



# Paycheck Protection Program Data Analysis

**Team 29**

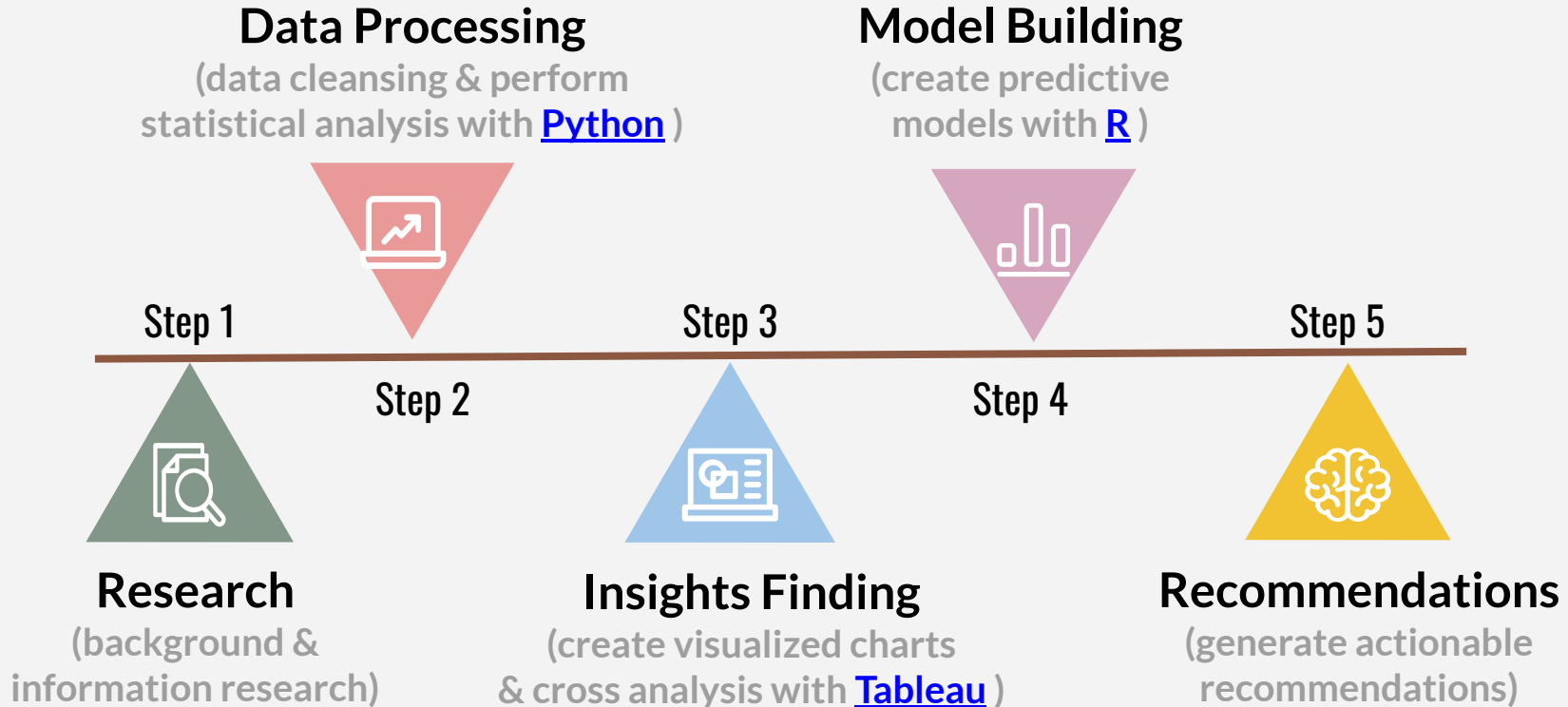
**Po-Han Yen, Hsin Chen, Jinping Guo, Jing Lin**

# ABOUT TEAM 29

We are current MS students in the UMD Robert H. Smith School of Business pursuing our degrees in Business Analytics, and we're so excited to join this competition!



# ANALYSIS PROCESS



# BACKGROUND INTRODUCTION

## Introduction of Paycheck Protection Program (PPP)

- Originated from the Coronavirus Aid, Relief, and Economic Security Act.
- Provide American small businesses with eight weeks of **cash-flow assistance to keep their workers on payroll.**
- The loans are backed by the Small Business Administration.

## Introduction of Data sets

- **PPP data** for the state of **Georgia**
- **Additional: CBP data** (from the US Census Bureau)
  - a. Include total wages for a given industry and region in Georgia.
  - b. Data resource: <https://www.census.gov/data/datasets/2019/econ/cbp/2019-cbp.html>



# MISSION OBJECTIVES

## Find

Insights to understand why certain loan applications have been removed from the program.



## Predict

Whether or not a loan application would be removed based on available information.



# Insights & Findings

- What characteristics make these applications removed?
- How do the characteristics of loans that were removed compare to loans that were not removed?

➡ **Removal Rate =** 
$$\frac{\text{number of removed applications}}{\text{total number of applications}}$$

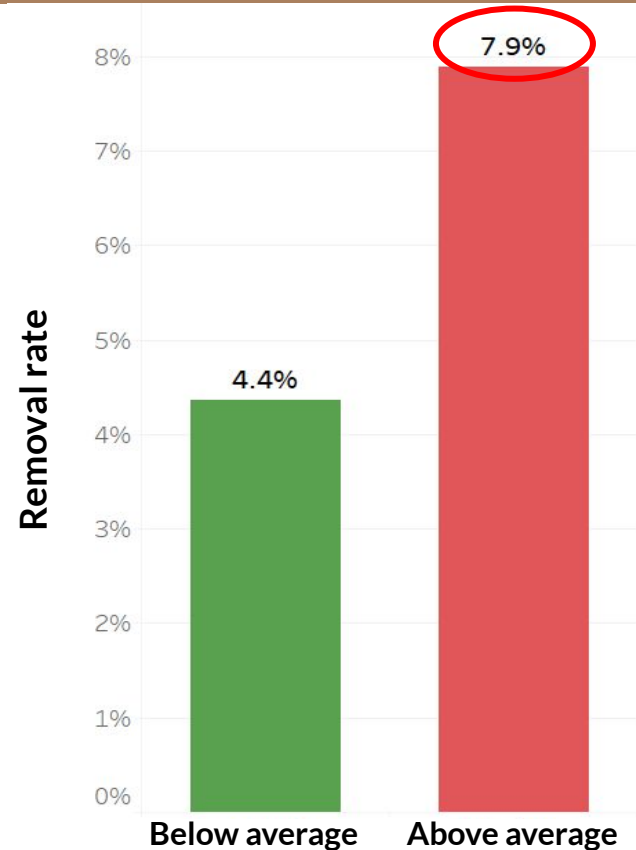
# Applied Amount

- Business applied amount of PPP loan is based on businesses' self-claimed payroll.
- Applied amount per person V.S. Industry average payroll.

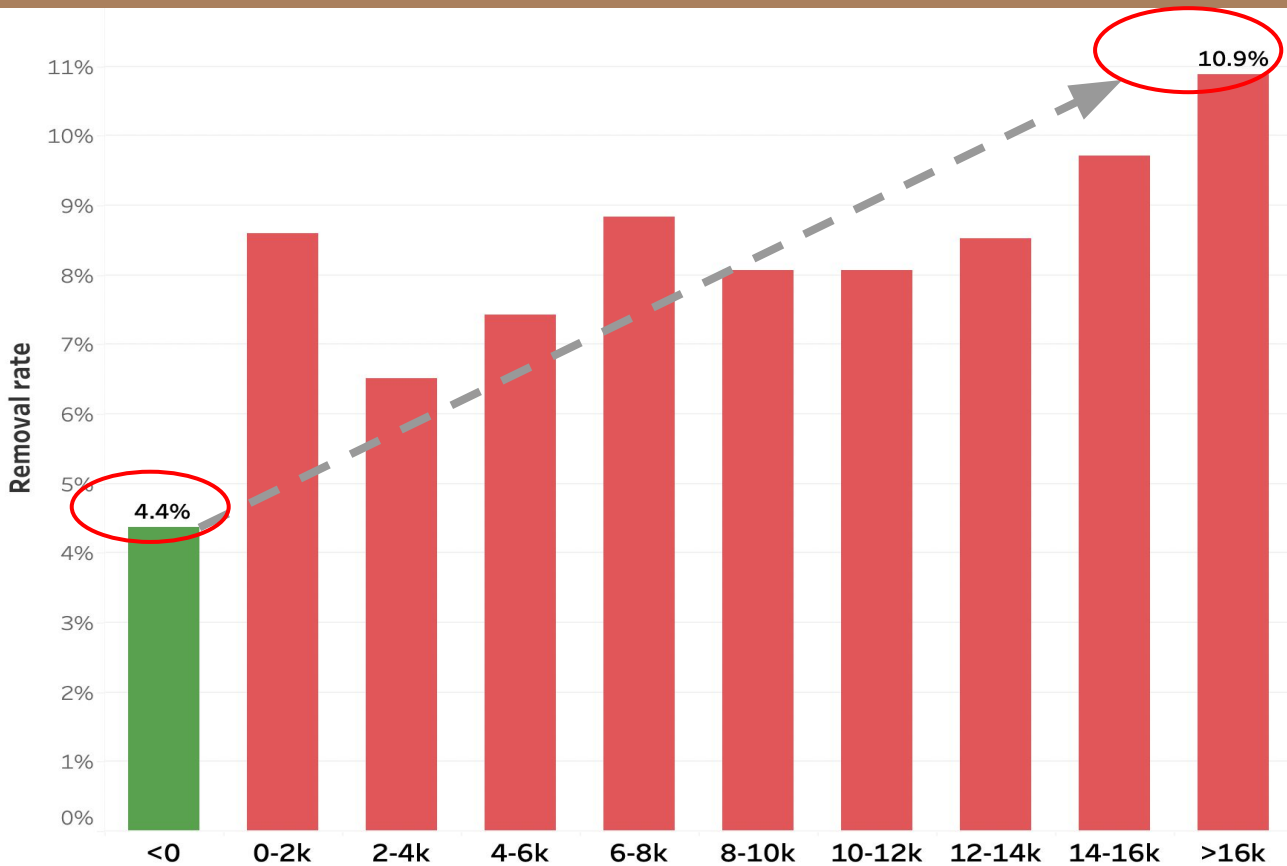
## Findings:

Businesses that claimed employees' payroll is more than their industry's average payroll have higher remove rate.

Applied amount per person below or above industry's average v.s. Removal rate



## Difference between self-proclaimed business payroll and industry's average payroll



## Applied Amount

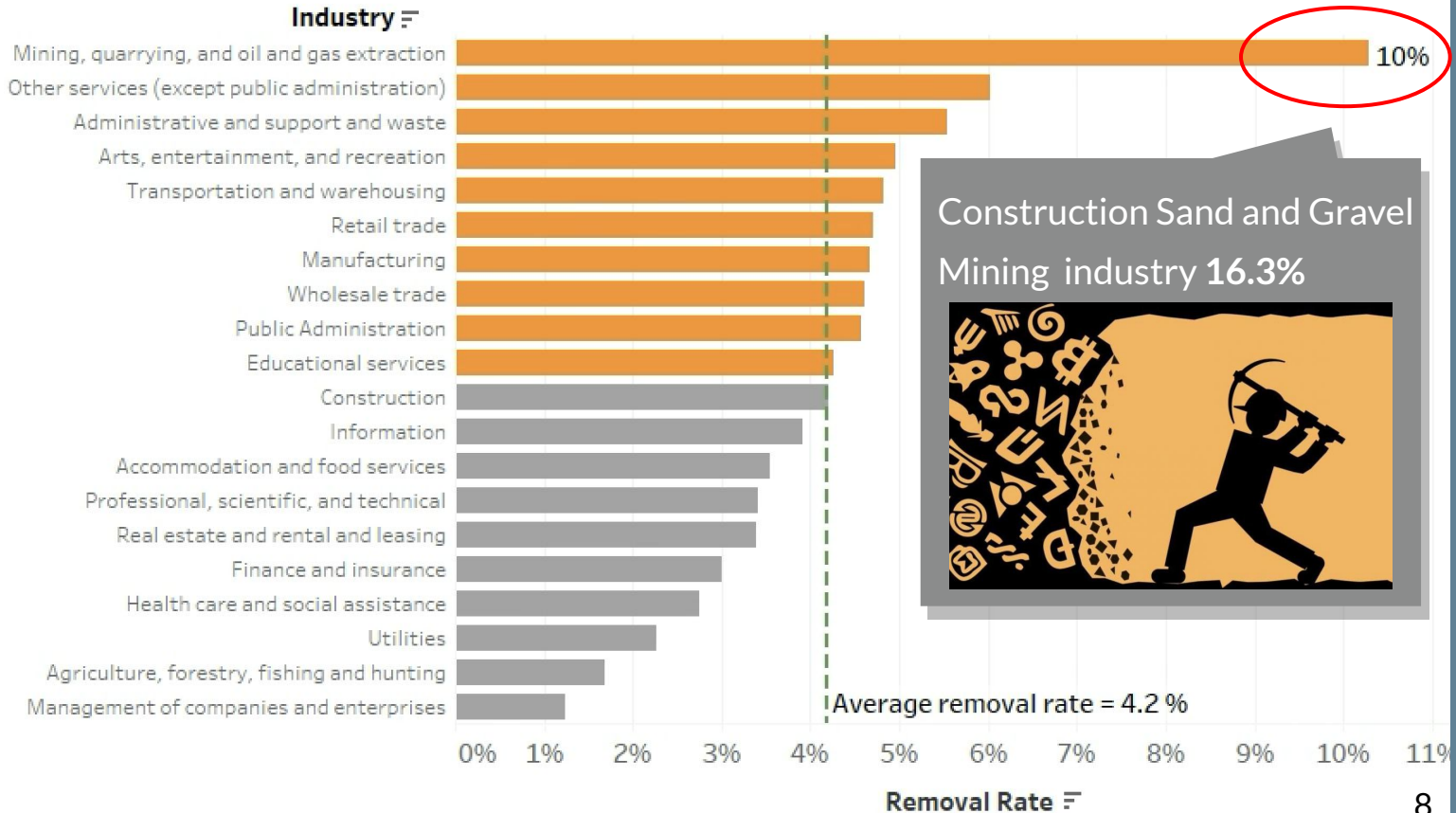
### Findings:

When the difference increases, **the removal rate increases.**

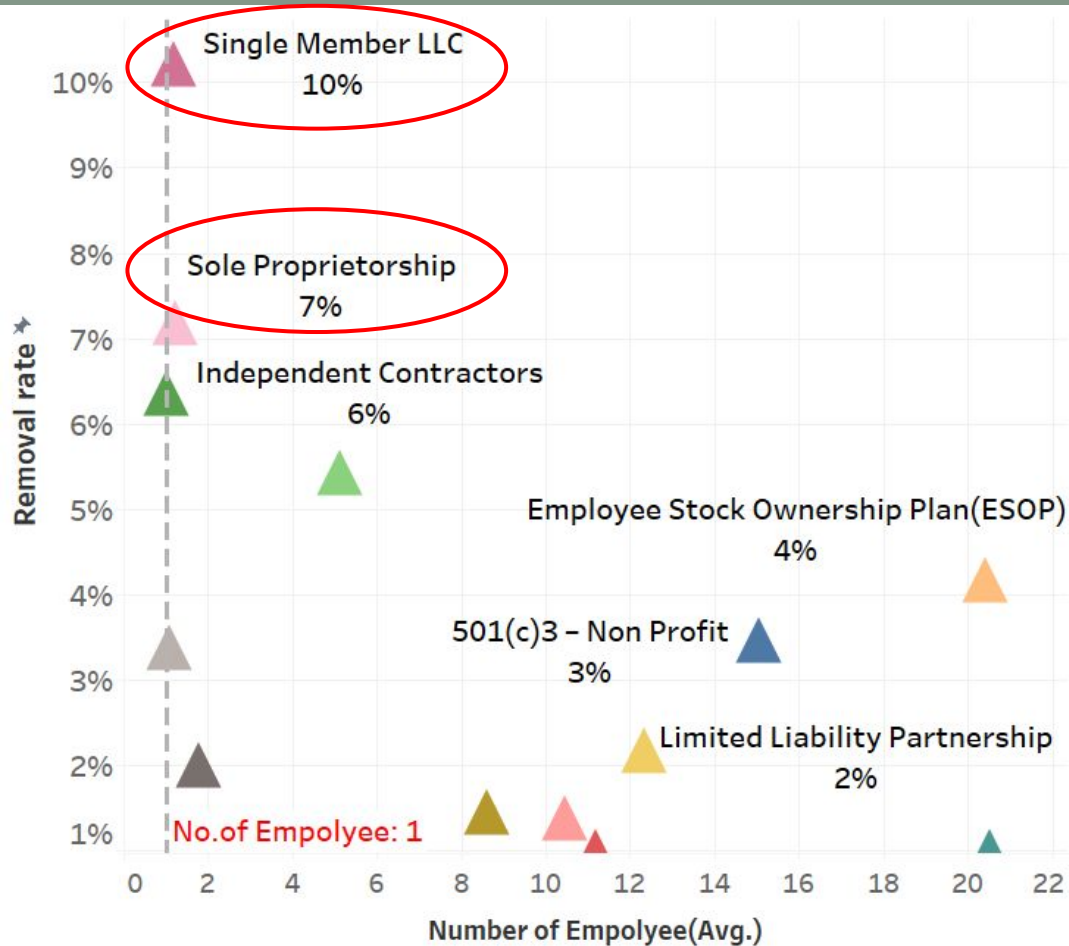


## Different industry v.s. Removal rate

Industry  
(NAICS Code)



## Business Type v.s. Removal Rate



## Business Type

- Single Member LLC and Sole Proprietorship
- Businesses that have **fewer employees** have **higher removal rate**.



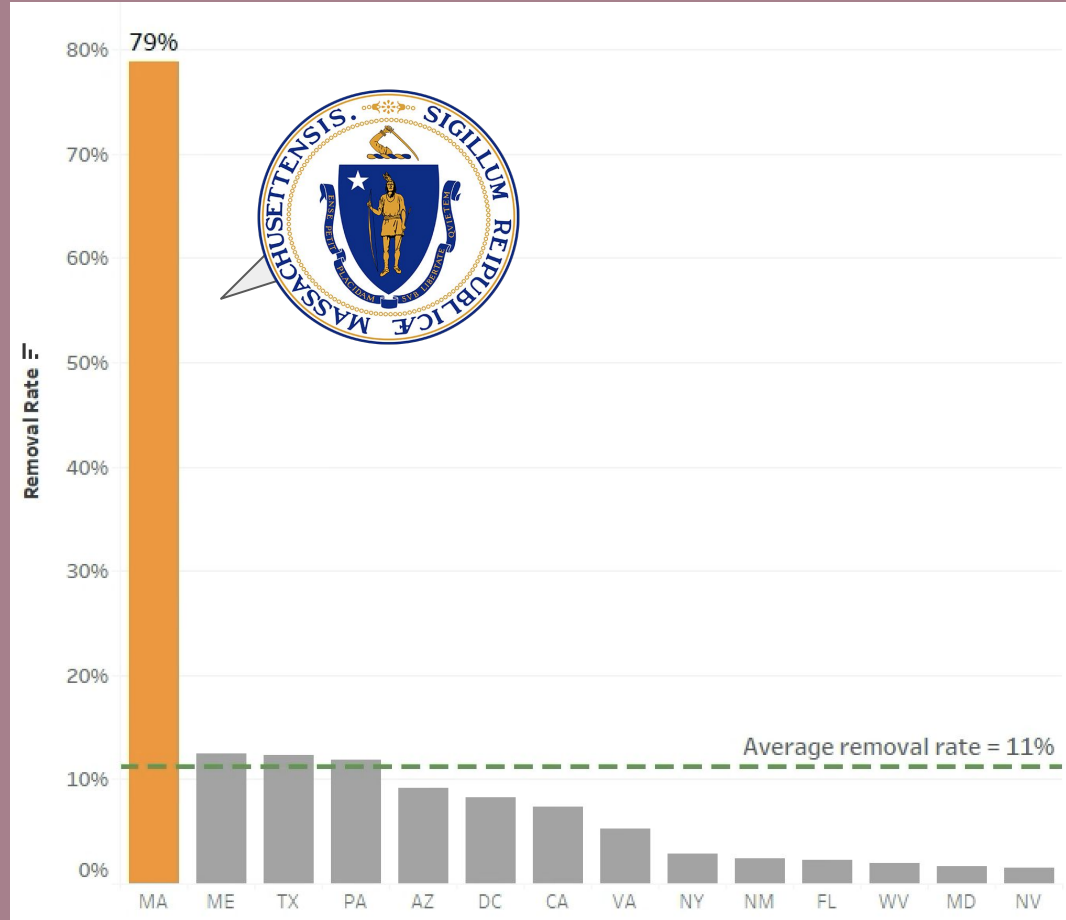
<https://prakhrajain2244.wordpress.com/2017/02/13/one-person-company-opc-understanding-this-unique-and-interesting-entity/>

# Location

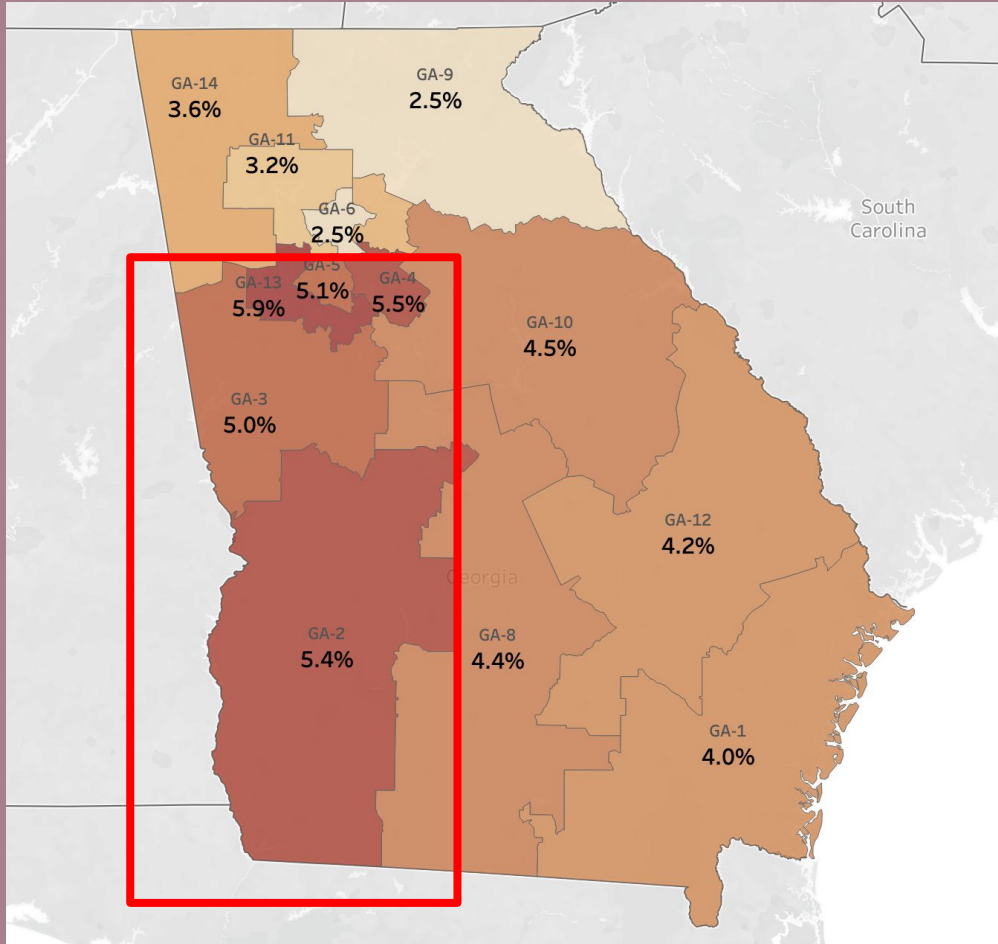
- Servicing lenders located in **Massachusetts** has the highest removal rate
- Lender name: **Leader Bank**
  - Applicants:  
**7 / 1930 Approved**



## Servicing lenders' removal rate in different states



## Removal rate v.s. District of Georgia



## Location

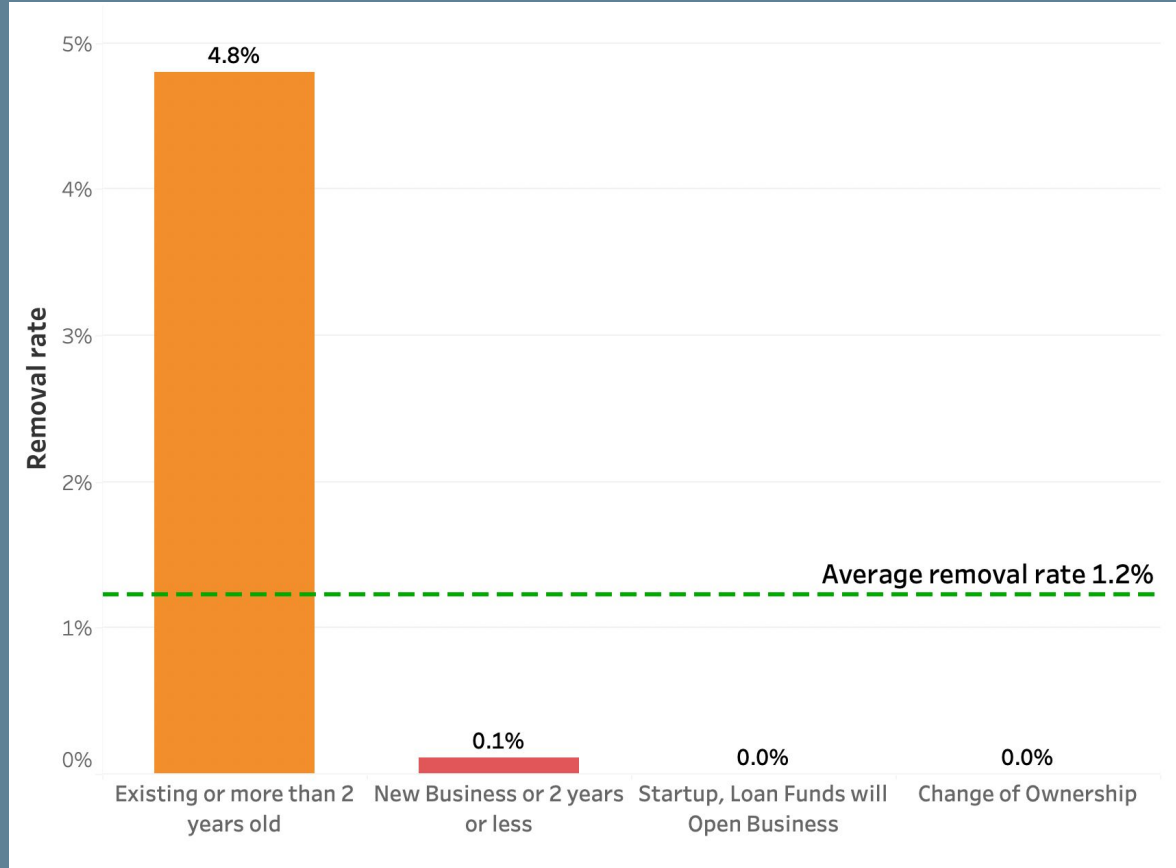
### Georgia's District Map

- **South-West** Districts (2, 3, 4, 5, 13) have relatively higher removal rates (5% - 6%), **twice** more than those of the **northern** districts.

# Business Age

- New businesses has **extremely low removal rate** than long-established businesses.

## Business age v.s. Removal Rate





# Predictive Model with R

Predict whether a new application would be removed from the program or not based on available data.

# Challenges

Too many **similar features**  
(*Multicollinearity*)



Removing features with  
very high correlation

Too many **unique values**  
**Skewed data**



Grouping, Transforming  
& Removing outliers

**Imbalanced classes**  
(550000 v.s. 25000)



Assigning different weights  
to each class

**False positive** &  
**False negative** trade off



Testing different combinations  
of cutoffs & weights

# Logistic Models Comparison



	Accuracy	Sensitivity	Specificity	Cutoff	Weight
Model 1	80%	67%	81%	0.4	(8:1)
Model 2	76%	75%	76%	0.4	(10:1)
Model 3	72%	<u>81%</u>	72%	0.4	(12:1)
Model 4	88%	48%	89%	0.5	(8:1)
Model 5	84%	58%	85%	0.5	(10:1)
Model 6	80%	67%	81%	0.5	(12:1)
Model 7	91%	34%	<u>94%</u>	0.6	(8:1)
Model 8	89%	42%	92%	0.6	(10:1)
Model 9	88%	48%	89%	0.6	(12:1)

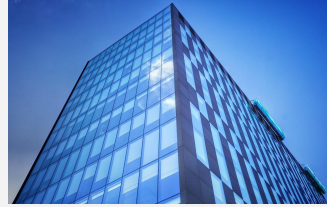
```

> ER2
[1] 0.19891651 0.24106843 0.27722348 0.12495510 0.16256446 0.19893389 0.08638971 0.10575931 0.12490874
> ER3
[1] 0.6668794 0.7512444 0.8116146 0.4784939 0.5843012 0.6658583 0.3361838 0.4165922 0.4763242
> ER4
[1] 0.8074656 0.7592971 0.7185518 0.8939031 0.8494735 0.8074960 0.9410701 0.9169555 0.8940548
    
```



# RECOMMENDATIONS

## Companies



- Have more than 1 employee and be founded less than 2 years.
- Immediately apply once the program started.
- Avoid lenders in Massachusetts, especially Leader Bank.
- Applied amount per employee should not exceed the industrial & regional average salary.

## Government



Larger amount → Use model with higher sensitivity

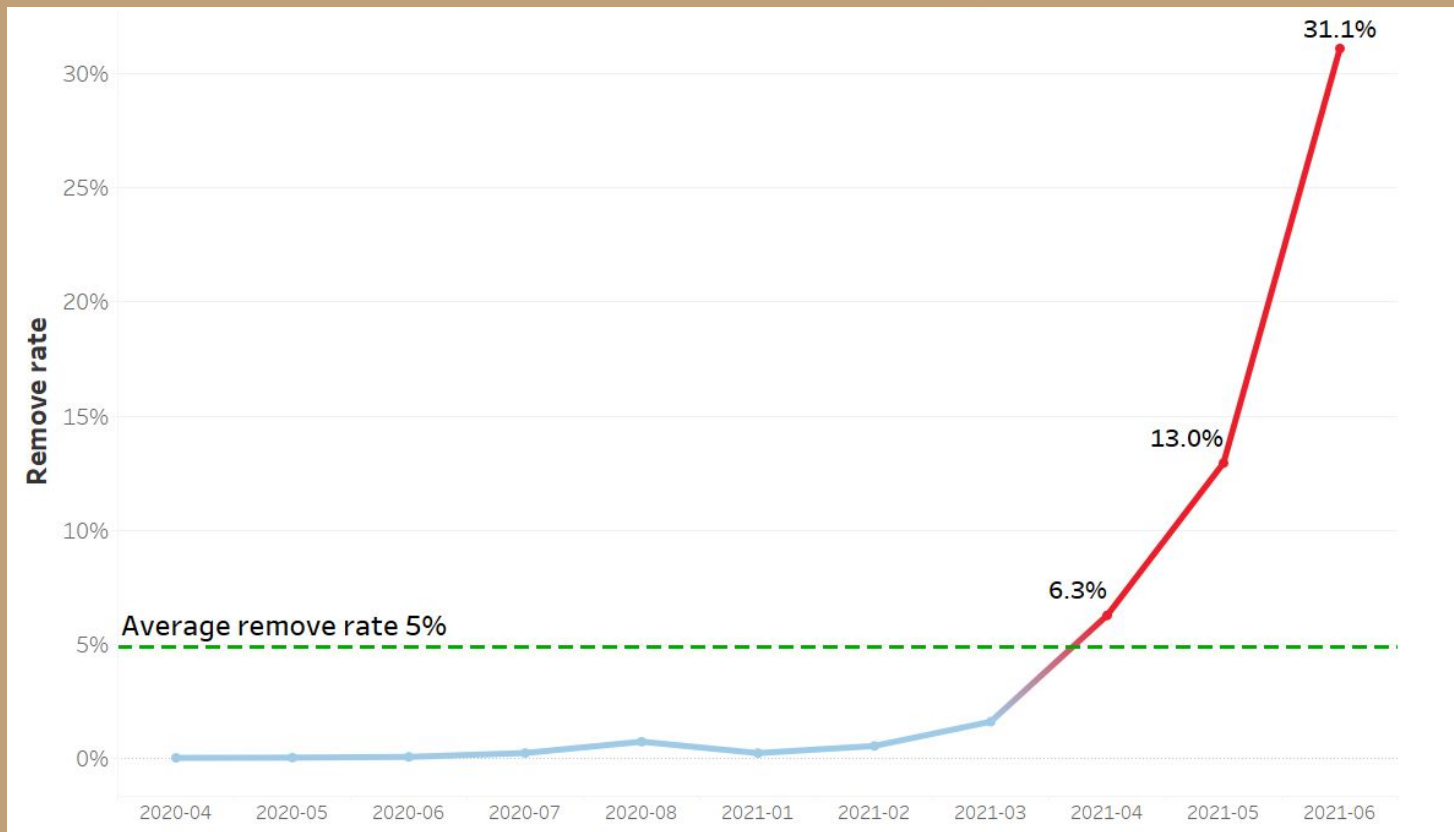
Smaller amount → Use model with higher specificity



**Thank you for listening !  
Any questions?**

Date  
Approved

## Approved Data Vs. Removal Rate



# Data Processing with Python

