The Black Dollar

Did the Paycheck Protection Program Exclude the Black Community: An Analysis of Loans Issued in the State of Maryland

Amina Lampkin
Philip Merrill College of Journalism
University of Maryland
College Park, Maryland, United States of America
alampkin@terpmail.umd.edu

ABSTRACT

In this article, you'll learn about the Paycheck Protection Program that was developed in response to the coronavirus pandemic. Issuing out loans for small businesses to apply for, the program was run by the United States Small Business Administration. The SBA's Office of the Inspector General found that small businesses owned by those from underserved communities were not prioritized through the distribution of funds.

Demographic data from the American Community Survey and loan records for the state of Maryland were analyzed to learn more about the differences that exist between loans issued in predominantly Black and predominantly white communities.

KEYWORDS

African-American, Black-owned businesses, lending discrimination, Paycheck Protection Program, CARES Act, coronavirus

1. INTRODUCTION

1.1 BACKGROUND

Since March 2020 our lives in the United States have been changed by the coronavirus pandemic. Our economy especially took a hit due to the various lockdowns and our inability to conduct business as usual. On March 27th, 2020 the Coronavirus Aid, Relief, and Economic Act (CARES Act) was passed to provide approximately $2 trillion in economic relief in response to the pandemic. The law provisioned the Paycheck Protection Program, a loan program to help small businesses continue to pay their employees, as well fund other essential business expenses (Lampkin, 2020).

Under the guidance of the United States Small Business Administration and Department of the Treasury, qualified small businesses across the country were able to apply for loans from local lenders. State by state data for these loans was released in early July. After accusations of the fund failing to benefit those businesses that need it most and instead being issued to companies owned by federal employees, there was clearly a need to investigate the data a bit closer.

For the last semester, I have cleaned and analyzed the data to learn more about the impact of the Paycheck Protection Program in the state of Maryland. With my work I hope to learn more about the consequences of instances of possible lending discrimination.

1.2 MOTIVATION

The United States has a history of creating programs that sound good in theory, but end up overlooking the communities that need the help the most. In the case of the Paycheck Protection Program, a report from the SBA's Office of the Inspector General found that business owners in underserved and rural markets were not prioritized as intended (U.S. Small Business Administration Office of the Inspector General, 2020). This population of business owners included those from socially and economically disadvantaged backgrounds, veterans, and women (ibid.). Following the release of this report on May 8, 2020, concerns arose that the funds were not being disbursed equitably. After pressure from news publications and frustrated lawmakers, the SBA and Treasury released data of the loans in early July (Lampkin, 2020).

After finding that firms owned by foreign companies and large corporations backed by Wall Street applied for and received loans from the program, the concerns gained more traction and import to the issue of how one of the largest economic programs in the history of the country was operating. As a data journalist with a dedication to making information more accessible and using data to reveal disparities that exist within our communities, I decided to analyze the Paycheck Protection Program within the state of Maryland.

1.3 SPECIFIC CONTRIBUTIONS

The contributions to this project are that of which you'd expect from a data journalist. I began with doing some research into the legislation behind the data. From there, I began to use reliable manipulation techniques to clean and analyze the data. From running several analyses, I had to decide how I would present the story of the program and the Maryland dataset. To supplement the story, I also integrated the data to build interactive maps and tables for embedding. My contributions to the body of research are focused in trying to make the information digestible and accessible to the average reader.

2. LITERATURE REVIEW

The gap that my research attempts to fill is under the umbrella of understanding how Black Americans are affected by discrimination in the lending industry. The body of research on this topic is typically centered on lending discrimination in relation to mortgages and homebuying. For investigating lending for small businesses, the body of research is not as vast.

The problem of lending discrimination is related to the legacy of slavery and Jim Crow laws in the United States of America. Given the headstart that white Americans have being descendants of plantation owners who were able to leave behind inheritances that served as starting capital for other businesses, white Americans have a mean and median net worth of $188,200 and $983,400, respectively, according to the Board of Governors of the Federal Reserve System. In comparison, Black Americans have a mean and median net worth equivalent to less than 15 percent of their white counterparts at $142,500 and $24,100, respectively (Bhutta et al., 2020). Lack of access to starting capital is one of the leading reasons start-up ventures fail.

For prospective Black business owners, not being able to accumulate the necessary capital to launch was called the "discouraged entrepreneur" phenomenon by economists in the piece called *Unequal Access: Financial Institution Lending to Black - and White-Owned Small Business Start-ups* (Bates, 1997, 488). Even if Black business owners do end up taking the dive, they are 40 percent more likely to report withholding a credit application for fear that it will be rejected (Blanchflower et al., 2003, 932).

Once the application is submitted, researchers found that Black borrower applications are more than twice as likely to be denied a loan at rates of 65.9 percent compared to 26.9 percent for white applicants (ibid.). For those Black applicants who were approved, they typically paid an interest rate 1 percent greater than their white counterparts (Blanchflower et al., 2003, 931). Even controlling for all other variables like credit scores between the subjects, researchers found these disparities.

With this information setting the scene for understanding the loan industry as it relates to racial discrimination, my research into one of the largest federal fiscal polices in recent history will prove that even in emergency situations that this lending discrimination persists (Rubio, 2020).

3. METHODOLOGY

3.1 DATA COLLECTION

The Small Business Administration released the data from the Paycheck Protection Program in early July. The SBA stopped accepting loan applications from lenders on the program's last day, August 8, 2020. Since this project did not begin until September 2020, the dataset analyzed covered the duration of the program in its entirety. I was able to navigate to the SBA's website and find where the datasets were housed.

The first option was to download the dataset of those loans greater than or equal to $150,000. While this dataset amounted to nearly 15 percent of the loans issued in the program, it had already been analyzed in part by The Washington Post and other news publications. Additionally, I was more interested in learning about the widespread effects of this program throughout the state of Maryland. Consequently, I selected the Maryland dataset of loans issued less than $150,000.

As part of the data analysis, I would have liked to include data for the total number of businesses within each zip code in Maryland. However, requesting that information was becoming quite a hassle after being rerouted to several different departments. As a workaround, I decided to compare the loan data to demographic estimates about each zip code from the American Community Survey. The data I pulled uses ACS survey responses in the last 5 years to generate estimates for the entire population. Although these demographics do not control for the exact number of businesses, it gives readers some background into the surrounding communities being impacted by the loan program.

Using Python and the *tidyverse* library, I was able to pull several pieces of demographic information to use in conjunction with the loan data. To do this, I used the *load_variables* function to store the table names. The resulting object is a table with three columns, *name, label, concept*. The name column includes short codes of the data tables within the survey. If the short name has an underscore, the following values correspond to what row in the table you're accessing. The label column describes what the actual estimate is. The concept table provides you with the regular name of that table. For example, for the table named *B02001_003,* the label reads *Estimate!Total!!Black or African American alone* and the concept column *Race*, indicating that we are accessing the third row of the Race table which is about the total population of Black people. Using the Census Data Explorer, I was able to double check the tables and corresponding variables I pulled by searching for the short code.

For each variable I pulled from a table, I had to determine what geography I wanted the data to be grouped into. Still using our previous example, one option would be to group the Black population estimates by State. Because the loan data is more granular than states and there was no option to group the values by city, I decided to group the values by zip code. I will elaborate on this decision in the data cleaning section. The demographic variables I pulled and ended up using are listed in Table 1 in the Appendix.

3.2 DATA CLEANING

3.2.1 LOAN DATASET

The raw dataset began with 74,103 records. There were fourteen different columns. The columns of interest for my particular work are described in Table 2 in the Appendix. Using OpenRefine, I worked on trying to standardize some of the dataset. I first checked the dataset for any records that were not in the state of Maryland. The exact reason why there were businesses not located in Maryland in the dataset is unclear. But an educated guess would assume that the business owners live in Maryland, but own businesses in surrounding states like Virginia and Washington, D.C. Those records were removed from the dataset.

The next issue to tackle was the *City* column. This column was built off the borrower's applications, meaning that some of the cities listed were not areas with administratively decided boundaries, but rather a regional nickname. Additionally, some of the cities may have been misspelled by a letter, all lowercase, or all uppercase. Given that there were more than 1,500 unique city entries to go through, I decided I would not be standardizing the city names. If I decided to create a visual with the city names, it would be on a much more manageable scale of maybe 10-100 records to manually fix.

The demographic columns that asked the borrower's race, gender, and veteran status already had consistent and standardized responses so there was nothing I needed to check there.

3.2.2 DEMOGRAPHIC DATA

For each of the variables I pulled from the American Community Survey, I also calculated a percent margin of error for each variable's estimate. Again, because the American Community Survey demographic data are estimates for the population based off of a sample, when you pull each variable you're provided with an estimate for that particular zip code and a margin of error. Continuing with our Black population variable used earlier, the percent margin of error was calculated by taking the *moe* column and dividing it by the *estimate* column, then multiplying the quotient by 100 to get a percentage. This was done for every variable listed in **Table 1.** In addition, for the Black population, white population, male population, female population, and poverty variables I calculated the proportion of the zip code's population affected by that variable compared to the zip code's entire population. Within each pulled variable, the population estimate for that zip code is stored in the column *summary_est.* For the Black population variable that is calculated by dividing the *estimate* column by the *summary_est*, then multiplying the quotient by 100 to get a percentage. These values were stored in the columns named *pct_variablename*.

The population estimate for that variable is under *estimate.* The margin of error is in *moe*. The entire population estimate and its margin of error for a particular zip code are listed as *summary_est* and *summary_moe*. The last two columns are engineered to find the margin of error for the variable estimate and the percentage of the zip code's population that are affected by the variable.

After creating a CSV for each of the demographic variables, I combined them into one document using Microsoft Excel. Using R, I then used a left join between the loan data and demographic data on the columns with the zip codes. Using the zip code option in the American Community Survey meant data from every zip code tabulation area in the United States was pulled. The left join between the Maryland loan data and all of the ACS demographic data meant only the demographic data for the Maryland zip codes was kept. The result was a file of all the loan data where each record

	GEOID	NAME	variable	estimate	moe	summary_est	summary_moe	pct_moe_black	pct_black
1	00601	ZCTA5 00601	B02001_003	145	93	17242	314	64.137931	0.84096973
2	00602	ZCTA5 00602	B02001_003	1070	425	38442	150	39.719626	2.78341397
3	00603	ZCTA5 00603	B02001_003	1930	346	48814	749	17.927461	3.95378375
4	00606	ZCTA5 00606	B02001_003	149	130	6437	304	87.248322	2.31474289
5	00610	ZCTA5 00610	B02001_003	696	355	27073	205	51.005747	2.57082702
6	00612	ZCTA5 00612	B02001_003	2750	668	60303	1231	24.290909	4.56030380
7	00616	ZCTA5 00616	B02001_003	582	265	10765	1203	45.532646	5.40640966
8	00617	ZCTA5 00617	B02001_003	841	423	23974	316	50.297265	3.50796696
9	00622	ZCTA5 00622	B02001_003	35	44	6578	1216	125.714286	0.53207662
10	00623	ZCTA5 00623	B02001_003	951	402	42427	1216	42.271293	2.24149716

Figure 1: An example of what is stored in each variable object that is pulled from the American Community Survey. *GEOID* is the zip code. *NAME* is the official zip code tabulation area from the United States Census Bureau. *Variable* is the table code and row that the estimates are representing. In this instance it is from the RACE table and the third row (the total Black or African-American population).

included the corresponding demographic data for the zip code the loan was issued in.

3.2.3 FINAL CLEANING

Now that each record has the corresponding demographic data connected to it, the final step in cleaning was removing those zip codes and corresponding loan records with a population margin of error greater than 10 percent. Since the demographic data are only estimates, removing those values helped us increase the accuracy and integrity of the dataset as a whole.

Once this was done, the final dataset came out to 61,793 records. I also decided that the final demographic variables were going to be total population, Black alone population, white alone population, and the median income.

Following the data cleaning, I aggregated the records according to if it was a majority white or Black zip code. To do this, I filtered the entire dataset by if the percentage of white alone population percentage, *pct_white*, was greater than or equal to 50 percent. I then exported that subset into its own file. The same procedure was done with the Black alone population percentage (*pct_black)*. If a zip code was neither majority white or Black, it is not included in either subset dataset.

For the number of loans in each zip code, I made a pivot table with the *COUNT* function in Excel. Then, I set an *if statement* in Excel that checked if the population percentage of white people is greater than or equal to 50 fill in a "w" and vice versa for the population percentage of Black people with a "b".

4. EXPERIMENT

4.1 HYPOTHESES

To identify if there is actually a difference between the loans issued in predominantly white and predominantly Black communities, I conducted Welch's Two Sample T-Test in R Studio.

Research Question 1: Is there a difference between the loan dollar amounts issued to businesses in predominantly Black communities vs. businesses in predominantly white communities?

Research Question 2: Is there a difference between the number of loans issued to businesses in predominantly Black communities vs. businesses in predominantly white communities?

For each research question, I also developed a hypothesis. The null of each hypothesis states that there is no difference between the two community types.

Question 1 H_A: The average dollar amount of loans issued to businesses in predominantly Black communities is not equal to those issued in predominantly white communities.

Question 2 H_A: The number of loans issued to businesses in predominantly Black communities is not equal to those issued in predominantly white communities.

4.2 ASSUMPTIONS

For the Welch's Two Sample T-Test, there are five assumptions I have been taught to check for. The first three are related to the sample. There must be and can only be two independent random samples. The dependent variable must be interval-ratio. There can also only be two possible subgroups that the data can fall into.

For the random sample assumption, I used the random sample function without replacement to pull a random sample of records from those loans and zip codes that are majority white and majority Black. Each random sample is equivalent to 10 percent of the records. Money and the number of loans per zip code are ratio measurements as they both have a zero point. Each loan record and zip code can only be classified as a predominantly white or Black community.

The fourth assumption is that the sample distribution is normal. Unfortunately for both hypothesis questions, the data were skewed very left. I attempted to subtract the average loan amount from each loan as well as the average number of loans in each type of community from the number of loans. However, the data were still skewed so I ran the T-Test on the original data still. I am unsure of how this has impacted my analysis. However, I do know that failing this assumption brings into question the veracity of the results.

Because we're using the Welch's Two-Sample T-Test, the variances do not need to be equal.

4.3 HYPOTHESIS 1 RESULTS

Table 3: Descriptive statistics for the random samples of loan dollar amounts

Black Communities		White Communities	
n	954	**n**	4,382
Mean	$28,270.98	**Mean**	$35,362.49
Median	$19,329.00	**Median**	$20,833.00
Range	$148,200.00	**Range**	$149,292.00

After completing the Welch's T-Test, we reject the null hypothesis that there is no difference between the two communities by chance by an alpha of 0.05.

$$t(1626.2) = 7.98, p < 0.05$$

There is a statistically significant difference between the average loan amount issued in a predominantly white community vs. a predominantly Black community. Following Welch's T-Test, I also conducted Cohen's D test for effect size. The effect size of this hypothesis test was 0.252 indicating that the effect of this difference is small.

4.4 HYPOTHESIS 2 RESULTS

Table 4: Descriptive statistics for the random samples of number of loans

Black Communities		White Communities	
n	3	**n**	12
Mean	347	**Mean**	434.17
Median	307	**Median**	385
Range	244	**Range**	1,260

After completing the Welch's T-Test, we failed to reject the null hypothesis that there is no difference between the average number of loans for the two communities by an alpha level of 0.05.

$$t(9.3393) = 0.73359, p > 0.05$$

The p-value for this test was 0.4812. I did not calculate Cohen's D for this sample as it would overinflate the results since the sample size of hypothesis test was only 15.

4.5 FINAL FINDINGS

The hypothesis tests proved that there is a statistically significant difference between the dollar amounts in loans issued in white communities compared to Black communities. The effect of this difference is small in magnitude, statistically speaking. However, the T-Test did not reveal a statistically significant difference between the number of loans issued in white communities compared to Black communities.

In the following section, I will address some of the limitations involved with this research and how I'd like to improve upon it in future work.

5. LIMITATIONS

The biggest limitation is that the datasets for both hypotheses failed the assumption of normality. Given that this is my first time using the statistical methodologies I've learned in other information science courses here at the University of Maryland on real, imperfect data I am not completely sure how to resolve this. While I don't have a specific resolution for this, it does make me question the validity of the results and wonder about the implications of my data being so skewed. Even trying to rectify the situation by subtracting the average loan amounts and average number of loans for each community from each record in the datasets did not help.

Another limitation of the data was the lack of statewide data to compare the Paycheck Protection Program loans to. Being pressed for time and also working on this research as part of a longform news story made it difficult for me to get the information I desired from the state of Maryland. Without the statewide data, I could not analyze how the loans issued in Maryland compared to the actual businesses here and instead had to analyze the loans in the context of the communities they were issued in.

A final limitation of the demographic data was that the American Community Survey values are only an estimate. I would have preferred to use the 2020 Census.

6. CONCLUSION AND FUTURE WORK

In closing, I thoroughly enjoyed working with this dataset. I gained a lot of new skills as I worked through the data collection all the way through to producing the different map graphics for the story on this dataset published with *Capital News Service*.

The story I wrote for *Capital News Service* did not focus on the results of the hypothesis testing primarily because it would be difficult to explain and could prevent the average person from understanding the important aspects about the program. Instead, I focused on more general statistics like the average loan amount and number of loans in white communities being about 20% more than that of Black communities. I highlighted other points such as the most common industries for loan applications, as well some of the highest and lowest loans issued. Embedded within the story are layered maps made using Carto and presentation quality tables created using Datawrapper. I have included a link to the story in the references.

It's disappointing that Black communities are still impacted by the repercussions of the Mid-Atlantic Slave Trade and the history of discrimination in this country. Even in instances where Black people would like to build generational wealth for the next generation, there are hidden forces through the various institutions that command our lives.

For future research, I hope that as time progresses, more information is released about the true impact of the Paycheck Protection Program and that the field of analysis into the loans expands beyond discovering racial disparities, but gender and geographic disparities as well.

REFERENCES

[1] Bates, T. (1997). Unequal Access: Financial Institution Lending to Black - and White-Owned Small Business Start-ups. *Journal of Urban Affairs*, *19*(4), 487-495. Taylor & Francis Online. 10.1111/j.1467-9906.1997.tb00508.x

[2] Bhutta, N., Chang, A. C., Dettling, L. J., & Hsu, J. W. (2020, September 28). *Disparities in Wealth by Race and Ethnicity in the 2019 Survey of Consumer Finances*. Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/econres/notes/feds-notes/disparities-in-wealth-by-race-and-ethnicity-in-the-2019-survey-of-consumer-finances-20200928.htm

[3] Blanchflower, D. G., Levine, P. B., & Zimmerman, D. J. (2003). Discrimination in the Small-Business Credit Market. *The Review Economics and Statistics*, *85*(4), 930-943. JSTOR. NA

[4] Lampkin, A. (2020, November 19). *White and Black communities in Maryland experience a 20% difference in small business loans from CARES Act*. Capital News Service. https://cnsmaryland.org/2020/11/19/white-and-black-communities-in-maryland-experience-a-20-difference-in-small-business-loans-from-cares-act/

[5] Rubio, M. (2020, December 10). *Op-ed: The data is clear – the Paycheck Protection Program is a success*. CNBC. https://www.cnbc.com/2020/12/10/marco-rubio-op-ed-the-paycheck-protection-program-is-a-success.html

[6] U.S. Small Business Administration Office of the Inspector General. (2020, May 8). *Flash Report Small Business Administration's Implementation of the Paycheck Protection Program Requirements*. U.S. Small Business Administration. https://www.sba.gov/document/report-20-14-flash-report-small-business-administrations-implementation-paycheck-protection-program-requirements

Appendix

Tables

Table 1: List of Variables and their Table Names from the 2018 American Community Survey 5-Year Estimates

Bold variables represent those that were used for actual analysis in the project.

Variable	Table Origin	Table Short Code
Total population	**Race**	**B02001**
Black (alone) population		
White (alone) population		
Male population (total)	Sex By Age	B01001
Female population (total)	Sex By Age	B01001
Median income	**Median Income in the Past 12 Months (in 2018 inflation-adjusted dollars)**	**B19326**
Impoverished status (total)	Poverty Status in the Past 12 Months By Sex By Age	B17001

Table 2: Column Names and Descriptions from the Maryland Dataset of PPP Loans Less Than $150,000

Column Name	Description of the Data
LoanAmount	The amount of the loan that was issued to a business. The value is in US dollars ($).
City	Name of the city that the business is located in. Cities were not standardized (i.e. some are written in all uppercase/lowercase, others do not use conventional city names or reference a regional name)
Zip	Zip code that the business is located in. Column is important to join the demographic data onto.
NAICSCode	North American Industry Classification System Code defines the particular job industry that the business belongs to.
BusinessType	Stores the IRS designation of a business (i.e. Limited Liability Corporation, Corporation, Subchapter S Corporation, etc.)
RaceEthnicity	Stores the race or ethnicity of the business owner. The borrower has the option to share or withhold this information.
Gender	Stores the gender of the business owner. The borrower has the option to share or withhold this information.
Veteran	Stores the military status of the business owner. The borrower has the option to share or withhold this information.
JobsReported	Number of jobs reported at the establishment.