

CARES Act Economic Relief Plan

# Paycheck Protection Program



## Info Challenge 2022

Data Analytics L3

**Team 22007:**

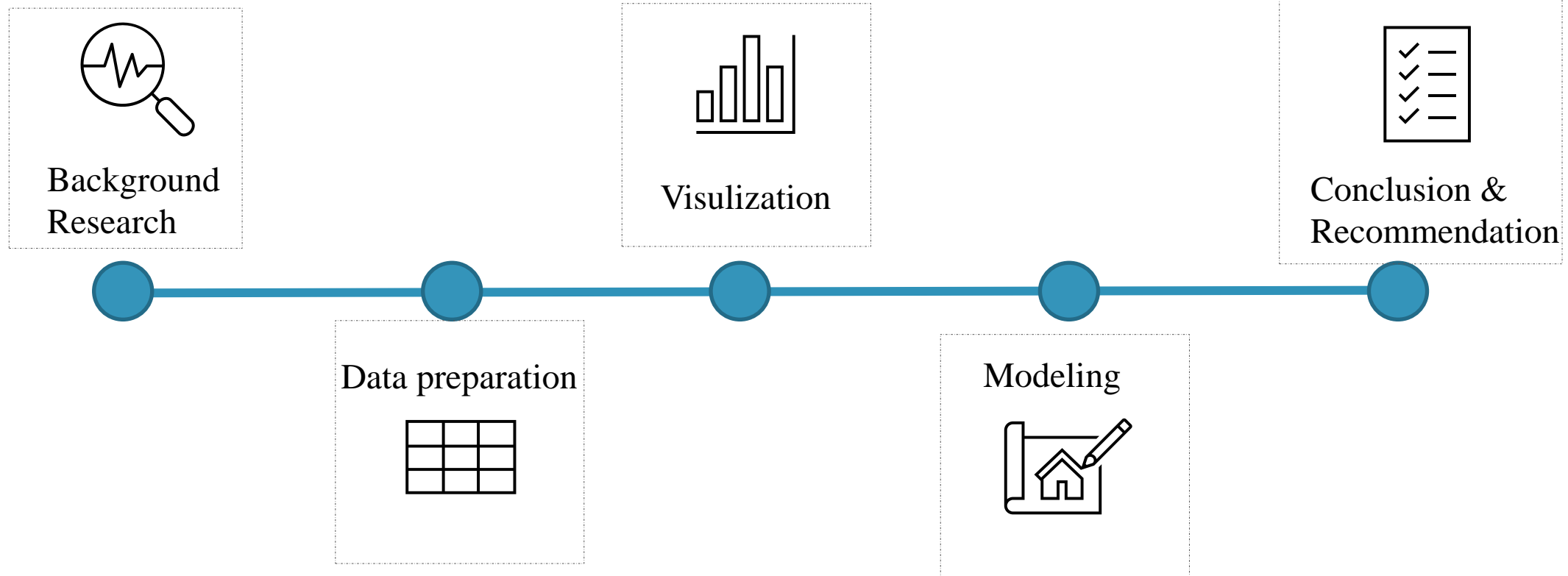
Tianli Ding

Rui Jin

Zihan Zhang

Jieqian Xiao

# Data Analytic Progress



# Paycheck Protection Program

- Uncollateralized
- Low-interest rate
- Potential to be forgiven
- Up to 2.5 times their monthly pre-COVID payroll
- **No “credit elsewhere” rule**
- **“First-come-first-served”**



U.S. Small Business Administration



**Administered by Banks**



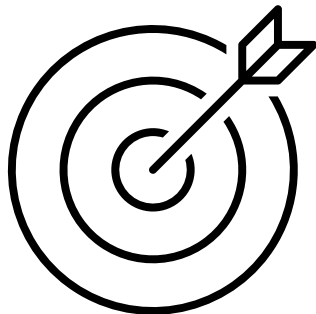
**Small and medium sized firms**

# Project Target

&

# Data Set

- Compare the **removed** and **unremoved** loans
- Find defining characteristics of the removed loans
- Predict the possible removed loans
- Discover the funds flow
- Give recommendations to both government and small business



- The information of removed and approved PPP loans in **GA**
- From 2020-04-03 to 2021-06-30
- Most variables are dummy variables
- Location information takes a big part

Removed	Approved	Variables
25836	553828	41

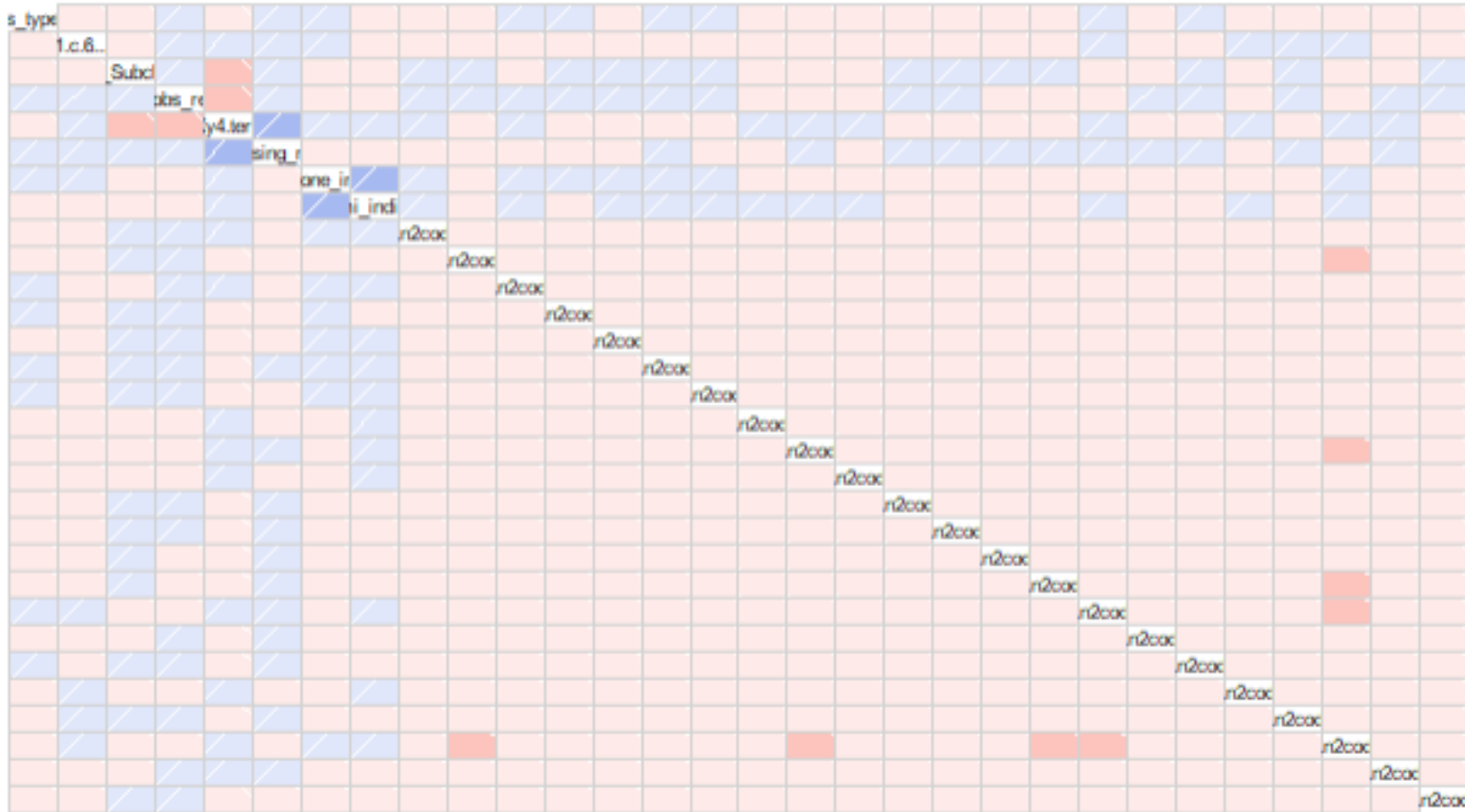


- **Amount of money**
- **Loan approval rate = approved / total**

# Correlation Heatmap

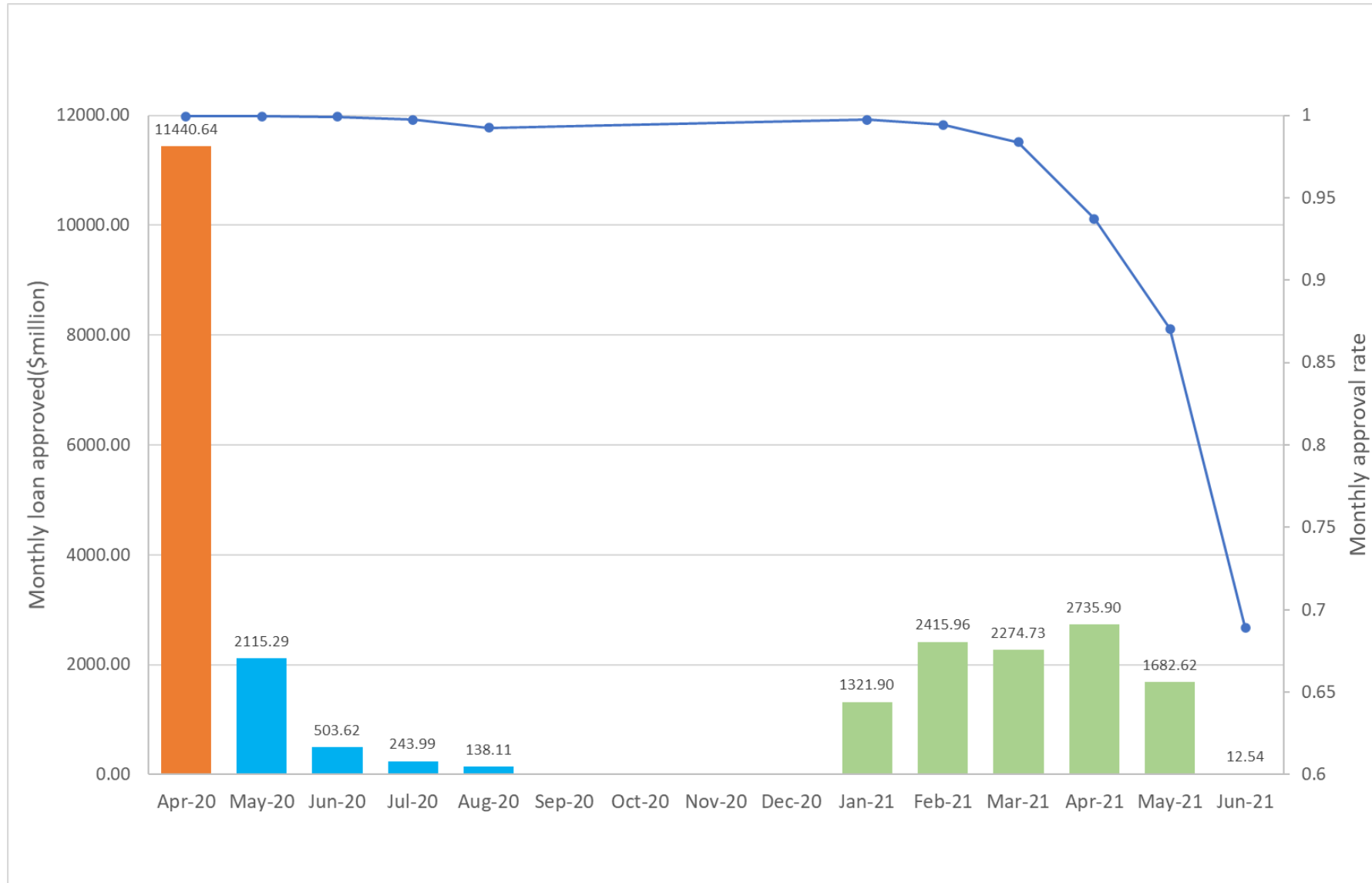
Correlation plot of *full processed dataset*.

Weak correlation between each element. Low risk of multicollinearity issue.



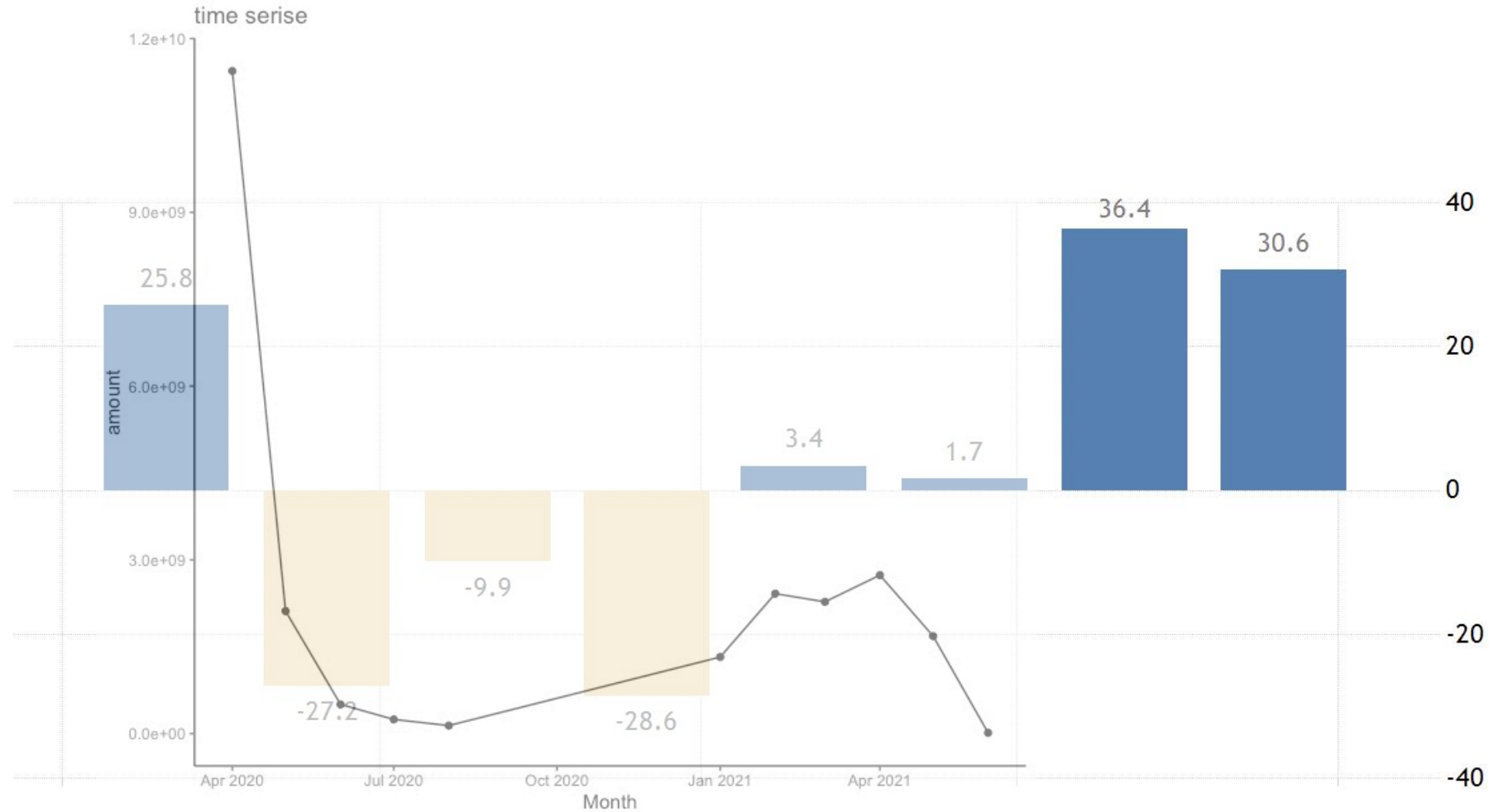
*full processed dataset: non-input data and redundant data has been removed already*

# Overview loan and approval rate



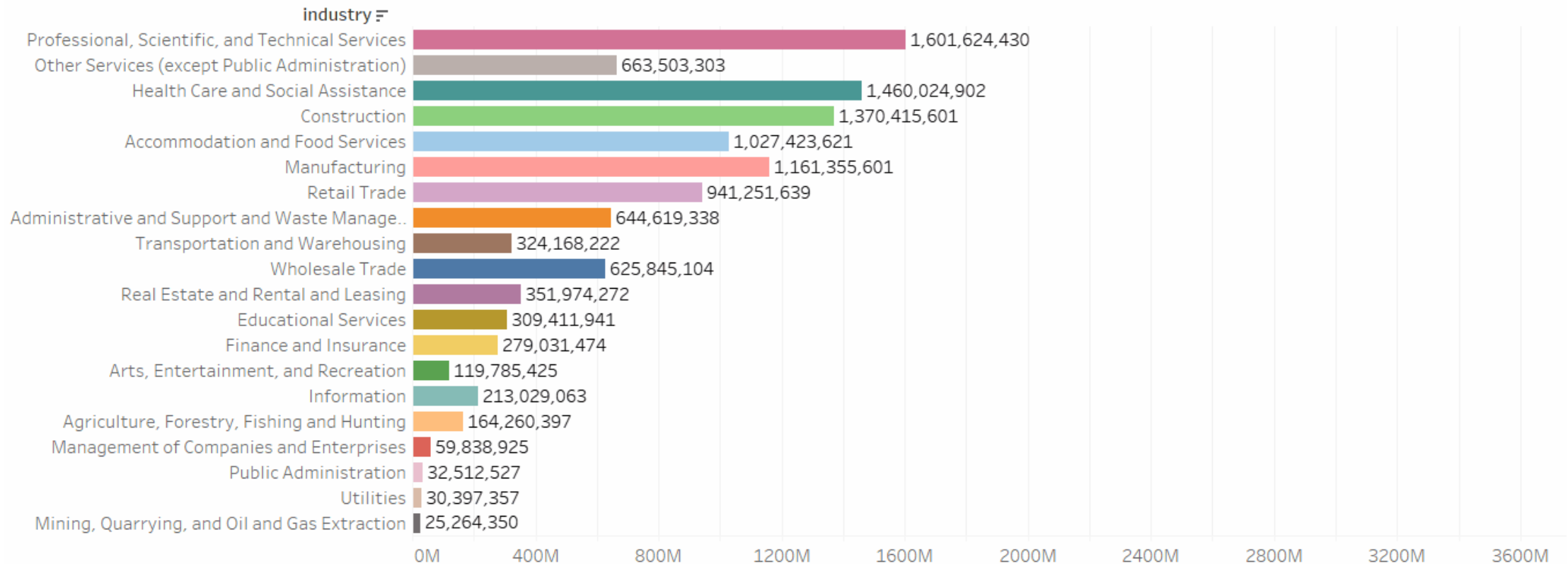
- Since Feb 2021, the loan approval rate decreased quickly.
- Before that, the approval rate is nearly 100%

# Business confidence Index



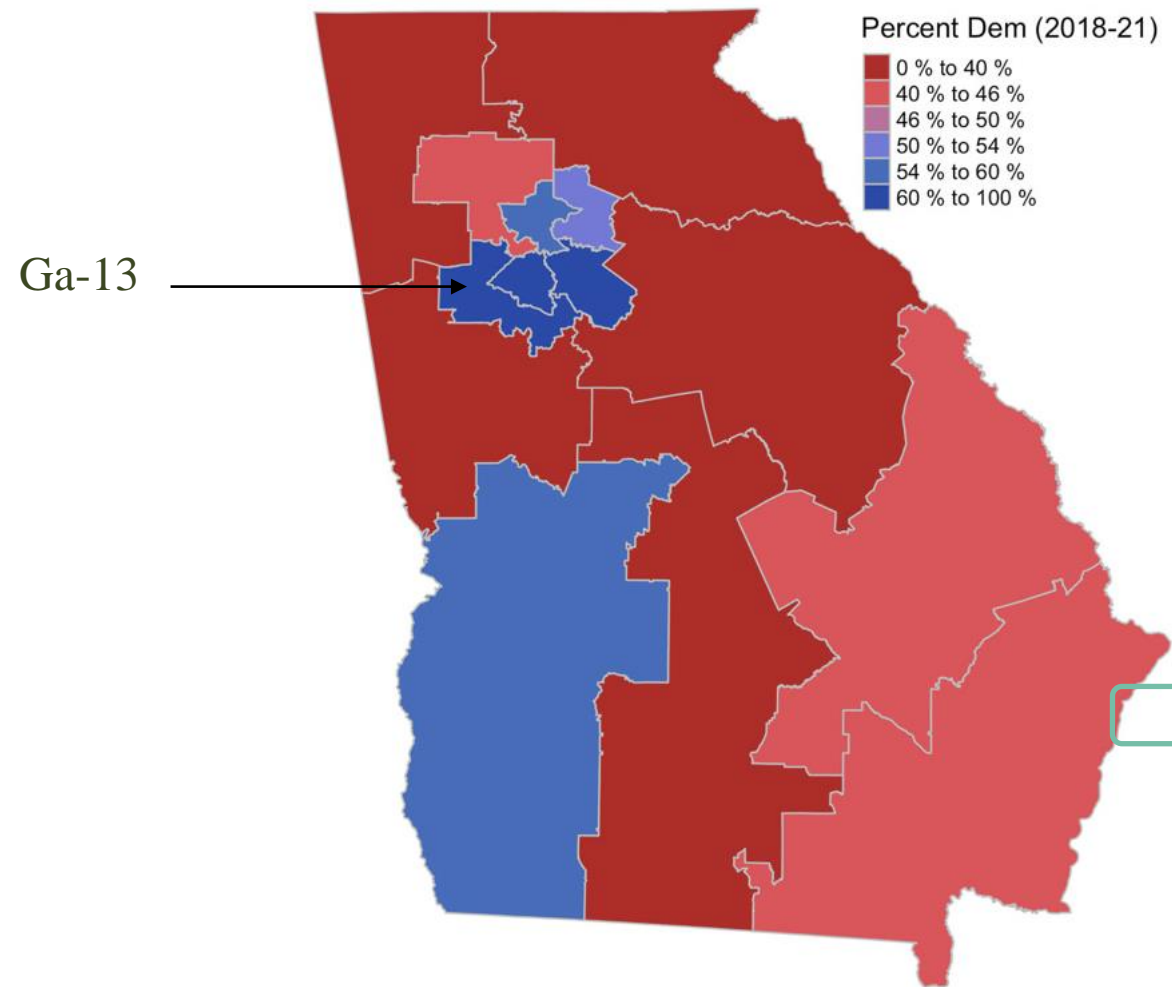
# Which industry money flows into most?

Culminative approved loan to different industries - April 2020





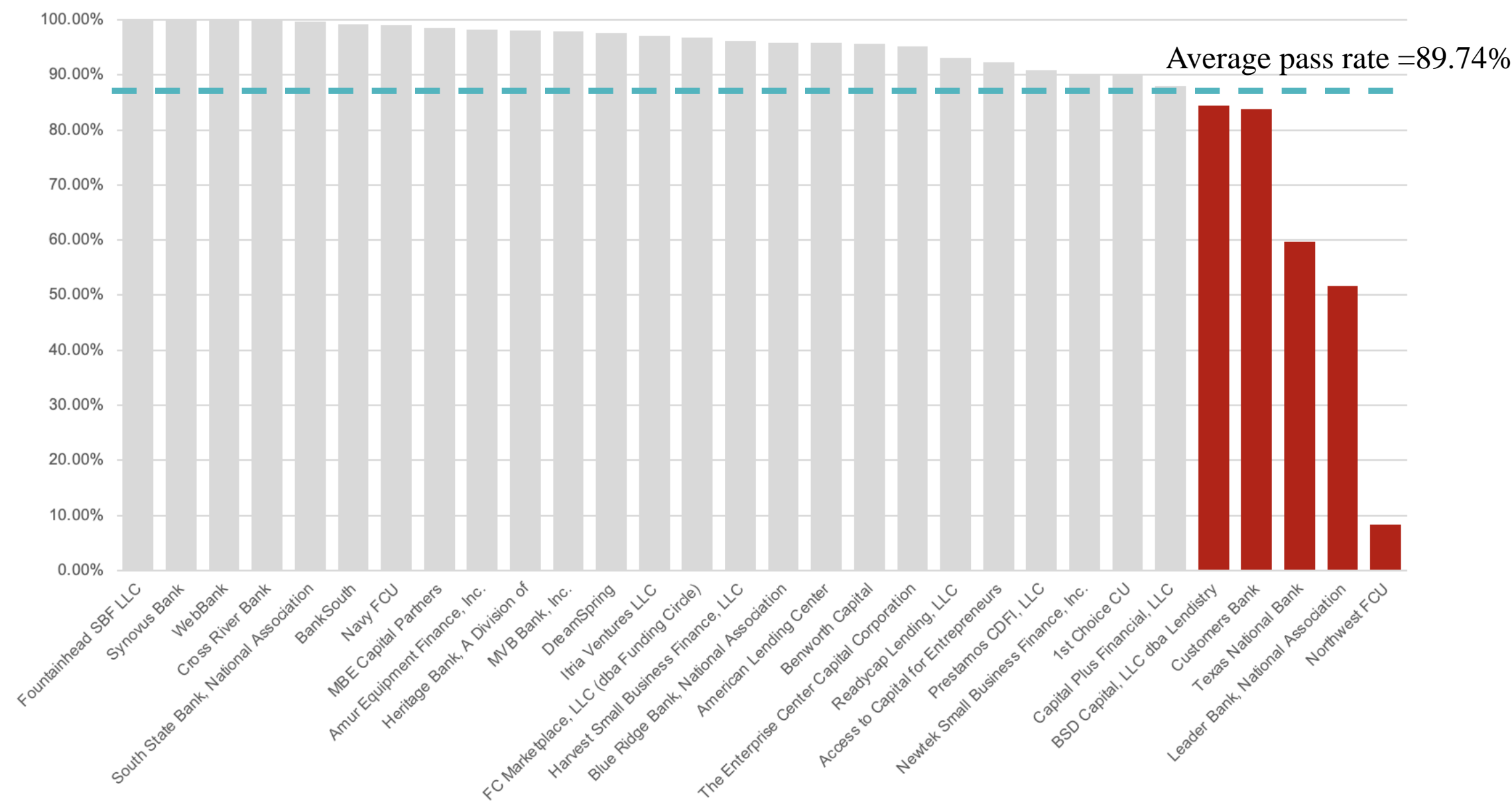
# Loan pass rates of congressional Districts



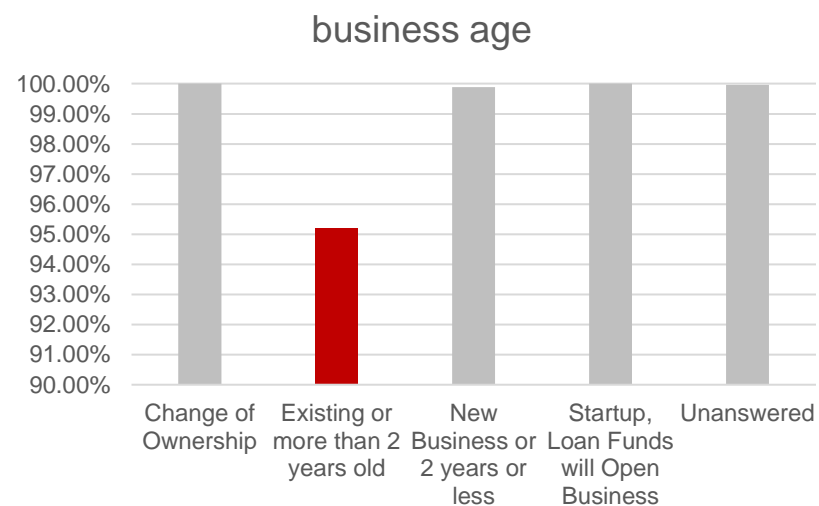
congressional_district	pass rate_amount	pass rate_companys
GA-01	98.59%	95.96%
GA-02	97.14%	94.63%
GA-03	97.33%	95.03%
GA-04	96.03%	94.53%
GA-05	97.79%	94.95%
GA-06	99.15%	97.53%
GA-07	98.78%	96.69%
GA-08	97.92%	95.60%
GA-09	99.23%	97.52%
GA-10	97.80%	95.47%
GA-11	98.99%	96.84%
GA-12	98.29%	95.81%
GA-13	95.15%	94.14%
GA-14	98.63%	96.42%

Democratic pass rate< Republican pass rate

# Lender

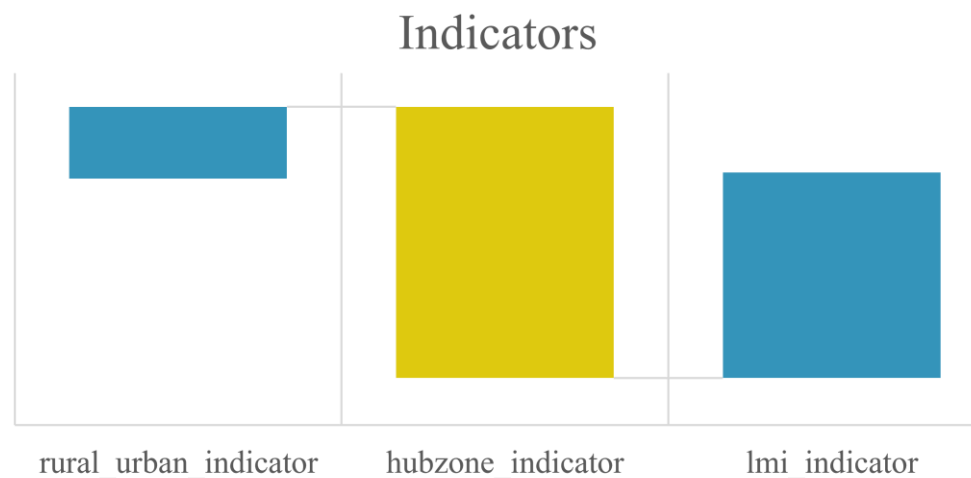


# Other factors

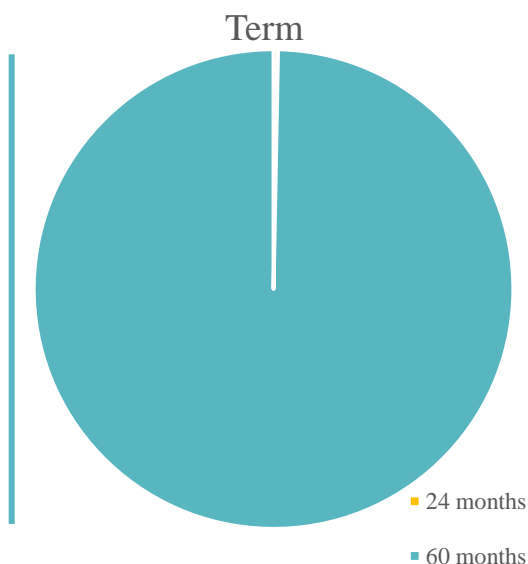


Existing or more than 2 years old < Others

Probably because they have a bad credit history with the bank

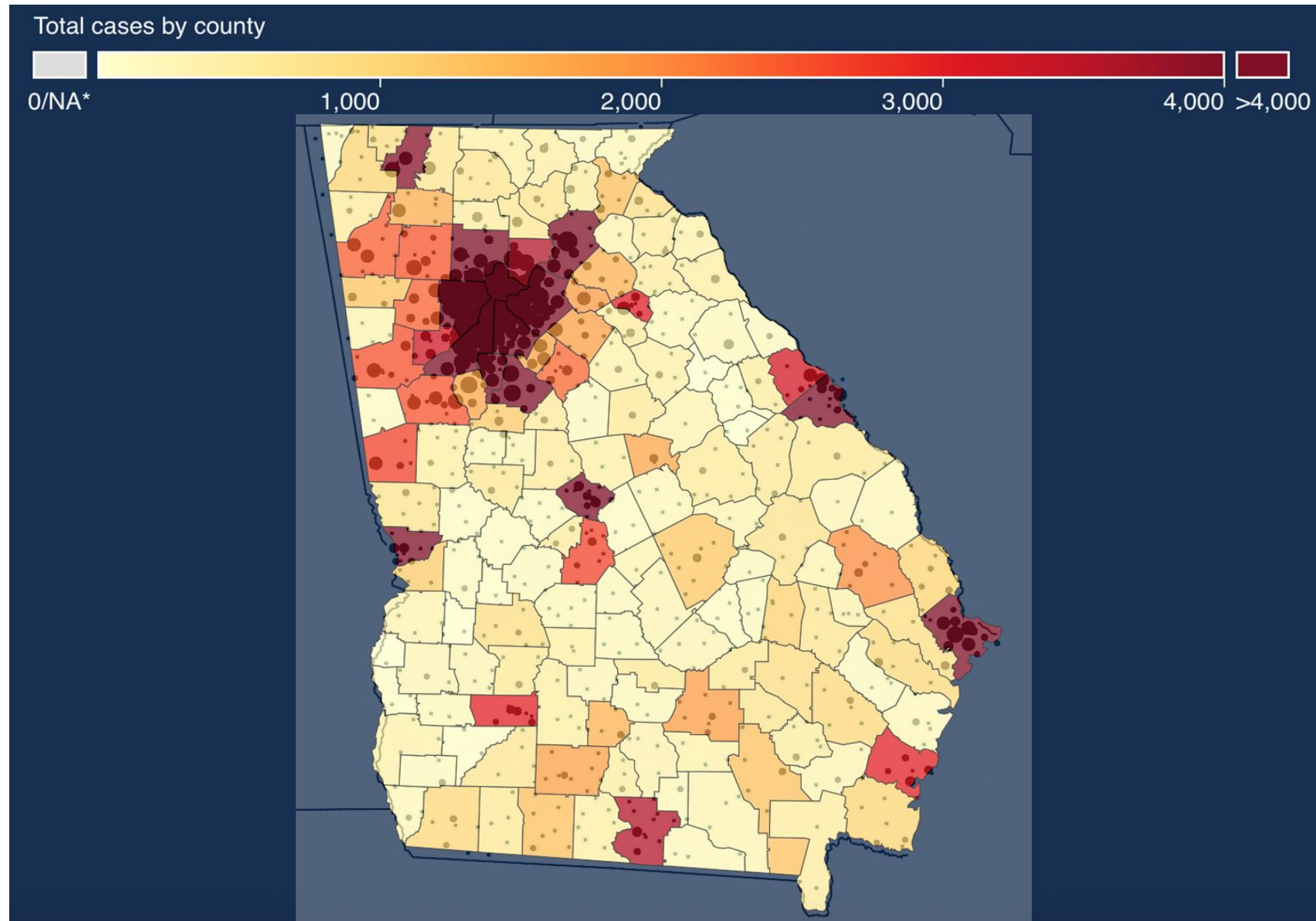


Rural > Urban  
Hubzone > No-hubzone  
No-lmi > lmi



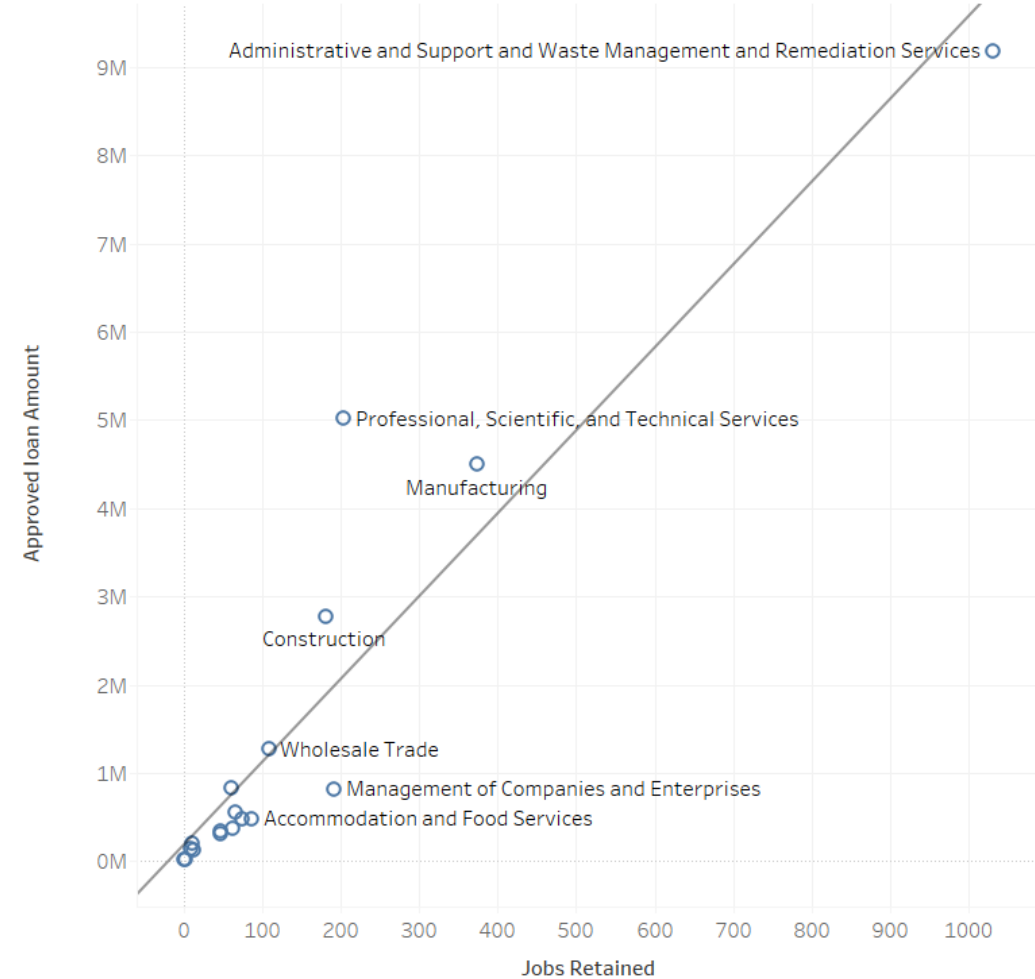
For removed set  
Term=60 takes 99.7%

# Did the PPP hit the target?



Is there any relationship between loan amount and company size for each industry?

A linear trend shows between job retained and approved loan amount through industries



Sum of Jobs Retained vs. sum of Amount. The marks are labeled by industry. Details are shown for industry. The data is filtered on Business Type, which keeps Housing Co-op, Joint Venture, NA and Trust.

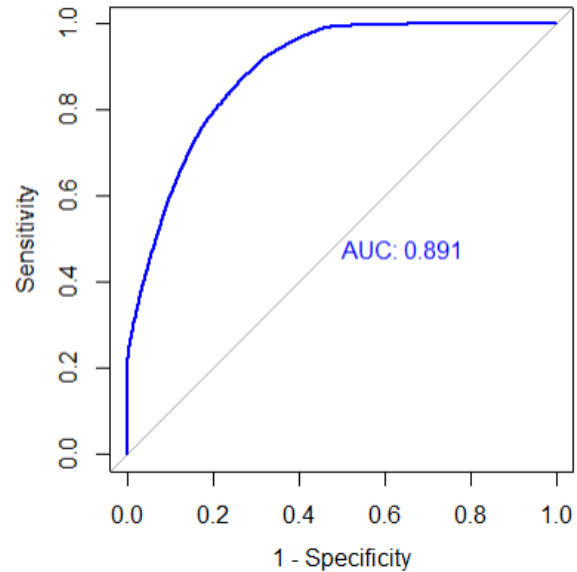
# Did the PPP hit the target?



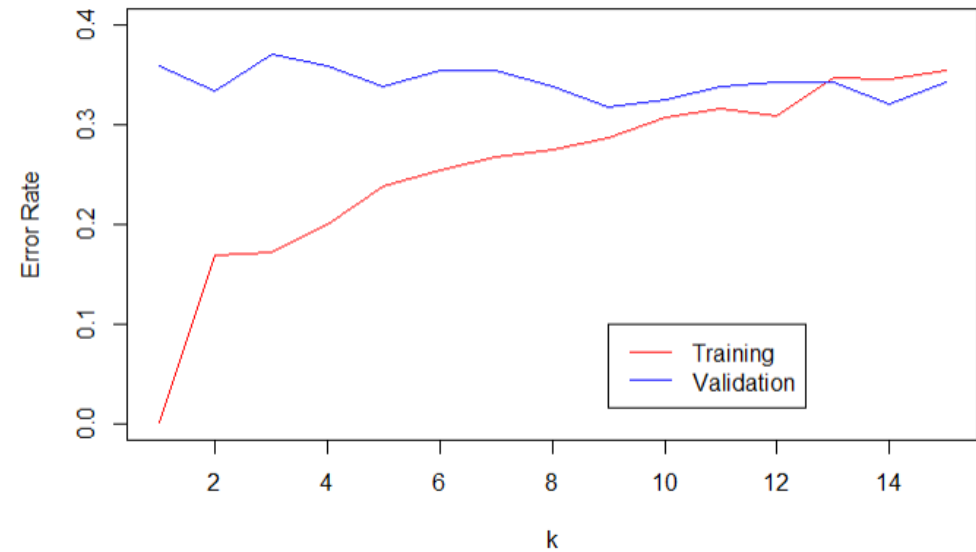
Size	Amount	offer jobs
<=100	80.33%	78.58%
100-200	9.20%	9.65%
200-300	4.22%	4.98%
300-400	2.25%	2.52%
<=500	4.00%	4.27%

# Logistic model and KNN algorithm

## Logistic regression model



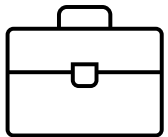
## K-nearest neighbors algorithm



	Logistic	KNN
Accuracy	0.8023107	0.69375
Sensitivity	0.8674717	0.8648649
Specificity	0.7371497	0.6422764
Dataset size used	25836*2(down sampling)	10000(random sample from down sampling set)

# Conclusion & Recommendation

- There is a lag of the effect of PPP program on SMEs.
- Banks do ensure the program carry on efficiently, but more banks or financial institutions involved will help the loan approval to be fairer.
- Business should be aware for following characteristics: amount applied, job retained, date applied, lender, hubzone, Low- and Moderate-Income Communities



- Larger Amount ↑
- More Job retained ↑
- Closer approach to the fund release date by SBA.
- Less likely to be removed



- Lenders
  - Avoid(BSD Capital, LLC dba Lendistry, Texas National Bank),
  - Prefer(Cross River Bank, Fountainhead SBF LLC)



- Hubzone business is less likely to be removed
- Low- and Moderate-Income Communities is more likely to be removed



# THANKS!

Mentor: Jeff Lang

**Paycheck Protection Program  
Borrower Application Form**

☐ Sole proprietor ☐ Partnership ☐ C-Corp ☐ S-Corp ☐ LLC  
☐ Independent contractor ☐ Eligible self-employed individual  
☐ 501(c)(3) nonprofit ☐ 501(c)(19) veterans organization  
☐ Tribal business (see 31(b)(2)(C) of Small Business Act) ☐ Other

**Business Legal Name**  
**Business Address**  
**Business Type**  
**Business Phone**  
**Primary Contact**  
**Number of Employees**  
**Business TIN (EIN, SSN)**

**Business Description**  
 (Please describe the business in detail, including the nature of the business, the products or services offered, and the market served.)

**Business Description**  
 (Please describe the business in detail, including the nature of the business, the products or services offered, and the market served.)

**Business Description**  
 (Please describe the business in detail, including the nature of the business, the products or services offered, and the market served.)

He is my fortress; I will trust.  
He shall deliver you from the snare of the fowler  
and from the perilous  
And under His wings  
refuge;  
His truth  
5

# Reference

- Bartik, Alexander W., Zoe B. Cullen, Edward L. Glaeser, Michael Luca, Christopher T. Stanton, and Adi Sunderam. 2020. “The Targeting and Impact of Paycheck Protection Program Loans to Small Businesses.” NBER Working Paper No. 27623.
- Rachel Atkins, Lisa Cook, Robert Seamans. 2021. “Discrimination in lending? Evidence from the Paycheck Protection Program.” *Small Bus Econ* (2022) 58:843–865.
- João Granja, Christos Makridis, Constantine Yannelis, Eric Zwick DID. 2020. “The Paycheck Protection Program Hit the Target?” NBER Working Paper 27095.
- R. Glenn Hubbard, Michael R. Strain. 2020. “Has the Paycheck Protection Program Succeeded?” NBER Working Paper No. 28032.

# Appendix 1: logistic model

```
glm.fit: fitted probabilities numerically 0 or 1 occurred  
Call:  
glm(formula = xy4$`ppp_full$removed` ~ ., family = binomial(),  
    data = xy4)
```

```
Deviance Residuals:  
    Min       1Q   Median       3Q      Max  
-3.0830  -0.3459   0.0000   0.7572   3.9277
```

Coefficients:

	Estimate
(Intercept)	-2.751e+01
amount	3.754e-06
business_type501(c)6 - Non Profit Membership	-5.074e-01
business_typeCooperative	-1.390e+01
business_typeCorporation	-1.271e+00
business_typeEmployee Stock Ownership Plan(ESOP)	1.674e+01
business_typeIndependent Contractors	-8.917e-01
business_typeJoint Venture	-1.441e+00
business_typeLimited Liability Company(LLC)	-1.208e+00
business_typeLimited Liability Partnership	-6.241e-01
business_typeNon-Profit Childcare Center	-1.260e+00
business_typeNon-Profit Organization	-9.363e-01
business_typePartnership	-9.008e-01
business_typeProfessional Association	-4.168e-01
business_typeQualified Joint-Venture (spouses)	-1.234e+00
business_typeSelf-Employed Individuals	-9.667e-01
business_typeSingle Member LLC	-7.788e-01
business_typeSole Proprietorship	-6.293e-01



# Appendix 1: logistic model

```
servicing_lender_nameD. L. Evans Bank
servicing_lender_nameDelta Community CU
servicing_lender_nameDouglas National Bank
servicing_lender_nameDreamSpring
servicing_lender_nameDrummond Community Bank
servicing_lender_nameDurden Banking Co., Inc.
servicing_lender_nameEast West Bank
servicing_lender_nameEastern Bank
servicing_lender_nameEconomic and Community Development Institute
servicing_lender_nameEmbassy National Bank
servicing_lender_nameEmory Alliance CU
servicing_lender_nameEmpower FCU
servicing_lender_nameEnterprise Bank & Trust
servicing_lender_nameEquity Bank
servicing_lender_nameEssex Bank
[ reached getOption("max.print") -- omitted 312 rows ]
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

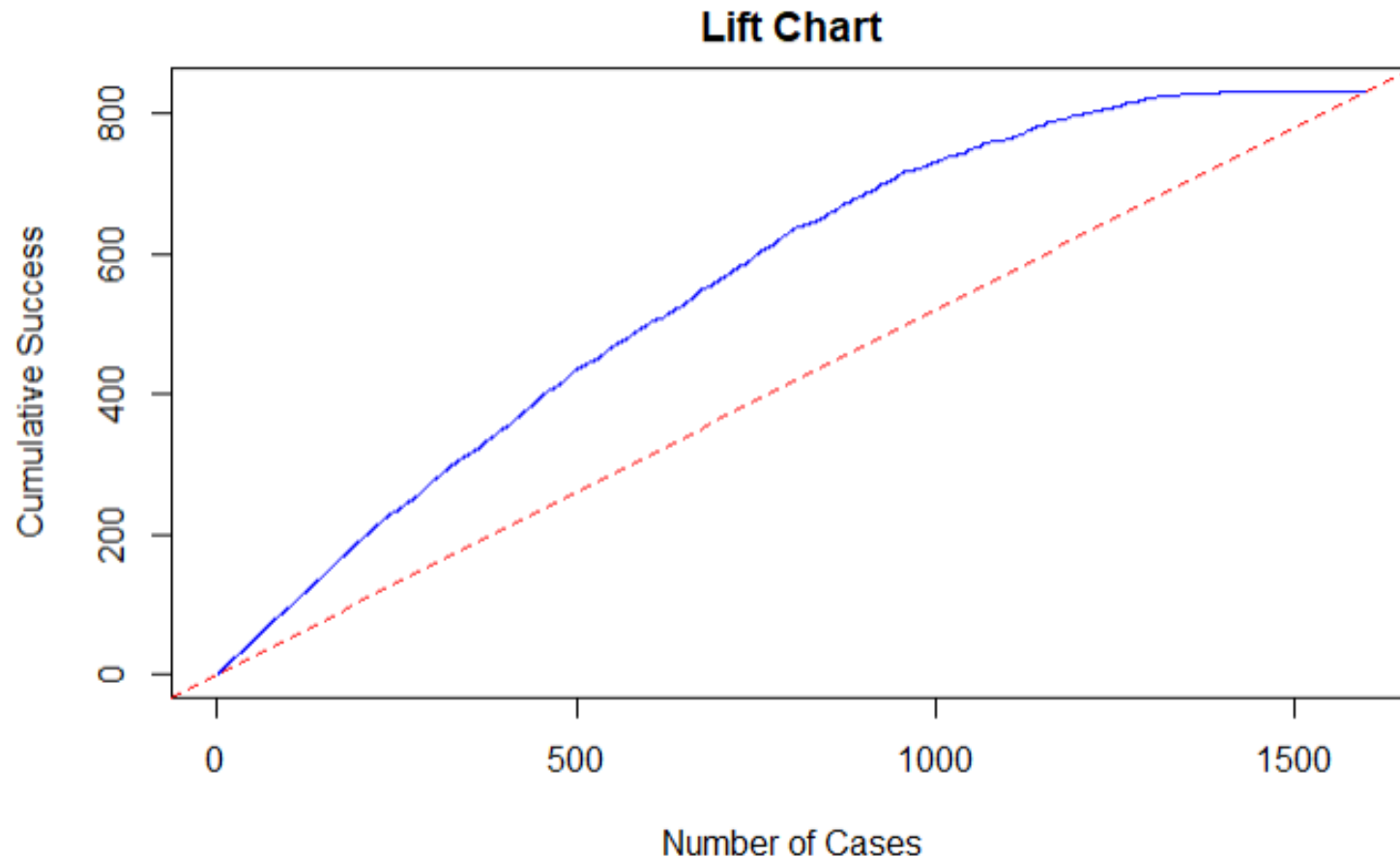
    Null deviance: 71633  on 51671  degrees of freedom
Residual deviance: 40184  on 51160  degrees of freedom
AIC: 41208

Number of Fisher Scoring iterations: 17
```

\*\*\*

For full model please check github code:<https://github.com/amatate/ppp-infochallenge1>

# Appendix2: KNN performance



Confusion matrix of train,  
validation, test

```
prediction    0    1
0 1949  240
1 1098 2713
```

```
prediction2    0    1
0   716  118
1   490 1076
```

```
prediction3    0    1
0   462   79
1   306  753
```

# Appendix 3: processed data for training

```
> summary(Xy4)
      ppp_full$removed      amount      business_type      jobs_retained      congressional_district
not removed:25836  Min.   :    56  Sole Proprietorship      :27507  Min.   : 0.000  GA-05 : 7487
removed      :25836  1st Qu.: 12500  Independent Contractors      : 6704  1st Qu.: 1.000  GA-13 : 7064
              Median : 20416  Limited Liability Company(LLC): 6027  Median : 1.000  GA-04 : 6616
              Mean   : 33301  Self-Employed Individuals      : 4817  Mean   : 3.118  GA-07 : 3888
              3rd Qu.: 20833  Corporation                     : 3631  3rd Qu.: 1.000  GA-03 : 3824
              Max.   :10000000  Subchapter S Corporation        : 1430  Max.   :500.000  GA-02 : 3776
                                   (Other)                        : 1556                (Other):19017

processing_method      term      rural_urban_indicator      hubzone_indicator
PPP:41467      Min.   : 6.00      R:11146      N:36979
PPS:10205      1st Qu.:60.00      U:40526      Y:14693
              Median :60.00
              Mean   :55.55
              3rd Qu.:60.00
              Max.   :60.00

      business_age_description      lmi_indicator      n2code      servicing_lender_name
Change of Ownership              :    1      N:32376      Length:51672      Length:51672
Existing or more than 2 years old :49711      Y:19296      Class :character      Class :character
New Business or 2 years or less   : 1253      Mode  :character      Mode  :character
Startup, Loan Funds will Open Business: 2
Unanswered                        : 705
```

# Appendix4: unused data comparison

