University of Maryland College Park
Information Challenge 2022
Paycheck protection program: Analysis

Team – IC22-052
Team members:
Manikanta Koneru
Jayasree Karthik
Nandula

Contents:

- Introduction

- Exploratory data analysis

- Model build:
  - Decision tree
  - Random forest
  - Feature selection
- References
- Conclusion

Introduction:

U.S. Small Business Administration:
A United States government agency that provides support to entrepreneurs and small businesses.

Paycheck protection program:

An SBA-backed loan that helps businesses keep their workforce employed during the COVID-19 crisis.

Periodical release of data by SBA on 11.5 million approved applications, of which some applications were later on removed from the database.

Goal: through data analysis, develop an understanding of why these loans might have been removed.

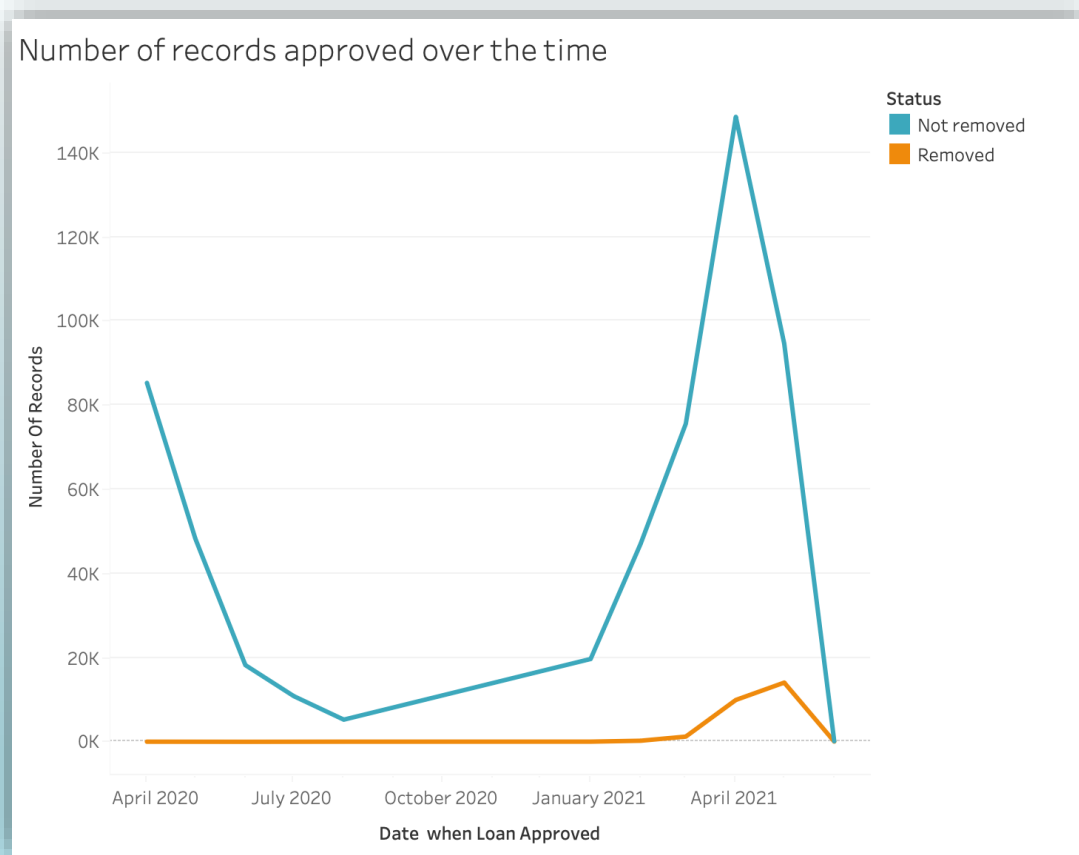Exploratory data analysis   Model build & validation   Further analysis   Scope of enhancement

Exploratory data analysis:

Plot: Loan approval date vs no. of applications (Time series graph)

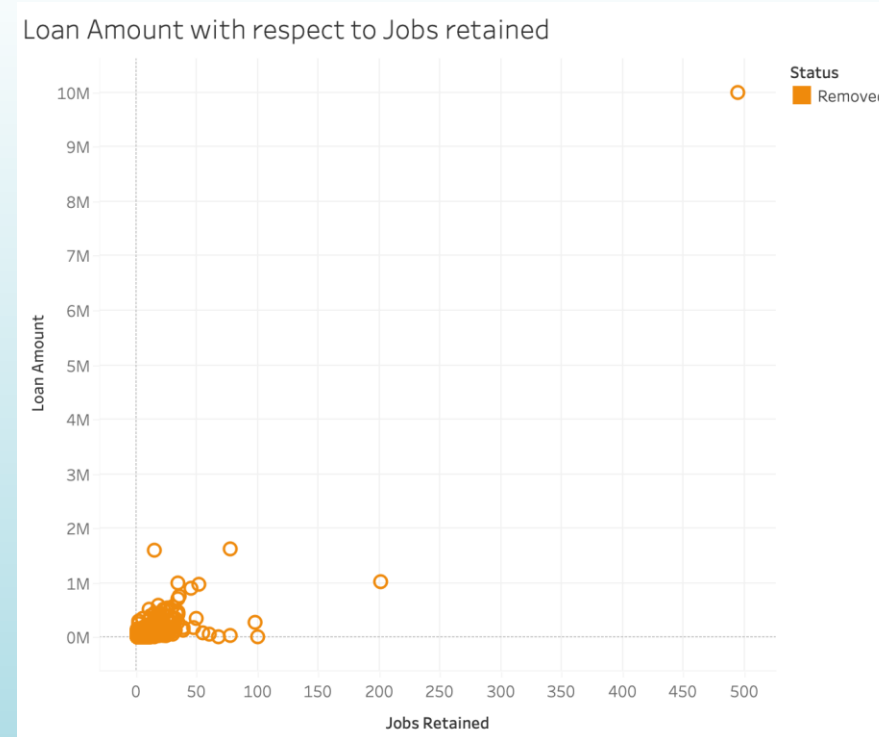Number of records approved over the time

The Consolidated Appropriations Act, 2021 (CAA) extended the Paycheck Protection Program (PPP) through **March 31, 2021**. However, the PPP funds have been exhausted, meaning the PPP program is no longer available as of May 31, 2021.

The amount of funds made available in the third round totaled $284 billion. Maximum loans of $10 million were available to first-draw borrowers, and loans up to $2 million were offered to second-draw, small business owners.

Reference: https://www.investopedia.com/your-guide-to-the-paycheck-protection-program-ppp-and-how-to-apply-4802195#:~:text=Round%20three%20funding%20of%20%24284,rounds%20of%20PPP%20loan%20funding.

Exploratory data analysis:

Plots: Jobs retained vs loan amount requested – Full data & removed data

- The Georgia loans of $14.4 billion as of June 27 — represent about 55% of the 278,000 active small businesses in the state, according to the National Federation of Independent Business.
- About 18,300 companies in the state received loans of between $150,000 and $10 million, and another 139,000 state-based businesses obtained loans of less than $150,000, according to the SBA.
- The loans helped companies retain more than 800,000 jobs.

Number of records for each county in Georgia
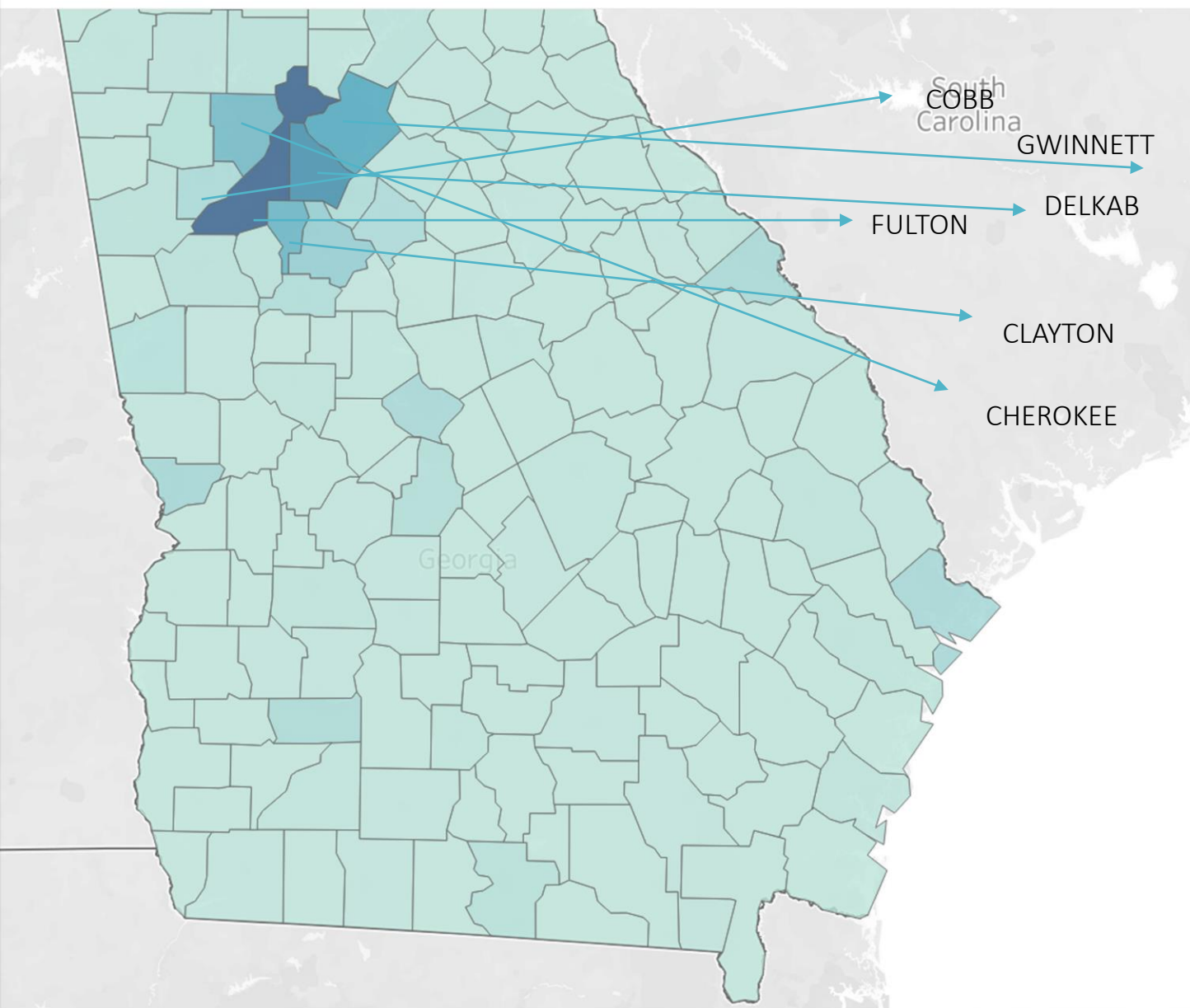
Georgia: Counties with most applications Historically Underutilized Business zones in pink

- Counties in Georgia with highest number of applications: Fulton, Dekalb, Gwinnett, Clayton, Cherokee, Cobb
- Counties characterized by many small businesses
- Highest no. of PPP applications from these counties as affected by COVID19
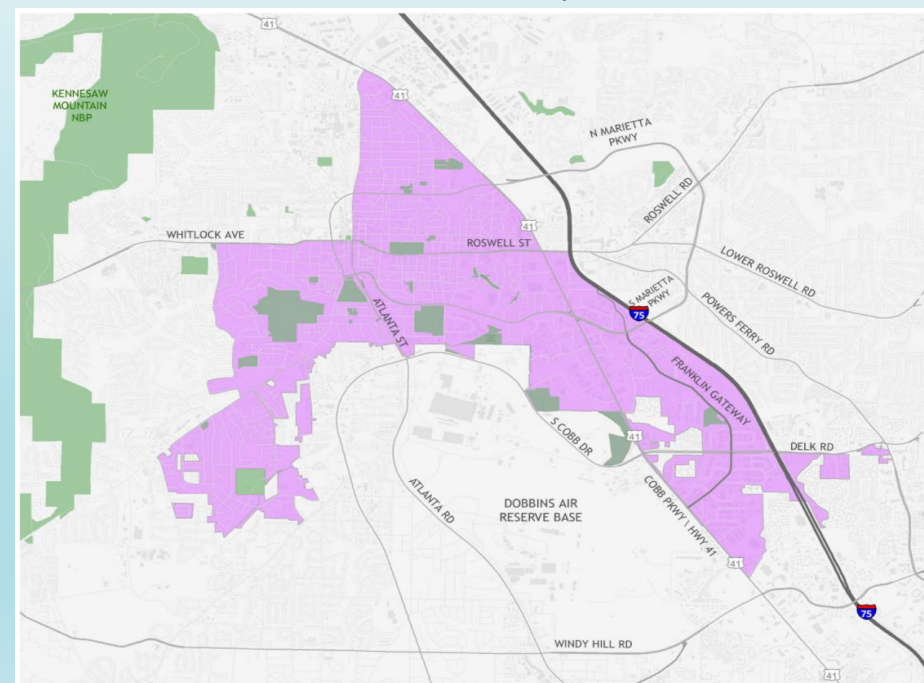
Applications breakdown w.r.t Loan status:

Status	Loan Status		
	Active Un-Disbursed	Exemption 4	Paid in Full
Not removed	344	273,238	280,243
Removed	23,331	2,505	

number of removed records for each industry type
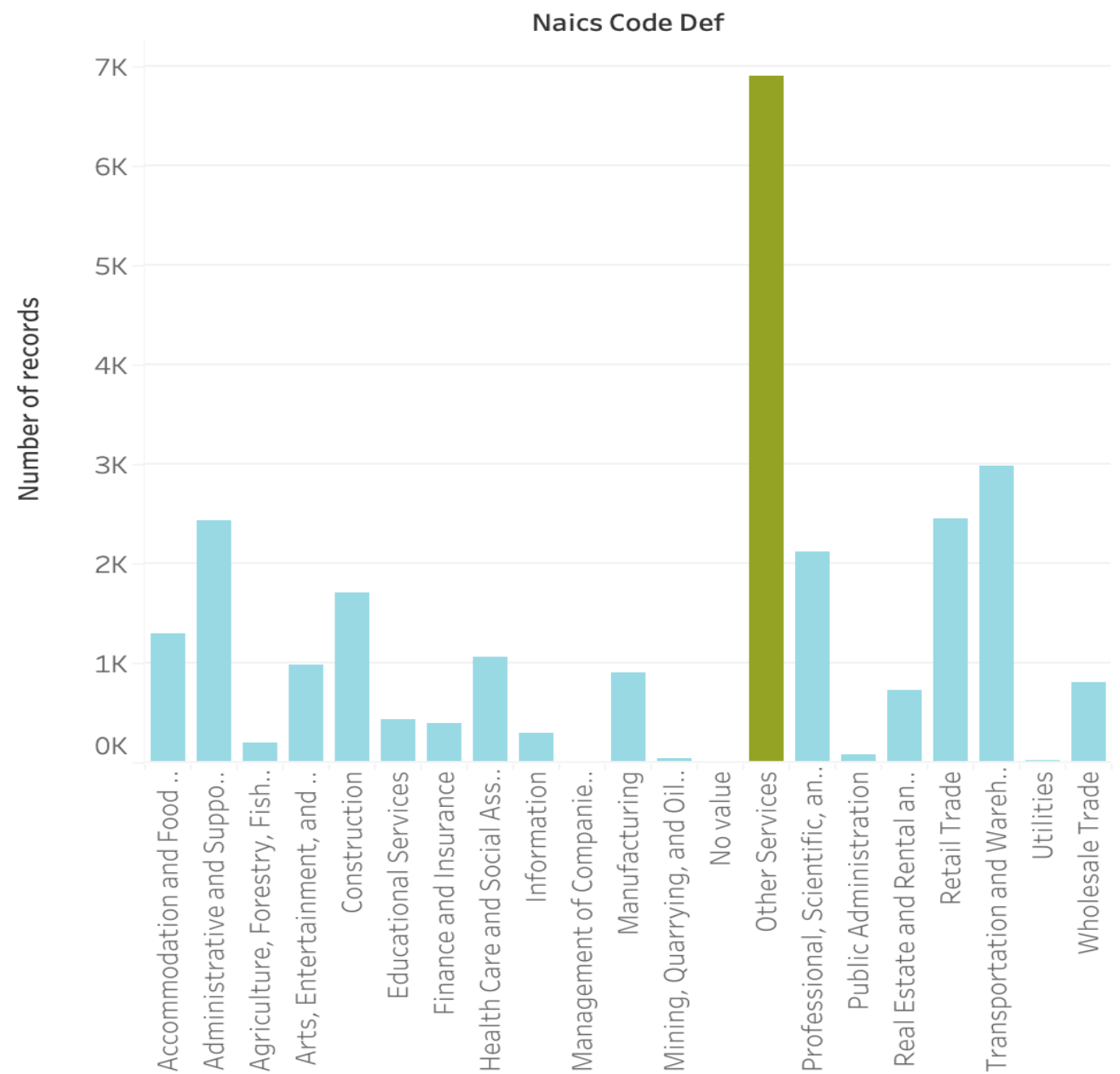
Analysis on industry type vs number of applications
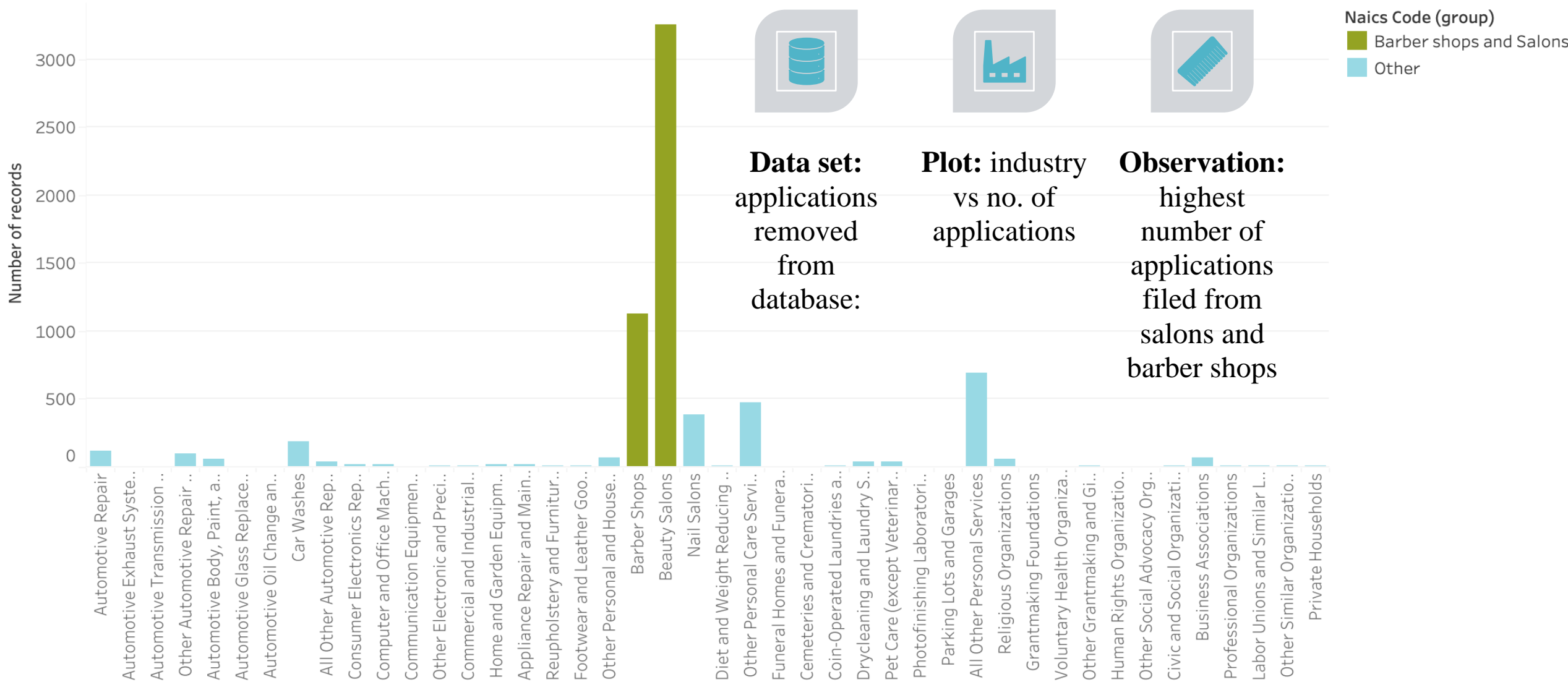
Data set: Applications removed from database:

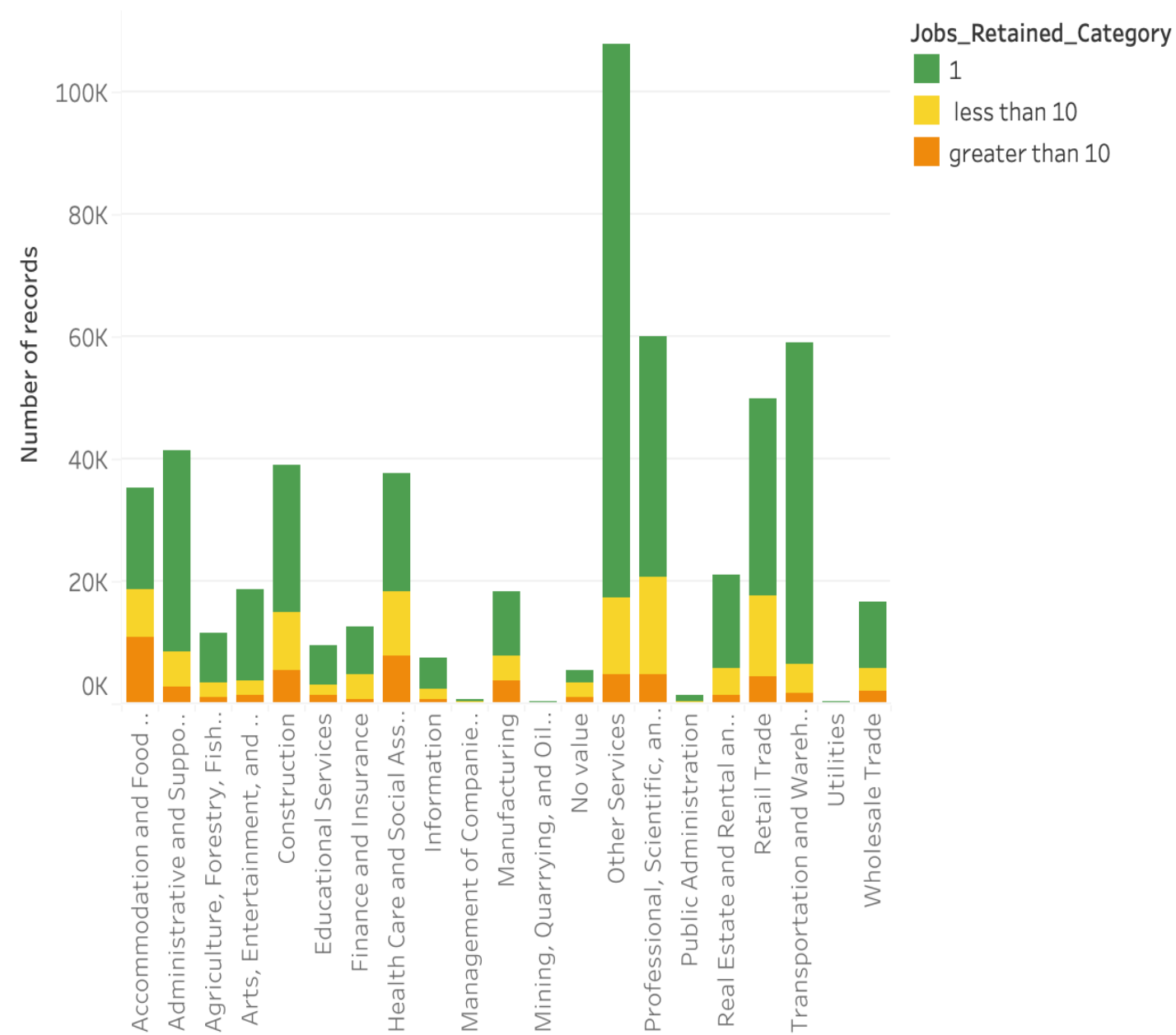**Plot:** Industry vs no. of applications

Observation: Highest number of applications filed from the industry 'Other services'

Analysis on industry type (continued):

Removed records for all businessess under Other Services

Data set: applications removed from database:

Plot: industry vs no. of applications

Observation: highest number of applications filed from salons and barber shops

Naics Code (group)
- Barber shops and Salons
- Other

jobs retained category proportion for each industry type for records not removed
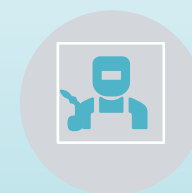
Analysis on industry and jobs retained:
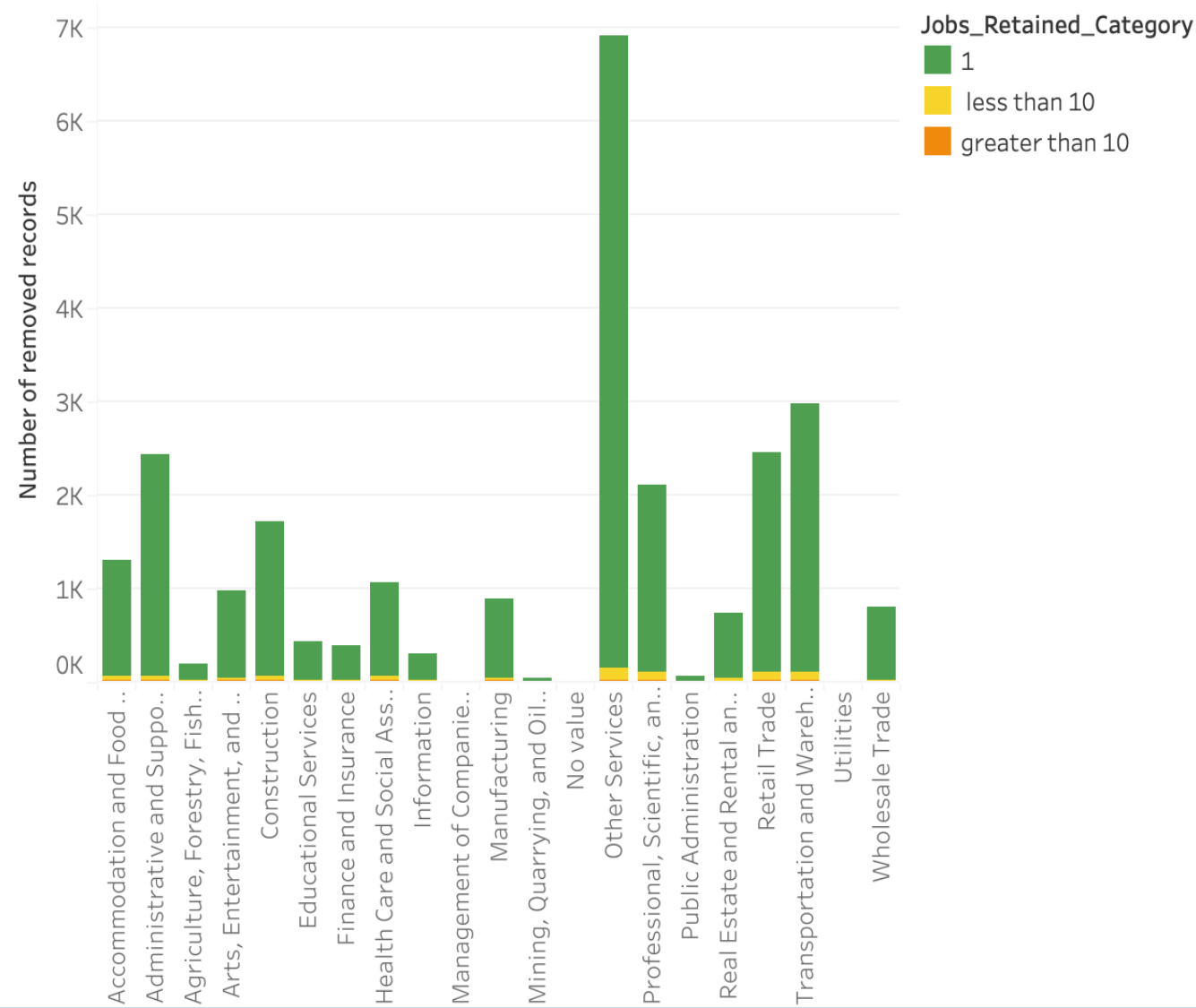
Data set: full data set

Plot: industry vs no. of applications, category as no. of jobs retained

Observation: no. of jobs retained are 1 majorly from the applications from 'other services' industry. other industry applications are also considerably high.

jobs retained category proportion for each industry type for removed records

Analysis on industry and jobs retained:

Dataset: Removed applications from database

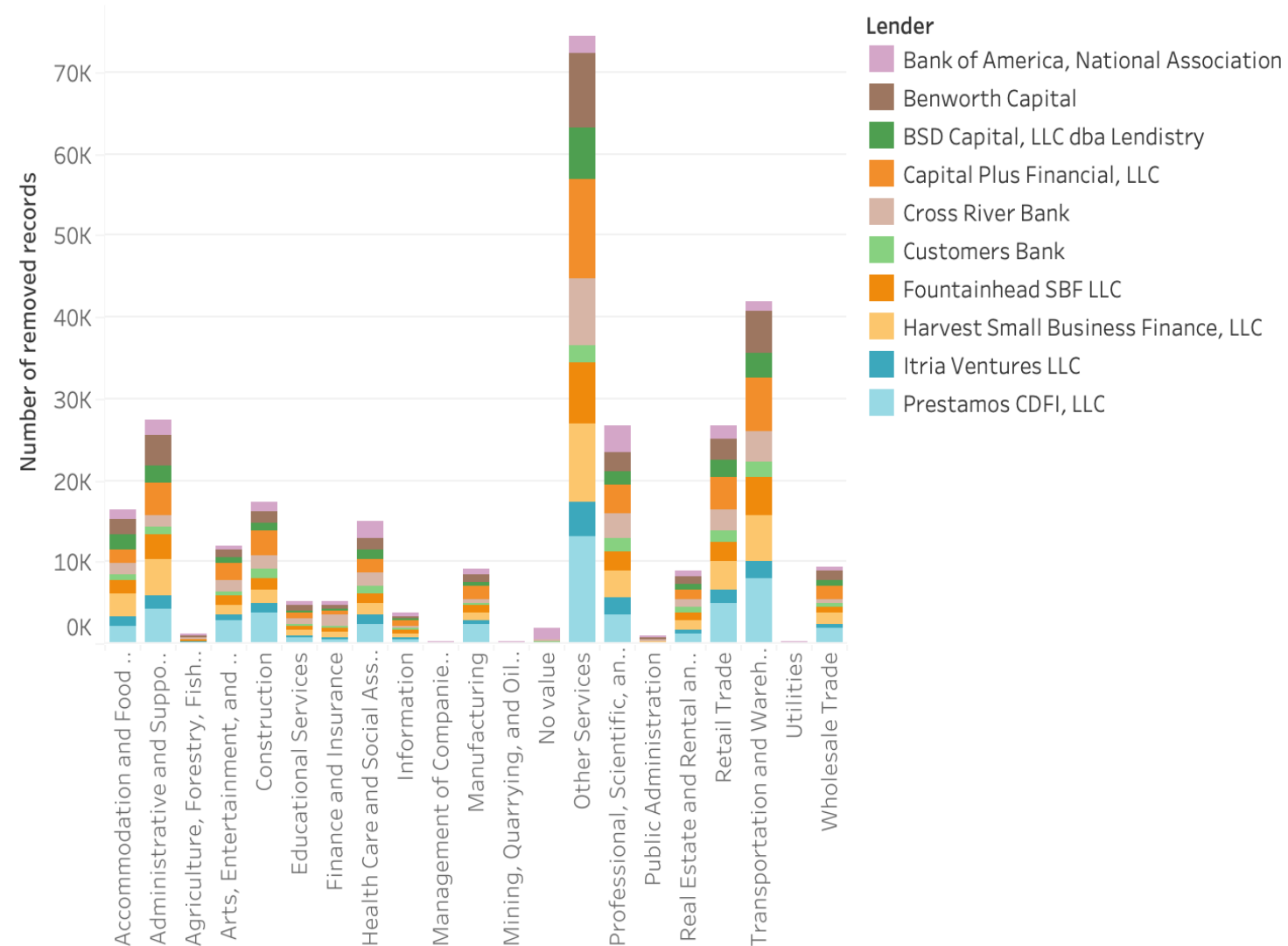**Plot:** Industry vs total number of applications with category as no. of jobs retained.

Observation: Across all of the industries, most of the number of jobs retained =1.

One job is saved for each of the applications in 'Other services' industry

Analysis on Industry vs lender:

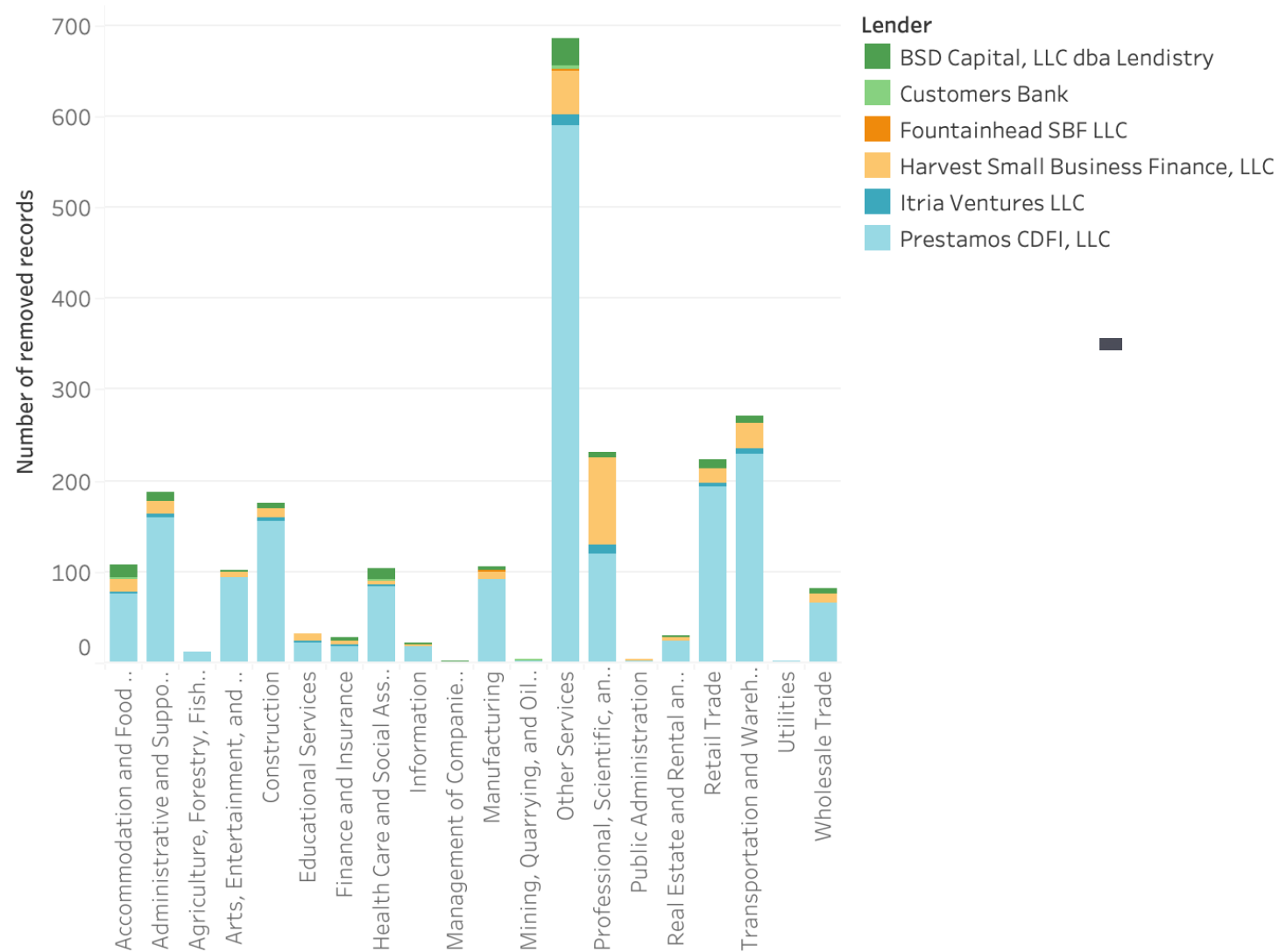Top 10 lenders for each of the industry for records not removed

- **Dataset:** Full dataset
- **Plot:** Industry vs no. of applications, categorized by lender
- **Filter:** Loan status: Exempt 4 (disbursed but not paid in full or charged off) and paid in full
- **Observation:** Top 10 lenders are as listed.
- **Major lenders:**
    Benworth Captial
    Capital Plus Financial LLC
    Prestamos CDFI LLC

Analysis on Industry vs lender (continued):

Top 10 lenders for each of the industry for Removed records

Lender
- BSD Capital, LLC dba Lendistry
- Customers Bank
- Fountainhead SBF LLC
- Harvest Small Business Finance, LLC
- Itria Ventures LLC
- Prestamos CDFI, LLC

Dataset: Applications removed from database

Plot: Industry vs no. of applications categorized by lenders

Filter: Loan status as Exempt 4 (disbursed but not paid in full or charged off)

Major lender: Prestamos CDFI LLC

*in Newswire*   *October 5, 2021*

Class Action Alleges Prestamos CDFI Never Gave Borrowers Approved Paycheck Protection Program Funds

by **Corrado Rizzi**

Machine learning: Decision tree

- Predictive model to traverse from observations (branches) to conclusions about the item's target value(leaf node).

- White-box model, works well on large datasets, little data preparation, in-built feature selection

- Accuracy & error rate computed through confusion matrix

- Gini impurity, information gain, variance reduction, measure of goodness

Test data	Predicted		
Actual	Category	Not removed	Removed
	Not removed	165059	341
	Removed	755	6994

Receiver Operating Characteristic

AUC = 0.95
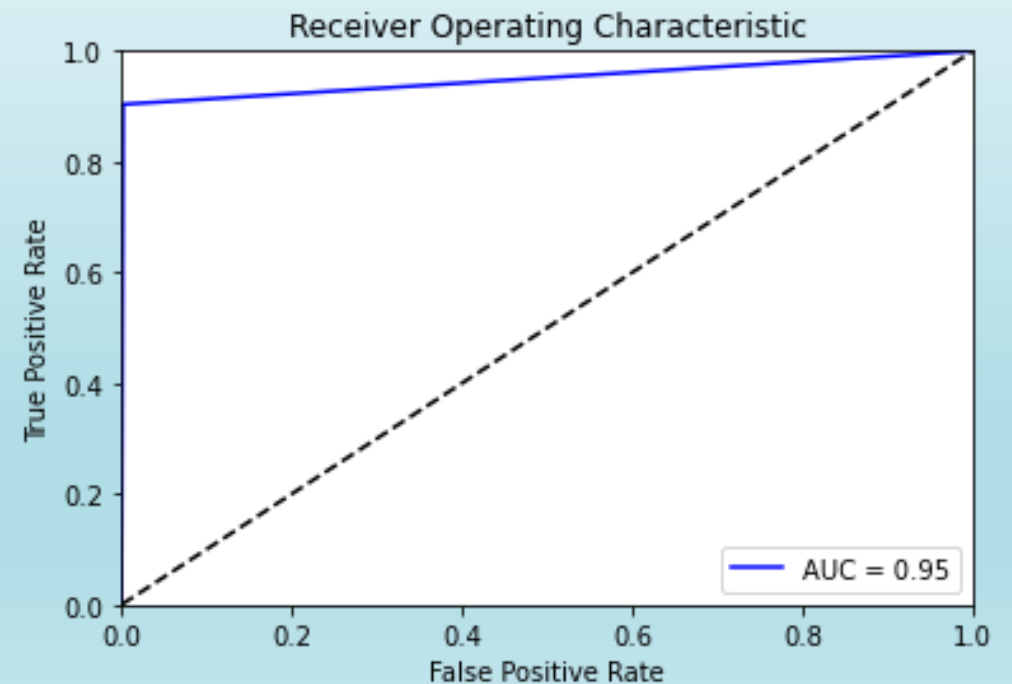
Machine learning: Random forest

- Ensemble learning method – builds multitude of decision trees at training time.

- Output for classification -> class selected by most trees.

- Bagging(reduces overfitting), ranking of variable importance, reduces variance, improves accuracy

Test data	Predicted		
	Category	Not removed	Removed
Actual			
	Not removed	165154	246
	Removed	712	7037

Receiver Operating Characteristic — AUC = 0.95
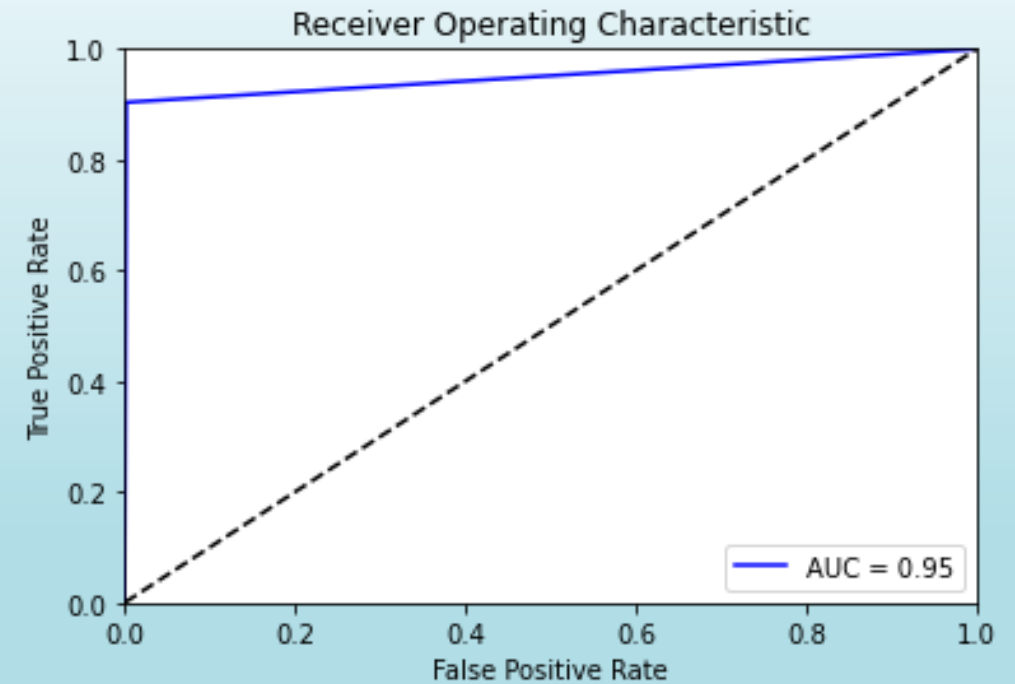
Feature importance:

Bias observed in the data due to the fields 'Loan Status', 'undisbursed amount':

		Loan Status	
Status	Active Un-Disbursed	Exemption 4	Paid in Full
Not removed	344	273,238	280,243
Removed	23,331	2,505	

feature
loan_status_Active Un-Disbursed
undisbursed_amount
loan_status_Exemption 4
loan_status_Paid in Full
amount
current_approval_amount
Other_Bank
Is_Prestamos_CDFI_LLC
Is_Capital_Plus_Financial_LLC
Is_Customers_Bank

Due to possible bias, explored model excluding top 3 features. Resulting model is just as good as random guessing.

Receiver Operating Characteristic — AUC = 0.53
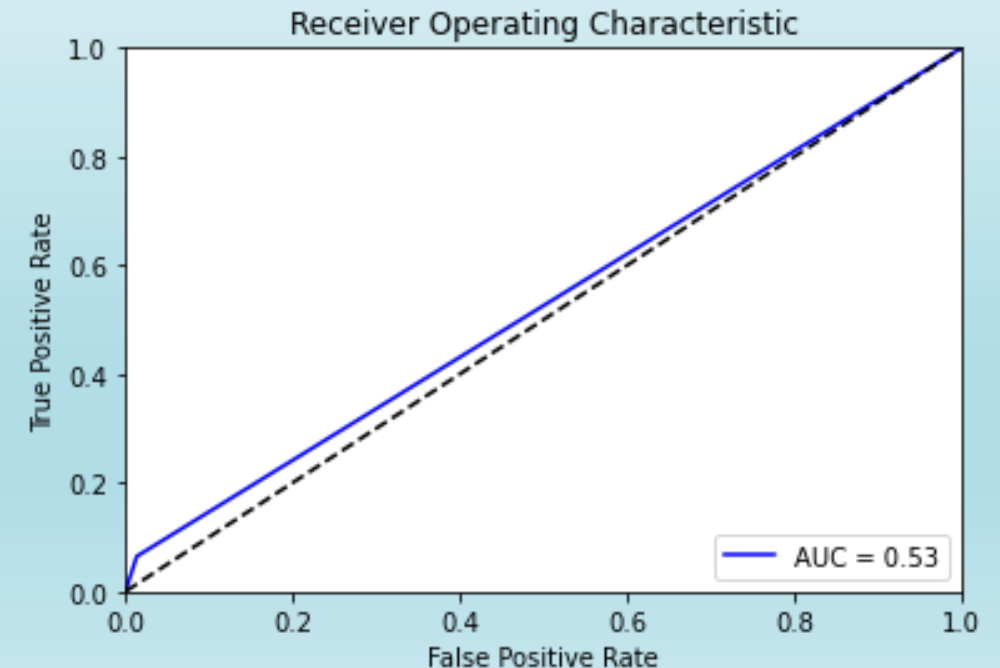
Literature survey - evidence

References:

https://data.onlineathens.com/paycheck-protection-program-loans/

https://www.cnn.com/projects/ppp-business-loans/states/ga

https://opendata.atlantaregional.com/datasets/ppp-loans-georgia-1/explore

https://www.sba.gov/sites/default/files/2021-03/BorrowerApplication2483ARPrevisions%20%28final%203-18-21%29-508.pdf

https://www.sba.gov/sites/default/files/2020-07/PPP%20Results%20-%20Sunday%20FINAL.pdf

https://www.womply.com/blog/are-ppp-loans-public-information-what-data-is-available/

https://www.bizjournals.com/atlanta/news/feature/on-leadership/2021

https://www.forbes.com/advisor/business-loans/how-many-jobs-were-saved-because-of-ppp-loans/

https://www.justice.gov/archives/oip/foia-guide-2004-edition-exemption-4

https://projects.propublica.org/coronavirus/bailouts/faq

https://www.forbes.com/sites/brockblake/2021/04/06/why-ppp-loan-applications-are-getting-stuck-and-how-to-keep-yours-moving/?sh=2e86468e4872

https://data.onlineathens.com/paycheck-protection-program-loans/

Conclusion:

Considering additional important demographic information from the PPP application form could help us uncover better insights.

Ex: gender, ethnicity etc.