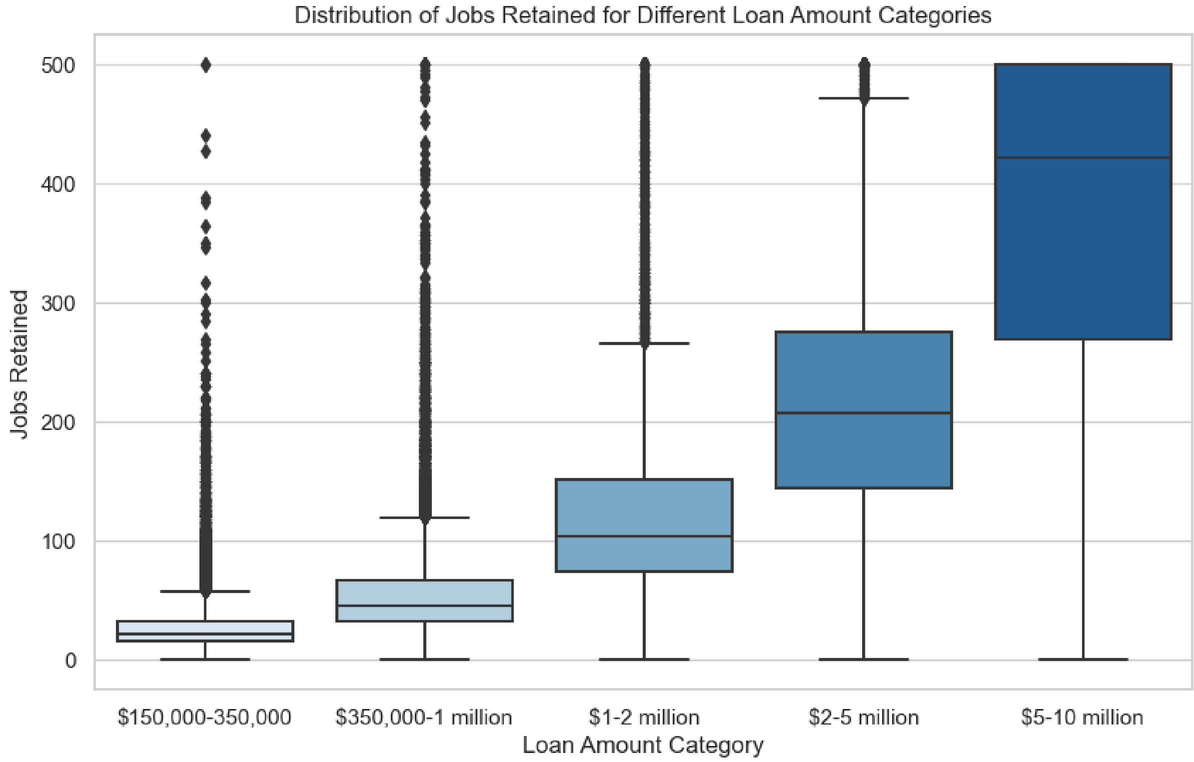

Figure 2: Distribution of Jobs Retained by Loan Amount Range.

Loan Amount vs. Business Type Next, to analyse the relationship between two categorical variables, namely Loan Amount and Business Type, we plot a stacked bar chart. This shows the frequency distribution of the intersection of different ranges of the Loan Amount variable and types of business. Figure 3 shows that the lowest Loan Amount range, Category 0, has the highest number of loans across most business types, especially in Corporations, Limited Liability Companies (LLCs), and Sole Proprietorships.

Categories 1 and 2 of Loan Amount also show significant counts, indicating diverse loan amounts for various business types. Categories 3 and 4 have lower counts, suggesting fewer loans in these higher amount ranges.

Corporations have the highest loan counts across all loan amount categories, indicating a prevalent involvement of corporations in obtaining PPP loans (verifying previous findings). LLCs are also prominent, with substantial counts in most loan amount categories. Non-Profit Organizations have a noticeable presence, particularly in categories 0, 1, and 2. Sole Proprietorships have significant counts, especially in category 0.

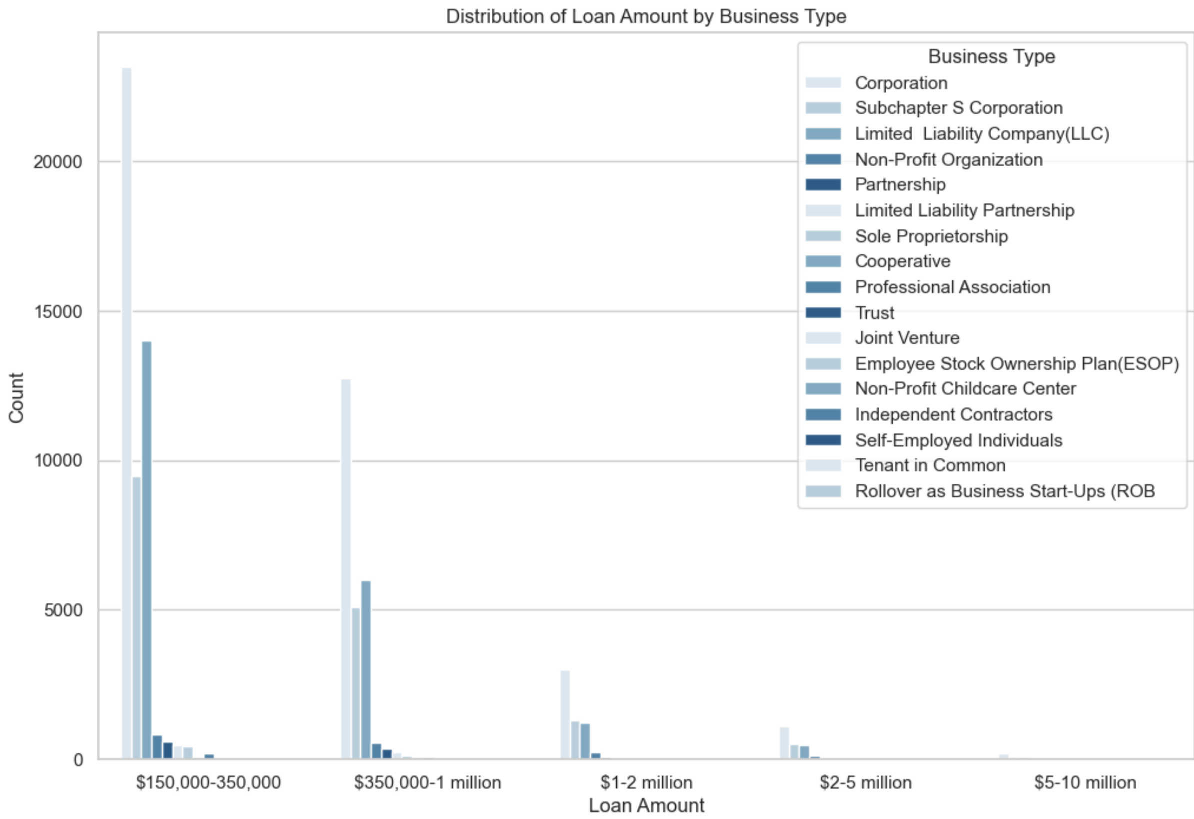

Figure 3: Distribution of Loan Amount Range by Business Type.

In summary, most businesses securing PPP loans fall into the categories of Corporations, LLCs, and Non-Profit Organizations, which confirms previous findings. It is also found that the lowest loan amount range caters to a diverse set of businesses, which also confirms the finding that this is the most frequent loan category. Similarly, smaller loan amounts (categories 1 and 2) are widespread, indicating the PPP program's reach to various business scales. Noticeably, Non-Profit Childcare Centers are more prevalent in category 0, potentially reflecting the need for support in the childcare sector. These insights help to understand the distribution of PPP loans across different business types and loan amount categories, providing valuable information for further analysis.

2.4.3 Jobs Retained vs. Business Type

To understand the relationship between Jobs Retained and Business Type, we plot a bar plot with error bars. This plot calculates the mean jobs retained for each business type. This means that, for example, Corporates retain just under 50 jobs on average.

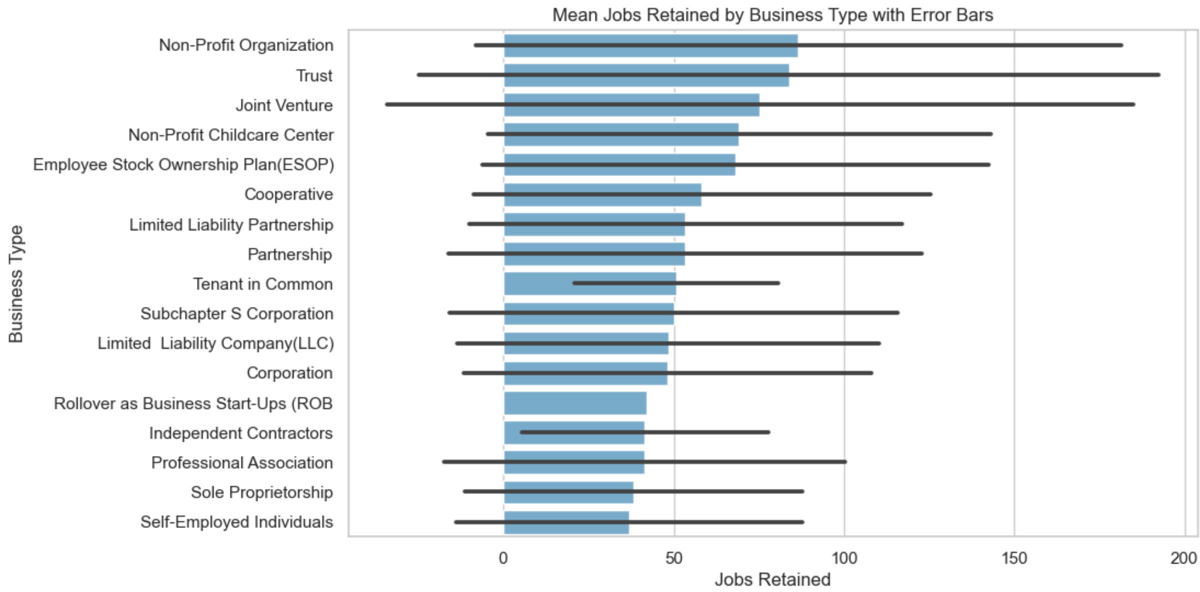

Figure 4: Average number of Jobs Retained by Business Type.

Figure 4 shows that most business types retain around 50 jobs, which is approximately equal to the average Job Retention statistic found earlier. However, the error bars show great variation. Each error bar dictates how much individual data points typically deviate from the mean. Most error bars in this plot are quite long, which suggests (and verifies from earlier) higher variability in the number of jobs retained within that category of business type. This highlights that the mean value may not be highly representative of typical job retention.

It is interesting to note that Non-Profits seem to have the highest average number of jobs retained of any business type. Additionally, Self-Employed Individuals, Independent Contractors, and Sole Proprietorships, which share similarities in their business structures as they are all forms of businesses where an individual operates and manages the business independently, tend to have the lowest job retention, on average.

2.4.4 Ethnicity-Gender vs. Loan Amount

Using the variable created, Ethnicity-Gender, the relationship between Ethnicity-Gender and Loan Amount is studied using a clustered bar chart.

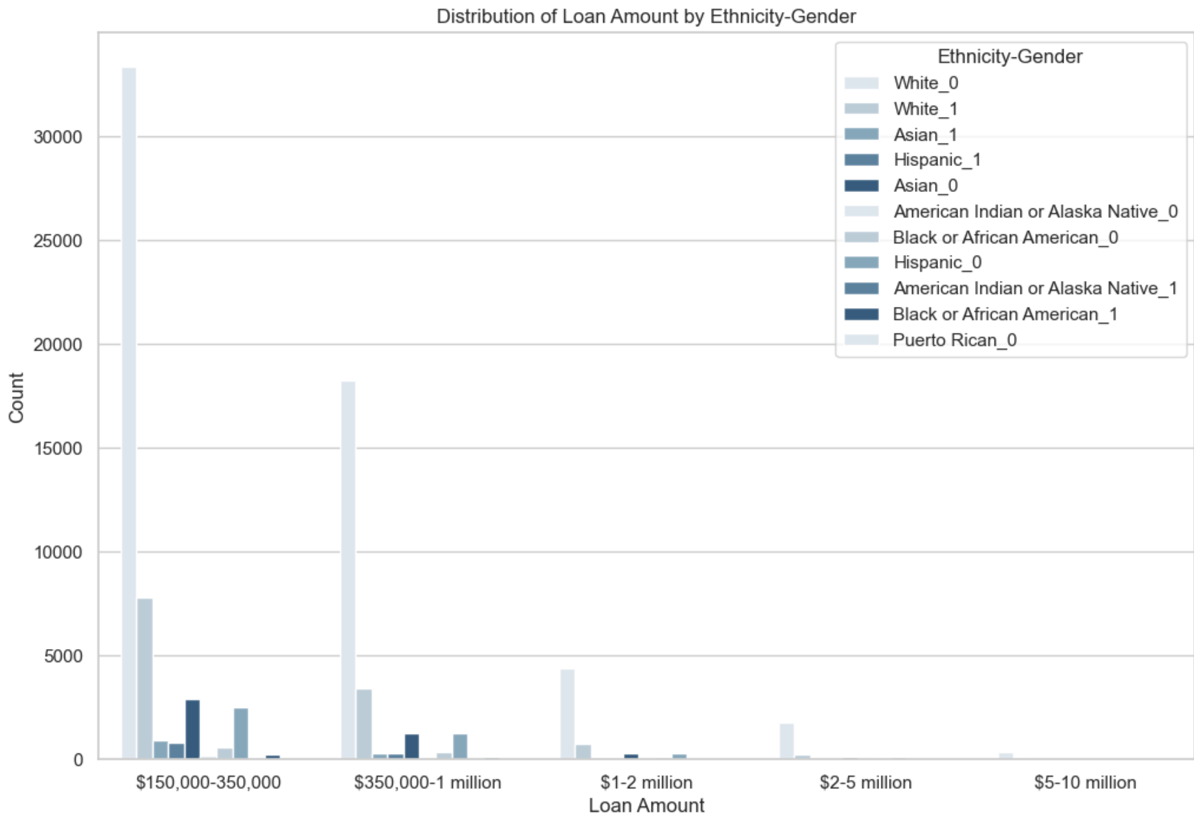

Figure 5: Distribution of Loan Amount Range by Ethnicity-Gender.

It is shown that the owner archetype receiving the most loans, across all loan amount categories, is dominantly White_0 or owners who are both White and Male.

2.5 Time Series Analysis by Approval Month

A point of interest in this analysis is whether there is a relationship between the Loan Amount and the Date when the PPP loan was approved. Exploring this relationship might indicate underlying trends or seasonality, which could be incredibly interesting in understanding patterns in job retention throughout a given year. The relationship between the month data and loan amounts is plotted using a line graph.

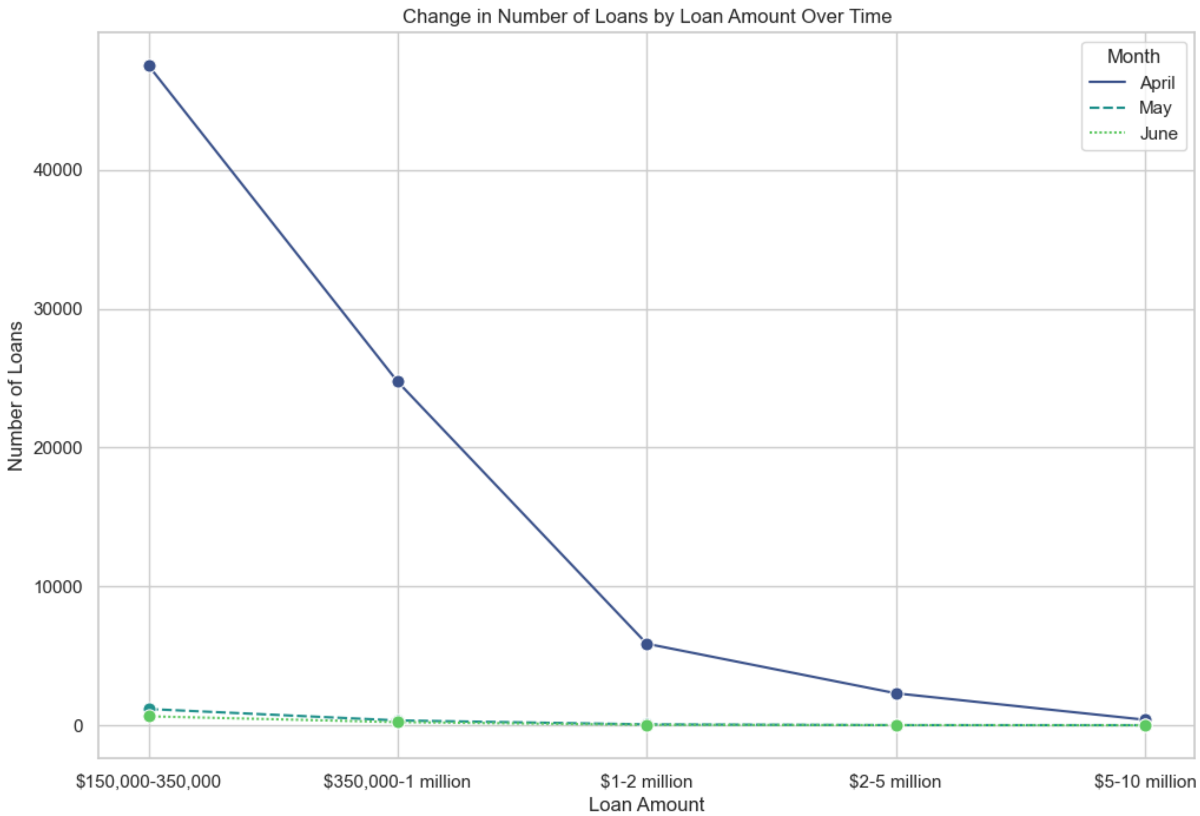

Figure 6: Change in the number of loans provided by Loan Amount Range over the three months covered by the dataset.

There is a huge drop in the number of loans overall from April to May, especially in the lowest loan amount category. The change in the number of loans provided across categories is negligible from May to June in comparison to the change from April to May, as is shown by the near overlap of the lines representing May and June. From this, it is learned that the number of loans provided did indeed change drastically over time and was not consistent over the three months that the dataset provides. This finding leads to the question of how Job Retention changed over time.

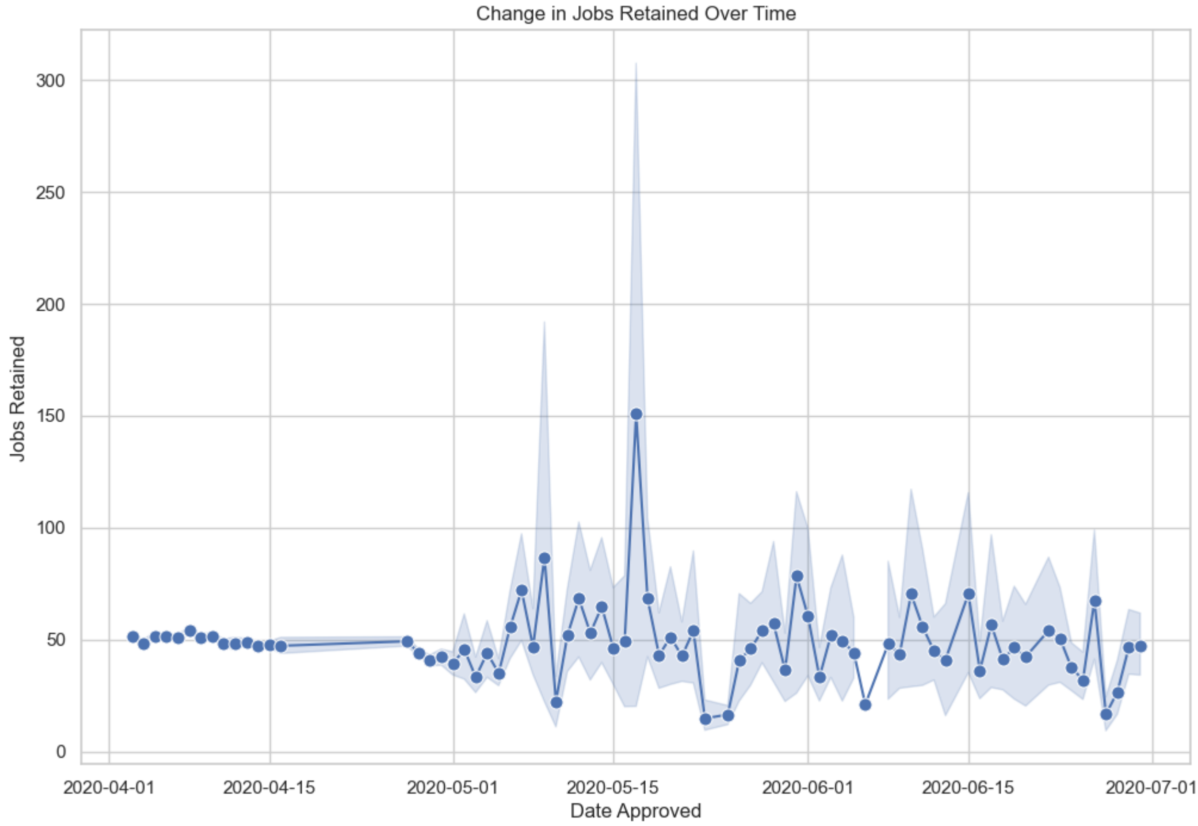

Figure 7: Change in the number of Jobs Retained over the three months covered by the dataset.

The line graph depicts Jobs Retained over Date Approved. This offers valuable insights into changes in employment associated with the loan approval timeline. Notably, the line appears relatively stable during April, suggesting a consistent level of jobs retained throughout this period. Connecting this to the previous figure showing Number of Loans, it is interesting that the number of Jobs Retained did not fluctuate as much during the month where the number of loans provided was highest. This might indicate a lag-time of the impact of the loan, where the date it was approved is not the same as the date upon which the loan is used by the business or makes an impact on the business's job retention.

There exists a noticeable increase in variability towards the end of April and early May, indicating a period of fluctuation in job retention. This might be the delayed impact of the loans that were more generously given throughout April than in May or June. The most striking observation is the sharp spike in mid-May, where the number of jobs retained reaches its peak. This spike suggests a specific event that significantly influenced job retention during this time frame, perhaps the collective "cashing-in" of PPP loans of businesses that received support in April. Subsequently, there is a notable decline in jobs retained, dropping below the levels observed in April by the end of May. The fluctuation around similar values in June, particularly around 50 jobs, indicates a certain level of stability but at a lower magnitude compared to the peak observed in mid-May. However, the large confidence intervals outside of April indicate that the spike may not be anything significant.

The economic intuition behind these observations could be multifaceted. The stability in April may indicate the initial impact of the loans on job retention. The increased variation at the end of April and early May might be attributed to external economic factors or policy changes affecting businesses and their ability to retain jobs. The spike in mid-May could be linked to a specific economic event, such as the implementation of supportive policies or a surge in business activities. The subsequent drop may signify a temporary boost in job retention rather than a sustained trend. The fluctuation in June, around levels observed in early May, may suggest a new equilibrium in job retention after the initial impact of the loans. This analysis helps guide where the literature review might be most interesting for this study. The line graph directly addresses the question of how jobs retained change concerning the date of loan approvals. It provides a visual representation of the temporal patterns in job retention, allowing for the identification of critical timeframes and trends associated with the loan approval and use process.

3 Integrating Macro Data

In order to provide a visualization of the main message of this paper, unemployment rate data and layoff rate data are merged with the original data described thus far. These new data are relevant and interesting because this study hopes to understand the effectiveness of the PPP in reducing unemployment during times of economic hardship. In addition to using the Jobs Retained data that already exists, merging with unemployment rates and layoff data will help to strengthen the analysis by providing additional detail at the state level.

Both the unemployment rate data and the layoff data, at the national level, are sourced from Federal Reserve Economic Data (FRED). The FRED database is maintained by the Research Division of the Federal Reserve Bank of St. Louis. To emphasize, merging with FRED's comprehensive datasets, the study gains the contextual backdrop necessary for a thorough analysis. The unemployment rate data provides a macroeconomic perspective, allowing for a comparison between jobs retained through PPP loans and the overall employment situation in different regions. Layoff data contributes essential information about the severity of job losses, offering insights into the broader economic impact. The state-level granularity afforded by the FRED data adds a nuanced examination of variations in economic conditions, enabling the identification of regional patterns and trends.

Additionally, the inclusion of these variables allows for a comprehensive evaluation of the PPP's effectiveness, considering not only job retention but also its role in mitigating widespread layoffs. As the study delves into various business-related variables, understanding how they interact with unemployment and layoffs provides a holistic view of the program's impact. Ultimately, the incorporation of FRED data enhances the statistical power of the analysis, offering valuable economic insights that can inform future policy decisions and recommendations.

Data is merged from Statistics Canada on the industry names for NAICS Codes. This is added so that a more comprehensive and easily understandable industry analysis may be conducted using the NAICS Codes in the existing data. The integration of monthly unemployment rates by state from the US Bureau of Statistics offers a dynamic and state-specific perspective on employment trends. Unlike aggregate national unemployment data, the inclusion of state-level unemployment rates allows for a more nuanced analysis of regional variations and economic conditions. This additional layer of information is crucial for understanding how employment dynamics vary across different states, contributing to a more detailed and localized assessment of the dataset.

Additionally, the merging of data from the US Census Bureau on population by state in 2020 is instrumental in creating per capita measures. This demographic information provides the necessary population context for calculating per capita ratios, enabling a fair and normalized comparison across states. The per capita measures offer insights into the relative impact of various economic indicators, such as jobs retained or unemployment rates, on individual residents within each state. Overall, the integration of population data enhances the precision and relevance of the analysis by accounting for differences in state population sizes. Below we include plots to begin analysis using the newly merged data.

3.1 National Layoff Rate vs. Average Number of Jobs Retained

Comparing the monthly layoff rate to the average number of jobs retained due to the approval for PPP loans provides a focused examination of the program's impact on employment stability. The Paycheck Protection Program (PPP) was designed to assist businesses in retaining their workforce during economic challenges, particularly amid the uncertainties caused by events such as the COVID-19 pandemic. Understanding how the monthly layoff rate aligns with the average number of jobs retained through PPP loans is crucial for evaluating the program's effectiveness in achieving its intended goals.

The monthly layoff rate, in this context, becomes a dynamic metric reflecting the ebb and flow of employment disruptions within specific periods. Contrasting this with the average number of jobs retained through PPP loans offers a direct correlation to the program's ability to counterbalance layoff trends. A lower layoff rate coupled with a higher average number of jobs retained suggests a successful intervention by PPP loans in stabilizing employment, demonstrating the program's efficacy in protecting jobs during challenging economic phases.

This analysis not only serves as an evaluation of the PPP's impact on job retention but also provides actionable insights for policymakers and businesses. It helps identify whether PPP loans are effectively mitigating layoffs and fostering a resilient job market. Additionally, it offers guidance on potential areas for improvement or adjustments in the program to optimize its contribution to overall employment stability, and ultimately, provides valuable data in answering the research question.

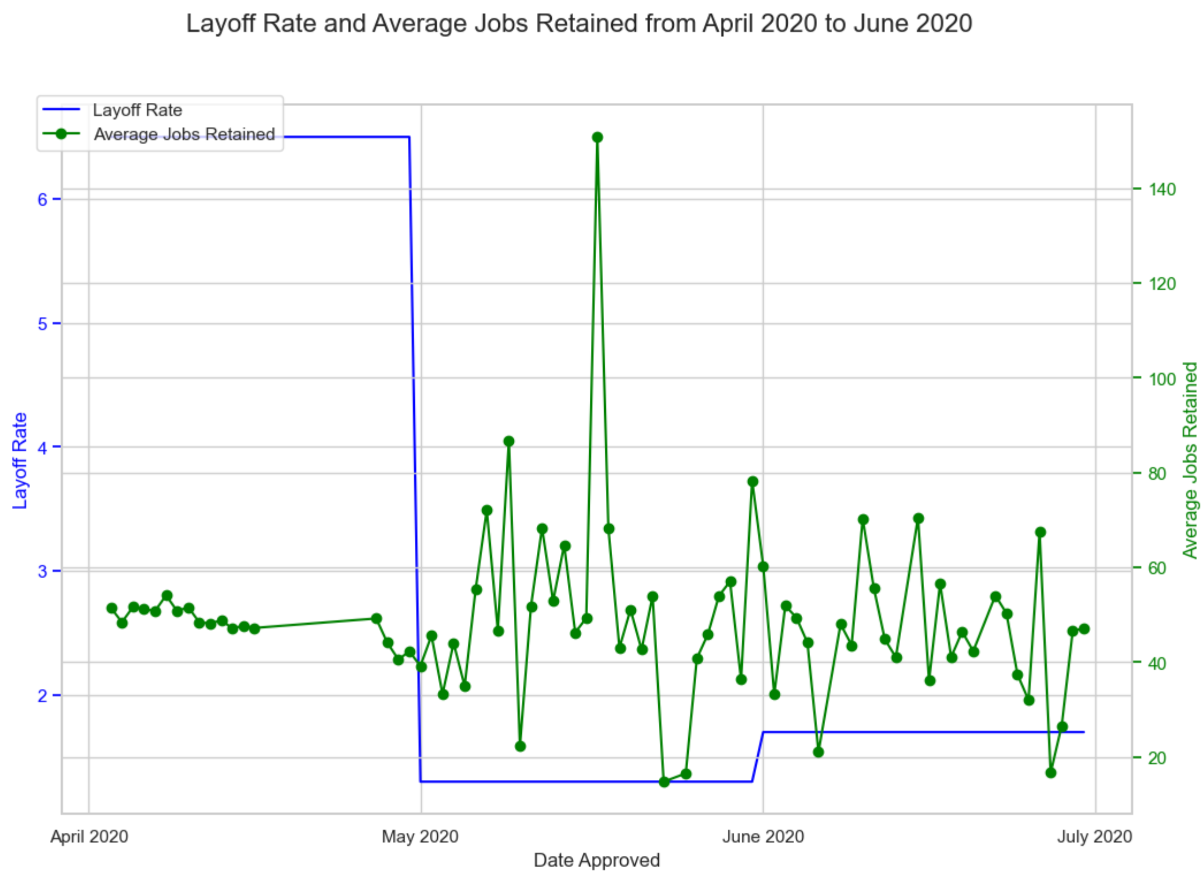

Figure 8: Change in the average number of Jobs Retained from April to June 2020 compared with the change in the national layoff rate.

From this plot, it is shown that the layoff rate drops drastically between April 2020 and May 2020. During this time, the average number of jobs retained remains relatively constant at about 50 per month. The drastic drop in the layoff rate suggests a potential positive impact of economic interventions such as the PPP during the early stages of the COVID-19 pandemic. This decline in layoffs aligns with the program's objective of preserving jobs and supporting businesses through financial assistance.

The stability in the average number of jobs retained around 50 per month during this period indicates that businesses, with the aid of PPP loans, were successful in maintaining a consistent level of employment. This could signify the effectiveness of the PPP in providing financial relief to businesses, enabling them to retain a steady workforce despite the economic uncertainties and challenges posed by the pandemic.

From May 2020 to June 2020, the layoff rate changes minimally. At the same time, the average number of jobs retained fluctuates greatly, from below 20 to above 150. This variability may reflect varying responses among businesses to the evolving economic landscape. Some entities may have experienced challenges in sustaining employment, leading to fluctuations in the average number of retained jobs, even as the layoff rate remains stable

In the context of the PPP, this dynamic pattern suggests that while the program might have initially stabilized employment, subsequent months brought diverse challenges and responses within the business landscape. Understanding the economic significance of these fluctuations is crucial for policymakers to refine interventions like the PPP to address uncertainties in the labour market. Adjustments to the program or complementary measures may be considered to ensure sustained economic recovery.

3.2 National Unemployment Rate vs. Average Number of Jobs Retained

The PPP was implemented as a crucial economic relief measure to counteract the adverse effects of events like the COVID-19 pandemic, with a primary focus on preserving jobs within businesses. The next plot studies the monthly unemployment rate and the average number of jobs retained through PPP to investigate the program's efficacy in stabilizing employment levels during challenging economic periods.

The monthly unemployment rate serves as a key indicator of the overall economic health and labour market dynamics. It reflects the proportion of the workforce actively seeking employment but unable to secure jobs, offering a macro-level perspective on job availability. Contrasting this with the average number of jobs retained through PPP provides a micro-level insight into the specific impact of the program on individual businesses and industries. A significant decrease in the monthly unemployment rate, accompanied by a consistent or increasing average number of jobs retained, suggests a positive correlation with the PPP's intended outcomes. This scenario signifies that businesses, supported by PPP loans, are successfully weathering economic uncertainties and maintaining their workforce. It highlights the program's effectiveness in mitigating unemployment risks and contributing to overall employment stability.

Conversely, if the unemployment rate remains high despite PPP implementation, it prompts a closer examination of potential challenges or gaps in the program's impact. Understanding these nuances is vital for policymakers, as it informs the refinement of economic relief strategies. It allows for targeted adjustments to the PPP or the introduction of complementary measures to address specific areas where employment stabilization may require additional support.

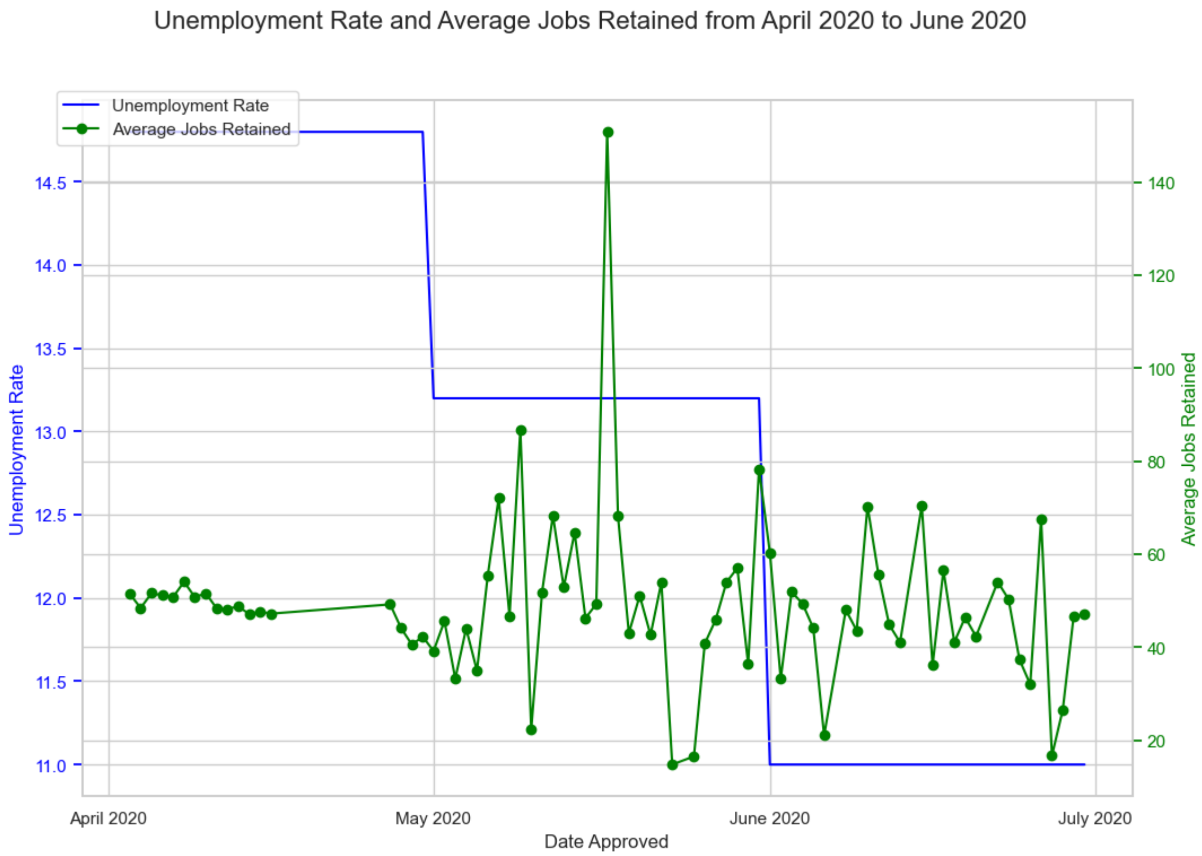

Figure 9: Change in the average number of Jobs Retained from April to June 2020 compared with the change in the national unemployment rate.

The consistent drop in unemployment rates from April to June, coupled with the highest average number of jobs retained in May, suggests a positive impact of the Paycheck Protection Program (PPP) on the overall national job security landscape. This trend aligns with the intended goals of the PPP, indicating its potential contribution to stabilizing employment and mitigating the adverse effects of economic uncertainties, particularly during the challenging period of the COVID-19 pandemic.

The decline in unemployment rates signals that the PPP may have played a role in supporting businesses and preserving jobs across the nation. As businesses received financial assistance through PPP loans, they likely had the means to retain their workforce, contributing to the observed reduction in unemployment. The program's success in fostering job security is reflected in the downward trajectory of unemployment rates over the period.

3.3 Industry Analysis

Different industries experience varied impacts during economic downturns or crises. Understanding the distribution of retained jobs by industry helps identify sectors that are more resilient or vulnerable. If certain industries show a higher proportion of retained jobs, it indicates that the implemented policies have been successful in supporting those sectors. Conversely, lower job retention in specific industries may signal the need for targeted interventions.

Policymakers can use the data to formulate targeted plans for economic recovery. By identifying industries that have retained a significant number of jobs, they can prioritize support and stimulus measures for sectors that play a crucial role in overall employment stability and economic growth.

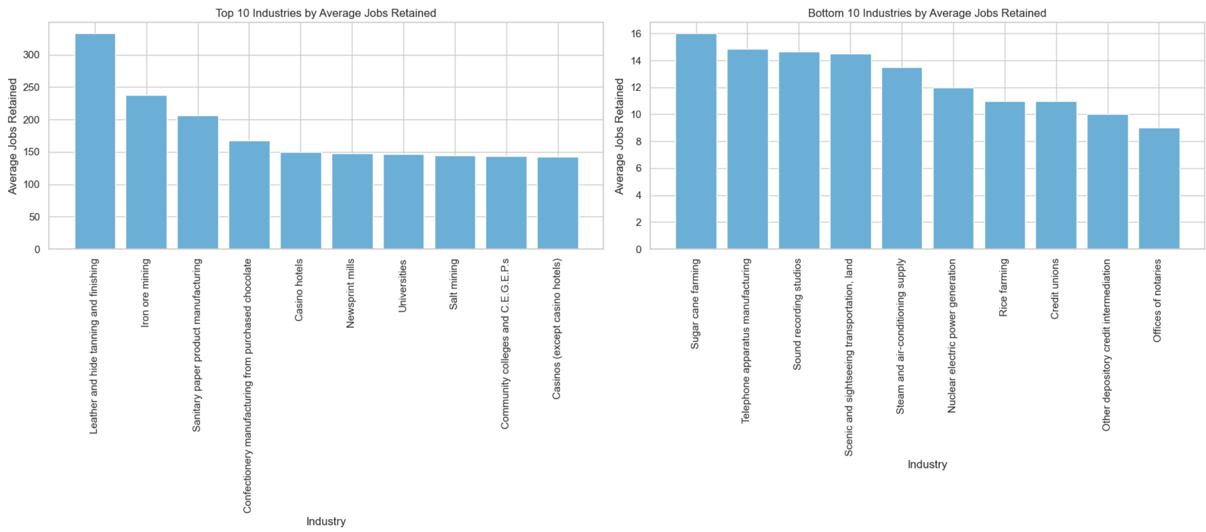

Figure 10: Top 10 and Bottom 10 industries by average number of Jobs Retained due to the approval of a PPP loan.

The industries with the highest jobs retained through PPP loans exhibit a diverse range of sectors, showcasing the program's broad impact on preserving employment across various segments of the economy. Notable industries include leather and hide tanning and finishing, iron ore mining, sanitary paper product manufacturing, confectionery manufacturing, casino hotels, newsprint, universities, salt mining, community colleges, and casinos. This diversity suggests that the PPP has been successful in providing support not only to traditional sectors but also to industries with unique characteristics, such as those related to hospitality, education, and manufacturing. This diversity aligns with the program's intention to provide a lifeline to a wide range of businesses, ensuring that jobs are retained in both large-scale manufacturing and service-oriented industries.

The lower number of jobs retained in specific industries may indicate the unique challenges these sectors encountered in leveraging PPP support. Sectors like sugar cane farming and nuclear power generation might have faced structural or operational difficulties that impacted their ability to retain jobs despite PPP assistance. The inclusion of telephone manufacturing and credit intermediation suggests that industries experiencing significant shifts, such as technological changes or alterations in financial services, may have struggled to retain jobs even with PPP support.

4 Visualizing and Evaluating PPP Outcomes

To determine the effectiveness of the PPP as a policy, we consider its ability to preserve jobs for those most in need during times of economic hardship, such as the COVID-19 pandemic during which it was introduced. This study aims to identify whether factors such as specific business types, owner traits, loan amounts, or lenders play a role in determining the success of PPP loans in job retention for small businesses. This study also seeks to understand if the loans actually targeted the places that needed the money most or not by incorporating data on layoffs, unemployment rate and monthly local GDP growth.

4.1 PPP Relief Targeting

To describe the main message visually, a plot is created to contrast the change in Jobs Retained per capita and Unemployment Rate per capita at the state level, across the period of April 2020 to June 2020 (i.e. the period covered in the dataset).

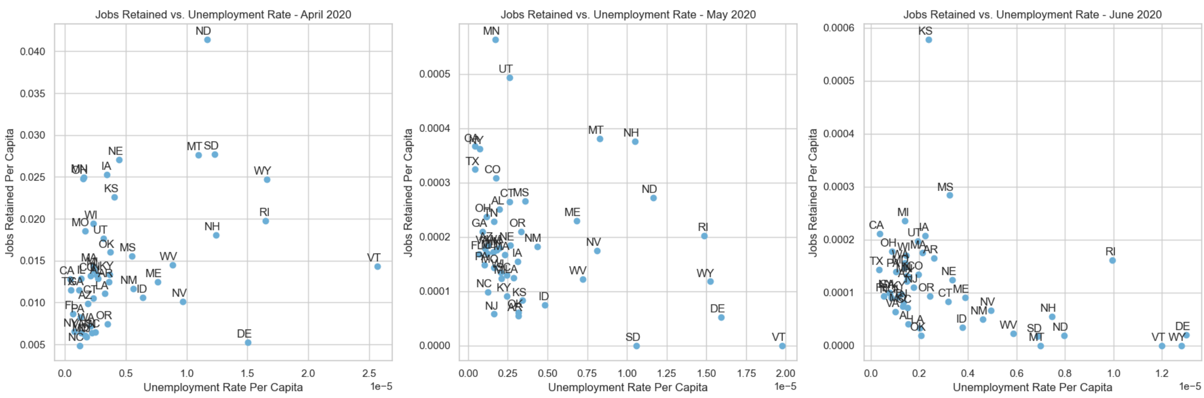

Figure 11: States exhibiting the most need for the PPP, measured by high unemployment rate per capita, tend to see moderate to low jobs retained per capita while those states with low unemployment per capita tend to see higher jobs retained per capita.

Importantly, the difference in units across plots is intentional. Using the same axis across plots rendered the differences illegible. This indicates an important feature; the unemployment rate per capita changed drastically across months. This is corroborated by the plot provided earlier showing the change in national unemployment over the period of interest. By presenting the changes graphically, the plot provides an immediate and intuitive understanding of the PPP's impact on preserving jobs in the face of economic challenges posed by the COVID-19 pandemic.

The main message of the paper revolves around assessing the effectiveness of the PPP in reducing unemployment during economic hardships. The contrast plot directly supports this objective by visually comparing the two crucial indicators – jobs retained and unemployment rates – at the state level. It would be most important to notice that states with the highest unemployment per capita, meaning areas with the largest need for the PPP program, also have the highest job retention. This would indicate the effectiveness of the PPP program; are those most in need receiving support?

The examination of state-level data spanning the tumultuous period from April to June 2020 reveals notable variations in unemployment rates and jobs retained, offering a nuanced perspective on regional economic dynamics. Vermont, Delaware, Rhode Island, and Wyoming emerge consistently in the top 4 as states with the highest unemployment rate per capita. Of these states, Wyoming and Rhode Island tend to achieve higher jobs retained per capita than Vermont and Delaware over the period. States tending to achieve the highest number of jobs retained per capita include North Dakota, New Hampshire and Montana over the months of April and May. Each of these states dropped significantly in their jobs retained in June.

States like North Carolina, California and New Jersey tend to populate the bottom left of the plots, meaning they experience low unemployment rates per capita and low job retention per capita. Each plot includes a rather populous bottom left but we also notice upwards spread on the left. This indicates that the PPP tends to improve job retention per capita in states where the unemployment rate per capita is already relatively low.

In contrast, each plot is devoid of points in the top right. This would indicate that states experiencing high unemployment per capita also were able to retain more jobs per capita due to the PPP. Points in the top right would thus indicate the efficacy of the policy. It can be noted that most states that exhibit higher unemployment, at least in April and May, experience at minimum moderate job retention. By June, however, these states begin to experience less job retention. Notably, the absence of states concurrently experiencing both high unemployment and high job retention suggests a complex interplay of economic factors influencing regional outcomes.

Overall, these findings pave the way for a deeper exploration of the policies and contextual factors shaping these dynamics, providing crucial insights for policymakers and researchers alike. States facing persistent challenges with high unemployment rates and lower job retention may indicate potential gaps or limitations in the program's reach or effectiveness in certain regions. The clustering of states with low unemployment rates but high job retention highlights potential areas for improvement in the program to address challenges faced by those states in greater need of the PPP

4.2 State Distribution of Loan Amount Per Capita

The creation of a map illustrating the distribution of loan amounts per capita across states is highly relevant to the research question and the central message of the paper, which seeks to assess the effectiveness of the Paycheck Protection Program (PPP) in preserving jobs during economic hardship induced by the COVID-19 pandemic. This visual representation provides a geographic lens through which to examine how PPP funds were allocated and distributed across different states, shedding light on potential disparities and variations in the program's impact.

By mapping the distribution of per capita loan amounts, the research gains valuable insights into the regional patterns of PPP utilization. Understanding how states received and distributed these funds is crucial for evaluating the program's effectiveness in supporting businesses and preserving employment opportunities. The map serves as a visual tool to identify any concentration or disparities in loan disbursements, helping to uncover trends and disparities that may be indicative of the program's success or areas for improvement.

Furthermore, the distribution of loan amounts is a key variable in assessing the broader economic impact of the PPP. It provides a foundational understanding of how financial resources were mobilized across states, offering insights into the scale and scope of businesses that benefited from the program. This, in turn, contributes to the overarching goal of investigating the PPP's efficacy and its role in reducing unemployment.

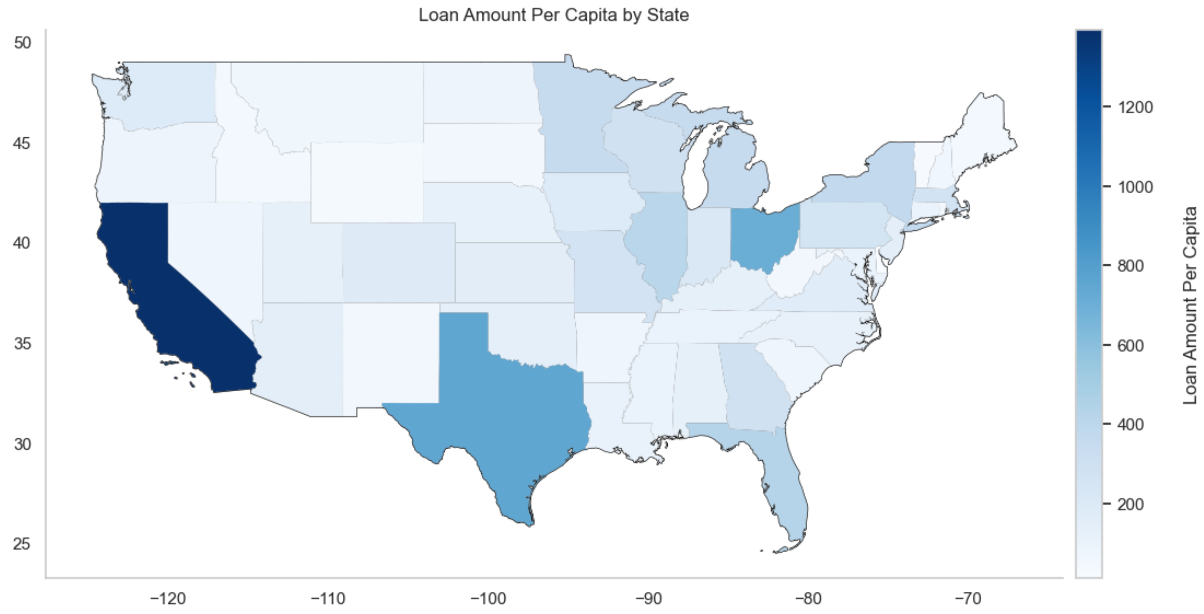

Figure 12: Distribution of Loan Amount Per Capita by State.

The map depicting the distribution of per capita loan amounts serves as a visual representation of the economic impact of the Paycheck Protection Program (PPP) at the state level. The colour gradient on the map reflects the per capita loan amounts, with darker shades indicating higher loan amounts and lighter shades indicating lower amounts.

Due to the nature of the loan amount variable being categorical and discrete (i.e. takes on values 0-5 to represent ranges of loan values), the midpoint of each loan amount range is used to calculate a midpoint. This midpoint, a continuous quantity, is then used to determine the loan amount per capita by state. This provides a more granular representation of the relationship between loan amounts and population.

Notably, California and Texas emerge as the states with the darkest shades, signifying that businesses there received higher loan amounts through the PPP. The upper east states also appear darker than the west and center, suggesting a consistent trend of relatively higher loan amounts in that region. The west, save California, exhibits almost entirely lighter shades, indicating lower average loan amounts.

Analyzing these variations, it becomes apparent that certain states, particularly those with darker hues, might have experienced a more substantial economic impact from the PPP. This could be attributed to factors such as the prevalence of small businesses in the state or its overall entrepreneurial landscape. California, being the most populous state and a major economic hub, likely saw a higher demand for PPP loans, resulting in a darker shade on the map. Similarly, the darker shades along the Upper East may be indicative of the economic significance and higher entrepreneurial activity in that region.

Conversely, the lighter shades in certain areas may suggest lower demand or a different economic structure. Understanding the regional disparities in loan distribution can provide insights into the varied economic challenges and successes across states, ultimately contributing to a more comprehensive evaluation of the PPP's impact.

4.3 State Distribution of Jobs Retained Per Capita

By mapping the distribution of per capita jobs retained, we gain valuable insights into the geographical impact and efficacy of the PPP in different states. Recall that the variable Jobs Retained Per Capita refers to those jobs saved by a business due to the approval of a PPP loan.

The colour gradient on the map signifies variations in the number of jobs retained per capita, with darker hues representing higher job retention and lighter shades indicating lower job preservation. The relevance of this map to the research question lies in its ability to showcase the disparities and successes in preserving employment across states during a period of economic uncertainty. Understanding which states have higher or lower job retention per capita can unveil patterns and factors contributing to the PPP's effectiveness.

This visual representation becomes a critical tool for identifying regions where the PPP has had a more significant impact on preserving jobs, aligning with the paper's goal of investigating the program's overall efficacy. The map can help discern potential correlations between job retention, economic characteristics of states, and the implementation and outcomes of the PPP.

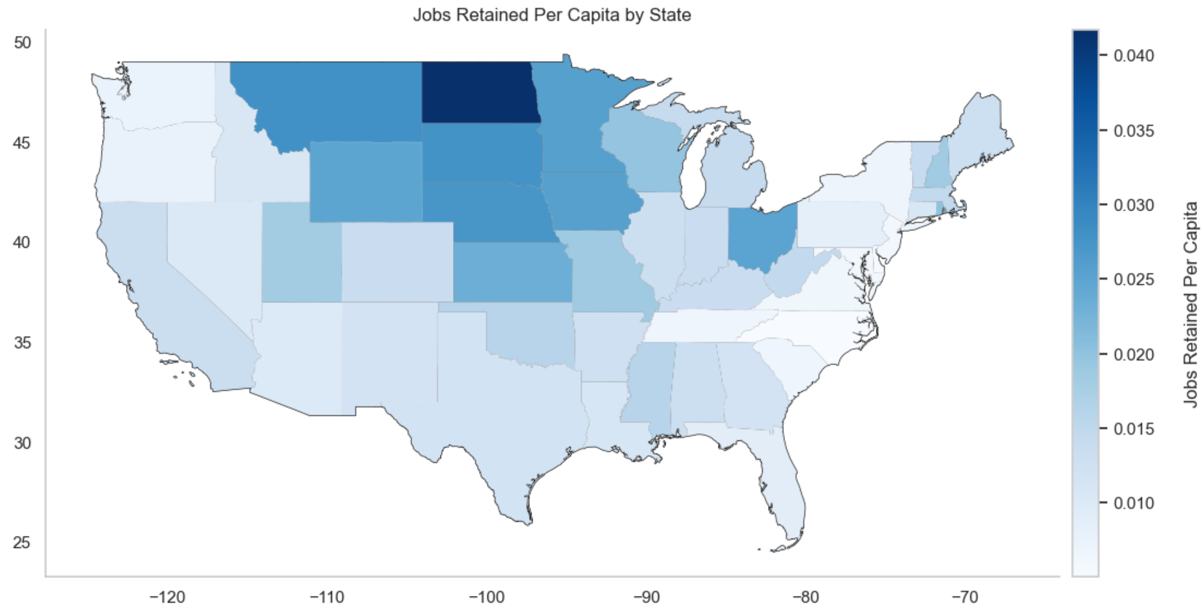

Figure 13: Distribution of Jobs Retained Per Capita by State.

The analysis of the distribution of jobs retained amounts across states gains heightened significance when juxtaposed with the map depicting per capita loan amounts. The comparative examination of these two visuals offers compelling insights into the intersection of financial support received through the Paycheck Protection Program (PPP) and the resultant preservation of employment across different regions.

Notably, the contrast between the two maps reveals intriguing patterns. The entire northern region appears notably darker on the per capita jobs retained map compared to its counterpart displaying per capita loan amounts. This divergence suggests that, despite potentially receiving lower loan amounts, the northern states have effectively retained a substantial number of jobs. This observation prompts further inquiry into the efficiency of job retention strategies implemented in these states or the economic resilience of their industries.

Pockets of darker hues in the Upper East Coast on the jobs retained map highlight specific states or regions where the PPP has had a pronounced impact on preserving employment, even if the loan amounts might not have been exceptionally high. This localized success underscores the importance of examining the nuanced outcomes of the PPP at a finer geographical scale.

The most intriguing revelation emerges from the darkest shade in the state of North Dakota on the per capita jobs retained map. Despite being a single state among many, this distinctiveness signals North Dakota's exceptional success in maintaining jobs relative to the loan amounts received. This outlier prompts an in-depth exploration of the state's economic landscape, business response strategies, or potential regional factors that contributed to such a substantial impact on job preservation.

This comparative analysis of the jobs retained and loan amount maps enriches the narrative around PPP effectiveness. It highlights regional variations, challenges preconceptions about the direct correlation between loan amount and job retention, and underscores the importance of socialized factors in shaping employment outcomes during economic hardships.

4.4 Shannon Diversity Index by State

In conducting an economic analysis of Jobs Retained across states, an innovative approach is employed, using the variable constructed earlier, Ethnicity-Gender, to describe the nature of diversity in firm ownership for those businesses participating in the PPP.

The methodology begins by calculating relative frequencies for each state, discerning the proportion of Jobs Retained within specific combinations of Ethnicity and Gender. This involves determining the relative frequency of Jobs Retained for each Ethnicity-Gender category in relation to the total Jobs Retained in that state.

The subsequent step involves the application of the Shannon Diversity Index, a robust formula that considers the relative frequencies of each Ethnicity-Gender category. By assigning a diversity score to each state based on the calculated Shannon Diversity Index, the analysis yields insights into the diversity landscape, with higher scores indicating greater diversity within the Jobs Retained.

The diversity scores are normalized to ensure consistency and comparability. Normalization guarantees that the scores fall within a specific range, often normalized to a scale of 0 to 1. This step aids in mitigating the impact of scale differences between states, ensuring a fair and meaningful comparison.

The visualization phase entails colour-coding the states on a map using the diversity scores, employing a choropleth map for clarity. Darker colours represent higher diversity, allowing for a straightforward interpretation of the distribution and composition of the Jobs Retained across Ethnicity-Gender categories in different states.

This map quantifies but also visually represents the diversity of Jobs Retained, providing valuable insights into the composition and distribution of the workforce across Ethnicity-Gender categories in various states. The Shannon Diversity Index, a key component of this analysis, encapsulates the richness and variety within the workforce, offering a nuanced understanding of diversity beyond mere headcounts.

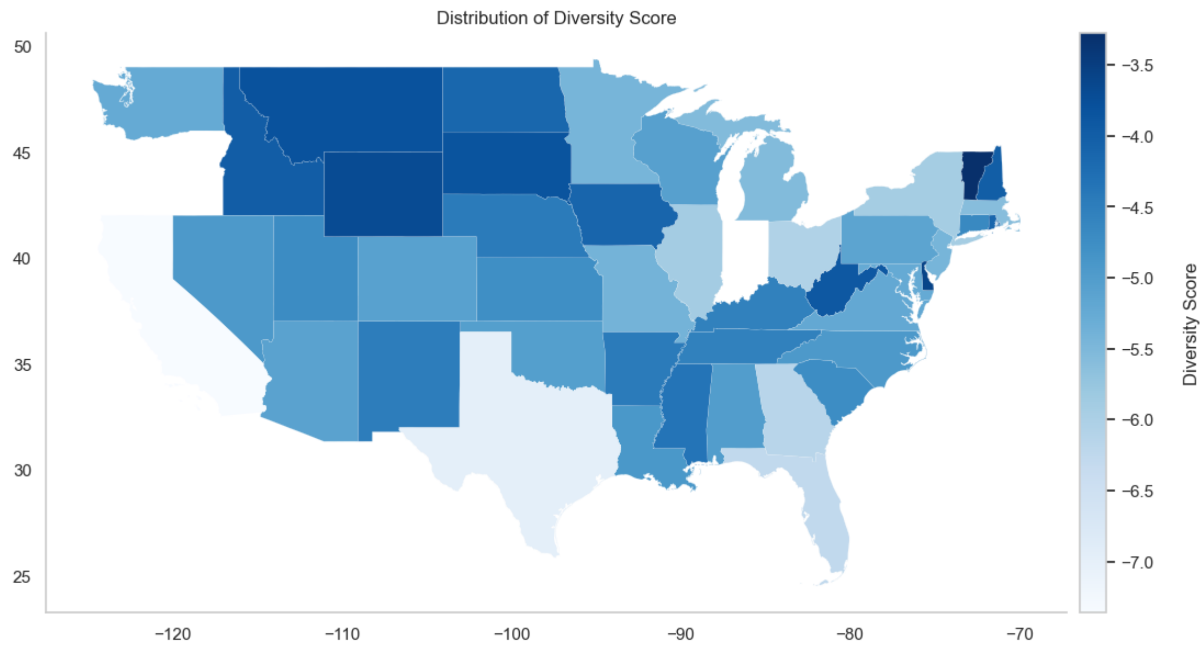

Figure 14: Distribution of Jobs Retained Per Capita by State.

The geographic distribution of diversity scores reveals intriguing patterns that correlate with previous observations of loan amounts and jobs retained across states. Notably, states situated in the center-north region, historically characterized by lower loan amounts and jobs retained, exhibit the highest diversity scores. This juxtaposition suggests a noteworthy shift in the workforce landscape, indicating that despite lower economic indicators, these states are fostering greater diversity in businesses participating in the program. This could be indicative of targeted efforts or unique regional dynamics that prioritize inclusivity.

In contrast, certain pockets along the upper East Coast exhibit both high loan amounts and relatively high jobs retained, aligning with elevated diversity scores. This consistency in the correlation between economic performance and diversity highlights the potential synergies between a robust economy and a diverse business landscape. It implies that states with flourishing economic indicators also tend to foster diverse business participation in the program.

An unexpected finding emerges in the case of California, where the diversity score is surprisingly low concerning businesses participating in the program. Despite its reputation as a hub for innovation and diverse industries, the state appears to lag in terms of diversity in the context of the program. This discrepancy prompts further investigation into the underlying factors contributing to California's lower diversity score. Potential explanations could range from industry concentration to specific demographic dynamics within the participating businesses.

In conclusion, the geographic distribution of diversity scores not only reflects historical economic trends but also introduces novel insights into the evolving landscape of businesses participating in the program. The unexpected patterns, such as the high diversity scores in traditionally lower-performing states and the lower diversity score in California, underscore the complexity of factors influencing diversity in the context of government programs. Further analysis is warranted to uncover the nuanced dynamics that contribute to these patterns and to inform targeted strategies for fostering diversity in various regions.

5 Assessing Complementary Sources of Government Aid

The main question of this paper is to investigate the effectiveness of the PPP in its ability to minimize unemployment during the COVID-19 pandemic. So far, it has been found that states with lower per capita unemployment rates tend to see higher per capita jobs retained due to the PPP as opposed to states with higher per capita unemployment rates. This begs further investigation into whether those states most in need of the PPP received aid.

Examining other government policies and interventions besides the PPP that may have influenced job retention rates, such as unemployment benefits, eviction moratoriums, and other economic stimulus packages, may foster a more comprehensive understanding of the broader policy landscape. This information, if added to the dataset, can help contextualize the impact of the PPP on job retention in each state. In particular, the CARES Act, a federal legislation passed by the United States Congress in March 2020 in response to the COVID-19 pandemic, aimed to provide economic relief and support to individuals, businesses, healthcare providers, and state and local governments affected by the pandemic.

Investigating how federal relief funds from the CARES Act were distributed among state and local governments during the initial stages of the pandemic response would be a highly relevant and useful additional source of data. This information is valuable for understanding the flow of resources and support to different levels of government during a critical period of economic uncertainty and public health crisis.

It is necessary to scrape this data because other sources that report on government federal relief tend to report data for each state separately. That is, there are separate datasets available for download that provide information on a state by state basis. This scraped data is a helpful aggregation of this information. Additionally, datasets that might include this information across states tend to require some calculation to reach the numbers provided by this website (noted by the author).

This data is different from the data that is already in the dataset because this study has thus far investigated only one source of financial aid and its impact on job retention. However, in the real world, the support available to businesses during COVID-19 extended beyond just the PPP, such as the support provided by other streams from the CARES Act.

It will fill holes in the dataset by adding external validity to the results. Considering other sources of funding that may have been available to the same businesses participating in the PPP (and which are under study in this paper) will allow for a more comprehensive understanding of the factors impacting job retention. For example, we may find that states with high job retention had both high PPP loan amounts and high CARES Act Funding. Incorporating this information will thus make any conclusions from causal analysis conducted more likely to be applicable to other relief programs in the future.

The address of the website that can be used to scrape this addition to the dataset is: https://taxfoundation.org/blog/federal-coronavirus-aid-to-states-under-cares-act/.

This data can improve the results of this study by adding another dimension to this analysis, namely, uncovering whether other factors were at play in improving job retention in each state that may have influenced the state's overall per capita job retention. Identifying any confounders or omitted variables may lead to more conclusive patterns in the PPP's effectiveness in targetting states more in need of the policy by demonstrating which states might have had additional measures in place compared to others.

By incorporating information on the allocation and use of federal relief funds from the CARES Act, researchers can gain insights into how states distribute resources to support businesses and workers. This additional context allows for a more thorough assessment of the PPP's impact, as it considers the extent to which states supplemented federal assistance with their own initiatives. Disparities in the distribution of CARES Act funding among states may also shed light on variations in job retention outcomes, helping identify factors beyond PPP loans that influence employment stability.

Consequently, analyzing state shares of CARES Act funding alongside existing data enriches the understanding of the multifaceted dynamics shaping job retention rates across different states, ultimately contributing to a more robust evaluation of the PPP's effectiveness in mitigating economic hardship.

Incorporating this additional data not only strengthens the paper's empirical analysis but also enhances its originality by delving into more contextual dimensions of the PPP's effectiveness. By uncovering potential confounders and omitted variables, this study can provide more nuanced insights into the relationship between government interventions and job retention, ultimately contributing to a more robust understanding of the policy landscape during times of economic crisis.

Data on the estimated state share of the CARES Act funding will be merged with existing results by comparing the percent of the total CARES Act funding received by the state with the jobs retained per capita due to approval of the PPP loan by state. A comparison of these variables will help answer the research question as it provides a comprehensive view of how CARES Act funding distribution correlates with job retention rates attributed to PPP loans across different states. The study will then be able to assess the relationship between the magnitude of federal assistance received by each state and its corresponding success in preserving jobs through PPP loans.

Understanding how variations in CARES Act funding allocation correspond to disparities in job retention rates attributed to PPP loans provides valuable insights into the impact of federal relief efforts on mitigating unemployment during the COVID-19 pandemic.

To scrape the desired information about the state shares of CARES Act funding from the Tax Foundation website, HTML scraping is used with the library BeautifulSoup. This is because the website does not provide an API to use to directly access the state funding shares data in a structured format such as XML.

5.1 Distribution of CARES Act State Funding

To normalize the state funding totals under the CARES Act, the state allocation of the total national funding provided under the Act is calculated as a per capita total for each state. These per capita funding allocations are then mapped across states with a colour gradient indicating differences in the total per capita CARES Act allocation to that state. The gradient legend indicates that lighter shades are smaller per capita totals while darker hues are larger.

The purpose of plotting this map is to be able to compare differences in CARES Act funding across states with previous maps showing the distribution of both Loan Amount under the PPP and Jobs Retained as a result of PPP loan approval. Comparing these maps may indicate patterns in states with high job retention and high CARES Act funding. This would indicate an additional source of support to consider in being able to isolate the true effect of the PPP on addressing unemployment during times of economic hardship.

Mapping the distribution of CARES Act funding by state alongside PPP metrics offers a powerful lens through which to assess the effectiveness of federal relief efforts in combating unemployment during the COVID-19 pandemic. Normalizing state funding totals allows for a clear visualization of disparities in resource allocation, highlighting states that received a larger share of federal aid. By juxtaposing this information with maps showing PPP loan amounts and jobs retained, patterns and correlations between funding allocation and employment outcomes can be discerned. States demonstrating high job retention alongside significant CARES Act funding may suggest successful relief efforts.

This comparative analysis not only aids in identifying effective support strategies but also facilitates targeted interventions by revealing which areas have benefited most from federal assistance. Such insights are invaluable for policymakers seeking evidence-based approaches to navigate economic recovery and mitigate unemployment challenges in times of crisis.

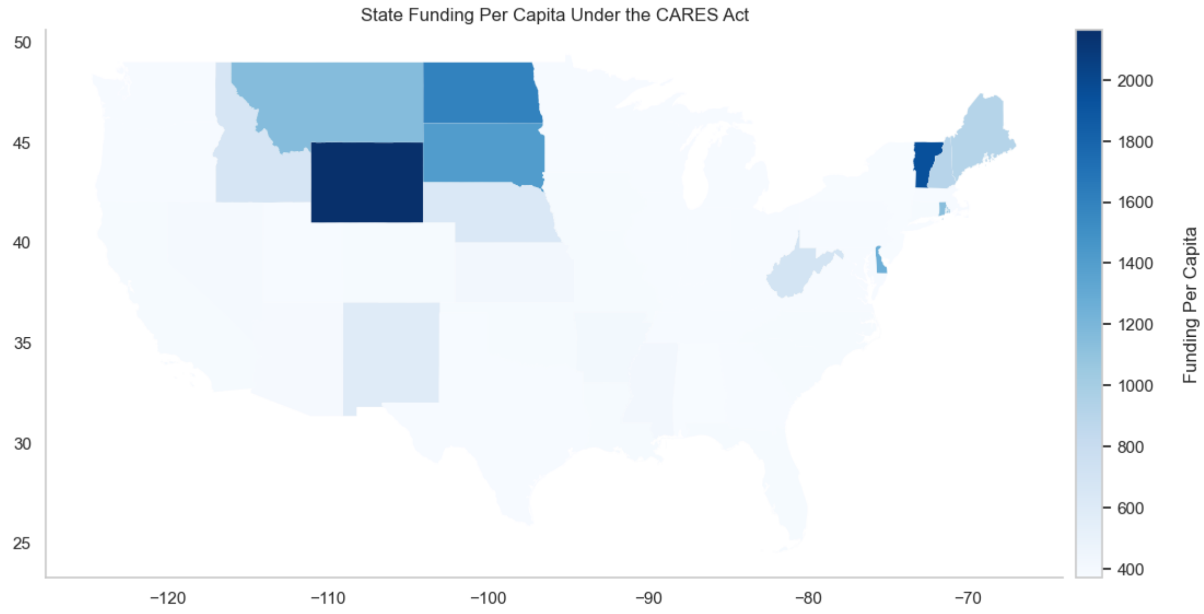

Figure 15: Distribution of CARES Act Funding per capita by State.

The map above shows a relatively similar distribution across states as seen in the map titled Jobs Retained Per Capita by State. This indicates that those states in the central and upper North tended to receive higher CARES Act Funding per capita than other states and the same states saw higher jobs retained per capita due to the PPP. This could either indicate that those states received sufficient funding relative to their populations to be able to retain jobs while other states, despite receiving higher shares of the total allocation (as is discussed further below) still did not receive an adequate amount for their populations, and thus saw lower jobs retained per capita. Ultimately, the above map and the Jobs Retained Per Capita by State map indicate a correlation between higher amounts of funding per capita and higher jobs retained per capita.

To compare the above with a map depicting the share of the total allocation from the CARES Act, we plot another map below. We find ultimately that differences in the share of the allocation are higher in states that likely have more small businesses or higher populations. Comparing the below map with the above, we see that states receiving a larger share of the relief (i.e. California, Texas, Florida) are not shown above as those states receiving more funding per capita. This confirms that the above graph is likely a better comparator than the one below.

We include the analysis of the state share of the total funding provided under the CARES Act for completeness below. Put simply, this is the percent of all CARES Act Funding dispensed that went towards each state.

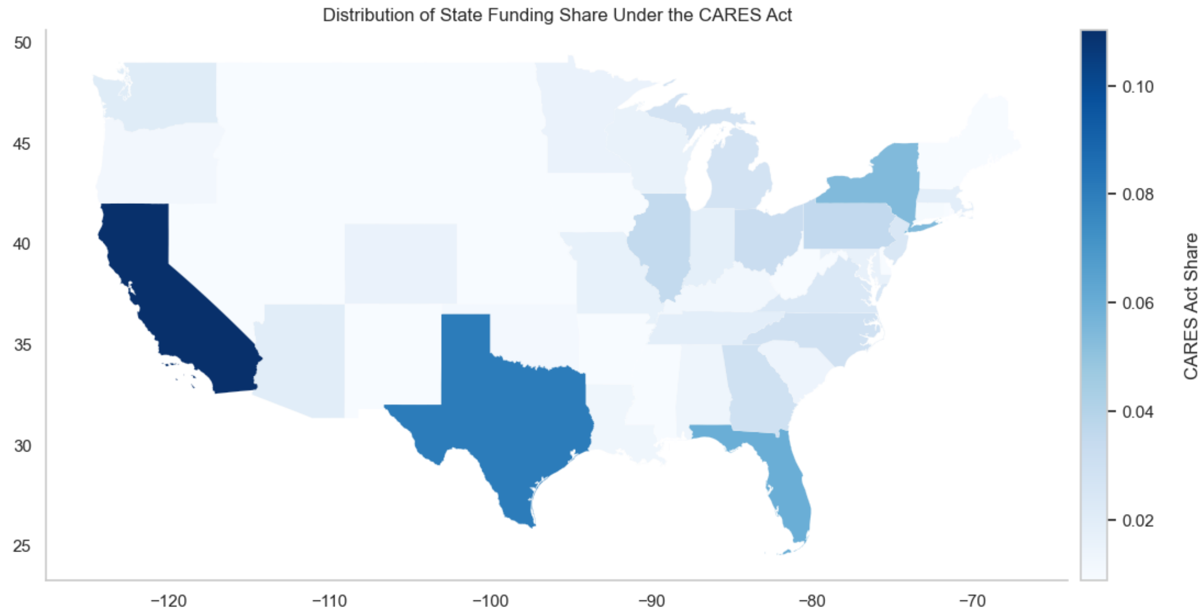

Figure 16: State Share of total CARES Act Funding.

This map visualizes the percent of the total funding provided by the CARES Act that went towards each state.

We find that the highest share of CARES Act funding was allocated to California, Texas, Florida, and New York. This is consistent with the distribution seen earlier in the Loan Amount Per Capita by State map, as we saw California, Texas, and Florida as states exhibiting higher per capita loan amounts as well.

The consistent presence of California, Texas and Florida among the states with the highest loan amounts per capita under the Paycheck Protection Program (PPP) and the greatest share of CARES Act funding signifies several noteworthy implications. Firstly, it suggests that these states hold significant economic importance within the United States, likely characterized by larger populations, extensive business activity, and heightened economic significance compared to other regions.

Secondly, the high loan amounts per capita and share of CARES Act funding indicate a substantial demand for financial relief among businesses in these states, likely stemming from the severe economic impacts of the COVID-19 pandemic and the necessity to preserve jobs and stabilize local economies.

Thirdly, the consistent allocation of significant funding underscores the effectiveness of federal relief programs, such as the PPP, in targeting areas with the most pressing economic needs. This targeted assistance may have played a crucial role in mitigating unemployment and bolstering economic resilience in these key states during the pandemic.

Lastly, recognizing the unique economic circumstances of these states is essential for policymakers when designing future relief initiatives. Tailoring policies to address the specific challenges faced by high-impact states like California, Texas and Florida can enhance the effectiveness of federal assistance programs and foster more equitable economic recovery efforts nationwide.

Overall, the prominence of these states in both PPP loan amounts per capita and CARES Act funding allocation underscores their pivotal role in economic relief efforts and highlights the importance of targeted support in addressing regional disparities and promoting sustainable recovery.

We find similar differences between Jobs Retained Per Capita by State and CARES Act funding by state as we did between the former and the Loan Amount Per Capita by state plot. We again see that the states with the highest funding, in this case from the CARES Act, tend not to be among the states with the highest jobs retained per capita, which are those in the central and upper North. From this plot, we see that states in the central and upper North tend to exhibit the lowest shares, while there exist pockets of higher share distribution amongst the Eastern states.

In other words, while certain states received substantial financial support from the CARES Act, they may not have been as successful in retaining jobs relative to their population size. Conversely, states in the central and upper North regions, despite potentially receiving lower funding, managed to retain a higher number of jobs per capita.

This observation suggests that factors beyond funding allocation may influence job retention rates. It could imply variations in the effectiveness of economic policies, business resilience, or the severity of the pandemic's impact across different regions. Understanding these dynamics is crucial for policymakers to tailor interventions effectively and address the specific challenges faced by different states in preserving employment and supporting economic recovery. We confirm that states with larger populations or more small businesses did indeed receive more funding but likely not enough to exhibit higher jobs retained. States with lower populations and smaller entrepreneurial hubs received enough funding in comparison (and higher per capita amounts overall) leading to higher jobs retained per capita. This begs the question of whether increased funding per capita in states like California or Texas would allow for similar success in jobs retained per capita as seen in North Dakota or Montana.

5.2 Comparison of CARES Act Funding, PPP Loan Amounts and Jobs Retained

Understanding how federal relief funds (CARES Act) are distributed relative to the loan amounts provided through programs like PPP and the resulting job retention provides insights into the effectiveness of resource allocation. To better compare these three metrics, a scatterplot is created along three dimensions by state. The x-axis represents the CARES Act Funding Per Capita, the y-axis represents the PPP Loan Amount Per Capita, and the size of each point on the plot corresponds to the number of Jobs Retained Per Capita.

Firstly, this plot will indicate whether there is a connection between those states that receive higher per capita funding from the PPP and the CARES Act. If we see states that receive higher per capita amounts from both, this indicates certain distribution criteria used by the federal government. If we see that these differ, it might point to how relief programs aim to support states in conjunction. That is, it may suggest a level of coordination between relief programs to provide the most equitable and effective amounts of relief to each state as possible.

Secondly, this plot will allow for a greater understanding of the impact of relief funding on Jobs Retained per capita due to PPP loan approval. If we see that states that received higher CARES Act funding had higher job retention, this might indicate that Jobs Retained is not an isolated causal relationship with just PPP funding but also is a function of other sources of support. This helps answer the research question by discerning the true impact of the PPP on unemployment when considering other relief programs as well.

Comparing these metrics highlights any disparities between states or regions in terms of funding allocation, loan disbursement, and job retention. These disparities may indicate areas that require additional support or where existing relief measures are not effectively addressing economic challenges.

Regional variations in per capita CARES Act funding, PPP loan amounts, and job retention provide valuable insights into the unique economic challenges and strengths of different states or regions. Understanding these dynamics is essential for tailoring interventions to address local needs and promote inclusive economic recovery.

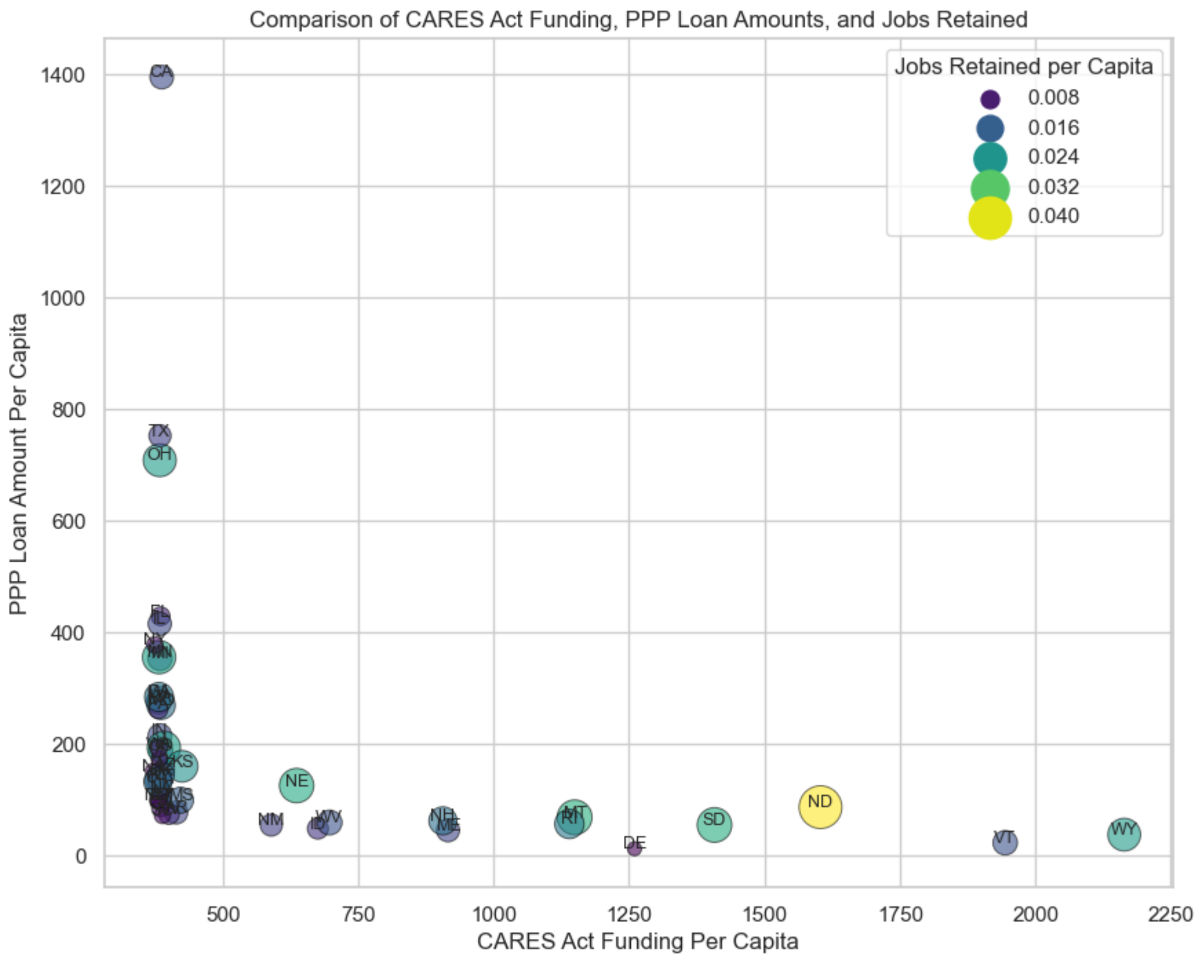

Figure 17: Comparison of CARES Act Funding per capita, PPP Loan Amount Range per capita, and Jobs Retained per capita.

The scatterplot provides several key insights. First, we notice that those states with higher per capita PPP support tend to have lower per capita CARES Act support. The opposite is also true, where those states receiving higher CARES Act support per capita tend to have lower per capita PPP support.

There is no state that receives high per capita support from both programs, as we notice no points in the top right of the plot. Instead, we see the opposite. A large cluster of states in the bottom left indicates that many states are receiving low per capita support from both programs.

The absence of states in the top right of the plot, where both per capita supports are high, indicates a lack of states simultaneously benefiting from substantial support from both programs, highlighting potential limitations or trade-offs in the implementation of federal relief initiatives.

It is possible that states with higher PPP per capita had a greater demand for financial assistance specifically tailored to small businesses, which the PPP provided. This targeted support may have been more effective in addressing the immediate needs of businesses in those states, leading to a lower relative need for CARES Act funding.

Alternatively, lower CARES Act per capita in states with higher PPP per capita could suggest that those states had unmet needs or challenges that were not adequately addressed by the CARES Act. This could be due to various factors such as eligibility criteria, distribution mechanisms, or limitations of the CARES Act funds.

The high vs. low dichotomy between each support program, as seen by plots along each axis and in the corners such as Wyoming (high CARES, low PPP) or California (high PPP, low CARES) can suggest a potential trade-off between the two forms of federal economic relief.

This phenomenon could reflect a reallocation of resources across states, where states prioritize one form of support over the other based on their specific economic needs and circumstances. States opting for higher PPP support may prioritize preserving jobs and supporting small businesses through direct loans, potentially at the expense of receiving lower CARES Act support, which provides broader financial assistance. Conversely, states with higher CARES Act support may have chosen to allocate resources towards broader economic relief efforts beyond the scope of the PPP, resulting in lower PPP support per capita.

The concentration of states in the bottom left of the plot, receiving low per capita support from both programs, underscores the widespread challenges and economic impacts experienced across many states during the COVID-19 pandemic, calling for a comprehensive reassessment of federal relief strategies to address the diverse needs of states and regions effectively.

Lastly, we note that those states with the highest jobs retained per capita due to the PPP such as North Dakota, Montana, and South Dakota also received relatively high per capita CARES Act funding. This corroborates the findings from the maps we saw earlier and also points to the possibility that larger, more populous states simply did not receive as much aid as they needed to exhibit similar levels of success in job retention.

6 Merging Socioeconomic Indicators

Recall that the analysis of the original data involved merging population data to get per capita amounts. It also required merging data for formatting purposes such as getting state acronyms, Federal Information Processing Standards (FIPS) code for each state, and filtering by contiguous states only for the purposes of mapping. Data on layoff rates, NAICS code industry names, and unemployment rates were also merged.

The scraped data added the amount of funding allocated to each state through the CARES Act, which was another relief program funded by the U.S. Government during COVID-19. Analysis of this scraped data involved converting it into a share of the total allocation to get the percentage of the funding allocated to each state. It also involved merging the scraped data with population data to get the per capita CARES Act funding by state.

The original dataset included 19 variables while the most recent iteration, following several rounds of merging new data and creating new variables, includes 29 variables. In order to add even more data points, in part to prepare for robust analysis via machine learning methods, this section merges an additional dataset containing information on employment, income and benefits, and worker type at the state level.

This study aims to investigate the effectiveness of the PPP as a policy. The new dataset is relevant to this topic because it offers crucial insights into the broader socioeconomic context surrounding employment dynamics. While the unemployment rate was merged earlier, this rate was a national average over the three months that this dataset covers, yielding only three data points overall. By incorporating employment data at the state level, we can analyze how PPP funding allocation and use impacts job retention rates within different states more granularly. Understanding variations in employment levels across states provides valuable context for interpreting the program's effectiveness in mitigating unemployment.

Additionally, merging income and benefits data enables a more complete assessment of the level of wealth in each state. Given that unemployment and jobs retained are directly related to the level of wealth and purchasing power of civilians, merging this data will to assess the program's impact on vulnerable populations and economically disadvantaged areas. Analyzing how PPP funds correlate with changes in income levels and poverty rates can elucidate whether the program effectively targets regions most in need of support. Moreover, merging these datasets allows for a more nuanced examination of how state-level economic factors interact with PPP implementation, providing policymakers with evidence-based insights to optimize future relief efforts and address unemployment challenges comprehensively.

Analyzing state-level income and benefits data is crucial for gaining a comprehensive understanding of the effectiveness of the PPP in mitigating unemployment. Higher levels of income generally lead to greater consumer spending, which drives demand for goods and services. By providing financial assistance to businesses to retain employees, the PPP aims to maintain consumer purchasing power and sustain demand in the economy. This, in turn, supports business operations and prevents layoffs or job losses. Additionally, if the level of income and benefits in a state is higher relative to other states, it might indicate a stronger baseline economic condition and greater resilience to economic shocks. States with higher income levels may have businesses and households that are better equipped to weather downturns, potentially leading to more effective utilization of PPP funds and a quicker recovery in employment levels. Therefore, examining income and benefits data at the state level provides valuable insights into the local economic context and can inform targeted policy interventions to address unemployment challenges effectively.

Integrating data on worker type at the state level can provide valuable insights into patterns of industry resilience and shed light on the effectiveness of the PPP in mitigating unemployment. By examining the relationship between states with higher per capita jobs retained and specific worker types, we can identify industries that are less susceptible to economic downturns. For instance, if certain states consistently retain more jobs in industries dominated by skilled workers or essential services, it may indicate that these sectors have a greater capacity to weather economic shocks. Understanding these patterns can help isolate the impact of the PPP on unemployment reduction, independent of industry dynamics. Moreover, identifying industries that exhibit higher job retention rates can inform targeted policy interventions and resource allocations to support vulnerable sectors and bolster overall economic resilience. Therefore, integrating worker type data at the state level offers a nuanced perspective on the efficacy of the PPP and contributes to a more comprehensive understanding of its impact on unemployment levels nationwide.

Overall, integrating employment, income and benefits, and worker type data enriches the analysis of PPP effectiveness by contextualizing program outcomes within the broader socioeconomic landscape, thereby informing more targeted policy interventions aimed at minimizing unemployment and promoting economic recovery. It may also help add colour to previous analyses by better understanding the socioeconomic context of states exhibiting traits such as higher per capita jobs retained, loan amounts, or CARES Act funding. The new dataset is sourced from the United States Census Bureau for the year 2020.

6.1 Employment vs. Loan Amount by State

Using the newly merged data, a new variable is created to measure employment per capita by state. Analyzing the relationship between 'Employment Per Capita' and 'Loan

Amount Per Capita' by state through a scatter plot provides valuable insights into the effectiveness of the PPP in minimizing unemployment. The scatter plot allows us to visualize how the amount of loan funding received per capita by each state relates to the employment rate per capita.

States positioned higher on the y-axis, indicating higher 'Employment Per Capita', suggest a stronger job market and lower unemployment levels. Meanwhile, states positioned further to the right on the x-axis, indicating higher 'Loan Amount Per Capita', received more substantial PPP loan funding per capita.

By examining the distribution of data points on the scatter plot, we can discern whether there is a correlation between the amount of PPP loan funding provided to each state and its subsequent employment levels. A positive correlation would imply that states receiving more substantial PPP funding experienced better job retention or employment outcomes, supporting the program's effectiveness in preserving employment during economic challenges. Conversely, a lack of correlation or a negative correlation may suggest inefficiencies or disparities in the distribution and utilization of PPP funds, highlighting areas for further investigation or policy refinement. A correlation could also indicate that states that tend to have higher employment per capita receive lower PPP support.

Therefore, analyzing 'Employment Per Capita' versus 'Loan Amount Per Capita' by state provides a perspective on the distribution of PPP funding by state.

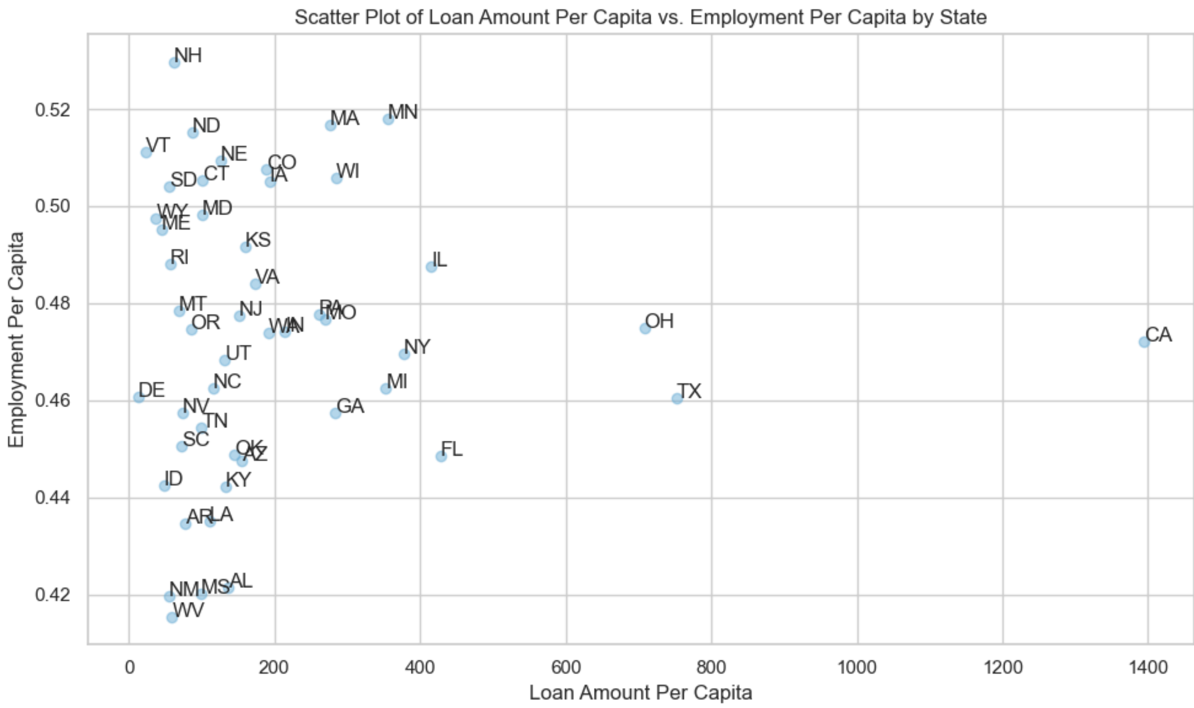

Figure 18: Comparison of per capita values of Loan Amount and Employment by state.

From this scatterplot we can see that states receiving higher per capita loan amounts, such a California, Texas, and Ohio, tend to have relatively moderate per capita employment levels. We also see that there is a large cluster of states that tend to receive lower per capita loan amounts but exhibit a range of low to high per capita employment levels.

This makes it difficult to discern whether there is a true connection between the two. Some states, such as West Virginia or Mississippi, receive low per capita loan amounts and also exhibit low per capita employment. This may be a function of population density or lack of entrepreneurial hubs.

On the other hand, states such as New Hampshire and Minnesota exhibit low per capita loan amounts but higher per capita employment. This spread of low to high per capita employment at similar levels of per capita loan amount makes it difficult to ascertain whether there is truly a connection between the two variables.

It may be the case that states like California or Texas were better able to use the government support they received. It is also possible that these states started with higher resilience to economic downturns. Conversely, the cluster of states receiving lower per capita loan amounts but exhibiting varying employment levels suggests that factors other than PPP funding, such as state-level economic conditions, industry composition, or access to alternative sources of financial support, may have influenced employment outcomes. Overall, the relationship between per capita loan amounts and per capita employment levels underscores the importance of considering regional economic disparities and policy responses when evaluating the effectiveness of the PPP in sustaining employment across different states.

6.2 Distribution of Worker Type vs. Jobs Retained by State

Using the data on worker types by state from the new dataset, new variables are created for each worker type per capita. Below, this study plots a comparison of the makeup of worker types per capita by state, showing the ten states with the highest jobs retained per capita, highest to lowest. Plotting this allows us to identify whether there is a particular worker type that tends to make up most of the working population per capita in states that had higher success with using the PPP to minimize unemployment (i.e. higher jobs retained per capita).

Identifying trends in worker types helps to answer the question of how effective the PPP is as a policy by breaking out the success of different subgroups of workers. If a specific worker type consistently dominates the workforce in states with high jobs retained per capita, it suggests that the PPP may have been particularly successful in supporting that particular subgroup of workers. This understanding helps in evaluating the policy's impact and effectiveness in preserving employment across different segments of the workforce.

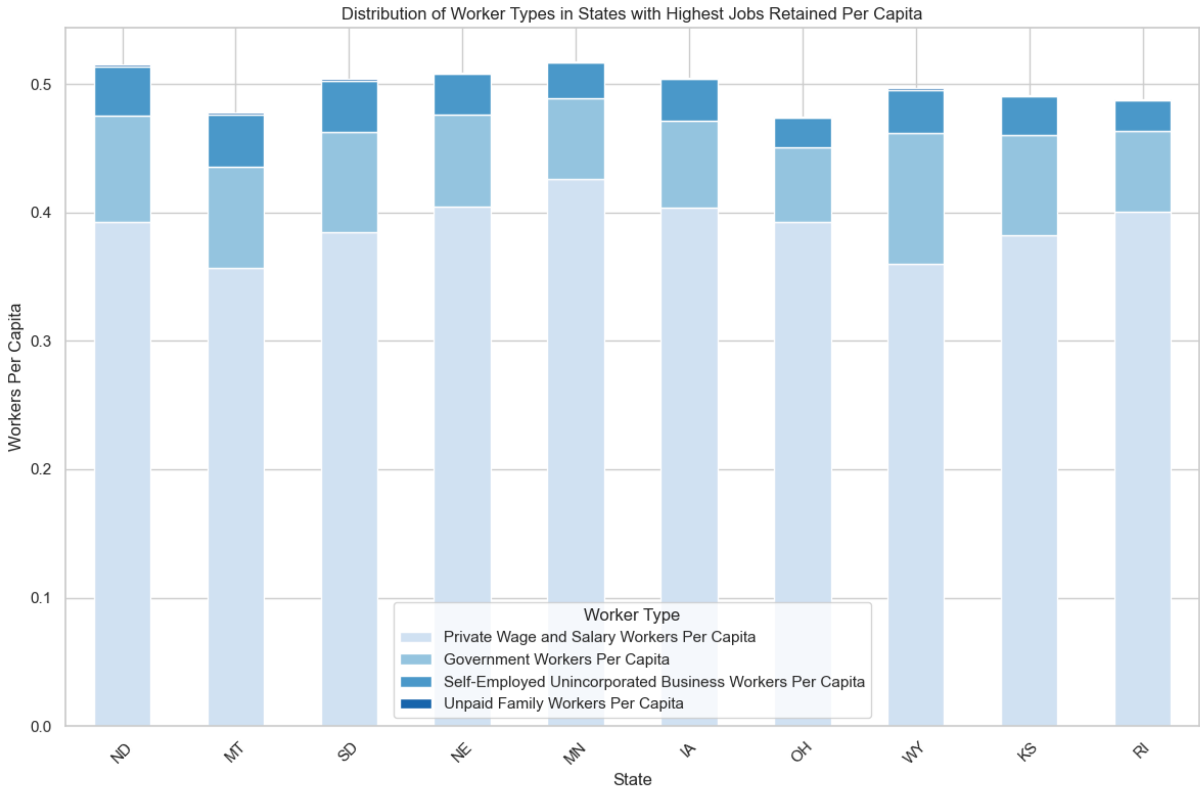

Figure 19: Distribution of worker types across states with the highest per capita Jobs Retained.

This plot shows that a dominant category of worker type does exist among states with the highest per capita jobs retained. This category is Private Wage and Salary Workers. The second largest majority is Government Workers, though it is quite clearly a much smaller portion of the worker types.

This suggests that the PPP might have been most effective in preserving jobs in the private sector than jobs in government or the self-employed. This finding is consistent with earlier analyses on the most common business type invovled in the PPP program, where the top 3 were all related to jobs in the private sector.

Economically, this could mean that aid programs like the PPP must consider their ability to minimize other forms of unemployment outside of the private sector. It could also mean that aid programs like the PPP should prioritize or tailor their support mechanisms to address the needs of different sectors of the economy more effectively. For instance, if private wage and salary workers constitute the majority of the workforce in states with higher per capita jobs retained, then policies targeting this demographic may yield a greater overall impact in terms of employment retention.

Moreover, the relatively smaller proportion of government workers in states with high per capita jobs retained suggests that public sector employment may not have been as significantly impacted by the PPP compared to the private sector. This could have implications for future policy decisions regarding the allocation of resources and support measures during economic crises.

7 OLS Regression

7.1 Preliminary Analysis

Recall that the dependent variable determined for this study is Jobs Retained, which is defined as the number of jobs preserved by a business as a result of obtaining approval for a PPP loan. The research question of this study is to assess the efficacy of the PPP in minimizing unemployment during times of economic hardship.

The primary independent variable of interest for this study is Loan Amount, which provides the range of the PPP loan value approved for a business. In order to answer whether the economic relationship between the outcome variable and primary predictor variable of interest is linear or non-linear, we begin by conducting a visual inspection to draw on the data.

Using per capita variables allows for normalization; this makes the relationship comparable across states and any patterns identified more meaningful. In other words, standardizing the variables by per capita values removes the influence of population size on the relationship being studied between the two variables and essentially prevents confounding due to differences in population size. If we inspect that the relationship appears to be a straight line, this suggests linearity. If the points form a curve or some other pattern that deviates from a straight line, this may indicate non-linearity.

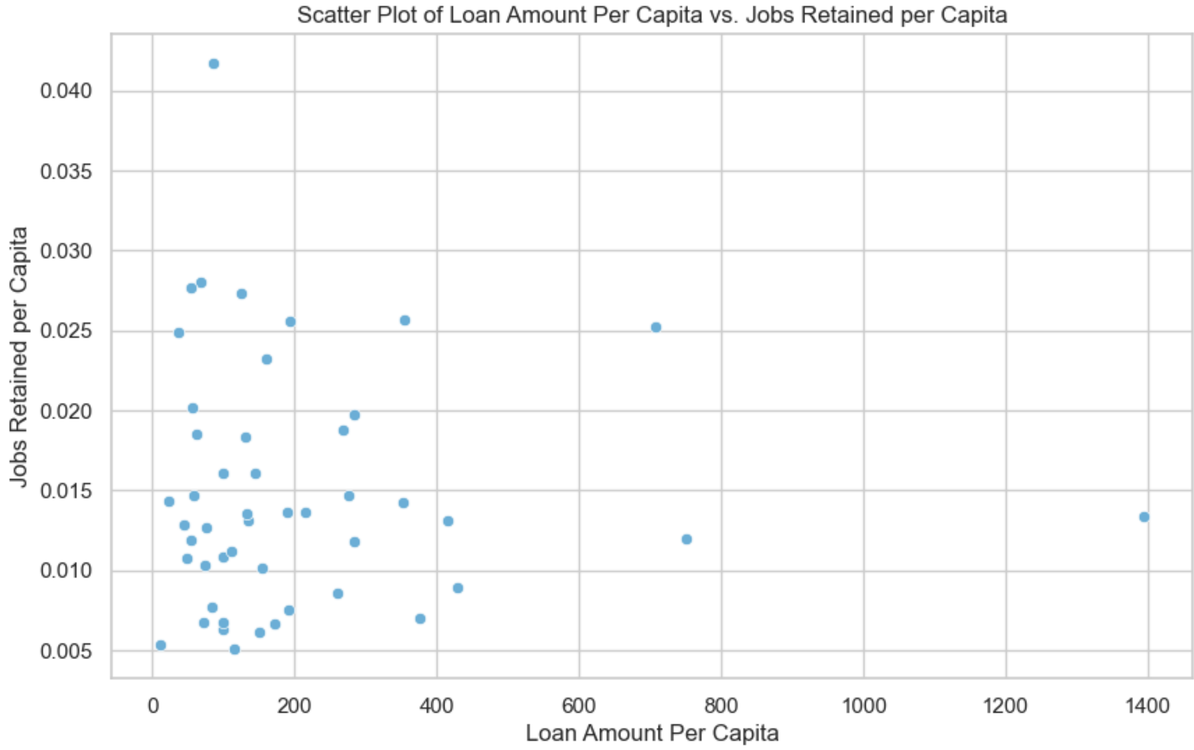

Figure 20: Loan Amount Per Capita vs. Jobs Retained Per Capita by State.

The scatterplot appears to have more of a vertical spread for lower loan amounts, with most points concentrated on that end. This makes sense as we saw earlier that most loans provided were in the lowest range. We do see that for higher loan amounts, however, the jobs retained per capita tend to be moderate.

To get a better understanding of this pattern, we use Seaborn's regplot() function and omit The order parameter. This causes Seaborn to automatically select the most appropriate polynomial degree for a polynomial of best fit based on the data. This will provide some further intuition about what the relationship between the variables may be.

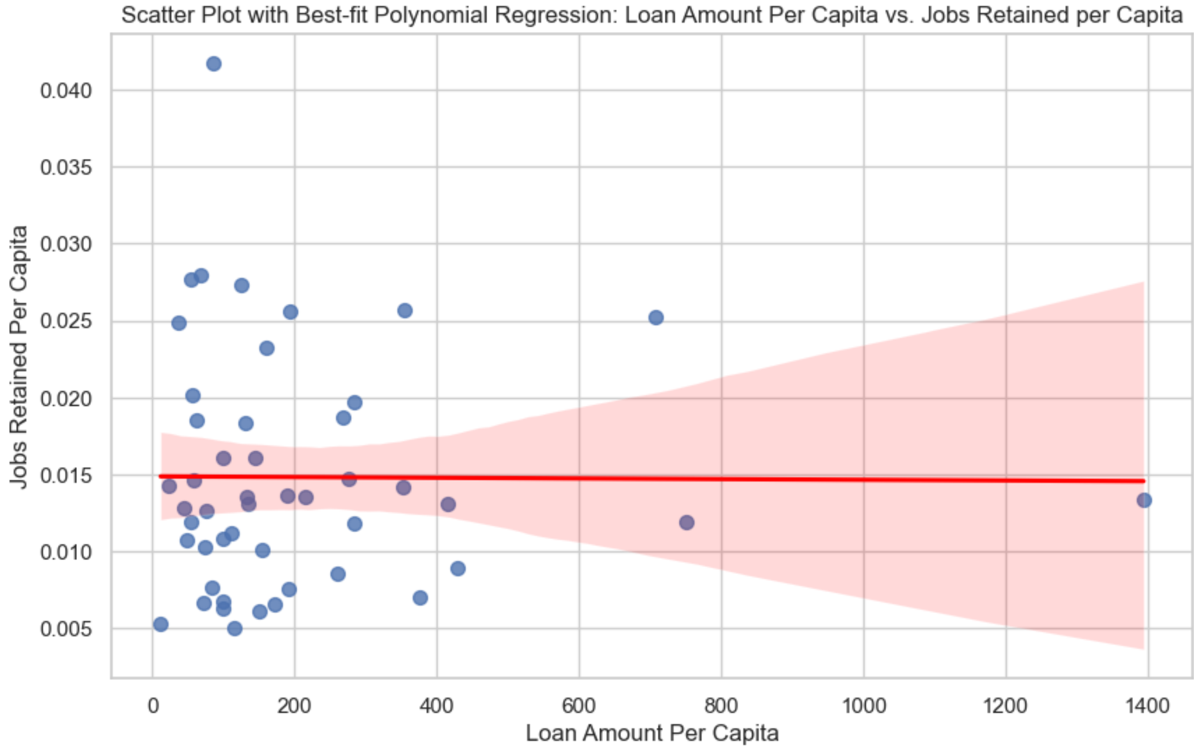

Figure 21: Best-fit Polynomial Regression for Loan Amount Per Capita vs. Jobs Retained Per Capita by State.

Given the large confidence intervals towards the right of the plot due to the outliers and the extremely weak correlation coefficient, we consider removing them. These points are associated with Texas, Ohio, and California. As these states were also highlighted as outliers in previous analyses, we remove them and refit the line.

We also multiply the per capita values by a constant to improve interpretability in subsequent analysis. In this case, we multiply by 100,000 to effectively scale up the values and make them more interpretable.

Figure 22: Best-fit Polynomial Regression for Loan Amount Per 100,000 People vs. Jobs Retained Per 100,000 People by State (Outliers Removed).

We replot and, using statistical methods, find a correlation coefficient of -0.08. This indicates a weak negative linear relationship between the two variables. Although, a low correlation coefficient doesn't necessarily rule out linearity, this would indicate, in terms of economic practicality, that there is almost no linear association between the two variables. This might indicate instead a non-linear relationship

Drawing from economic theory, it is possible that saturation effects or threshold effects are at play. In economic contexts, this refers to situations where the relationship between two variables changes non-linearly as one variable reaches a specific level or threshold. In other words, increasing one variable beyond a certain point does not lead to a proportional increase in the other variable. This might occur when markets become saturated, or resources limited. In this context, the latter is quite possible. Given that the majority of loan amounts were initially in the lower range, it's plausible that higher loan amounts could have led to greater jobs retained per capita. However, due to resource limitations or other constraints, these higher loan amounts may not have been feasible or available. Consequently, lower loan amounts may have resulted in less-than-optimal job retention, as the resources necessary to maximize job preservation may not have been fully accessible.

Drawing from evidence in the literature, as discussed earlier in the paper, many studies have found that the PPP did save millions of jobs but at higher costs than other government relief programs during the same period. Given this data and evidence, it would appear that the relationship between jobs retained per capita and loan amount per capita is plausibly linear. However, this must be further tested via regression analysis and statistical tests such as ANOVA and hypothesis testing. The regression analysis that follows uses Ordinary Least Squares (OLS) to estimate the parameters (β's) of the linear model.

7.2 Selecting Regressors for MLR

The existing theories relevant to this study are those set out by the literature on government relief programs, Keynesian economics, and labour market theory. Below are justifications for the inclusion of each of the predictors in the multivariate regression analysis that follows.

Rationale: Loan Amount and Employment Keynesian economics emphasizes the role of government intervention in the form of fiscal policy such as the PPP. The rationale behind such programs is to stabilize the economy during periods of economic downturns. In this view, the PPP is a classic Keynesian measure that aimed to boost aggregate demand by providing financial assistance to businesses. This would thereby prevent layoffs, reduce unemployment, and sustain economic activity.

The theory would suggest that, in a regression analysis examining the impact of the PPP on jobs retained, the loan amount can indeed serve as a predictor variable. The rationale is that higher loan amounts provided through the PPP should be associated with greater support for businesses, leading to higher job retention. From this perspective, there is an expectation that the loan amount would be positively related to the number of jobs retained due to this stimulus.

Keynesian economics also suggests that changes in aggregate demand, influenced by factors such as consumer spending, investment, and government policies, can impact employment levels. Higher income levels tend to lead to increased consumer spending, as individuals have more disposable income to allocate toward goods and services. This increased spending stimulates economic activity, leading to higher demand for goods and services, and potentially, greater job retention. Therefore, income levels indirectly affect job retention through their influence on consumer spending and aggregate demand.

Generally, individuals with higher income levels tend to have higher levels of disposable income. According to the marginal propensity to consume (MPC), people with higher incomes typically spend a larger proportion of their income on consumption. Therefore, an increase in income levels is associated with increased consumer spending, which can stimulate economic activity and potentially lead to greater job retention. So, we might expect to observe a positive relationship.

Rationale: Population Over 15 The size of the working-age population can affect the pool of potential workers available for employment in each state. A larger population over 15 may indicate a larger labour force and potentially more job opportunities.

The theory of labour supply and demand would support this connection. The working-age population constitutes the potential labour supply in an economy. As this population grows, it implies there are more individuals able to participate in the labour market. This increase in labour supply can influence the number of people seeking employment. In contrast, employers create jobs to meet the demands of production. The demand for labour is derived from the demand for goods and services. When higher demand is expected, firms tend to hire more workers to meet that demand.

Including this variable in the regression will analyze the relationship between the size of the working-age population and the number of jobs retained. If there is a positive correlation, it suggests that an increase in the working-age population tends to lead to a higher number of jobs retained, ceteris paribus.

Rationale: Unemployment Rate and Layoff Rate These variables reflect the overall health of the labour market. A higher unemployment rate and layoff rate may indicate greater job insecurity and lower job retention. Labour market dynamics support the inclusion of these variables. The unemployment rate measures the proportion of the labour force that is actively seeking employment but is unable to find work. A higher unemployment rate indicates greater job scarcity and higher levels of unemployment. In a competitive labour market, high levels of unemployment and layoffs can create downward pressure on wages and increase job insecurity. Employers may become more cautious.

The layoff rate represents the rate at which workers are involuntarily let go by their employers. A higher layoff rate suggests increased job instability. Similar to the unemployment rate, the layoff rate can be included as a predictor in regression analyses to assess its influence on job retention.

High unemployment and layoff rates can also cause increased job turnover as workers become more willing to accept job offers or switch jobs to secure employment in times of economic downturn. This churn in the labour market can impact the stability of the workforce and impact job retention as well. Ultimately, if there is a negative correlation between either of these rates and job retention, it suggests that higher levels of unemployment and layoffs are associated with lower job retention, ceteris paribus.

Rationale: Owner Ethnicity and Gender These variables capture demographic characteristics that might influence job retention outcomes or PPP allocation. Theories supporting the inclusion of these variables draw from labour market discrimination and inequities in access to financial resources.

Both of these variables reflect the diversity within the ownership structures of participating businesses. They may influence job retention due to factors like discrimination or network effects. Network effects can influence access to resources and opportunities, including access to financial support programs like the PPP. In some cases, certain ethnic or gender groups may have stronger networks or connections within specific industries or communities, which could enhance their access to information about PPP funding, assistance in navigating the application process, or access to financial institutions facilitating PPP loans.

If there are significant coefficients associated with ethnicity and gender variables, it suggests that these demographic characteristics are associated with differences in job retention outcomes among businesses that received PPP loans. It may also indicate pre-existing minorities in entrepreneurial spheres in the U.S.

Discrimination based on ethnicity and gender can affect access to financial resources, business networks, and opportunities. Minority-owned and women-owned businesses may face systemic barriers that impact their ability to access PPP loans and implement effective job retention strategies. Discrimination based on ethnicity and gender can affect access to financial resources, business networks, and opportunities. Minority-owned and women-owned businesses may face systemic barriers that impact their ability to access PPP loans and implement effective job retention strategies.

Ultimately, including these variables in regression analysis will help in identifying the role of demography in shaping job retention outcomes following PPP loan approval, shedding light on potential disparities and opportunities for policy intervention to promote inclusive economic recovery in the future.

Rationale: Worker Types Including worker type in the regression would require the creation of dummy variables for each category of employment. This would allow for an assessment of how changes in the composition of employment types impact job retention outcomes. The coefficients associated with each dummy will indicate the impact of being in a particular category of employment on the number of jobs retained, all else equal.

Economic theory suggests that labour markets can be segmented into different sectors or categories, each with its distinct characteristics and dynamics. For example, private wage and salary workers may have different job retention strategies and responses to economic shocks compared to self-employed individuals or government workers.

The effectiveness of PPP funding in retaining jobs may vary across different types of employment. For instance, businesses employing private wage and salary workers may be more likely to use PPP funds to maintain their workforce, whereas self-employed individuals or unpaid family workers may have different financial needs or employment arrangements that influence their response to PPP funding.

If certain categories of employment are found to have a significant impact on job retention outcomes, it suggests that the composition of the workforce plays a role in determining the effectiveness of PPP funding in preserving jobs. The economic theory supports this connection through the concepts of labour market segmentation and differential responses to economic policies and interventions across different segments of the workforce.

Rationale: Business Type Agency theory suggests that the legal structure of a business can influence the incentives and behaviour of its owners and managers. For example, in a corporation, where ownership is separated from management, managers may have different priorities and incentives compared to sole proprietors or LLC owners. This can affect how businesses utilize PPP funds and their decisions regarding job retention.

As cited earlier, Autor et al. (2022) found that PPP dollars often got caught up with shareholders or creditors rather than workers. This might differ based on business structure.

The legal structure of a business can also affect its access to financial resources and its ability to weather economic downturns. For instance, small businesses structured as sole proprietorships or partnerships may face greater financial constraints compared to larger corporations with access to capital markets.

It was found earlier that sole proprietorships, for example, tended to see less job retention. This means that the legal structure of a business may influence its reliance on PPP funding and its capacity to retain jobs.

Organizational behaviour may also play a role in this. For example, corporations may have more formalized HR policies and procedures in place for retaining employees compared to smaller sole proprietorships. If certain business types are found to have a significant impact on job retention outcomes, it suggests that the legal structure of a business plays a role in determining its response to PPP funding and its ability to retain jobs during economic uncertainty.

To narrow the list of business types in the data, we consider the four most heavily represented in the data and group the rest into an "Other" category. The four include Corporation, Subchapter S Corporation, Limited Liability Company (LLC), and Non-

Profit Organization.

Rationale: Industry Different industries may experience varying degrees of disruption and economic impact from external shocks, such as the COVID-19 pandemic and associated lockdown measures. For example, industries such as hospitality, tourism, and retail may have been more severely affected by lockdowns compared to sectors like healthcare or information technology. Therefore, the availability of PPP funding may have a more significant impact on job retention in heavily affected industries.

Likewise, Industries with more elastic demand may experience larger fluctuations in employment levels in response to changes in economic conditions. For instance, industries producing non-essential goods or services may see a sharper decline in demand during economic downturns, leading to greater pressure on businesses to reduce their workforce.

If certain industries are found to have a significant impact on job retention outcomes, it suggests that the industry's economic characteristics and response to external shocks play a role in determining its ability to retain jobs with PPP funding.

For a more focused analysis on the efficacy of the PPP on job retention and considering the need to narrow down the list of industries in the data, we prioritize based on sectors most directly impacted by shutdowns and those critical to immediate economic stability. The following is a condensed list reflecting both high impact from COVID-19 and potential for significant PPP impact:

Industry_Accommodation and Food Services: This sector was among the hardest hit by COVID-19 due to lockdowns and restrictions on dining.

Industry_Retail Trade (focused on essential goods and services): Small retail businesses, especially those providing essential goods like food, were crucial in maintaining community stability.

Industry_Health Care and Social Assistance: Essential for workers, especially healthcare and other essential service workers, to remain employed during the pandemic.

Industry_Arts, Entertainment, and Recreation: Significantly impacted by COVID-19 due to public gathering restrictions.

Industry_Personal and Laundry Services: These services faced significant downturns due to reduced demand with more people staying home.

This streamlined list focuses on sectors that are varied enough to represent different aspects of the economy but are all significantly impacted by the pandemic and therefore likely candidates for PPP impact analysis. Including these industries in the regression model will help to capture a broad spectrum of PPP effects across diverse economic activities, focusing on areas where job retention would be most critical.

Rationale: CARES Act Funding Per Capita Fiscal federalism refers to the division of fiscal responsibilities and resources between different levels of government, such as federal, state, and local governments. Under the CARES Act, the federal government allocated funding to states to support various programs, including the PPP. Analyzing state shares of CARES Act funding helps to understand how these resources were distributed across different states and how they were used to support job retention efforts.

If there is a significant relationship between these variables, it suggests that federal support provided through programs outside of just the PPP influences job retention outcomes at the state level. This essentially helps to understand how other sources of support impact job retention.

7.3 Multivariate Regression Analysis

Before fitting a multivariate regression, we check each continuous variable of interest for violations of regression assumptions such as normality and linearity. If violations exist, we consider correcting them using transformations. We find that all variables except for Unemployment Rate and Layoff Rate are right-skewed. Logarithmic transformations are applied to correct the right skew in these continuous variables. These include Employed, Population Over 15, and CARES Act Per 100,000.

As such, any mention of these three variables hereafter refers to their log value. For brevity, "log" is omitted in reference to these variables in written interpretation.

Having addressed violations, we now run the first few specifications and generate Table 1 to compare estimates from the regressions.

7.4 Model Description for Table 3

Before presenting the results of the regression analysis, a brief description is provided of each regression model that is estimated including why certain predictors were grouped together.

Since the outcome variable is Jobs Retained Per 100,000 People, each regression model will aim to explain variations in this outcome based on different sets of predictors. The base model is the simple linear regression fit earlier with just Loan Amount Per 100,000 as a predictor on the outcome. Each model includes the Loan Amount variable.

Regression 1: Base Model

$$JobsRetai\widehat{ned}Per100,000_i = \hat{\beta}_0 + \hat{\beta}_1\ LoanAmountPer100,000_i$$

Regression 2: Basic Economic Factors: To understand how basic economic indicators relate to jobs retained.

$$\begin{aligned}\text{Jobs Retai}\widehat{\text{ned }}\text{Per 100,000}_i = \hat{\beta}_0 & \\ & + \hat{\beta}_1 \times \text{Loan Amount Per 100,000}_i \\ & + \hat{\beta}_2 \times \text{Employed}_i \\ & + \hat{\beta}_3 \times \text{Unemployment Rate}_i \\ & + \hat{\beta}_4 \times \text{Layoff Rate}_i\end{aligned}$$

Regression 3: Impact of CARES Act Funding: Includes interaction of CARES Act Funding x Unemployment Rate. Aims to see how the effect of CARES Act funding on job retention varies with the unemployment rate.

$$
\begin{aligned}
\widehat{JobsRetainedPer100,000_i} = \hat{\beta}_0 \\
&+ \hat{\beta}_1 \times \text{Loan Amount Per 100,000}_i \\
&+ \hat{\beta}_2 \times \text{CARES Act Funding Per 100,000}_i \\
&+ \hat{\beta}_3 \times \text{Population Over 15}_i \\
&+ \hat{\beta}_4 \times \text{Unemployment Rate}_i \\
&+ \hat{\beta}_5 \times \text{Unemployment x CARES}_i
\end{aligned}
$$

Regression 4: Demographic Focus: Dummies for Ethnicity and Gender to explore the impact of owner demographics on jobs retained, controlling for employment levels.

$$
\begin{aligned}
\widehat{\text{Jobs Retained Per 100,000}_i} = \hat{\beta}_0 \\
&+ \hat{\beta}_1 \times \text{Loan Amount Per 100,000}_i \\
&+ \hat{\beta}_2 \times \text{Employed}_i \\
&+ \hat{\beta}_3 \times \text{Ethnicity\_American Indian or Alaska Native}_i \\
&+ \hat{\beta}_4 \times \text{Ethnicity\_Asian}_i \\
&+ \hat{\beta}_5 \times \text{Ethnicity\_Black or African American}_i \\
&+ \hat{\beta}_6 \times \text{Ethnicity\_Hispanic}_i \\
&+ \hat{\beta}_7 \times \text{Gender\_Male}_i
\end{aligned}
$$

Regression 5: Industry and Business Type Influence: Dummies for Industry and Business Type to investigate how industry and business type relate to job retention.

$$
\begin{aligned}
\widehat{JobsRetainedPer100,000}_i = \hat{\beta}_0 \\
&+ \hat{\beta}_1 \times \text{Loan Amount Per 100,000}_i \\
&+ \hat{\beta}_2 \times \text{Industry\_Restaurant}_i \\
&+ \hat{\beta}_3 \times \text{Industry\_Amusement Park}_i \\
&+ \hat{\beta}_4 \times \text{Industry\_Dry Cleaning and Laundry}_i \\
&+ \hat{\beta}_5 \times \text{Industry\_Child Day Care}_i \\
&+ \hat{\beta}_6 \times \text{Business\_Corporation}_i \\
&+ \hat{\beta}_7 \times \text{Business\_Subchapter S Corporation}_i \\
&+ \hat{\beta}_8 \times \text{Business\_LLC}_i \\
&+ \hat{\beta}_9 \times \text{Business\_Non-Profit}_i
\end{aligned}
$$

Comparison of Regression Estimates in Table 3 The table provided showcases results from five different Ordinary Least Squares (OLS) regression models, each employing various independent variables to explore the effectiveness of the PPP in minimizing unemployment. This analysis is situated within the broader economic context of understanding how different factors, including loan amounts, employment rates, layoff rates, and demographic characteristics, influence unemployment levels. By examining these relationships, we can gauge the PPP's impact on employment during the economic downturn prompted by the COVID-19 pandemic.

The decision to run multiple regression models stems from the need to understand the program's effectiveness under various economic conditions and across different demographics and industries.

Model 1 examines the basic relationship between PPP loans and jobs retained without controlling for any other factors besides the loan amount per 100,000. This model sets the baseline for comparison. Model 2 introduces controls for (log) employment levels and the unemployment rate, acknowledging that the local economic environment might influence

Table 3: OLS Regression Results for Models 1-5

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Loan Amount Per 100,000	-0.000*	0.000***	0.000***	0.000***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Employed		-0.777***		-0.772***	
		(0.004)		(0.004)	
Unemployment Rate		-0.004	-0.368		
		(0.011)	(0.304)		
Layoff Rate		0.021***			
		(0.006)			
CARES Act Per 100,000			-0.987**		
			(0.424)		
Population Over 15			-0.936***		
			(0.005)		
CARESxUnemployment			0.038		
			(0.029)		
Ethnicity_AI/AN				-0.062**	
				(0.031)	
Ethnicity_Asian				-0.228***	
				(0.010)	
Ethnicity_Black				-0.063***	
				(0.016)	
Ethnicity_Hispanic				-0.129***	
				(0.009)	
Gender_Male				0.005	
				(0.006)	
Industry_Restaurant					-0.097***
					(0.009)
Industry_Amusement Park					-0.111
					(0.112)
Industry_Dry Cleaning/Laundry					-0.170***
					(0.059)
Industry_Child Day Care					-0.121***
					(0.023)
Business_Corporation					0.000***
					(0.000)
Business_Subchapter S Corp.					0.000***
					(0.000)
Business_LLC					0.000***
					(0.000)
Business_Non Profit					0.000***
					(0.000)
const	1.484***	12.624***	25.539***	12.630***	1.498***
	(0.004)	(0.140)	(4.498)	(0.057)	(0.005)
Observations	48138	48138	48138	48138	48138
R2	0.000	0.448	0.503	0.455	0.003
Adjusted R2	0.000	0.448	0.503	0.455	0.003
F Statistic	2.761*	9769.224***	9732.485***	5737.901***	29.147***

Notes: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors in parentheses.

the effectiveness of PPP loans. Model 3 adds variables related to layoffs and the (log) CARES Act funding, exploring if prior layoffs or additional government assistance modify the PPP's impact. Model 4 further includes demographic variables for owners involved in the program. The interaction between (log) CARES Act funding and unemployment considers that the PPP's effectiveness might vary across different populations and in response to combined governmental interventions. Model 5 extends the analysis to the industry level and business types, acknowledging that the PPP's impact might differ across sectors and organizational structures.

Of the models in Table 3, Model 3 emerges as the preferred specification for several reasons. Firstly, it incorporates economic indicators (layoff rates and (log) CARES Act funding) that directly relate to the conditions under which PPP loans were intended to operate. Secondly, its addition of the CARES Act as a separate variable allows for an examination of how direct financial support interacts with loan-based support in job retention. Finally, its reasonable complexity balances between Model 2's simplicity and Models 4 and 5's granularity, which might introduce overfitting or obscure broader trends.

7.4.1 Evaluating Table 3 Regressions

To assess the performance of these regressions, several measures can be considered:

- R-squared and Adjusted R-squared: These metrics indicate the proportion of variance in the dependent variable explained by the model. Higher values suggest a better fit. Adjusted R-squared accounts for the number of predictors in the model, providing a more accurate assessment for models with different numbers of variables. Model 3's R-squared values indicate a good fit, especially when compared to the simpler Model 1, suggesting that including economic conditions significantly improves our understanding of PPP effectiveness.

- Standard Errors and P-values: They assess the precision of the coefficient estimates and the statistical significance of each predictor, respectively. Variables with low p-values (¡ 0.05) are considered to have a statistically significant impact on the dependent variable. In Model 3, the CARES Act funding and layoff rate have significant p-values, highlighting their importance in job retention.

- F-statistic: It tests the overall significance of the regression model. An F-statistic below 12 might suggest that the model is not a good fit for the data, potentially indicating that key variables are missing or that a non-linear model might be more appropriate. It makes sense then that the F-statistic is below 12 for Model 1 (it is missing key variables) but not the case for Models 2, 3, 4, or 5.

7.4.2 Interpretation of Table 3 Results

The regression results provide insights into the PPP's effectiveness and its interaction with other economic factors. For example, the significance of the (log) CARES Act funding in Model 3 suggests that direct financial support plays a critical role in job retention alongside PPP loans. The negative coefficient for the unemployment rate in Model 2 hints that in areas with higher unemployment, PPP loans were less effective, possibly due to broader economic challenges.

The analysis of PPP's impact on job retention through OLS regressions reveals complex interactions between loan assistance, economic conditions, demographic factors, and industry characteristics. The chosen regression models, particularly Model 3, shed light on the multifaceted nature of economic recovery efforts and the pivotal role of combined governmental interventions. The statistical evaluations and tests suggest that while the PPP was an essential component of the response to the economic downturn, its effectiveness was nuanced and contingent upon a constellation of factors.

7.5 Model Description for Table 4

Similar to Table 3, Loan Amount Per 100,000 People is included in each regression model.

Regression 6: Worker Types and Economic Health: To assess how different types of employment influence job retention amid varying economic conditions.

$$
\begin{aligned}
\widehat{JobsRetainedPer100,000_i} = \hat{\beta}_0 \\
&+ \hat{\beta}_1 \times \text{Loan Amount Per 100,000}_i \\
&+ \hat{\beta}_2 \times \text{Worker\_Gov Workers}_i \\
&+ \hat{\beta}_3 \times \text{Worker\_Self Employed}_i \\
&+ \hat{\beta}_4 \times \text{Worker\_Family Workers}_i \\
&+ \hat{\beta}_5 \times \text{Unemployment Rate}_i \\
&+ \hat{\beta}_6 \times \text{Layoff Rate}_i
\end{aligned}
$$

Regression 7: Loan Amount Effectiveness: Includes interaction Loan Amount x Unemployment Rate to evaluate how the effectiveness of loan amounts in retaining jobs interacts with the local unemployment rate.

$$
\begin{aligned}
\widehat{JobsRetainedPer100,000_i} = \hat{\beta}_0 \\
&+ \hat{\beta}_1 \times \text{Loan Amount Per 100,000}_i \\
&+ \hat{\beta}_2 \times \text{Unemployment Rate}_i \\
&+ \hat{\beta}_3 \times \text{CARES Act Per 100,000}_i \\
&+ \hat{\beta}_4 \times \text{Loan x Unemployment}_i
\end{aligned}
$$

Regression 8: CARES and demography: To understand the nuanced connections between owner demographics and federal funding.

$$
\begin{aligned}
\widehat{JobsRetainedPer100,000_i} = \hat{\beta}_0 \\
&+ \hat{\beta}_1 \times \text{Loan Amount Per 100,000}_i \\
&+ \hat{\beta}_2 \times \text{CARES Act Per 100,000}_i \\
&+ \hat{\beta}_3 \times \text{Ethnicity\_AI/AN}_i \\
&+ \hat{\beta}_4 \times \text{Ethnicity\_Asian}_i \\
&+ \hat{\beta}_5 \times \text{Ethnicity\_Black}_i \\
&+ \hat{\beta}_6 \times \text{Ethnicity\_Hispanic}_i \\
&+ \hat{\beta}_7 \times \text{Gender\_Male}_i \\
&+ \hat{\beta}_8 \times \text{Pop Over 15}_i \\
&+ \hat{\beta}_9 \times \text{Unemployment Rate}_i \\
&+ \hat{\beta}_{10} \times \text{CARES x Unemployment}_i
\end{aligned}
$$

Regression 9: Comprehensive Model: All covariates to build a comprehensive model incorporating lessons learned from the previous regressions to see which factors have the most significant impact on job retention.

$$
\begin{aligned}
\widehat{\textit{Jobs Retained Per}}\ 100,000_i = \hat{\beta}_0 \\
&+ \hat{\beta}_1 \times \text{Loan Amount Per 100,000}_i \\
&+ \hat{\beta}_2 \times \text{Employed}_i \\
&+ \hat{\beta}_3 \times \text{Unemployment Rate}_i \\
&+ \hat{\beta}_4 \times \text{Layoff Rate}_i \\
&+ \hat{\beta}_5 \times \text{CARES Act Per 100,000}_i \\
&+ \hat{\beta}_6 \times \text{Population Over 15}_i \\
&+ \hat{\beta}_7 \times \text{CARES x Unemployment}_i \\
&+ \hat{\beta}_8 \times \text{Ethnicity\_AI/AN}_i \\
&+ \hat{\beta}_9 \times \text{Ethnicity\_Asian}_i \\
&+ \hat{\beta}_{10} \times \text{Ethnicity\_Black}_i \\
&+ \hat{\beta}_{11} \times \text{Ethnicity\_Hispanic}_i \\
&+ \hat{\beta}_{12} \times \text{Gender\_Male}_i \\
&+ \hat{\beta}_{13} \times \text{Industry\_Restaurant}_i \\
&+ \hat{\beta}_{14} \times \text{Industry\_Amusement Park}_i \\
&+ \hat{\beta}_{15} \times \text{Industry\_Dry Cleaning/Laundry}_i \\
&+ \hat{\beta}_{16} \times \text{Industry\_Child Day Care}_i \\
&+ \hat{\beta}_{17} \times \text{Business\_Corporation}_i \\
&+ \hat{\beta}_{18} \times \text{Business\_Subchapter S Corp.}_i \\
&+ \hat{\beta}_{19} \times \text{Business\_LLC}_i \\
&+ \hat{\beta}_{20} \times \text{Business\_Non Profit}_i \\
&+ \hat{\beta}_{21} \times \text{Worker\_Gov Workers}_i \\
&+ \hat{\beta}_{22} \times \text{Worker\_Self Employed}_i \\
&+ \hat{\beta}_{23} \times \text{Worker\_Family Workers}_i \\
&+ \hat{\beta}_{24} \times \text{Loan x Unemployment}_i
\end{aligned}
$$

Table 4: OLS Regression Results for Models 6-9

	Model 6	**Model 7**	**Model 8**	**Model 9**
Loan Amount Per 100,000	-0.000	0.000	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Employed				3.497***
				(0.038)
Unemployment Rate	0.000	0.051***	-0.352	-0.253
	(0.015)	(0.009)	(0.302)	(0.283)
Layoff Rate	0.036***			0.012**
	(0.008)			(0.005)
CARES Act Per 100,000		0.987***	-0.946**	-0.673*
		(0.010)	(0.422)	(0.395)
Population Over 15			-0.928***	-4.300***
			(0.005)	(0.037)
CARESxUnemployment			0.036	(0.023)
			(0.029)	(0.027)
Ethnicity_AI/AN			-0.057*	-0.041
			(0.029)	(0.027)
Ethnicity_Asian			-0.217***	-0.204***
			(0.010)	(0.009)
Ethnicity_Black			-0.046***	-0.034**
			(0.015)	(0.014)
Ethnicity_Hispanic			-0.089***	-0.067***
			(0.009)	(0.008)
Gender_Male			0.002	0.000
			(0.005)	(0.005)
Industry_Restaurant				-0.033***
				(0.006)
Industry_Amusement Park				0.158**
				(0.073)
Industry_Dry Cleaning/Laundry				-0.097**
				(0.038)
Industry_Child Day Care				0.009
				(0.015)
Business_Corporation				-0.000***
				(0.000)
Business_Subchapter S Corp.				0.000***
				(0.000)
Business_LLC				0.000***
				(0.000)
Business_Non Profit				0.000***
				(0.000)
Worker_Gov	0.312***			5.639***
	(0.043)			(1.049)
Worker_Self Employed	0.312***			5.639***
	(0.043)			(1.049)
Worker_Family	0.312***			5.639***
	(0.043)			(1.049)
LoanxUnemployment		0.000		0.000***
		(0.000)		(0.000)
const	0.312***	-9.848***	25.066***	5.639***
	(0.043)	(0.167)	(4.472)	(1.049)
Observations	48138	48138	48138	48138
R2	0.002	0.169	0.509	0.583
Adjusted R2	0.002	0.169	0.509	0.583
F Statistic	39.748***	2445.160***	4981.720***	3962.234***

Notes: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$. Standard errors in parentheses.

Comparison of Regression Estimates in Table 4 The table provided summarizes the outcomes of four OLS regression models that control for different covariates such as worker type and an interaction between loan amount per 100,000 and unemployment rate.

The decision to run these regressions stems from the need to dissect the multifaceted impact of the PPP alongside other factors such as demographic characteristics, government interventions, and industry-specific effects on job retention. Model 6 introduces worker type dummies and uses private wage workers as a reference category. This aims to highlight any patterns of connection between specific sectors and the effectiveness of PPP loan approval on jobs retained per 100,000. Model 7 introduces additional government variables and the interaction between loan amounts and unemployment rates, probing deeper into how various types of government support and economic conditions interact with the effectiveness of PPP loans. Model 8 expands the analysis of Model 3 by including ethnicity and gender dummies. Model 9 provides a compelte model, controlling for all covariates in the data the model.

Given the complexity of economic interactions and the diverse factors affecting job retention, Model 9 emerges as the preferred specification. It comprehensively includes demographic, economic, and industry-specific variables, offering a nuanced analysis of the PPP's effectiveness. Its robustness is evidenced by the highest R2 and Adjusted R2 values among the models, indicating that it explains a significant portion of the variance in jobs retained. Additionally, the inclusion of a wider range of variables allows for a detailed exploration of how various factors interact with each other in the context of job retention.

However, given that the R2 and Adjusted R2 of Model 8 is quite close in magnitude and uses singificantly fewer predictors, we can determine that the bulk of Model 9's predictive power comes from the subset of variables used in Model 8.

7.5.1 Evaluating Table 4 Regressions

The performance of these regressions can be evaluated using several key metrics:

- R2 and Adjusted R2: These metrics measure the proportion of the variance in the dependent variable that's explained by the independent variables in the model. A higher R2 value indicates a model that better fits the data. However, R2 alone doesn't penalize model complexity. The Adjusted R2 accounts for the number of variables and the sample size, making it a more reliable metric for model comparison. Model 9, with the highest Adjusted R2, is considered to have the best fit among the models presented.

- F Statistic: This tests the overall significance of the model. A higher F Statistic indicates that the model is statistically significant. The significantly high F Statistic values in all models, particularly Model 9, suggest that the variables collectively have a substantial impact on jobs retained.

- Residual Std. Error: This measures the average distance between the observed values and the values predicted by the model. Lower values indicate a better fit. Model 9 has the lowest Residual Std. Error, further supporting its selection.

7.5.2 Interpretation of Table 4 Results

The regression results provide insightful revelations about the PPP's impact on job retention across different demographics and industries. The significance of variables such as the (log) CARES Act Per 100,000 across multiple models highlights the positive impact of government interventions on job retention. The demographic variables in Models 8 and 9 suggest that the PPP's effectiveness varied significantly across different racial and ethnic groups. Industry-specific variables in Model 9 indicate that the impact of the PPP also varied markedly across different sectors.

The analysis demonstrates the nuanced efficacy of the PPP in sustaining employment across various demographics and industries during economic downturns. Model 9, with its comprehensive inclusion of variables and superior performance metrics, offers the most detailed insights, underscoring the complexity of economic recovery efforts and the importance of tailoring policy interventions to specific sectors and communities. The results underscore the PPP's role in job retention while highlighting the need for targeted policies to address disparities in its effectiveness across different groups and industries.

7.5.3 Final Preferred Specification

Model 9 from Table 2 is the standout model due to its highest R^2 and Adjusted R^2 values of 0.583, indicating it explains more than half of the variance in jobs retained per 100,000 people, which is considerably higher than the other models. It features a broad array of statistically significant predictors, including demographic factors, industry specifics, and policy measures like the CARES Act. This suggests a strong and diverse set of factors influencing job retention. The Fstatistic of 3962.234***, significant at the 1% level, further validates the model's overall statistical significance and its superior fit compared to the others.

In summary, Model 9 provides the most comprehensive and statistically robust framework for understanding the factors influencing job retention, making it the best model among those presented.

8 Causal Analysis

A research design that creates credible causal inference is difference-in-difference. This method posits that if outcomes across treated and untreated groups move in parallel in the absence of treatment, then the divergence of the post-treatment path from this trend is differential and indicates the treatment effect.

In order to use a DiD design, it is critical to meet the parallel trends assumption. It also requires that we have baseline data. DiD considers four groups; a control and treated group in a pretreatment period and a control and treated group in a post-period. Since this dataset does not provide any sort of baseline measure for pre-PPP loan approval numbers of jobs retained, we are unable to use a DiD design.

However, we are able to use an Instrumental Variable (IV). The IV approach is particularly useful when dealing with endogeneity issues in a regression model. Endogeneity can arise from omitted variable bias or reverse causality. In this data, it is possible that we experience both of these issues.

For example, it is possible that higher employment causes higher job retained per

100,000 but it is also possible that the opposite is true; that is, that higher jobs retained per 100,000 causes higher employment by state. This is an example of reverse causality.

As such, we introduce an IV. An IV is a variable that is correlated with the endogenous explanatory variable but uncorrelated with the error term. This is effective because endogeneity is caused by correlation between explanatory variables and the error term. Adding this exogeneity helps reduce the reverse causality.

The key challenge in using IV is finding a suitable instrument that satisfies these conditions. However, the right instrument can help you estimate causal effects even when a direct causal inference is complicated by these issues

8.1 Instrumental Variable Selection

It is suspected that that the loan amount is endogenous—perhaps because regions with higher initial unemployment rates may have received more funding. In this case, the IV needs to influence the loan amount without directly affecting the number of jobs retained (aside from its effect through the loan amount).

If loan allocations were influenced by political considerations, such as the political affiliation of local representatives or their involvement in the allocation process, this could serve as an instrument for the "Loan Amount Per 100,000 People," assuming these political factors do not directly affect job retention rates.

This implies that areas represented by certain political affiliations or with more direct involvement in the allocation process might receive more or fewer loans per 100,000 people.

8.1.1 First Stage of 2SLS (Two Stage Least Squares

In the first stage, we regress the endogenous explanatory variable on the instrument and all other control variables in the model. The goal here is to predict the values of the endogenous variable using the instrument.

As such, we first merge new data to add the variable Party Control which indicates which political party (Democrat or Republican) had control over the state during the period of the dataset. Data is scraped from Wikipedia: Political party strength in U.S.

states.

Using statistical methods, we find that the F-statistic is 9598.

8.1.2 Second Stage of 2SLS

In the second stage, we regress Jobs Retained Per 100,000 People on the predicted values of Loan Amount Per 100,000 People from the first stage, along with the other control variables.

Assuming the IV is valid, the coefficient from the second-stage regression provides an estimate of the causal effect of interest. In other words, the impact, on average, of a higher loan range midpoint value is approximately 0.3 percentage points jobs retained per 100,000.

The coefficient for predicted_loan_amount is positive (0.0003) and highly significant (p-value ¡ 0.000), suggesting a positive relationship between the loan amount and jobs retained. This implies that, on average and holding other factors constant, an increase in loan amount per 100,000 people is associated with an increase in the number of jobs retained per 100,000 people.

Many of the control variables are statistically significant with p-values ¡ 0.05, indicating that they have a meaningful association with the number of jobs retained. For instance, Employed, Unemployment Rate, Ethnicity_Asian, and Gov_Workers show significant effects. This underscores the importance of these factors in explaining the variation in jobs retained.

The R-squared value of 0.478 suggests that approximately 47.8% of the variability in the number of jobs retained per 100,000 people is explained by the model. This is a substantial proportion, which indicates that predictive power was not significantly impacted with the use of the IV.

The F-statistic is 2939, with a p-value of 0.00, indicating that the model is statistically significant. This means the variables collectively have a significant effect on the number of jobs retained. We also see that the F-statistic from the first stage is above 10, which means that the IV is not weak.

Table 5: OLS Regression Results Using IV

	Second Stage
predicted_loan_amount	0.000*
	(0.000)
Employed	-19.918***
	(0.247)
Unemployment Rate	-0.053***
	(0.011)
Layoff Rate	0.046***
	(0.006)
CARES Act Per 100,000	-14.364***
	(0.146)
Population Over 15	7.676***
	(0.131)
Ethnicity_AI/AN	-0.148***
	(0.030)
Ethnicity_Asian	-2.212***
	(0.023)
Ethnicity_Black	1.184***
	(0.020)
Ethnicity_Hispanic	0.327***
	(0.010)
Gender_Male	-0.024***
	(0.006)
Industry_Restaurant	0.366***
	(0.008)
Industry_Amusement Park	1.814***
	(0.083)
Industry_Dry Cleaning/Laundry	0.640***
	(0.043)
Industry_Child Day Care	1.730***
	(0.024)
Business_Corporation	0.000***
	(0.000)
Business_Subchapter S Corp.	0.000***
	(0.000)
Business_LLC	0.000***
	(0.000)
Business_Non Profit	0.000***
	(0.000)
const	81.450***
	(0.801)
Observations	48138
R2	0.478
Adjusted R2	0.478
F Statistic	2939.158***

Notes: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$. Standard errors in parentheses.

9 Machine Learning Methods

To decide which Xs to include in the regression tree, we refer to our model specifications. To best justify the inclusion of variables, we decide to run the regression tree using using all the X's from the preferred specification. From earlier, we identified the preferred specification as Model 9, which included all explanatory variables in the regression table.

Instead of using Loan Amount Per 100,000 People, we use the IV created in the previous section, predicted_loan_amount. The IV "predicted_loan_amount" should be chosen because it correlates with the potentially endogenous regressor but not with the error term, thus providing a cleaner estimation of the causal effect on the outcome variable.

First, we discuss the objective function and re-write the objective function for the regression tree we run.

Objective functions in regression trees aim to minimize the variance in each split of the tree. In other words, the goal is to group together observations that are similar to each other in terms of the outcome variable, Jobs Retained Per 100,000 People, by making splits on the predictor variables. The goal is to achieve leaves where the outcome variable values are as close to each other as possible, minimizing the prediction error within each leaf.

For a regression tree, this is measured by the Mean Squared Error (MSE) in each node, which the tree algorithm aims to minimize at each split. The MSE for a node can be represented as:

$$MSE_n = \frac{1}{N_n} \sum_{i \in N_n} (y_i - \bar{y}_n)^2$$

where:

- (N_n) is the number of observations in node (n)
- (y_i) is the actual value of the target variable for observation (i),
- The mean target value in node n is denoted as y-bar_n

For a regression tree with the variables below and as will be seen in the regression tree, the objective function aims to make splits on variables like ”predicted$_{l}oan_{a}mount,$”” $Employed,$”” $Unem$

$$
\begin{aligned}
\text{jobs retained per 100,000 people} = \frac{1}{N}\sum_{i=1}^{N}\Big(y_i - \Big(\hat{\beta}_0 \\
&+\hat{\beta}_1 \times \text{predicted\_loan\_amount}_i \\
&+\hat{\beta}_2 \times \text{Employed}_i \\
&+\hat{\beta}_3 \times \text{Unemployment Rate}_i\Big)\Big)^2
\end{aligned}
$$

Note: Recall from the note made earlier that the variable Employed was transformed so the value is log Employed.

Regularization parameters in regression trees include max_depth, min_samples_split, min_samples_leaf, etc. These parameters help prevent overfitting by restricting the size or depth of the tree.

max_depth limits the maximum depth of the tree. A deeper tree might capture more detailed patterns but risks overfitting. min_samples_split is the minimum number of samples required to split an internal node. Higher values prevent creating nodes that are too specific and might only apply to a few observations.

min_samples_leaf is the minimum number of samples required to be at a leaf node. It ensures that each leaf has enough observations to make a reliable prediction.

Adjusting these parameters affects model complexity and generalization. A more complex model might fit the training data well but perform poorly on unseen data (overfitting), while too simple a model might not capture important patterns (underfitting). More simply, overfitting reduces external validity

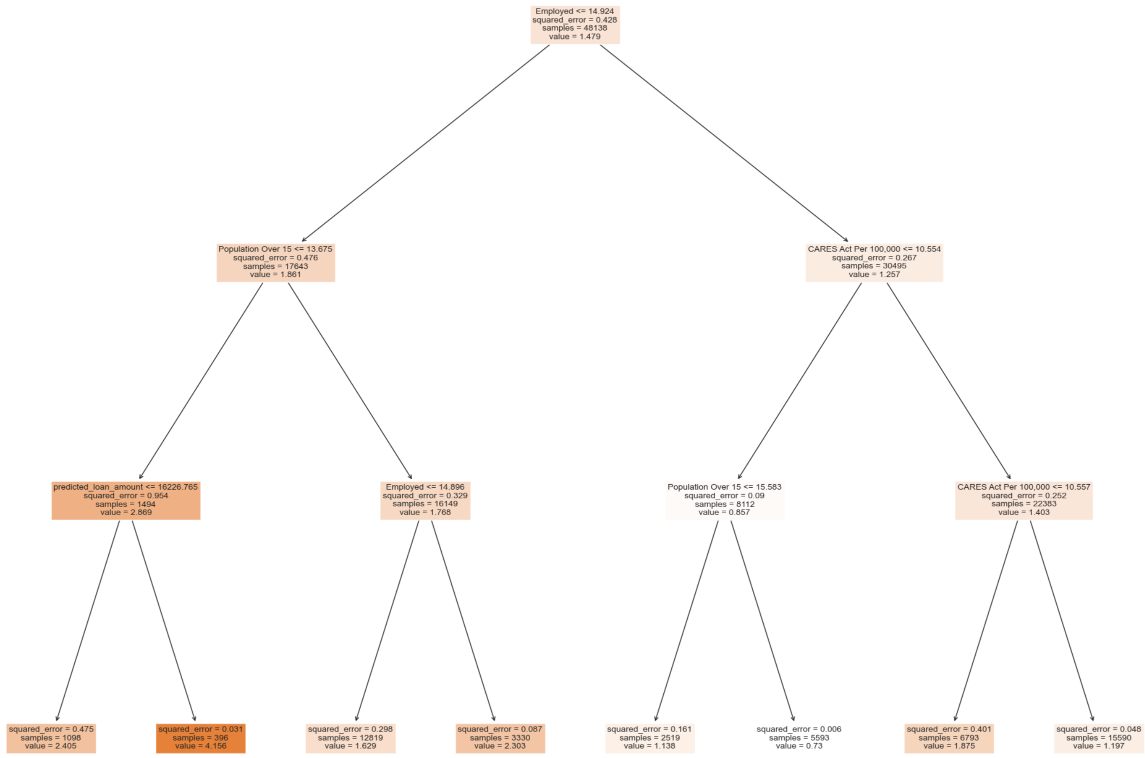

Figure 23: Visualization of regression tree.

9.1 Interpretation of Regression Tree

Having fit and plotted the tree, we can now interpret it by analyzing the splits it made. Each node in the tree represents a decision based on one of the explanatory variables, directing observations to the left or right child node. The leaf nodes provide the mean value of "Jobs Retained Per 100,000 People" for observations that end up in that leaf, based on the criteria defined by the path from the root.

Recall again that the variables Employed and Population Over 15 are transformed and are thus log Employed and log Population Over 15. As mentioned earlier, subsequent analysis omits "log" for the sake of brevity but all references to these variables is to their log value.

We see that the base node is Employed. This indicates that the initial and most significant split in the data is based on the "Employed" variable. This suggests that the number of employed individuals in a region is the most critical factor determining the "Jobs Retained Per 100,000 People." The choice of the base node reflects the algorithm's decision that, of all the variables considered, the "Employed" value for each state provides the most substantial differentiation in predicting the outcome variable.

From there, the tree branches out, with each level representing a further refinement based on another variable. The shading in the tree visualization serves as a guide to understanding the paths leading to different outcomes. Darker shades indicate the path through the tree that leads to higher values of "Jobs Retained Per 100,000 People." This visual cue helps identify which combinations of traits (i.e., values of the explanatory variables) are most associated with higher job retention rates.

The inclusion of this variable suggests that, among regions with similar numbers of employed citizens, the size or composition of the population over 15 can further differentiate job retention outcomes. This could reflect variations in workforce participation or differences in demographic pressures on employment.

Then the tree directs us to predicted_loan_amount. The use of this IV indicates that financial factors, as modeled through the predicted loan amount, play a crucial role in job retention after accounting for direct employment figures and demographic characteristics. This could suggest that access to financial support or the presence of robust financial activities is critical for sustaining employment. Ultimately, this helps answer the research question of the effectiveness of the PPP in minimizing unemployment during times of economic hardship.

Arriving at a leaf with a squared error of 0.031 indicates a relatively low level of error in the prediction of jobs retained per 100,000 people for observations that follow this path through the tree. The squared error metric here quantifies the average of the squares of the differences between the predicted values and the actual values in this leaf. A lower squared error in this context suggests that the combination of factors leading to this node—high or low levels of employment, specific demographic profiles, and certain ranges of predicted loan amounts— provides a comparatively accurate prediction of job retention rates. We also get a slightly higher R2 with the regression tree than with Model 9 from the OLS analysis at 0.585 compared to 0.583.

Ultimately, the path through the tree identifies a specific profile of regions characterized by their employment levels, demographic composition (specifically the population over 15), and predicted financial health (as approximated by predicted loan amounts) that correlate with certain job retention outcomes. The low squared error at the leaf indicates that for regions matching these criteria, the tree can predict job retention rates with relatively high accuracy. This sequence of variables highlights the importance of considering a multifaceted approach when analyzing job retention, recognizing the interplay between employment, demographics, and financial health.

9.2 Random Forest Model and Importance Matrix

We now run a Random Forest Model and interpret the importance matrix. We find the following.

9.2.1 Random Forest Model Results

Random Forest MSE (1.4220254152368643e-05): This very low MSE suggests that the Random Forest model has done an excellent job in predicting the "Jobs Retained Per 100,000 People". Random Forests work by building multiple decision trees (as specified by the n_estimators parameter) and averaging their predictions to reduce overfitting and improve predictive performance. The diversity among the trees, brought about by using random subsets of features and samples, often leads to more robust models compared to a single decision tree.

Regression Tree MSE (0.1777061127251885): This higher MSE, compared to the Random Forest, suggests that the single decision tree model is not predicting the target variable as accurately. Decision trees are prone to overfitting, especially if they are allowed to grow complex without restrictions. A single tree might capture noise in the training data that does not generalize well to unseen data, leading to larger errors.

The significantly lower MSE for the Random Forest model suggests it is a better fit for the data compared to the single regression tree. This is a common outcome because Random Forests are designed to overcome some of the key limitations of single decision trees, such as overfitting. However, an R2 of 0.999 for a random forest model is exceptionally high and could indeed indicate that the model is overfitting to the data. Overfitting occurs when a model learns the detail and noise in the training data to the extent that it performs poorly on new, unseen data. Given this, it might be the case that the regression tree is actually more useful and representative than the random forest model.

9.2.2 Importance Matrix Results

We can record the total amount that the reduction in mean squared error is due to splits over a given predictor, averaged over all trees. If splits on a given predictor (X) results in large reduction in mean squared error, that X is important.

We can therefore rank all of the Xs in the dataset based on how much they help reduce the error and create the importance matrix with this information below.

Ultimately, in the context of Random Forest and many tree-based models, feature importance is a way to understand which features contribute most to the model's predictions.

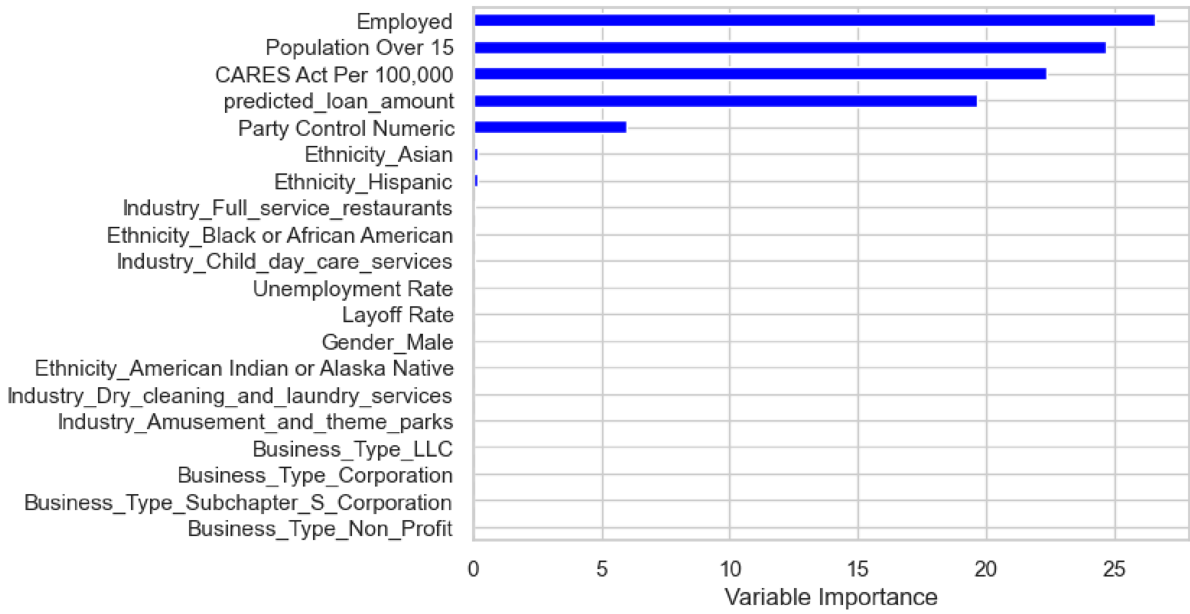

Figure 24: Importance matrix.

From the matrix we note that Employed, Population Over 15, CARES Act Per 100,000, predicted_loan_amount, and Party Control Numeric are more critical for the model's decisionmaking. This means they have more of an impact on the number of jobs retained per 100,000 people.

Note: Again, we remind the reader that the values for Employed, Population Over 15, and CARES Act Per 100,000 refer to their log values.

9.3 OLS Results vs. Regression Tree Results

Model 9 from the OLS regression outcomes offers a broader view of the determinants influencing job retention rates per 100,000 people. This model highlights significant predictors encompassing demographics, employment status, CARES Act funding levels, and industryspecific variables. Notably, it boasts an R^2 value of 0.583 and an Adjusted R^2 of the same magnitude, indicating a robust explanatory capacity. The model's F-statistic stands impressively at 3962.234***, affirming the collective impact of the variables and the model's overall statistical significance. These findings underscore the connection between government interventions, industry characteristics, and demographic elements in influencing job retention rates.

Conversely, the machine learning approach, particularly through the Random Forest model, exhibits a lower Mean Squared Error (MSE) of 1.4220254152368643e-05, suggesting a superior predictive accuracy for job retention rates. The feature importance matrix elucidates "Employed," "Population Over 15," "CARES Act Per 100,000," "predicted_loan_amount," and "Party Control Numeric" as pivotal predictors. This not only aligns with the OLS regression findings but also accentuates the critical role of employment levels and government support in job preservation.

The decision tree's initial node focusing on "Employed" suggests employment levels as the foremost determinant of job retention. Furthermore, the tree's branches reveal how combinations of variables, such as predicted_loan_amount and Population Over 15, interact to influence job retention rates. Such detailed insight into the conditional importance of variables, facilitated by the tree's splits, is beyond OLS regression's direct revelation capability.

Econometrically, OLS regression assumes a linear relationship between independent and dependent variables, striving to minimize the sum of squared residuals. It provides estimates of the independent variables' average effect on the dependent variable, assuming other factors remain constant. This method's economic intuition is direct: understanding the impact of variables like government support or employment levels on job retention within a linear context.

Regression trees, however, forgo linear relationship or homoscedasticity assumptions. They partition the predictor space into distinct regions to minimize variance within each, adeptly capturing complex interactions and non-linear effects. This model offers a nuanced understanding of various factors' conditional impacts on job retention. The economic intuition behind regression trees acknowledges the economy's complexity, recognizing that factors like financial support or employment may hold varying significance under different conditions.

While OLS regression elucidates the average effects of various factors, machine learning approaches like regression trees and Random Forest models expose complex interactions and conditional relationships. This comprehensive analysis not only highlights the PPP's pivotal role in job retention during economic downturns but also emphasizes the necessity for tailored policy interventions to mitigate disparities in its effectiveness.

10 Findings

This paper involved several forms of analysis, including exploratory data analysis, detailed geographic analysis, OLS regression analysis, and the use of machine learning tools such as a regression tree and random forest model.

Recall that this paper aimed to investigate the effectiveness of the PPP in minimizing unemployment in times of economic hardship. In particular, we measured the minimization of unemployment in the converse, using a measure of the jobs retained by a business due to the approval of a PPP loan.

At a baseline, we found patterns in participating businesses. More than half of the loans issued through the PPP were within the lowest range, 150,000 to 350,000 USD. We found that the most represented archetypes of participants included Corporations at 48% of all business types, White owners at 84% and male owners at 82%.

We also found that Non-Profit Organizations, the fourth most represented business type in the data, exhibited the highest average job retention while those working in more individual business structures such as independent contractors or sole proprietors retained the least.

From the geographic analysis, we determine that states in the central North tended to receive the lowest loan amounts per capita but had the highest number of jobs retained per capita. As such, there exists a clear divide between the North and South states in terms of jobs retained and loan amounts received. It also suggests that there are factors of great significance outside of financial aid that impact jobs retained.

California is an outlier as it demonstrates one of the highest loan amounts received per capita but the lowest jobs retained. Additionally, we note that states along the Upper East Coast maintain moderate standing across both jobs retained per capita and loan amount per capita received.

This analysis considers how other sources of government aid may have impacted the jobs retained by each state in addition to the PPP. In consideration of CARES Act Funding outside the PPP, we notice a positive correlation between states that received more from the CARES Act and their job retention. However, states receiving more support from other facets of CARES received less from the PPP, indicating potential trade-offs made at the federal level in terms of allocation strategies.

Furthermore, allocation seems to prioritize workers who fall under private wage and salary as opposed to those working in government or who are self-employed. This suggests a greater focus of the PPP on preserving employment in the larger private sector.

Taking all of this into consideration, we conduct an OLS regression analysis. After running nine different models, including the baseline model involving only loan amount as a predictor, we identify a preferred specification.

Model 9, which includes all predictors of interest, emerges with the highest R2 and

Adjusted R2, at 0.583. However, we note that Model 8 is an important contender as it includes far fewer predictors than Model 9 and maintains a comparable R2 and Adjusted R2 at 0.509.

This might indicate that the bulk of the predictive power in Model 9 is derived from the subset of explanatory variables used in Model 9. Despite this, we favour Model 9 due to its greater coefficient of determination.

With respect to causal inference and the ability of these models to determine any causal effect, we consider whether there might be an endogeneity concern. In concluding that there may exist reverse causality between loan amounts received by each state and their jobs retained, we introduce an instrumental variable.

The IV used is Party Control, which represents the party in power in each state. This is used to introduce further exogeneity and the variable created through this process is named predicted_loan_amount.

To further scrutinize the importance of explanatory variables of interest, we then run a regression tree. The model suggested by the algorithm indicates that the most important predictors are Employed, Population Over 15, and predicted_loan_amount. The model arrives at a prediction error of just 0.031, indicating that this combination of predictors provides a comparatively accurate prediction of job retention.

Finally, we consider the results of a random forest model and importance matrix. The random forest model provides an MSE of 0.00014 in comparison to 0.18 from the regression tree. The low value of the MSE from the random forest model indicates its high predictive power of jobs retained and that the single regression tree is not predicting as accurately. From the importance matrix, we see that the explanatory variables highlighted include Employed, Population Over 15, CARES Act Per 100,000, predicted_loan_amount, and Party Control Numeric.

We conclude by comparing the suggested models from the OLS regression, regression tree, and random forest model to answer our research question. In comparing the R2, we see that the regression tree provides a slightly higher power at 0.585 compared to Model 9's 0.583. The R2 from the regression tree indicates how well the combination of splits and variables used by the tree explains the variance in jobs retained across the entire dataset.

From the random forest model, we find an R2 of 0.99. This very much suggests overfitting, and so we return to comparing Model 9 and the regression tree.

In the end, we conclude that the most powerful predictors of the jobs retained per 100,000 people per state are the number of people who are employed, the number of people in the state that are over the age of 15 (and thus in the workforce), and the predicted loan amount received per 100,000 people by each state through the PPP.

Altogether, this evidence suggests some notable insights with regard to our research question. We find that the PPP's effectiveness varied across business types, potentially reflecting differences in operational costs, workforce size, and the nature of employment contracts across different ownership structures.

We also conclude that factors beyond financial aid significantly influence employment outcomes given the divide between the North and South in loan amounts per capita received and jobs retained per capita. This also indicates that the effectiveness of the PPP depended perhaps upon state-specific regulations that further affected the allocation of PPP funds or CARES Act funding.

We also find from our causal analysis that, on average and holding all other factors constant, an increase in the loan amount per 100,000 people is associated with an increase in the number of jobs retained per 100,000 people.

In summary, the PPP, as part of a broader suite of government interventions during the COVID-19 pandemic, has played a significant role in sustaining employment in various sectors and states. While the program effectively minimized unemployment in certain contexts, its impact is modulated by a range of factors including business type, geographic location, and the interplay with other government aid programs. The nuanced differences in job retention outcomes underscore the importance of tailored policy interventions that consider the unique economic landscapes and employment structures across states. This comprehensive analysis, therefore, underscores the PPP's role as a critical, albeit complex, mechanism in the government's effort to mitigate the economic fallout of the pandemic, highlighting areas for future refinement in policy design and implementation to enhance effectiveness in sustaining employment during economic crises.

11 Conclusion

In conclusion, this investigation into the effectiveness of PPP loans in preserving jobs for small businesses has already yielded valuable insights through the preliminary EDA. Notably, approximately 59% of loans fell within the range of 150,000 to 350,000 USD, with Corporations representing 48% of all business types. The dataset predominantly comprises White owners (84%), and predominantly male (82%). The Huntington National Bank emerged as the most frequent lender. It was found that average job retention is significantly influenced by a few businesses with exceptionally high job retention, making it a poor measure of overall Jobs Retained. Interestingly, businesses with higher loan amounts tended to retain more jobs.

Further analysis suggested that Non-Profit Organizations exhibit the highest average job retention, while Self-Employed Individuals, Independent Contractors, and Sole Proprietorships have the lowest. This will be a point of further study. The prevailing owner archetype receiving the most loans across all categories is White_0, representing owners who are White and male. The dataset also reveals substantial fluctuations in the number of loans over the three months studied. A geographic analysis introduced in Project Two reveals comparisons across distributions of loan amounts, jobs retained and the diversity score. Most interestingly, those states in the central North tend to have the lowest loan amount but the highest diversity scores and job retention. It is also found that California is an outlier in terms of higher job retention and loan amount but a low diversity score. The Upper East Coast tends to remain relatively high across categories, while the North and South are clearly divided in Jobs Retained.

Project Three provides insights into new data on CARES Act funding, employment, and worker classification. It reveals a clear relationship between higher CARES Act funding per capita and increased job retention per capita, particularly notable in states in the central and upper North regions. Conversely, states with higher per capita PPP support tend to have lower per capita CARES Act support, indicating potential trade-offs in how federal relief efforts are implemented. Moreover, the distribution of PPP loan amounts per capita underscores disparities in funding allocation, with states like California, Texas, and Florida receiving comparatively higher amounts. However, despite this discrepancy, states in the central and upper North regions, with lower funding levels, manage to retain more jobs per capita. This suggests that factors beyond funding play a role in influencing job retention rates. The predominant worker type among states with the highest per capita jobs retained is private wage and salary workers, highlighting the PPP's effectiveness in preserving employment within the private sector. This underscores the importance of tailoring support measures to meet the diverse needs of different sectors of the economy. However, there may be limitations in supporting public sector employment through such programs.

This research stands out in the literature by not only quantifying the PPP's impact on job retention across different states and business types but also by identifying the critical predictors of job retention, thereby offering a deeper understanding of the PPP's operational dynamics and suggesting a framework for evaluating aid relief program success.

As such, a pivotal aspect of this study is the robust analytical framework employed to dissect the PPP's effectiveness in the Final Project. By conducting nine different OLS regression models, the research delineates the complexity of factors influencing job retention. Model 9, with the highest R2 and Adjusted R2 scores, underscores the significance of including a comprehensive set of predictors to understand the PPP's impact fully. However, the comparable performance of Model 8, with fewer predictors, suggests that a subset of variables holds substantial explanatory power. This nuanced analysis is crucial for distinguishing the research from others, as it highlights the importance of selecting relevant predictors in assessing the PPP's effectiveness.

The study's introduction of an instrumental variable (IV) for addressing potential endogeneity issues marks another distinctive contribution. The use of Party Control as an IV to create a predicted loan amount variable adds a layer of sophistication to the causal inference analysis, enhancing the study's credibility in determining the PPP's true impact on job retention.

Further distinguishing this research is the employment of machine learning models, which provide additional insights into the PPP's effectiveness. The regression tree model's low prediction error and the random forest model's high predictive power, as indicated by a minimal Mean Squared Error (MSE), illustrate the utility of these methods in capturing complex, non-linear relationships that traditional regression models might miss. We find that the most powerful predictors are the number of people who are employed, the number of people in the state who are over the age of 15 (and thus in the workforce), and the predicted loan amount received per 100,000 people by each state through the PPP.

The importance matrix derived from the random forest model, highlighting variables like employment levels, population over 15, and CARES Act funding, among others, adds another layer of depth to the analysis. These findings not only corroborate the significant predictors identified in the OLS and regression tree models but also offer a granular understanding of the factors driving job retention.

Despite the rigorous analysis and significant findings, the research opens avenues for further exploration. Questions remain regarding the long-term impact of the PPP on job retention and the potential differential effects across various sectors and regions. Future research could delve into the program's sustainability and its role in economic recovery beyond the immediate crisis period. This data did not extend beyond a three-month period, so this would be an interesting area of further study. Additionally, exploring the interplay between PPP loans and other forms of government assistance in more detail could offer further insights into optimizing aid distribution in times of economic distress.

In conclusion, this study significantly advances our understanding of the PPP's effectiveness in sustaining employment during the COVID-19 pandemic. By employing a comprehensive analytical approach and highlighting the importance of selecting relevant predictors, the research offers valuable insights for policymakers and contributes to the ongoing discourse on economic recovery strategies. The identified need for further research underscores the complexity of economic interventions and the importance of continuous evaluation to refine and enhance policy responses to future economic crises.

11.1 Next Steps

Moving forward, a potential next step may involve a lender analysis. This would involve delving deeper into the characteristics of lenders and categorizing them into different types, such as large banks or community banks. This nuanced examination could uncover variations in loan distribution patterns and shed light on the role of different lenders in facilitating PPP loans. Obtaining data to define a threshold to define banks as large or small will be a critical first step in doing this.

In addition to the lender analysis, defining a counterfactual scenario, which represents businesses that did not receive PPP loans, would significantly enhance the study's validity in terms of causal inference. This comparative approach would provide a clearer picture of the causal impact of PPP loans on job retention, helping to isolate the specific effects of the program. Obtaining baseline data for businesses participating in the PPP before receiving the loan approval might even allow for a difference-in-difference design to support causal knowledge.

Expanding the scope of the research to explore patterns of success and standards in other economic relief programs would contribute to a broader understanding of the PPP system's context. Comparative analyses with similar initiatives could reveal common trends or unique features that influence job retention outcomes, providing valuable insights for policymakers and small business owners. This study began an examination of this by incorporating CARES Act Funding data outside of the PPP; however, this would be an area of great interest moving forward.

Additionally, a brief review of the existing literature shows the existence of great disputes with respect to defining the effectiveness of the PPP. Those in favour of it discuss how the PPP did well to prevent business closures and cannot be measured by jobs retained alone. On the other hand, economists argue that it did not save as many jobs as was necessary and that many businesses participating in the program were not actually in need of the program. These findings, alongside some of the competing arguments presented in the introduction, prompt further literature review to better compare across studies with similar model specifications, periods of study, and geographic focus.

Another next step in the research process could involve a deeper investigation into the factors influencing job retention rates among distinct types of businesses and demographic groups. Specifically, this study was able to identify that Non-Profit Organizations exhibit the highest average job retention but fails to be able to understand or further provide evidence for why this is the case. Looking into unique challenges and opportunities faced by each business type, including their access to financial resources, resilience to economic shocks, and ability to adapt to changing market conditions, may thus be an organic progression.

Exploring the interplay between PPP loans and other forms of government assistance in more detail could offer further insights into optimizing aid distribution in times of economic distress.

Lastly, despite the rigorous analysis and significant findings, the research opens avenues for further exploration. Questions remain regarding the long-term impact of the PPP on job retention and the potential differential effects across various sectors and regions. Future research could delve into the program's sustainability and its role in economic recovery beyond the immediate crisis period. This data did not extend beyond a three-month period, so this would be an interesting area of further study.

In summary, the proposed next steps encompass a detailed lender analysis, the incorporation of a counterfactual scenario, the exploration of other relief programs, industry-specific investigations, potential omitted variables influencing job retention, an analysis of entrepreneurial hubs across the U.S., and longitudinal analysis. By addressing these aspects, this research can build upon its findings thus far and offer nuanced key details about the dynamics of PPP loans and their implications for job retention in small businesses.

References

Autor, D., Cho, D., Crane, L., Goldar, M., Lutz, B., Montes, J., Peterman, W., Ratner, D., Vallenas, D. V., & Yildirmaz, A. (2022). The $800 billion paycheck protection program: Where did the money go and why did it go there? *National Bureau of Economic Research (NBER)*. https://doi.org/10.3386/w29669

Cortes, G. M., & Forsythe, E. (2020). *Impacts of the COVID-19 Pandemic and The Cares Act on Earnings and Inequality*. https://doi.org/10.17848/wp20-332

Cowley, S. (2022, February 1). *Little of the paycheck protection program's $800 billion protected paychecks*. The New York Times. https://www.nytimes.com/2022/02/01/business/paycheck-protection-program-costs.html

Dalton, M. (2021). Putting the Paycheck Protection Program into Perspective: An Analysis Using Administrative and Survey Data. *U.S. Bureau of Labor Statistics*, 1–39. https://doi.org/https://www.bls.gov/osmr/research-papers/2021/pdf/ec210080.pdf

FRED. (2024a, January 30). *Layoffs and Discharges: Information*. https://fred.stlouisfed.org/series/JTU5100LDR

FRED. (2024b, February 2). *Unemployment Rate*. https://fred.stlouisfed.org/series/UNRATE

Gelinas, N. (2023, March 25). *The post-covid states of america*. City Journal. https://www.city-journal.org/article/the-post-covid-states-of-america

Government of Canada, S. C. (2024, January 25). *North American Industry Classification System (NAICS) Canada 2022 version 1.0*. Government of Canada, Statistics Canada. https://www.statcan.gc.ca/en/subjects/standard/naics/2022/v1/index

Humphries, J. E., Neilson, C., & Ulyssea, G. (2020). The evolving impacts of covid-19 on small businesses since the Cares Act. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3584745

Mumford, S. W., Hutton, N. S., & Riegel, S. M. (2023). Contributions of the Paycheck Protection Program to nonprofit short-term sustainability. *Nonprofit and Voluntary Sector Quarterly*, *53*(1), 186–209. https://doi.org/10.1177/08997640231164538

SBA. "Paycheck Protection Program." U.S. Small Business Administration, May

2020. https://www.sba.gov/funding-programs/loans/covid-19-relief-options/paycheck-protection-program.

Selley, A. (2019). The Effectiveness of the Paycheck Protection Program on County-Level Employment in North Carolina. *UNC Carolina Digital Repository.* https://doi.org/https://doi.org/https://doi.org/10.17615/5m47-4503

Solovyeva, A., & Raphael Lam, W. (2023). How effective were job-retention schemes during the COVID-19 pandemic? A microsimulation approach for European countries. *IMF Working Papers*, *2023*(003), 1. https://doi.org/10.5089/9798400229985.001

Taulli, Tom. "Paycheck Protection Program: How to Make Things as Smooth as Possible." Forbes, April 5, 2020. https://www.forbes.com/sites/tomtaulli/2020/04/03/paycheck-protection-program-how-to-make-things-as-smooth-as-possible/?sh=7a6be102aca3.

Tremper , Nicholas, and Anuradha Raghuram. The Paycheck Protection Program, January 2021. https://www.jpmorganchase.com/institute/research/small-business/paycheck-protection-program-small-business-balances-revenues-and-expenses-weeks-after-loan-disbursement.

U.S. Bureau of Labor Statistics. (2020, May 1). *Unemployment rates down over the month in 38 states in May 2020.* https://www.bls.gov/opub/ted/2020/unemployment-rates-down-over-the-month-in-38-states-in-may-2020.htm

U.S. Bureau of Labor Statistics. (2021, June 1). *Unemployment rates lower in 49 states and D.C. from June 2020 to June 2021.* https://www.bls.gov/opub/ted/2021/unemployment-rates-lower-in-49-states-and-dc-from-june-2020-to-june-2021.htm

US Census Bureau. (2023, December 18). *State population totals and components of change: 2020-2023.* Census.gov. https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html

U.S. Department of the Treasury. "Paycheck Protection Program (PPP) SBA Loan Level Data." Northwest (NW): Washington, D.C., August 2020.

U.S. Department of the Treasury. "Paycheck Protection Program." U.S. Department of the Treasury, June 15, 2023. https://home.treasury.gov/policy-issues/coronavirus/assistance-for-small-businesses/paycheck-protection-program.