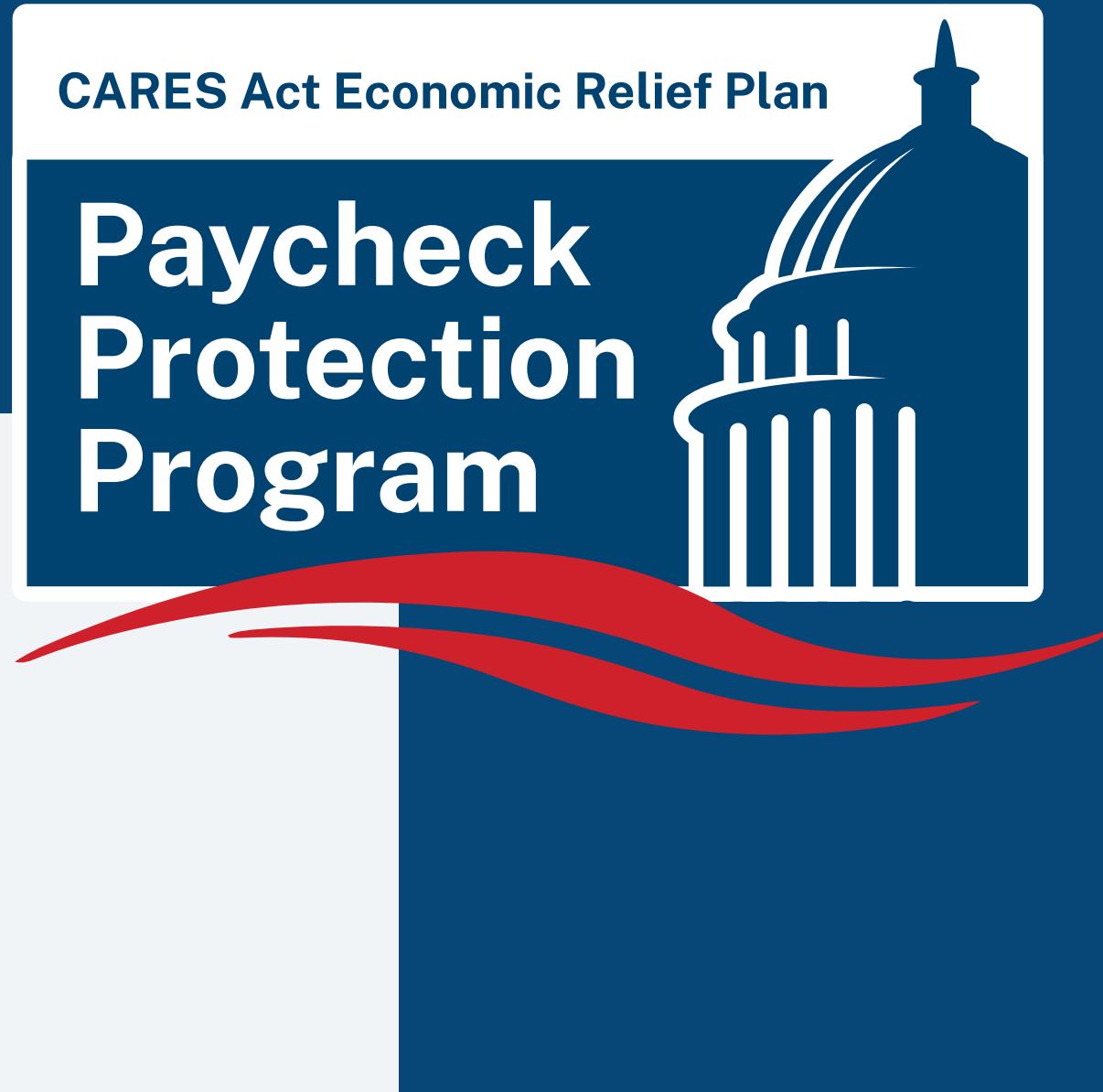


HIDDEN DRIVERS BEHIND THE
PAYCHECK PROTECTION PROGRAM

THE STORY OF GEORGIA

Presented by Dishant Vakte, Iju Lin, Naila Sharmin
Group: IC-22009





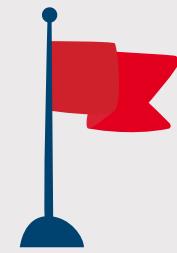
ABOUT THE PAYCHECK PROTECTION PROGRAM

The SBA introduced PPP to keep the workforce employed by providing businesses with uncollateralized, low-interest loans.

They have periodically released data on the more than **11.5 million approved applications**, but it **also has removed applications** that had been **previously present in the dataset**.

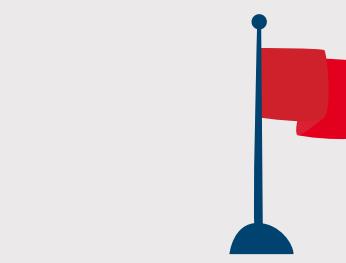
But how this money was allocated and **what types of businesses were actually benefited** from the program remains the question!

OUR GOAL



DETERMINING THE
DEFINING
CHARACTERISTICS OF THE
REMOVED LOANS

CHARACTERISTIC
COMPARISON
AMONG APPROVED &
REMOVED LOANS



FIGURING OUT THE
SUSPICIOUS
PATTERNS OF
LOAN
APPLICATIONS



PREDICT WHETHER
OR NOT A LOAN WAS
REMOVED FROM THE
DATA USING
PREDICTIVE MODELS

WERE THESE APPLICANTS FRAUD?

WHAT MIGHT BE THE POSSIBLE REASON?

WHY SO MANY OF THE APPLICATIONS WERE
REMOVED FROM THE DATAPool ?

WAS THERE NOT ENOUGH INFORMATION
FOR PROCESSING THE APPLICATIONS?

DID THESE APPLICANTS WITHDRAWED
THEMSELVES?



INSIGHTS

 PPP Loan Data — Anthony Williams, Decatur, GA

ALERT: There are 4 PPP loans for a total of \$83,332 in our database for businesses with the name "Anthony Williams" in Decatur, GA. This is typically due to the same business receiving both first and second-draw loans, but may also include similarly named but unrelated businesses, multiple branches of the same business, mistaken multiple applications, or potential fraud. Click each loan below to see full details.

\$20,833 Loan - Self-employed Individual - May, 2021 Decatur, GA 30032 Lender: Harvest Small Business Finan...	\$20,833 Loan - Sole Proprietorship - April, 2021 Decatur, GA 30034 Lender: Cross River Bank	\$20,833 Loan - Self-employed Individual - April, 2021 Decatur, GA 30032 Lender: Harvest Small Business Finan...
\$20,833 Loan - Sole Proprietorship - June, 2020 Decatur, GA 30032 Lender: Fundbox, Inc.		

 **Anthony Williams**

Entity: Self-employed Individual
Industry: General Freight Trucking, Local
Location: Decatur, GA

☆ GA ☆

• Search All PPP Data

Source: <https://www.federalpay.org/paycheck-protection-program/anthony-williams-decatur-ga>

SUSPICIOUS PATTERN...

In [229]: 1 Removed_loan[Removed_loan['name'] == 'ANTHONY WILLIAMS']													
Out[229]:													
	name	amount	state	address	city	zip	naics_code	business_type	jobs_retained	date_approved	...	originating_lender_state	loan_id
2751	ANTHONY WILLIAMS	20833.0	GA	111 Coffee Bluff Villa Rd N/A	SAVANNAH	31419-3109	448150	Sole Proprietorship	1	2021-03-25	...	PA	
9692	ANTHONY WILLIAMS	20520.0	GA	537 Northridge Crossing Dr	ATLANTA	30350-3286	561790	Independent Contractors	1	2021-04-04	...	TX	
10697	ANTHONY WILLIAMS	20832.0	GA	144 Allen St	IRWINTON	31042-3804	812112	Independent Contractors	1	2021-05-27	...	AZ	
14203	ANTHONY WILLIAMS	20833.0	GA	289 Wood Sage Dr	RIVERDALE	30274-1869	532412	Independent Contractors	1	2021-05-29	...	CA	
17894	ANTHONY WILLIAMS	17380.0	GA	1441 Woodmont Ln NW # 629	ATLANTA	30318-2866	454110	Sole Proprietorship	1	2021-04-04	...	TX	
20913	ANTHONY WILLIAMS	20420.0	GA	410 Darbyshire Rd	NORMAN PARK	31771-4577	333112	Sole Proprietorship	1	2021-05-13	...	AZ	
25304	ANTHONY WILLIAMS	20000.0	GA	3015 Kensington Ln	VALDOSTA	31602-4010	541430	Limited Liability Company(LLC)	1	2021-04-22	...	AZ	

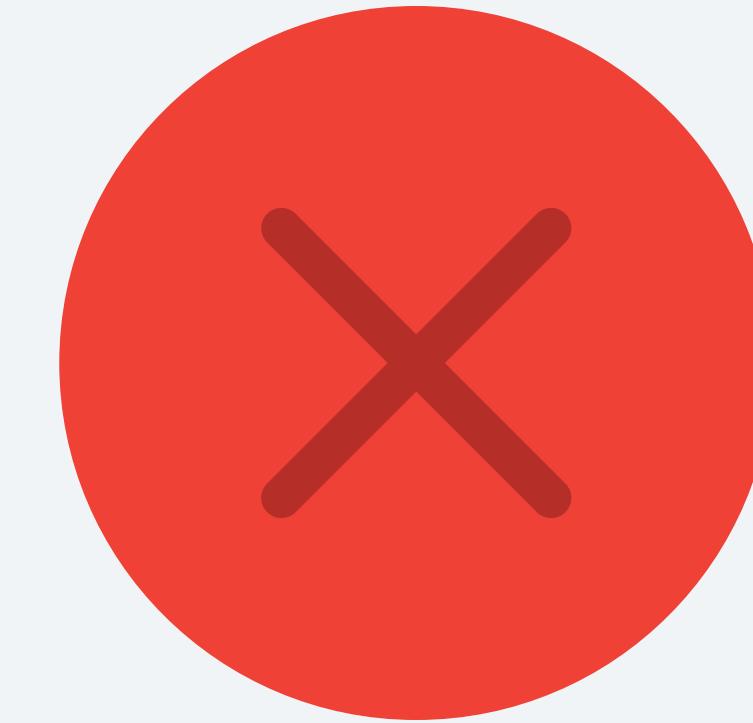
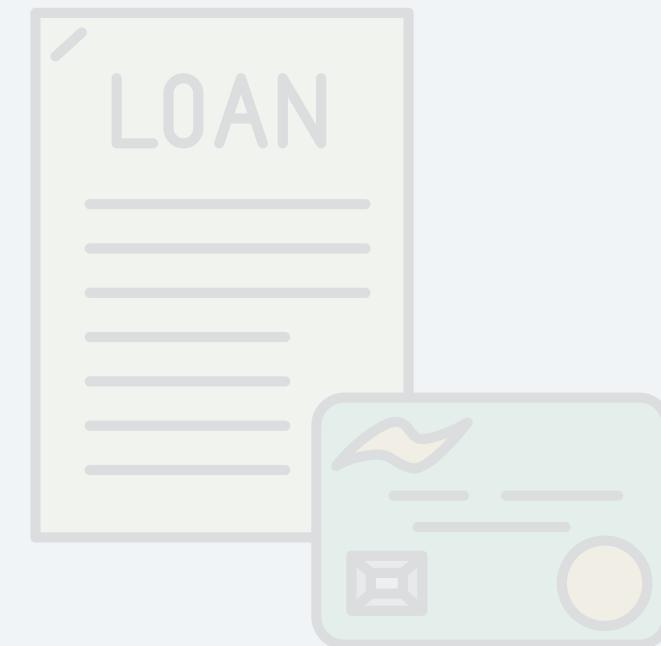
7 rows × 45 columns

In [232]: 1 Approved_loan[Approved_loan['name'] == 'ANTHONY WILLIAMS']													
	name	amount	state	address	city	zip	naics_code	business_type	jobs_retained	date_approved	...	originating_lender_state	loan_id
469120	ANTHONY WILLIAMS	20832.00	GA	1421 Edgerly Ave	ALBANY	31707-4301	812111.0	Sole Proprietorship	1	2021-05-02	...		
483301	ANTHONY WILLIAMS	4166.00	GA	5452 Pepperwood Ct	STONE MOUNTAIN	30087-5732	336412.0	Self-Employed Individuals	1	2021-04-01	...		
516913	ANTHONY WILLIAMS	20833.00	GA	6275 Baltusrol Trce	FAIRBURN	30213-5125	531190.0	Independent Contractors	1	2021-05-29	...		
524258	ANTHONY WILLIAMS	20832.50	GA	1421 Edgerly Ave	ALBANY	31707-4301	812111.0	Sole Proprietorship	1	2021-02-14	...		
529714	ANTHONY WILLIAMS	16666.00	GA	128 Northridge Dr	MACON	31220-6641	488410.0	Self-Employed Individuals	1	2021-03-23	...		

36 rows × 45 columns



WHAT ARE THE DIFFERENCES BETWEEN APPROVED & REMOVED LOANS?



INSIGHTS

CONGRESSIONAL DISTRICT

congressional_district	Approved_count	Removed_count	removal_percent	total_amount	Party
GA-13	64945	4044	5.861804	68989	Democratic
GA-04	63122	3656	5.474857	66778	Democratic
GA-02	36658	2079	5.366962	38737	Democratic
GA-05	75046	3995	5.054339	79041	Democratic
GA-03	38385	2006	4.966453	40391	Republican
GA-10	28923	1372	4.5288	30295	Republican
GA-08	30761	1416	4.400659	32177	Republican
GA-12	26053	1140	4.192255	27193	Republican
GA-01	24420	1028	4.03961	25448	Republican
GA-14	16546	614	3.578089	17160	Republican
GA-07	48446	1658	3.309117	50104	Democratic
GA-11	38271	1250	3.162875	39521	Republican
GA-09	19144	486	2.475802	19630	Republican
GA-06	43078	1092	2.472266	44170	Democratic

HUBZONE

hubzone_indicator	Approved_count	Removed_count	removal_percent
N	374604	19511	4.950585
Y	179224	6325	3.408803

INSIGHTS

BUSINESSES THAT APPLIED TO LOCAL LENDER TEND TO HAVE THEIR LOANS APPROVED

Value	Count	Frequency (%)
GA	118554	21.4%
CA	69660	12.6%
FL	69375	12.5%
AZ	52081	9.4%
TX	48224	8.7%
NJ	39511	7.1%
NC	37918	6.8%
NY	20324	3.7%
PA	17146	3.1%
UT	16020	2.9%
Other values (43)	65015	11.7%

APPROVED

Value	Count	Frequency (%)
TX	6713	26.0%
CA	5750	22.3%
AZ	5283	20.4%
PA	2797	10.8%
MA	1942	7.5%
FL	1575	6.1%
NY	688	2.7%
NJ	589	2.3%
GA	279	1.1%
VA	98	0.4%
Other values (14)	122	0.5%

REMOVED

INSIGHTS

BEAUTY SALONS ARE THE MOST COMMON AND FREQUENT INDUSTRY TYPE (NAICS TITLE-WISE) IN BOTH APPROVED & REMOVED APPLICATION POOL.

Value	Count	Frequency (%)
Beauty Salons	40609	7.3%
All Other Personal Services	16601	3.0%
Taxi Service	13289	2.4%
Barber Shops	12406	2.2%
Janitorial Services	11671	2.1%
General Freight Trucking, Local	10533	1.9%
Landscaping Services	9944	1.8%
All Other Miscellaneous Store Retailers (except ...)	9680	1.7%
Full-Service Restaurants	9606	1.7%
Offices of Real Estate Agents and Brokers	9347	1.7%
Other values (1030)	404407	73.0%

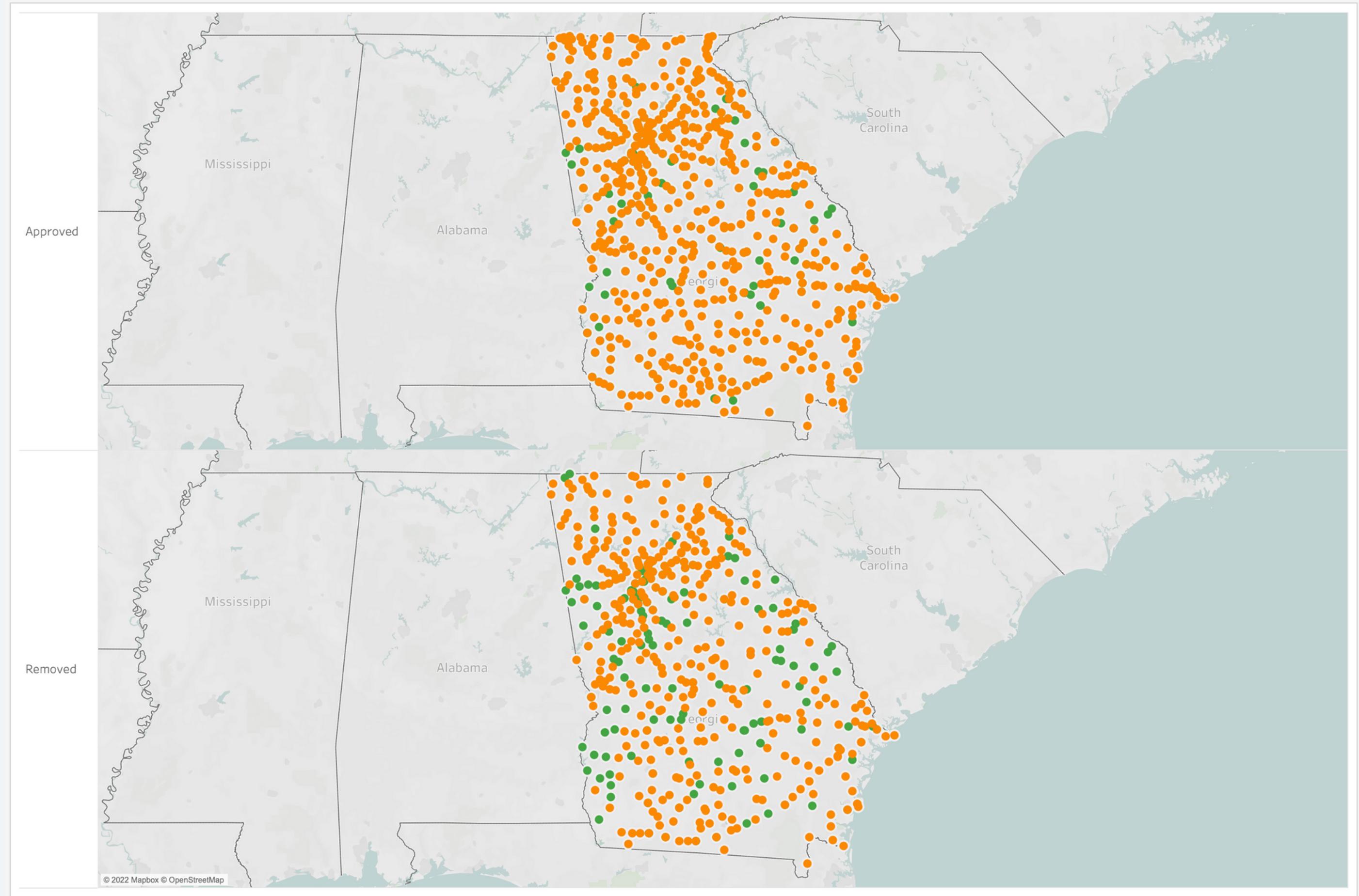
APPROVED

Value	Count	Frequency (%)
Beauty Salons	3249	12.6%
Barber Shops	1132	4.4%
Taxi Service	875	3.4%
Janitorial Services	763	3.0%
Landscaping Services	745	2.9%
All Other Personal Services	696	2.7%
Caterers	615	2.4%
Residential Remodelers	600	2.3%
All Other Miscellaneous Store Retailers (except Tobacco Stores)	556	2.2%
Independent Artists, Writers, and Performers	512	2.0%
Other values (688)	16091	62.3%

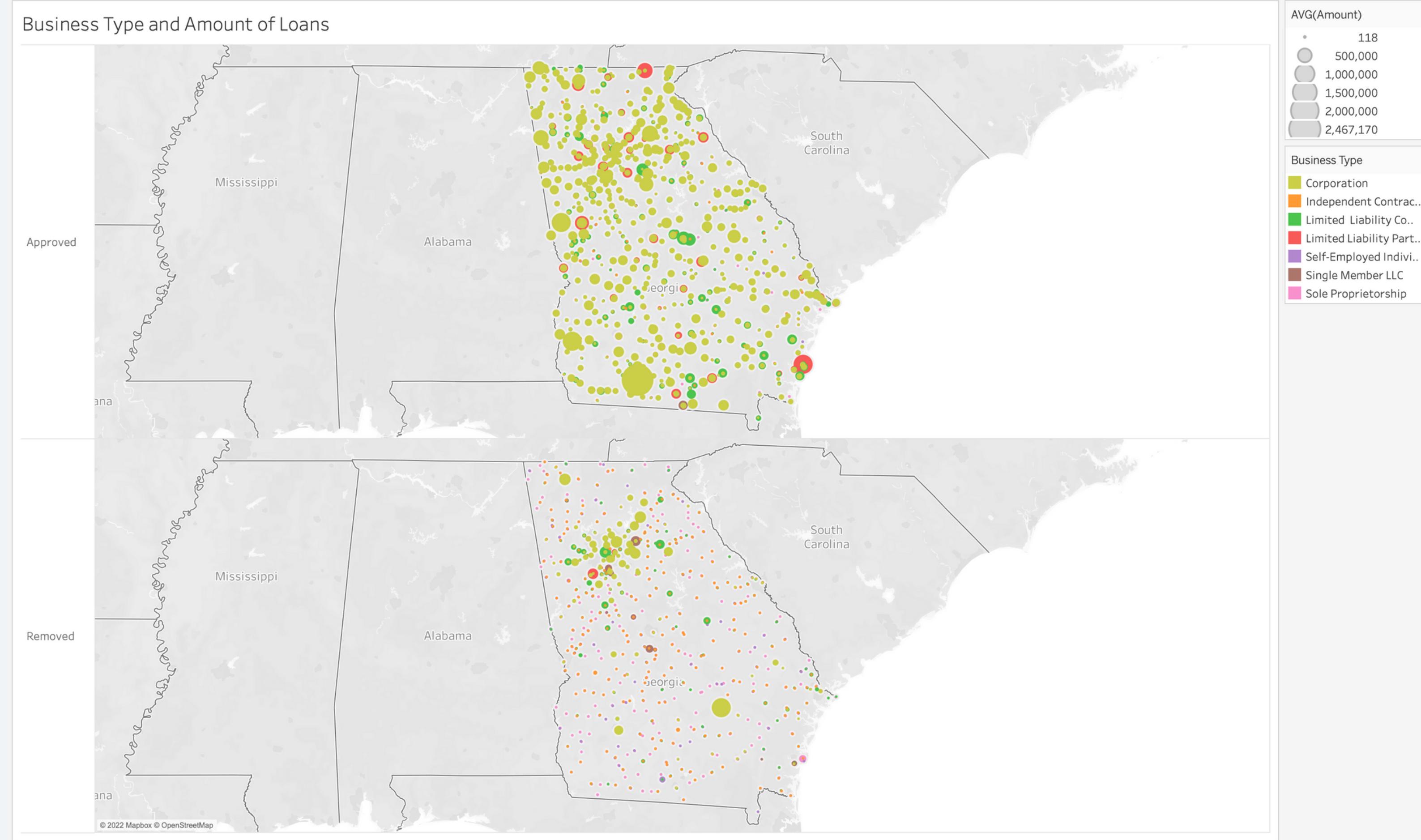
REMOVED



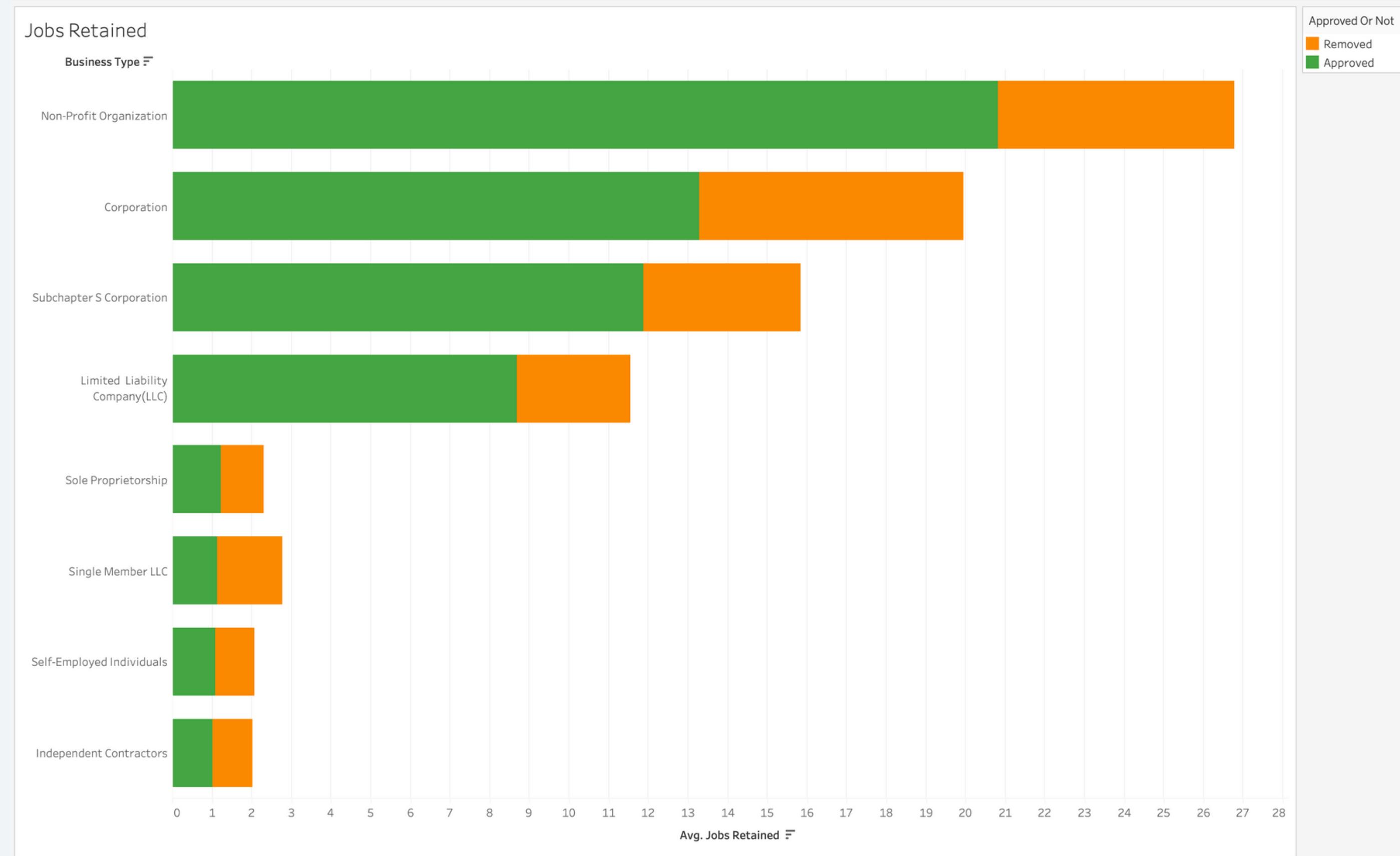
- MAXIMUM APPROVED BUSINESSES DO NOT COME FROM A LOW-MEDIUM INCOME AREA.



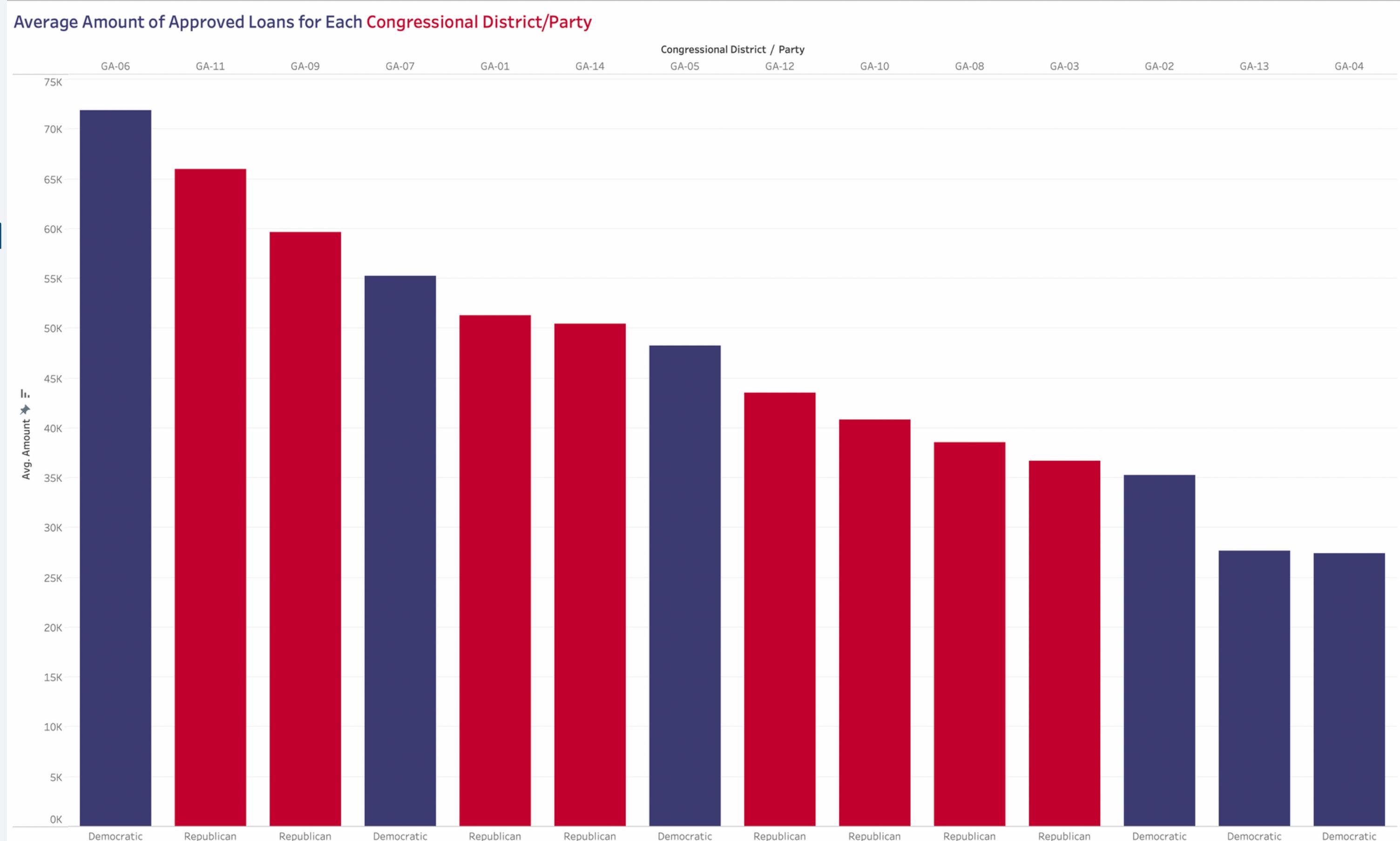
- THE CORPORATION BUSINESS CATEGORY TENDS TO GET A LARGER AMOUNT OF LOANS APPROVED. SMALL BUSINESSES ARE MORE PRONE TO GETTING REJECTED



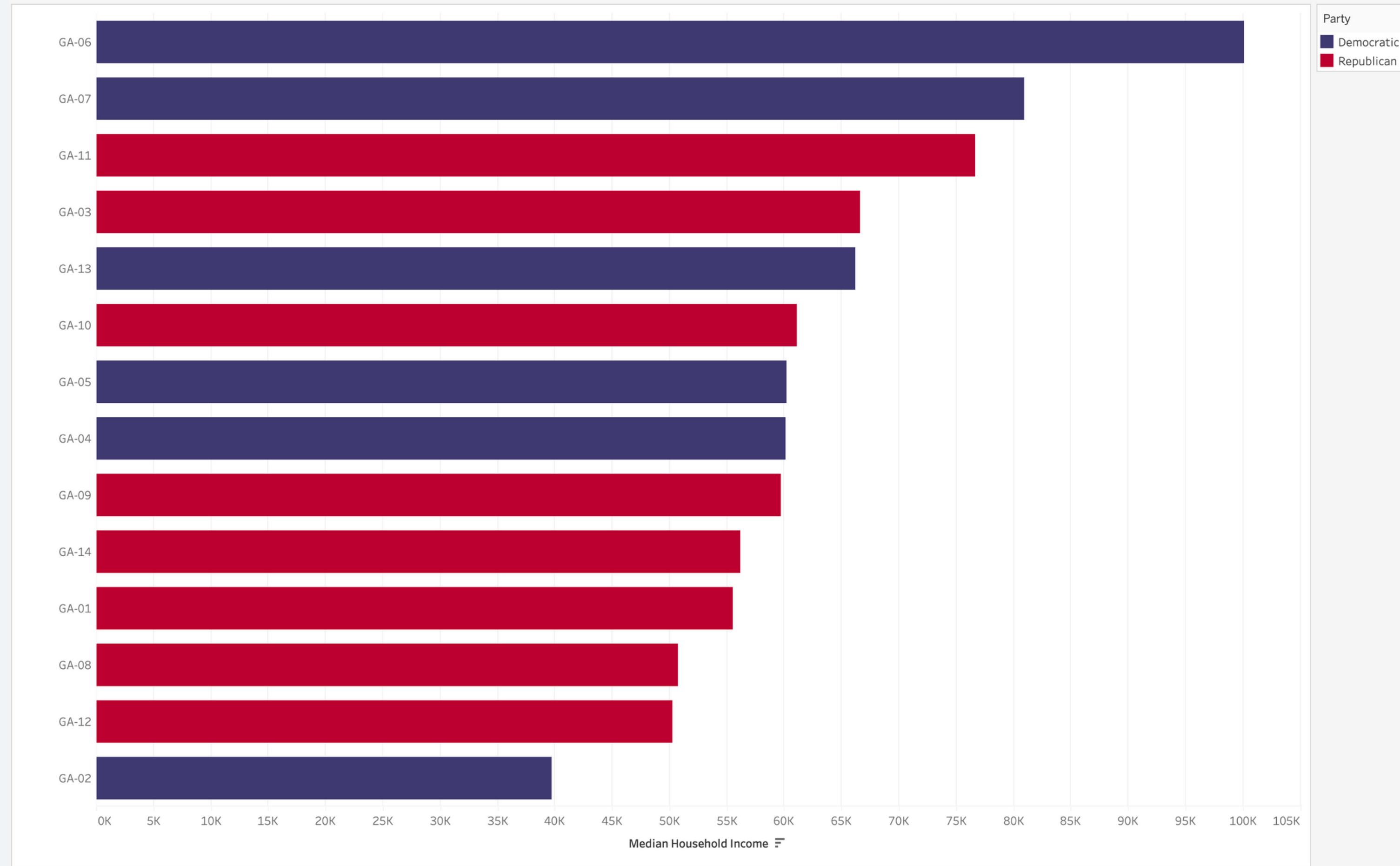
- APPLICANTS WITH THE HIGHEST JOB RETENTION ARE MORE LIKELY TO GET THEIR LOANS APPROVED.



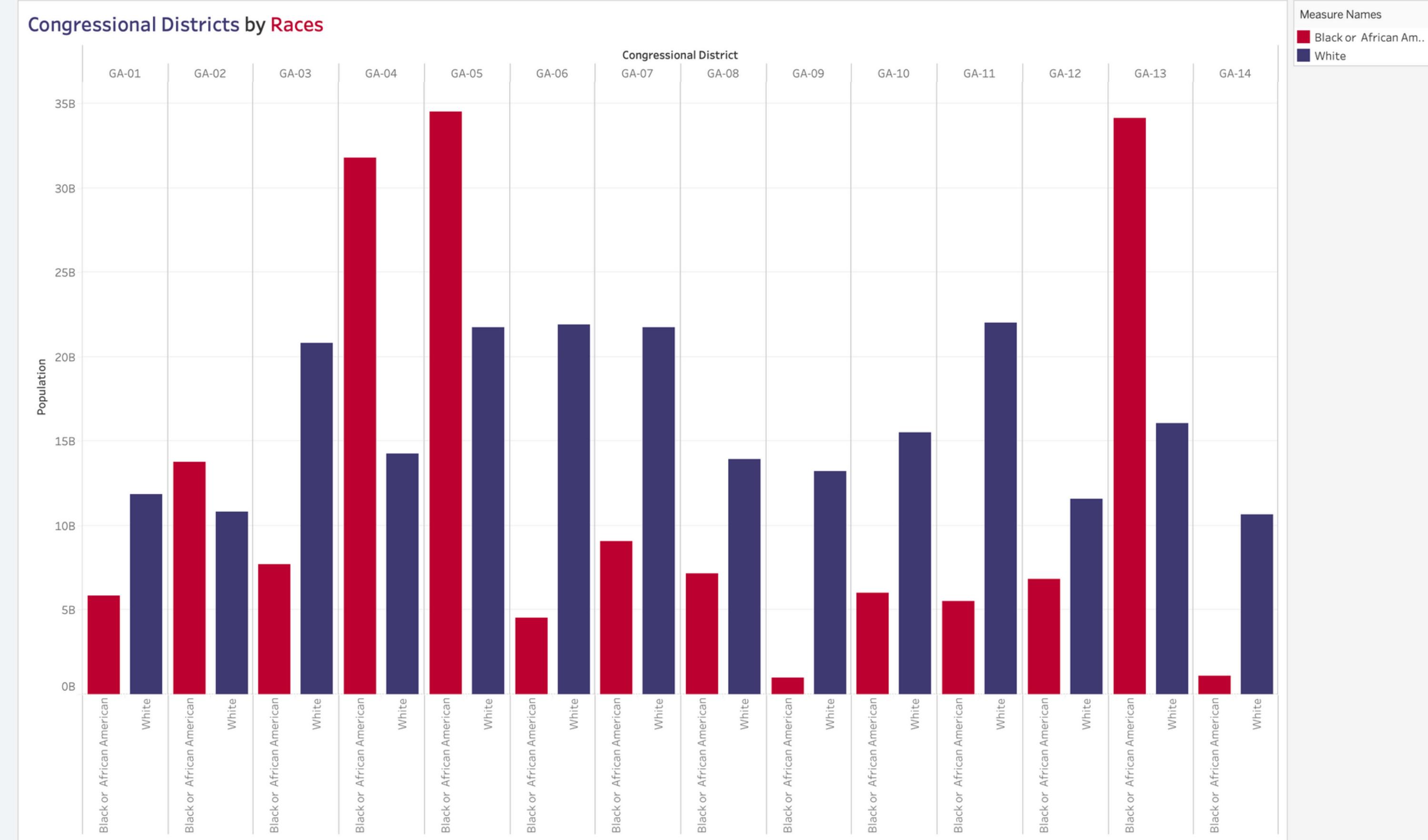
- HIGHEST REJECTION FROM BELOW CONGRESSIONAL DISTRICTS: GA-13, GA-04, GA-2
- GA-6 HAS THE LOWEST REJECTION



- MAXIMUM APPROVED BUSINESSES DO NOT COME FROM A LOW-MEDIUM INCOME AREA.



- HIGHEST REJECTION FROM BELOW CONGRESSIONAL DISTRICTS: GA-13, GA-04, GA-2.
- ALL THESE DISTRICTS ARE PRIMARILY INHABITED BY BLACK/ AFRICAN AMERICAN PEOPLE
- GA-6 HAS THE LOWEST REJECTION WITH THE HIGHEST WHITE POPULATION



REDLINING

" Refuse (a loan or insurance) to someone because they live in an area deemed to be a poor financial risk "

POTENTIAL FACTORS

THAT WE
CONSIDERED
AFTER ANALYZING
THE DATASET



BUSINESS
TYPE



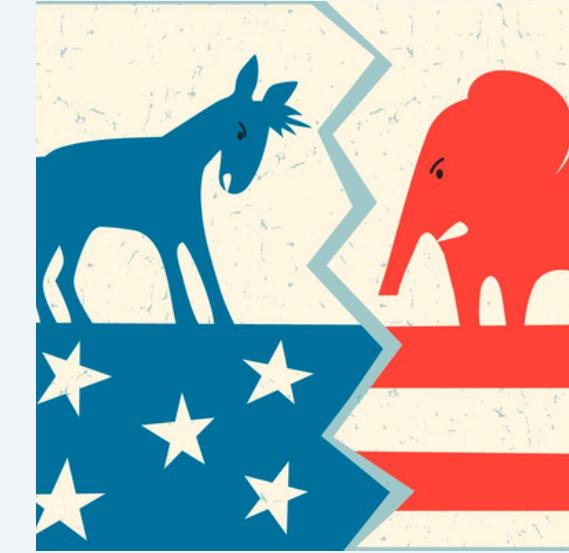
RACIAL
DISCRIMINATION



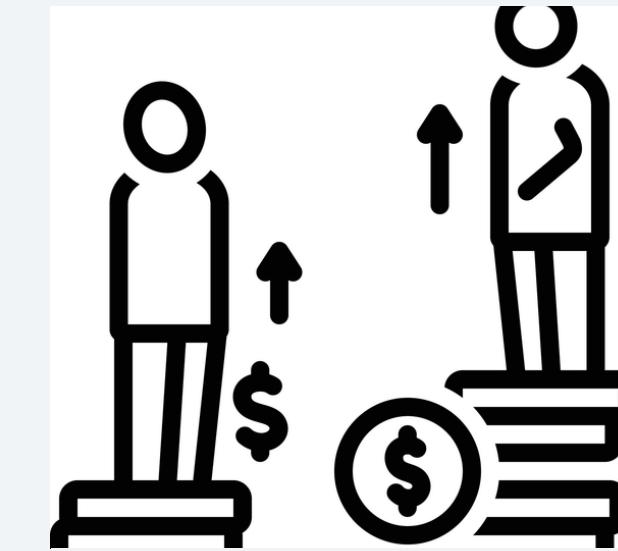
HUBZONE
INDICATOR



#JOBS
RETAINED



POLITICAL
AFFILIATION



INCOME
INEQUALITY



RURAL/URBAN
INDICATOR



LENDERS
LOCATION

MODEL INSIGHTS

```
1 ml_loans = loans[['amount', 'city', 'naics_title', 'business_type', 'jobs_retained',
2                   'congressional_district', 'processing_method', 'term', 'zipcode',
3                   'rural_urban_indicator', 'hubzone_indicator', 'business_age_description',
4                   'project_city', 'originating_lender_state', 'lmi_indicator', 'Party', 'Asian',
5                   'Black_or_African-American', 'White', 'Others', 'Median Household Income',
6                   'approved_or_not']].copy()
```

OVERSAMPLING

UNDERSAMPLING

MODELLING STAGE

```
clf = DecisionTreeClassifier()  
  
clf = clf.fit(X_train,y_train)  
  
y_pred = clf.predict(X_test)  
  
accuracy_score(y_test,y_pred)  
0.9277974783293932
```

```
cf_matrix = confusion_matrix(y_test, y_pred)  
print(cf_matrix)  
  
[[38258  496]  
 [ 8300 74770]]
```

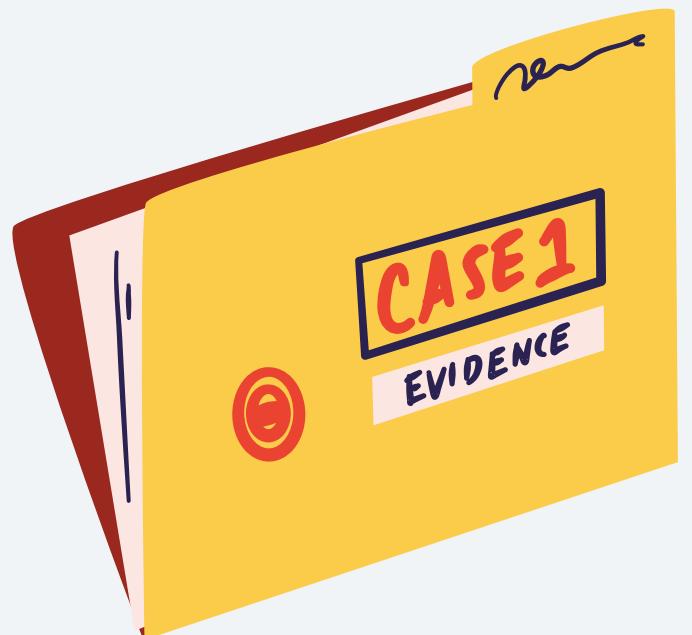
92%

ACCURACY
OF PREDICTING
WHETHER OR NOT A
LOAN WAS REMOVED
FROM THE DATA

SUGGESTION FOR IMPROVING THE PPP



- THE LENDERS SHOULD BE ABLE TO CHECK APPLICANT PORTAL IN REALTIME AND INTER-LENDER COMMUNICATION SHOULD TAKE PLACE BEFOR APPROVING ANY LOAN TO CROSSCHECK WHETHER THE APPLICANT HAVE SIMULTANEOUSLY THROUGH MULTIPLE CHANNELS.



- REGULATORY BENCHMARKS SHOULD BE IMPOSED UPON THE LENDERS. THEY MUST ENSURE THAT A CERTAIN % OF THE PPP MONEY IS BEING ALLOTTED TO HUBzone BUSINESSES & LMI AREAS NEIGHBORHOODS THAT NEEDS THE MONEY MOST). IN ORDER TO KEEP THEIR GOV. APPROVED LENDER STATUS

Thank
you

