

# DA2 Final Project

Peter Szilvasi

2023-12-22

## Introduction

In response to the COVID-19 crisis, the U.S. Federal Government, led by the Trump administration in 2020, initiated the Paycheck Protection Program (PPP). Designed to support a range of entities including businesses, self-employed workers, nonprofits, and smaller enterprises, this program offered low-interest loans. Its primary objective was to aid in covering essential expenses such as payroll, rent, interest, and utilities. Notably, businesses that maintained stable employee numbers and wages had the opportunity for partial or complete loan forgiveness. This report seeks to explore how these loans were distributed. My research question:

*How do business ownership characteristics, specifically ethnicity, gender, veteran status, and political affiliation of the state, along with the proportion of Black residents in a state, collectively influence the loan amounts received through the Paycheck Protection Program?*

## Data

To run my analysis, I collected my data from the [Black Wealth Data Center](#) website, and I enriched my data with political affiliation of state, by looking at the governor of the state in 2020 and his/her political party. I gathered this information from the website of [Ballotpedia](#).

## Data Cleaning

The original dataset includes 1,206,970 observations as the geographical grouping of the data is broken down to state, county, and zip code level which results the file to take up too much memory for me to be able to push it to GitHub. To fix this, I have filtered down to only state level observations. I also excluded those observations which were unanswered in certain categories. In the end I overwrote the original file to save space.

## Descriptive Statistics

Table 1: Descriptive Statistics

	N	Missing	Mean	SD	Min	Max
Loan Amount	807	0	1.625656e+08	5.263416e+08	3.010000e+02	8.835301e+09
Natural Logarithm of Loan Amount	807	0	16.36	2.70	5.71	22.90
Percentage of Black Population	807	0	11.26	10.41	0.60	44.70

Table 2: Descriptive Statistics for Binary Variables

	N	Missing	Yes	No	% of Total
White Owned	807	0	204	603	25.28%
Male Owned	807	0	405	402	50.19%
Non-Veteran Owned	807	0	408	399	50.56%
Republican	807	0	424	383	52.54%

The table summarizes the main characteristics of the dependent, explanatory, and control variables. For the regression analysis, I'll utilize the natural logarithm of the loan amount due to its skewed distribution. This logarithmic transformation aids in achieving a more symmetrical and normally distributed data. From the original dataset, I also created four binary variables, as shown in Table 2. For each measurement we have 807 observations and 0 missing values. White owned businesses are 25.28% of all businesses, while the remaining categories—male, non-veteran, and republican—each represent approximately half. This is a result of the structure of the dataset. Loan amounts are aggregated and grouped by state and different ownership characteristics. In the appendix I included detailed graphs about the distribution of each variable.

## Regression Models

Regression Table 1: Simple Regression with binary controlling variables

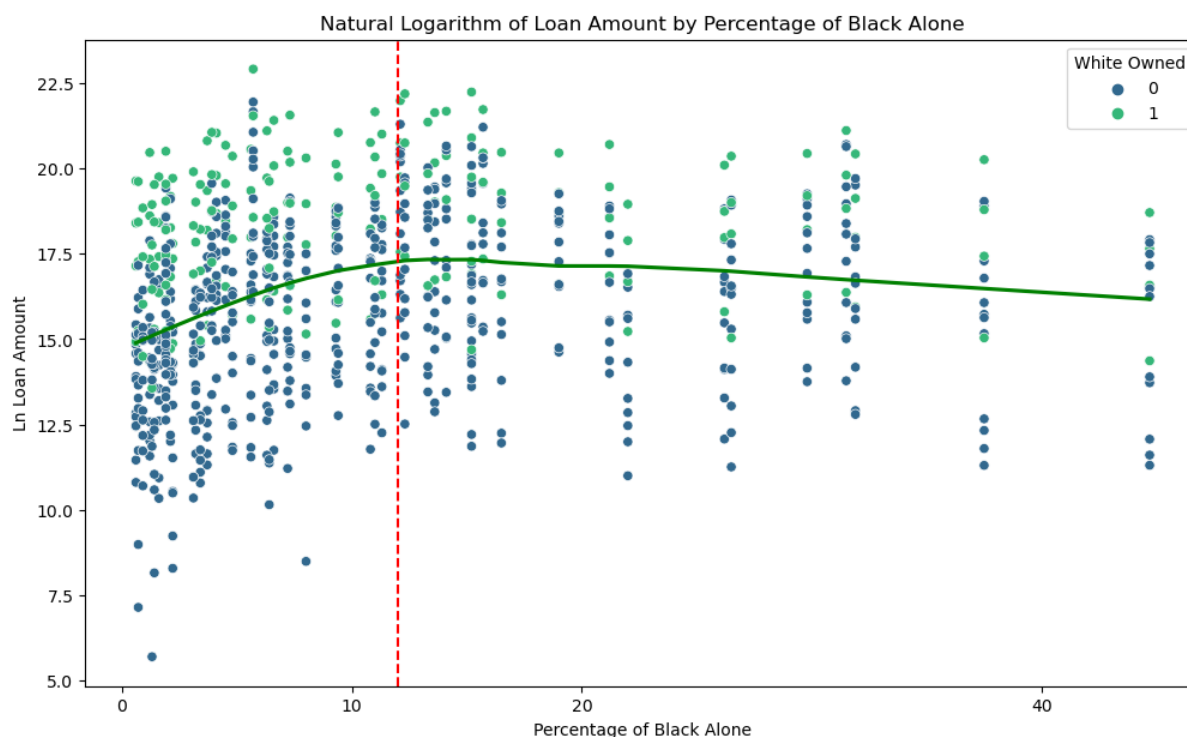
<b>Dependent variable: Natural Logarithm of Loan Amount</b>				
	(1)	(2)	(3)	(4)
White Owned	2.697*** (0.197)	2.700*** (0.190)	2.724*** (0.137)	2.728*** (0.134)
Male Owned		1.237*** (0.166)	1.249*** (0.119)	1.250*** (0.117)
Non-Veteran Owned			3.263*** (0.119)	3.268*** (0.117)
Republican				-0.700*** (0.117)
Observations	807	807	807	807
R <sup>2</sup>	0.189	0.242	0.608	0.625
Adjusted R <sup>2</sup>	0.188	0.240	0.607	0.623
Residual Std. Error	2.430 (df=805)	2.351 (df=804)	1.691 (df=803)	1.656 (df=802)
F Statistic	187.777*** (df=1; 805)	128.197*** (df=2; 804)	415.346*** (df=3; 803)	334.099*** (df=4; 802)

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

In Model (1), the constant indicates the average natural logarithm of loan amount for businesses not white-owned is 15.683. The "White Owned" coefficient suggests a 99% confidence that when the variable changes from 0 to 1, the ln loan amount rises by 2.697, equivalent to a 1383.31% increase. The R<sup>2</sup> indicates 18.9% variability in the dependent variable is influenced by ownership status. However, the model's Residual Error is considerably high, as we are looking at logarithmic values. To perfect the model, we can include controlling variables. In models, (2), (3), and (4), I included step-by-step one more variable to our model to see how each of them effect our dependent variable. All our controlling variables' coefficients are statistically significant at 99% confidence level. However, building on this model, I've further enhanced the regression to also include the Percentage of Black Population, and an interaction between White Owned and Male Owned. For Black Population Percentage, I incorporated a piecewise linear spline with a knot (or breakpoint) set at 12%. This decision is based on visual observations from Graph 1, specifically noting a change in direction of the loess smoothing line at 12%. Graph 1 also visualizes White Owned businesses, and it's visible that there are almost

no loans for these businesses below Ln Loan Amount of 15 and the highest values correspond to White Owned Observations.



Graph 1: Piecewise Linear Spline with knot at 12% for the Percentage of Black Population Variable

#### Model 2: Multiple Linear Regression with Splines and Interaction

##### **Dependent variable:** Natural Logarithm of Loan Amount

White Owned	2.410*** (0.126)
Male Owned	1.086*** (0.124)
Non-Veteran Owned	3.289*** (0.099)
Republican	-0.437*** (0.102)
Interaction: White Male Owned	0.664*** (0.179)
Black Population < 12%	0.239*** (0.013)
Black Population >= 12%	-0.053*** (0.007)
Observations	807
R <sup>2</sup>	0.735
Adjusted R <sup>2</sup>	0.732
Residual Std. Error	1.395 (df=799)
F Statistic	307.950*** (df=7; 799)

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

My final regression, Model 2, includes all previously discussed variables and shows key insights into how the loans were distributed. White, Male, Non-Veteran are in all positive correlation with the Loan Amounts. If we are looking at loans with the same characteristics, White Owned Businesses tend to get 1013.53%, Male 196.33%, and Non-Veteran Businesses 2581% more loans on average than those that are Non-white, female or veteran owned. Additionally White Male owned businesses got 94.20% more loans than those that are not White Male owned. Interestingly when looking at Republican and Democratic states, Republican ones got on average 35.37% less loans than Democratic states. As for Black Population, there

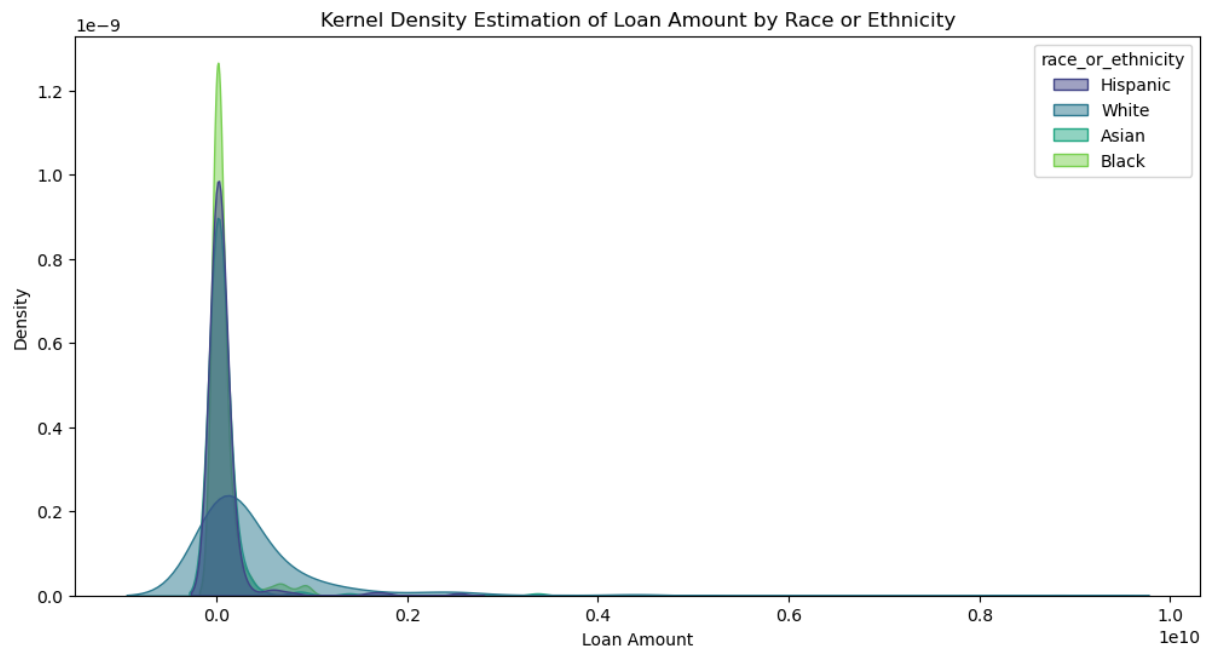
was a positive correlation until 12% as anticipated. On average, in states where the percentage of black population is less than 12%, 1% more black people means an increase of loan amounts by 27.04%. This trend changes above 12%, to -5.16%. Note that all these values are valid within a 99% confidence interval. The R-squared of the model is also quite high with 0.73 while our Residual Error is also lower than what we saw in Model 1 meaning that including splines and interaction was a good addition to the regression.

## Conclusion

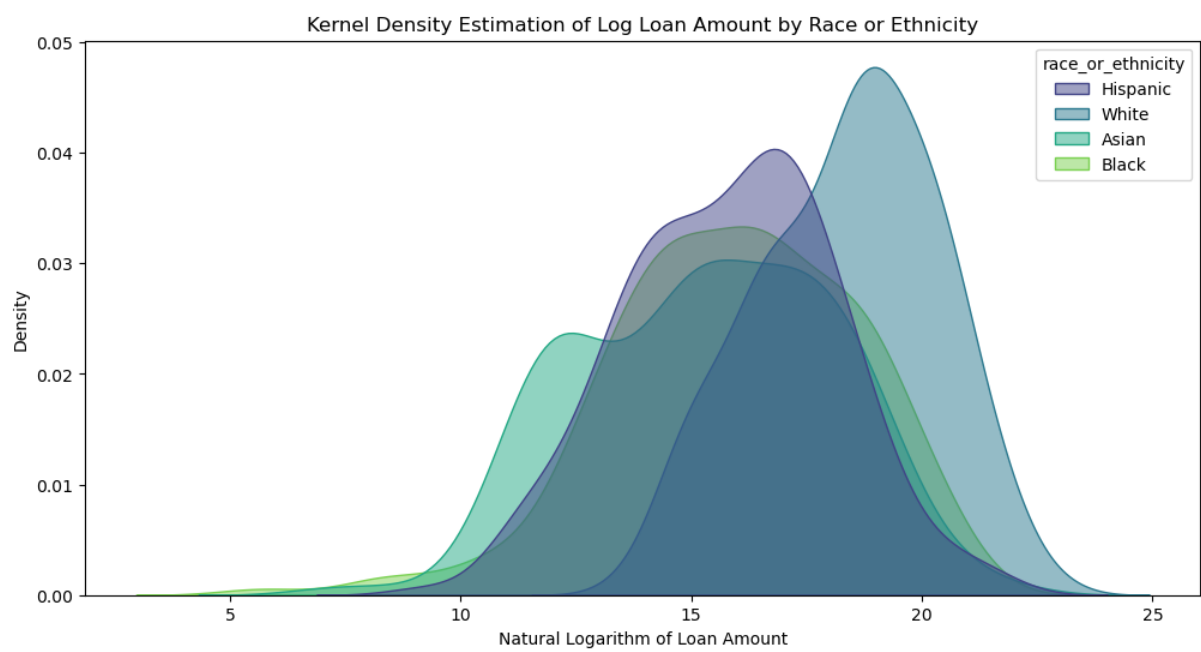
To address the question of how different business ownership traits affect the distribution of Paycheck Protection Program (PPP) loans, Model 2 provided significant insights. White, Male, and Non-Veteran owned businesses stood out, receiving notably higher loan amounts under the Trump administration's PPP. This points to a clear inequality within the dataset. Even though all the coefficients are 99% statistically significant and we should be able to generalize to the population our data represents, this is as far this dataset takes us. We do not have data about the percentage of white businesses, compared to other ethnic groups, just as we do not have information who applied for these loans and who got it. Nevertheless, the results show that grouped by states and ownership characteristics, White, Male, and Non-veteran businesses received an advantage in times of aid from the Trump Administration.

# Appendix

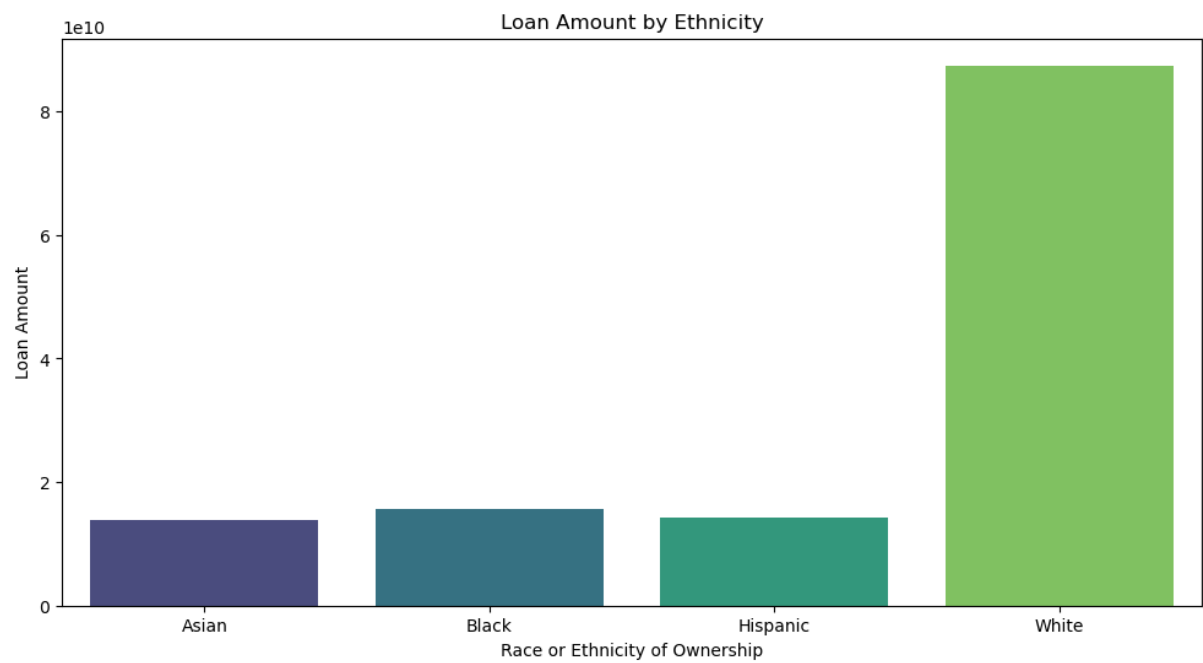
Graph 2: Loan Amount Before Ln Transformation:



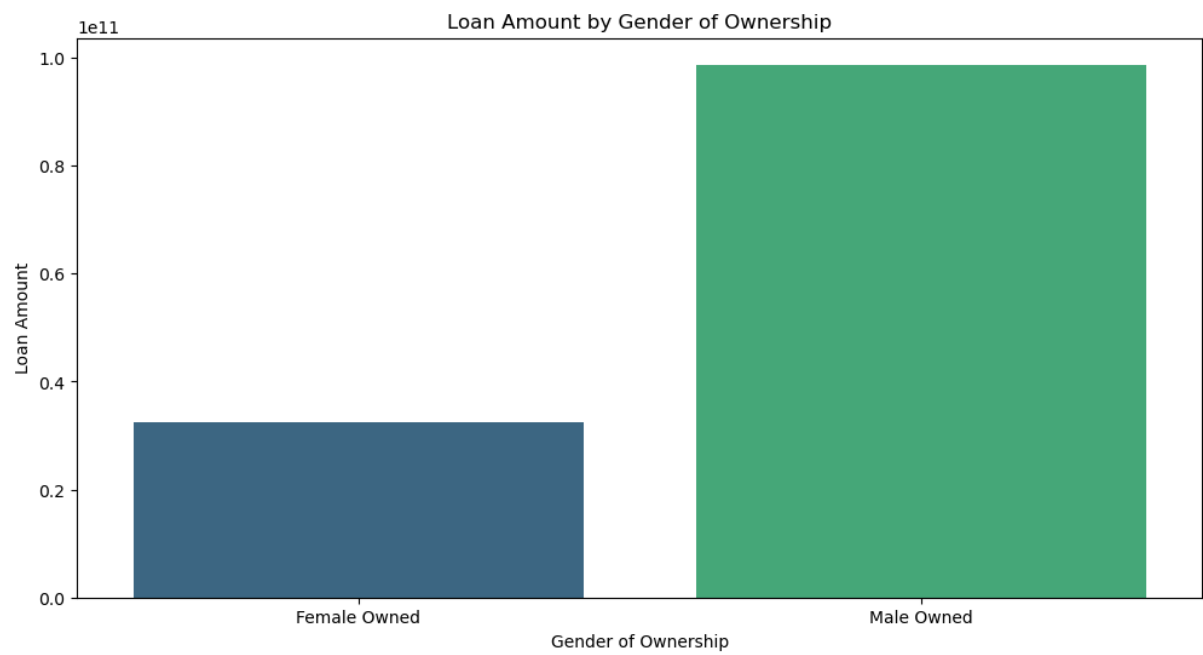
Graph 3: Loan Amount After Ln Transformation:



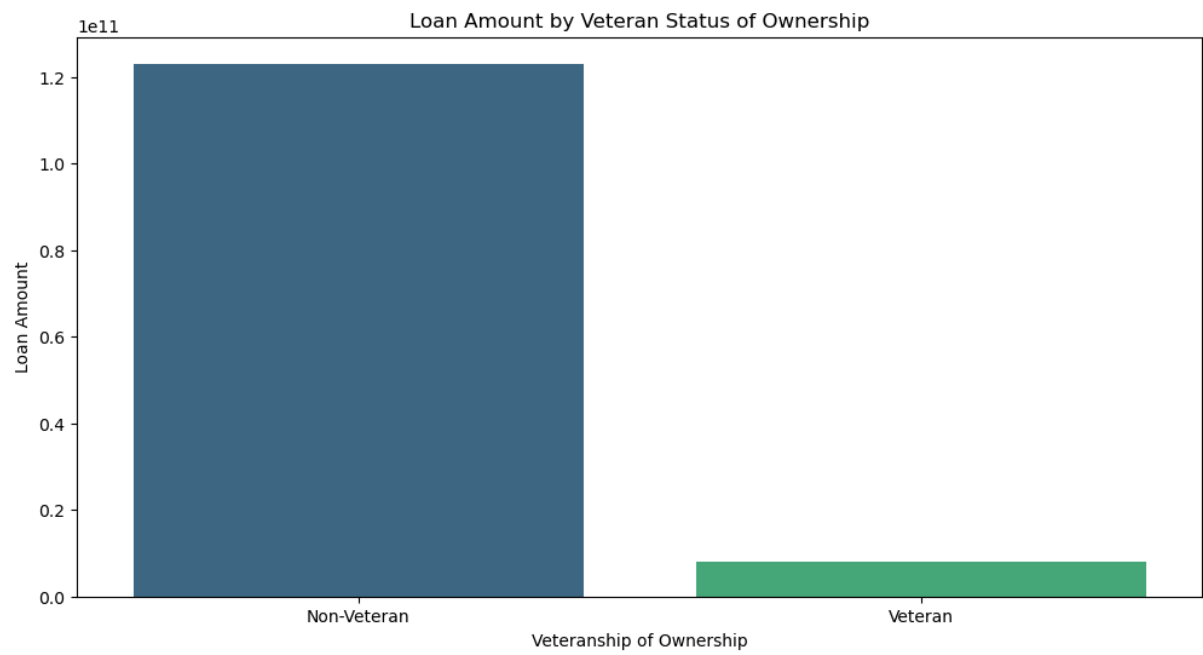
Graph 4: Loan Amount by Ethnicity



Graph 5: Loan Amount by Ethnicity



Graph 6: Loan Amount by Veteran Status



Graph 7: Loan Amount by Political Affiliation of State

