# U.S. Small Business Administration (SBA) Loan Default Analysis & Prediction

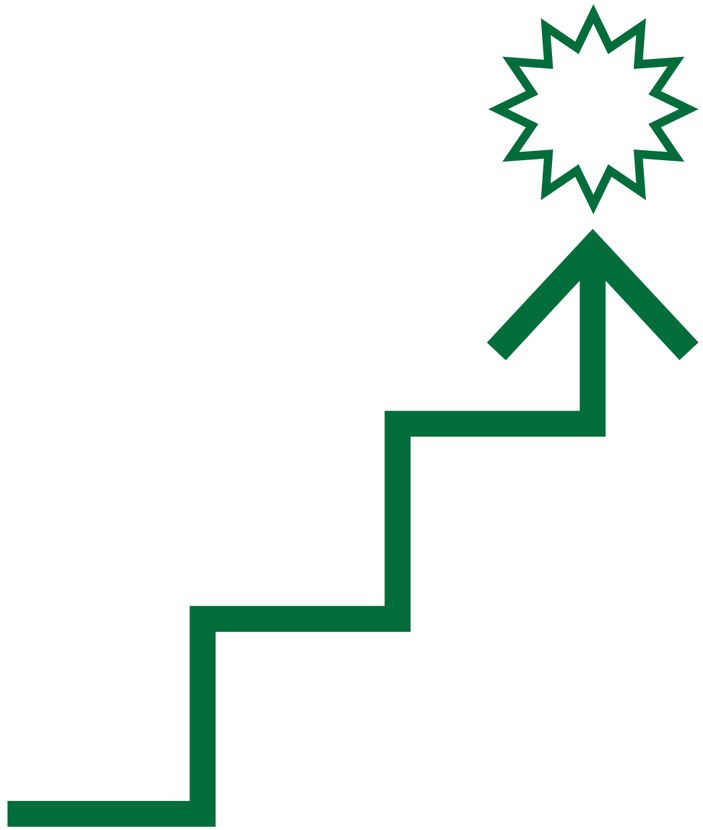
Julius Hermanto


Portfolio Project

# Business Problem

With the help of SBA, small businesses across USA are able to get a guarantee part of the loan the apply to, mitigating risk from the business owner(s) side and handing over to the SBA.

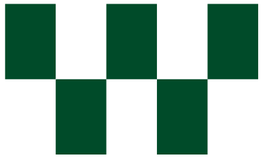
However, the default rate of 899,164 applied business loans is still high at **21.5%.**



# Objectives

## Risk Reduction

Identify and minimize the approval of high-risk loans to reduce the incidence of defaults and associated financial losses via model prediction.

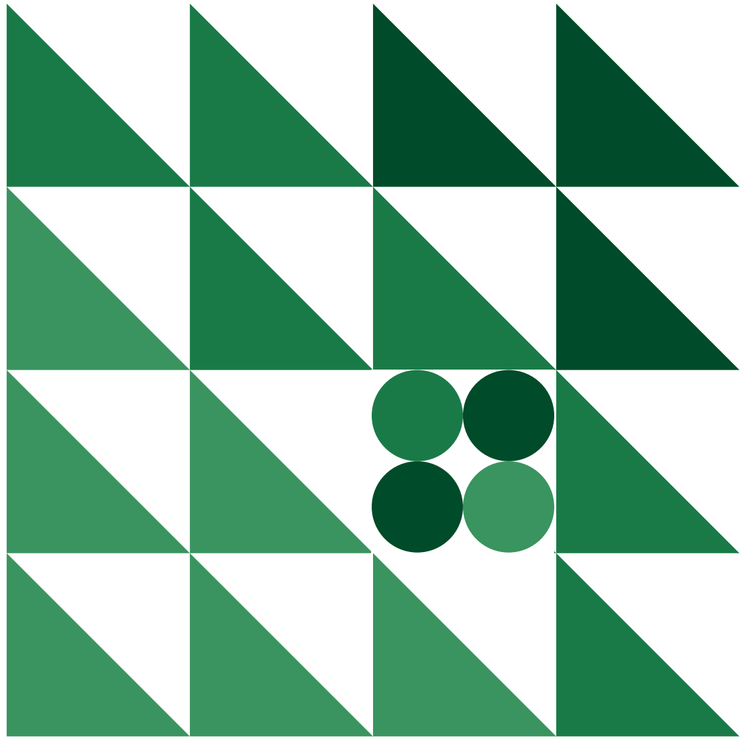


## Business Recommendations

Identify factors that highly correlates to the loan being defaulted and give potential business recommendations.

Section 01:

# Data Preprocessing



# Data Preprocessing (Highlights)

Missing & Duplicate Values:

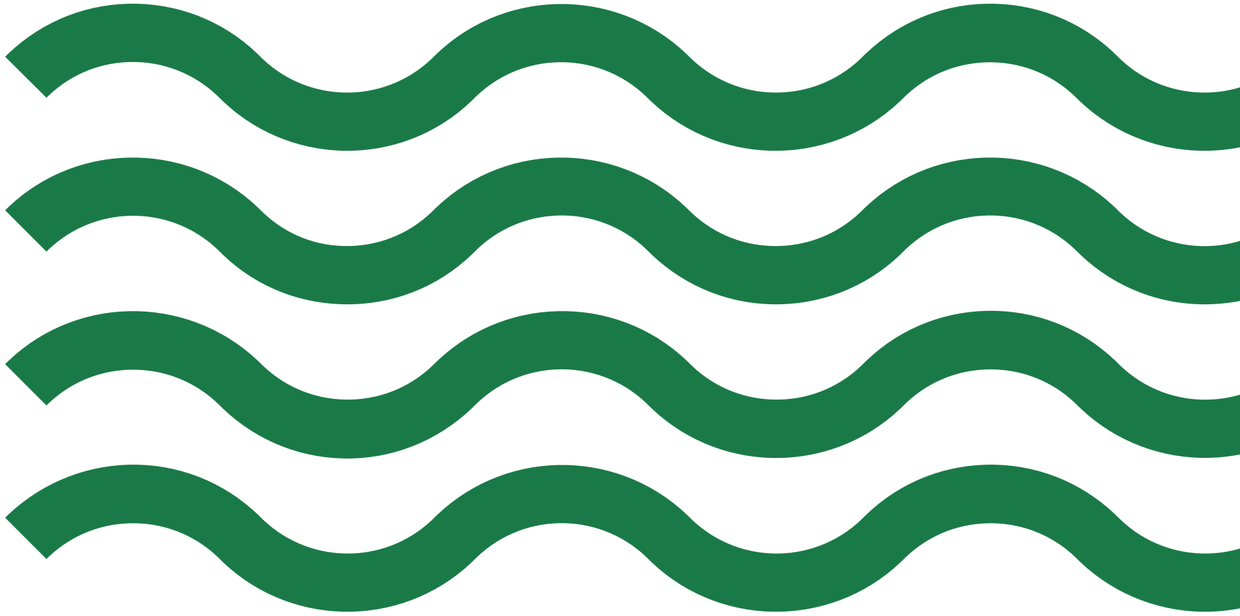
- Reduction of Data from **899,164** to **456,940**
- Most of the missing data comes from the **RevLineCr** and **LowDoc** fields.

Feature Engineering:

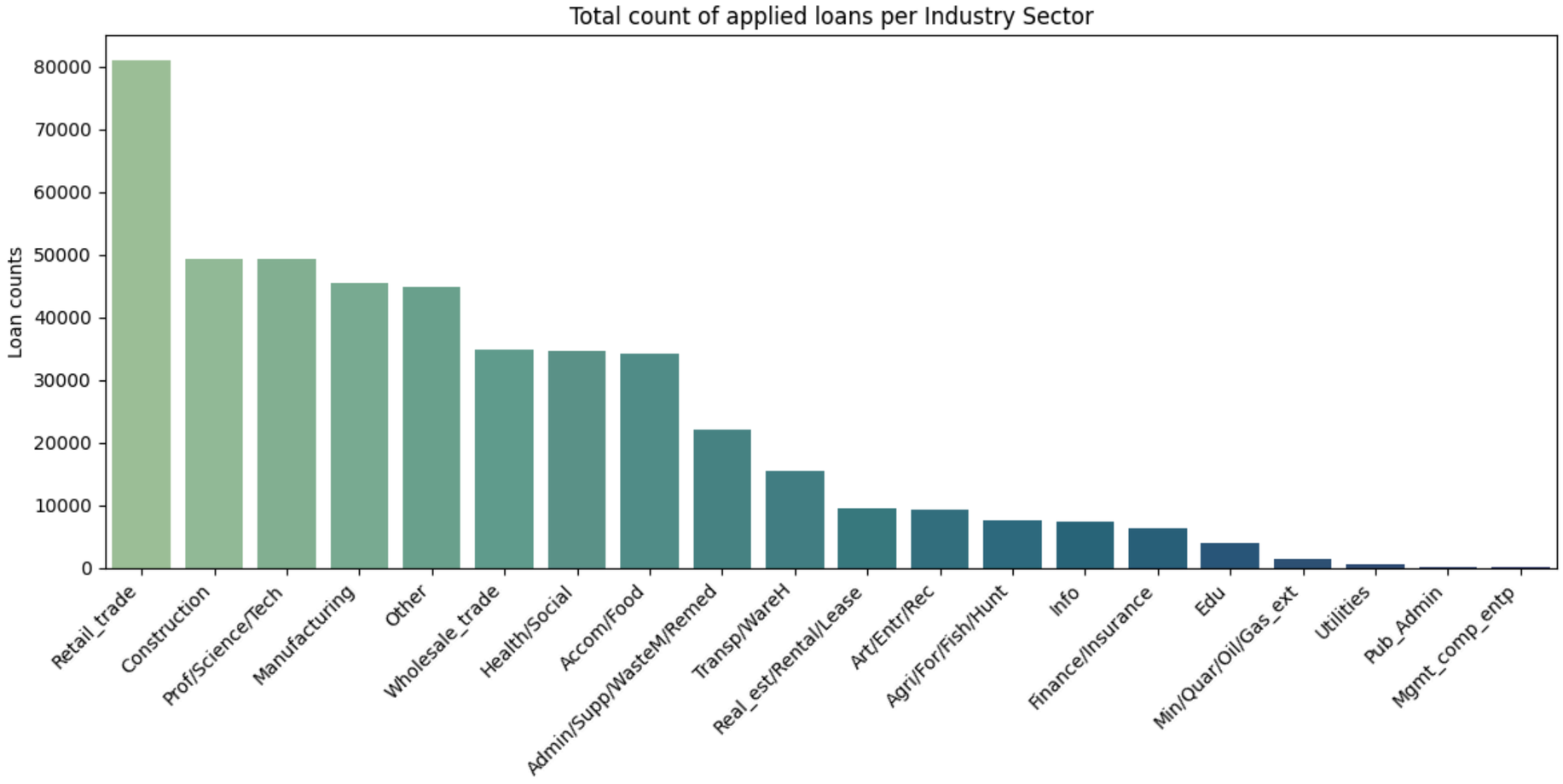
- **DaysToDisbursement**, how long did it take from the approval date to the disbursement date.
- **SBA_AppvPercent**, how much percentage-wise did the SBA approved.
- **Industry** / **IndustryCode**, converted from NAICS code.
- **StateSame**, if the applicant applied for their loan from the same state.
- **CompanySize**, categorize the number of employees.
- **Default**, converted from MIS_Status.

Section 02:

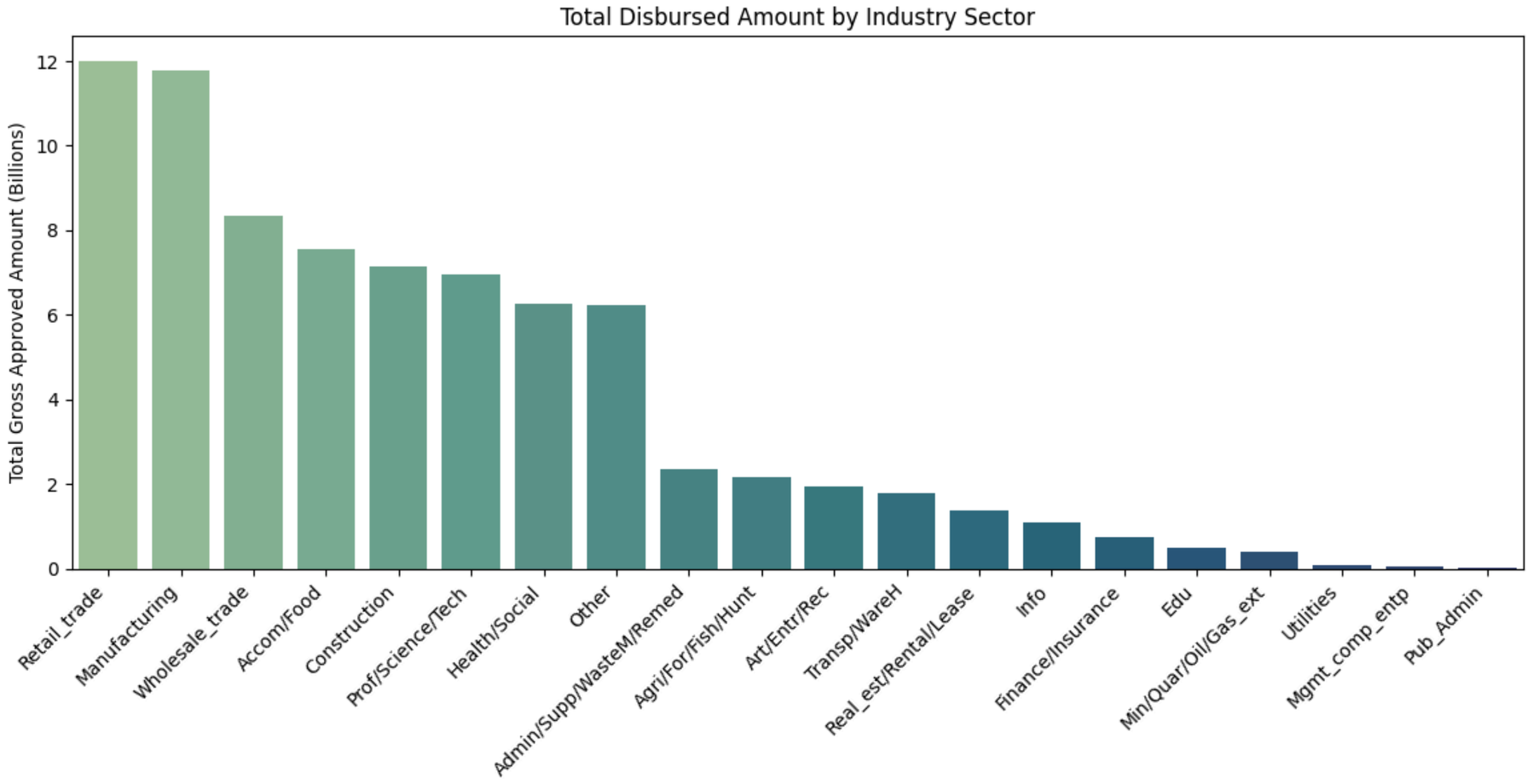
# Exploratory Data Analysis



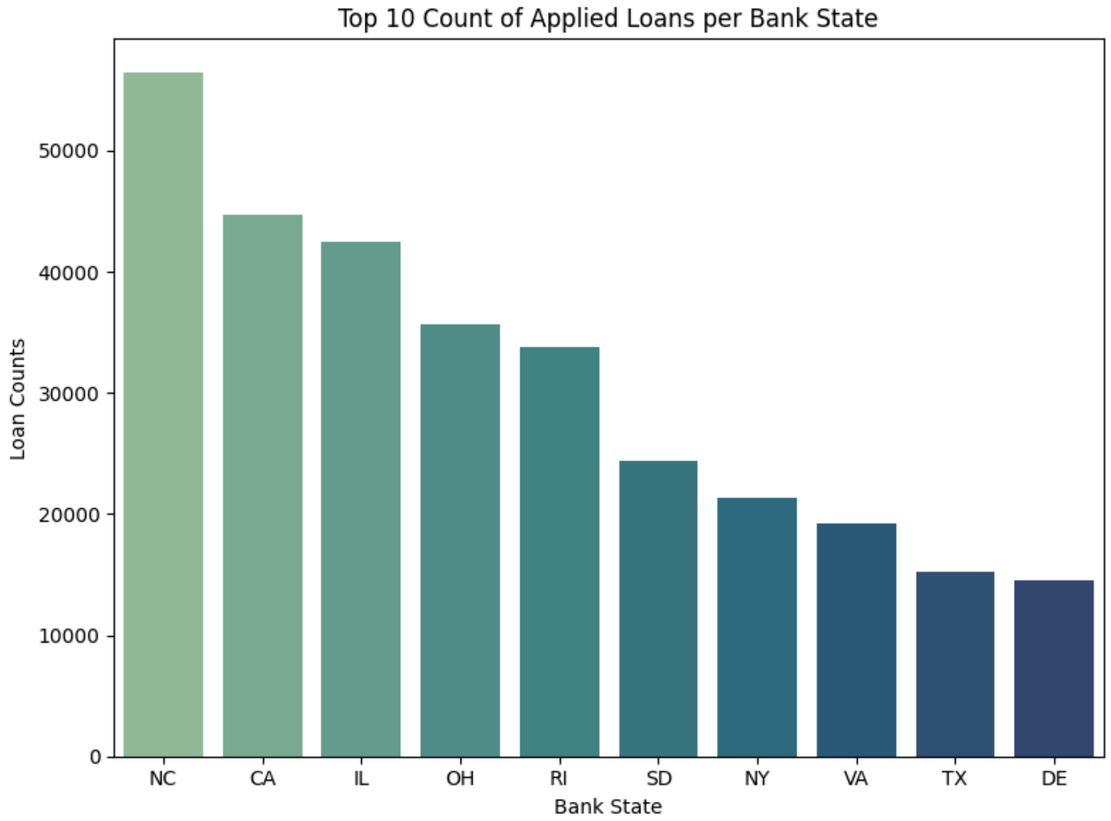
Top 3 industry sectors that applied for loans are from **Retail Trade, Construction,** and **Professional/Scientific/Tech**.

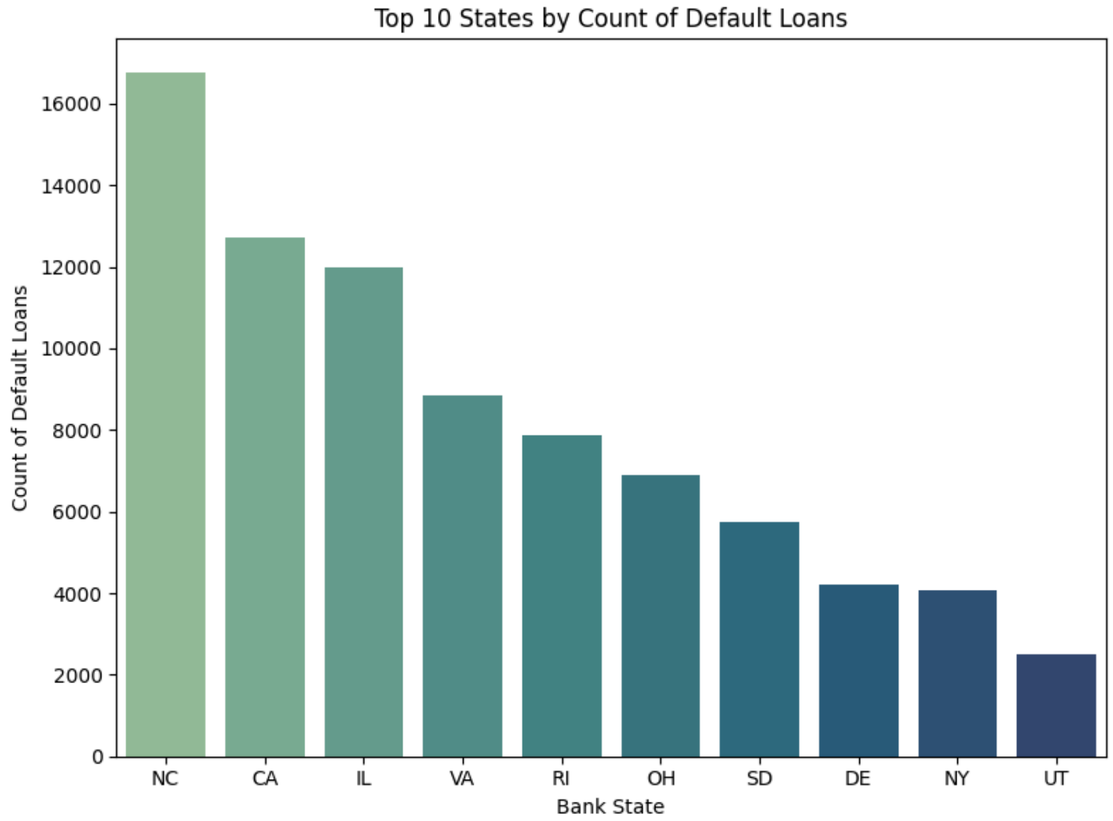


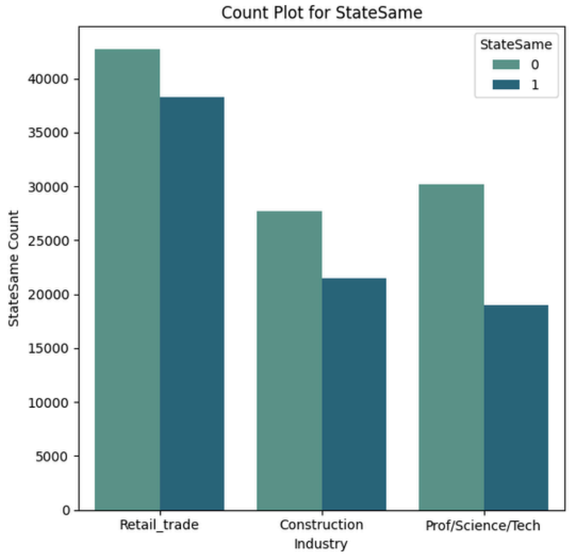
Top 3 industry sectors with the most disbursed loans are from **Retail Trade,** **Manufacturing,** and **Wholesale Trade**.

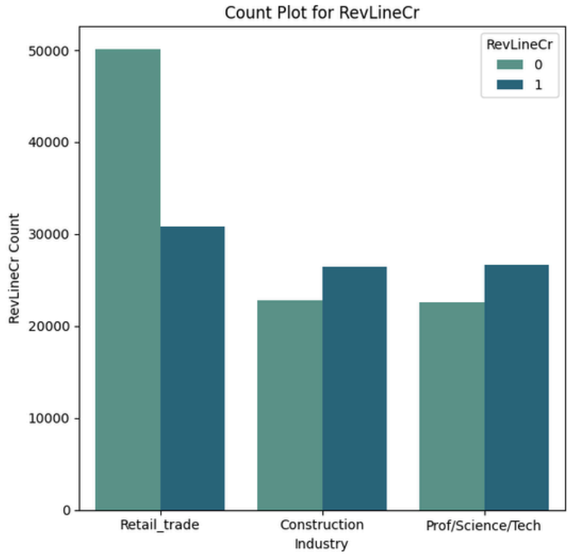


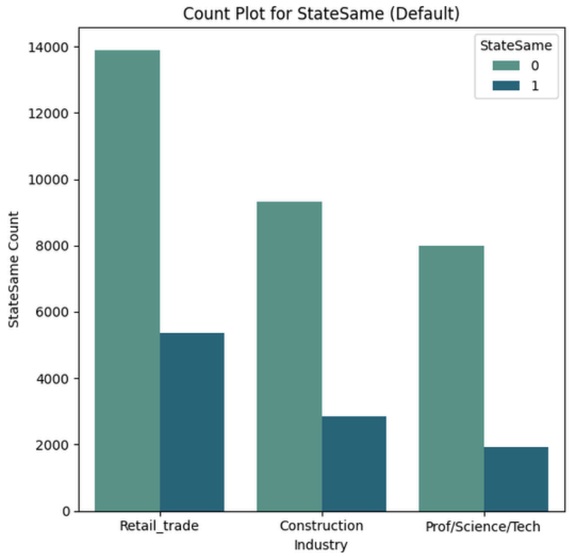
Most businesses applied for loans from banks in **North Carolina.** It also happens to be where the majority of defaults come from.

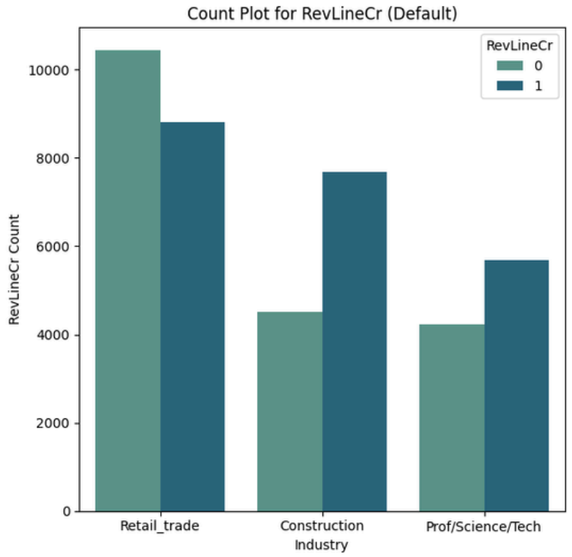





# Top 3 Industry Sectors:

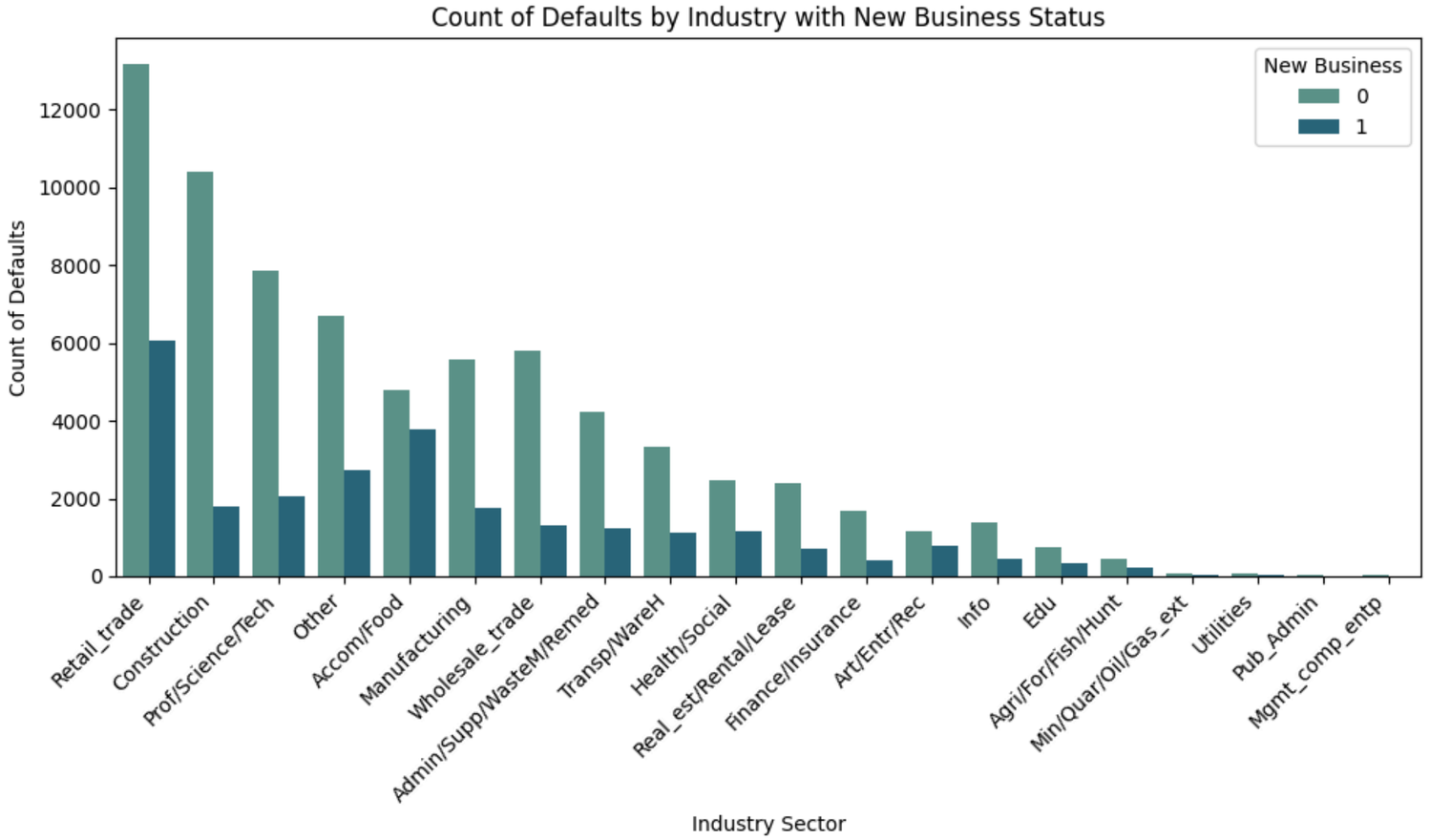
• Defaulting businesses often use a revolving line of credit, indicating higher credit utilization.

• Construction and Prof/Science/Tech industries often default due to heavy debt reliance and future sales.

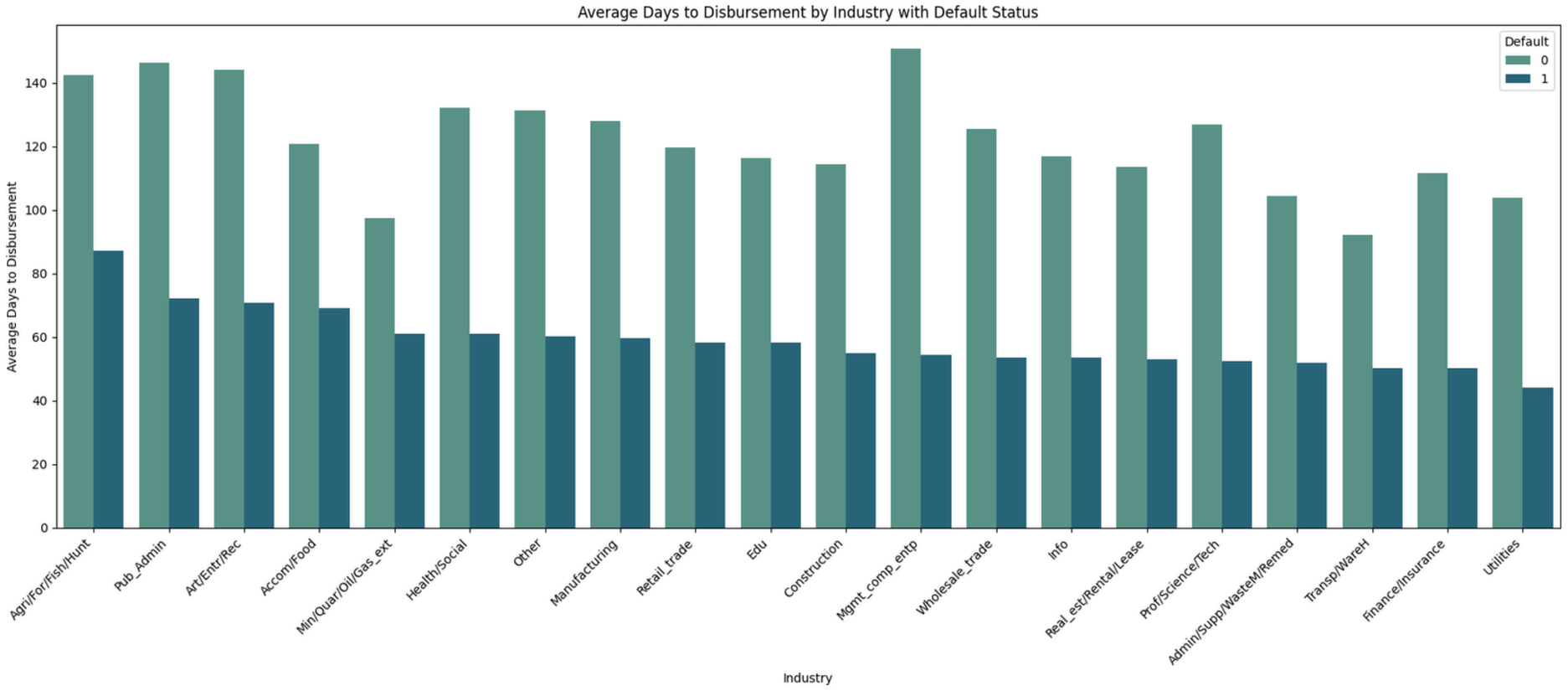
• Most businesses applying for loans are not from the same state, suggesting easier approval processes and larger loan amounts in certain states.

• This discrepancy suggests businesses strategically choose lenders based on approval likelihood and better loan terms,

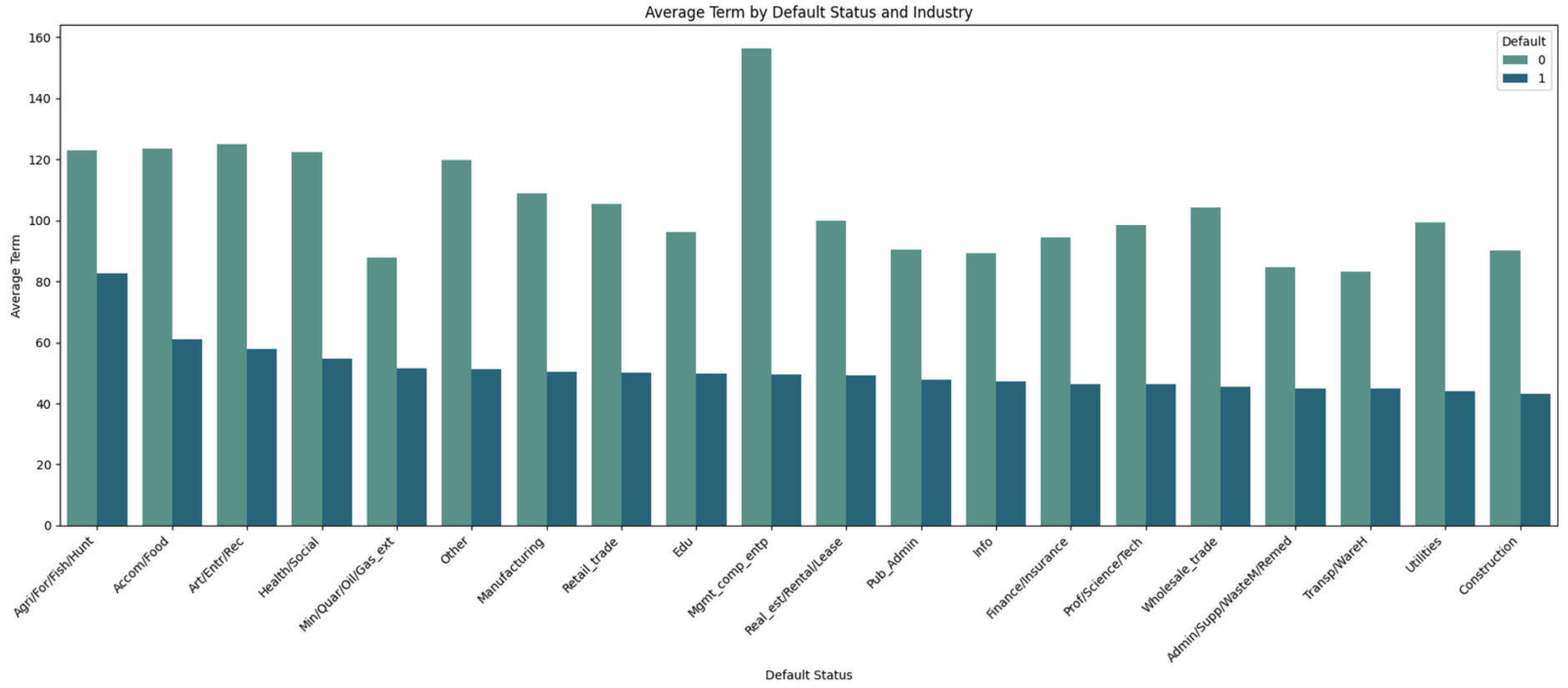
**Most business loans default are not new businesses**, but some sectors, such as Accommodation/Food and Arts/Entertainment/Recreation, **have new businesses nearly equal to old ones**, suggesting an economic failure may affect these industries and posing the highest risk.

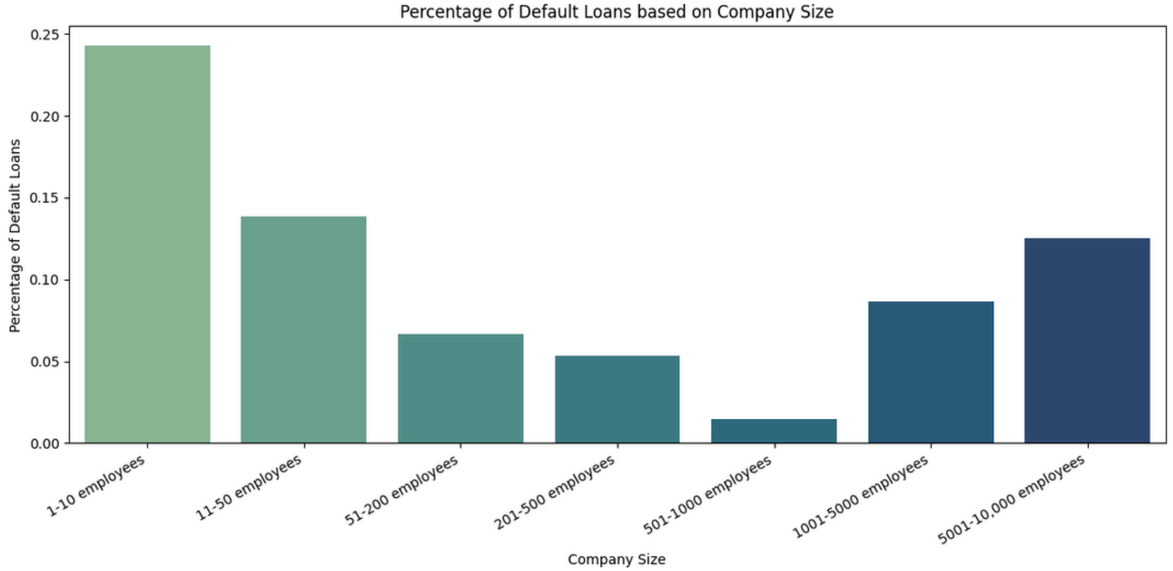


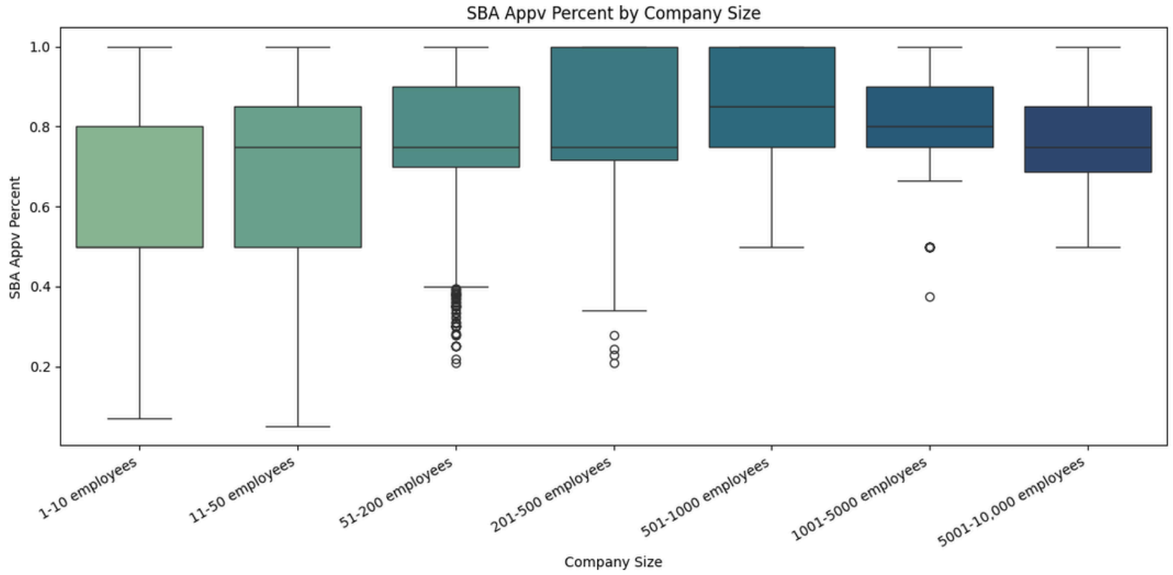
Most default loans have shorter disbursement days, suggesting **poor business checks** and **urgent loans**, where businesses may accept higher interest rates or face cashflow issues, indicating potential practices by banks.



Most defaulting loans are applied for shorter terms, indicating **higher monthly payments**, suggesting that businesses may have **miscalculated their stability** and are unable to repay.





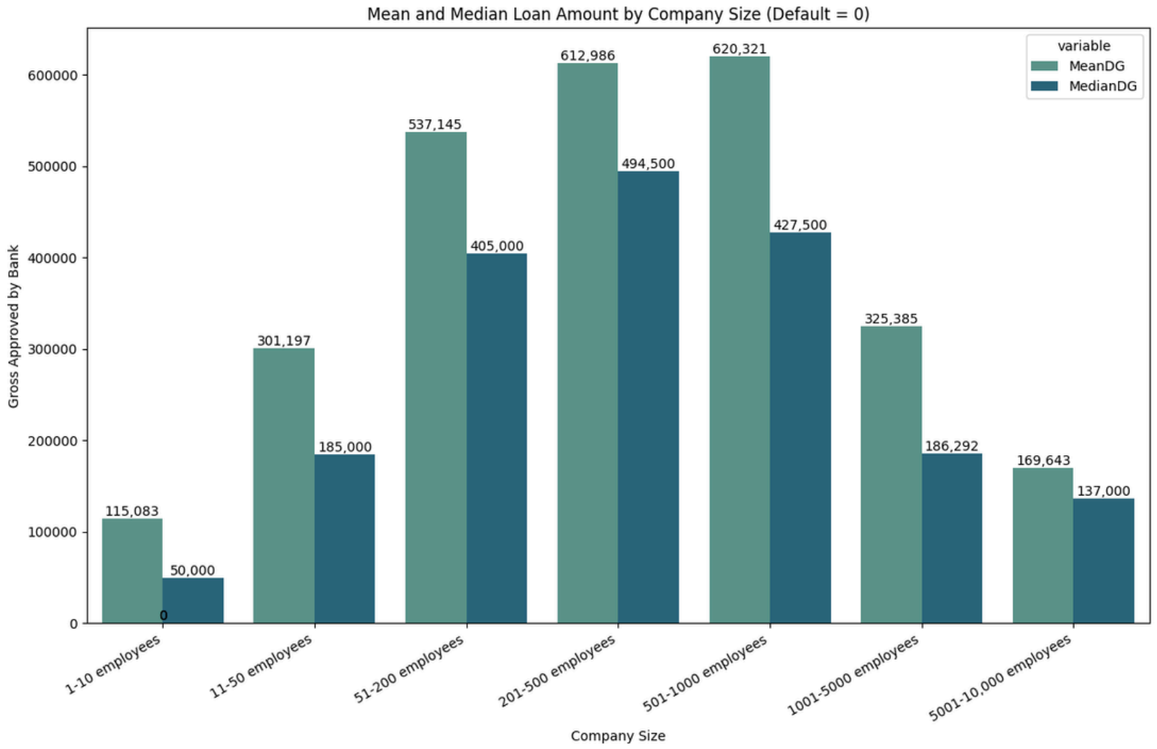



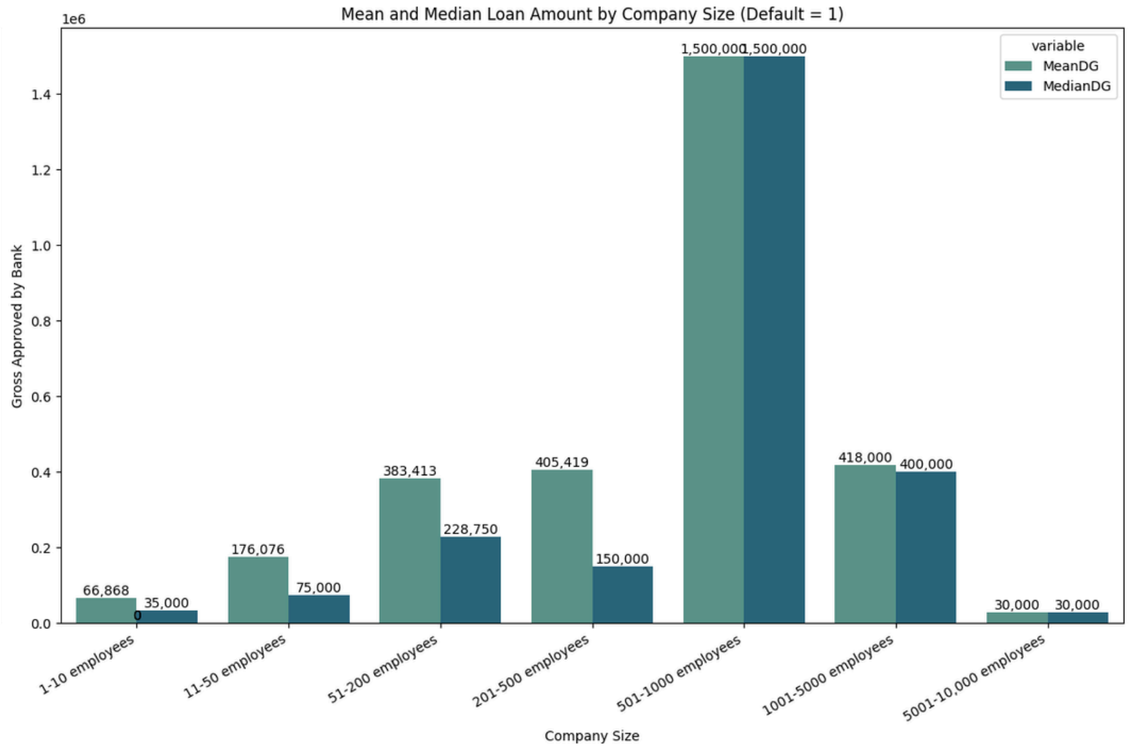

1. Businesses with mid-sized employees have lower default percentage, indicating that they have good cashflow.

2. The bigger the business, the more salaries and benefits they have to pay, insisting in applying for a loan.

3. The SBA's average approval percentage for larger businesses is higher, indicating a more thorough review of loan applications.

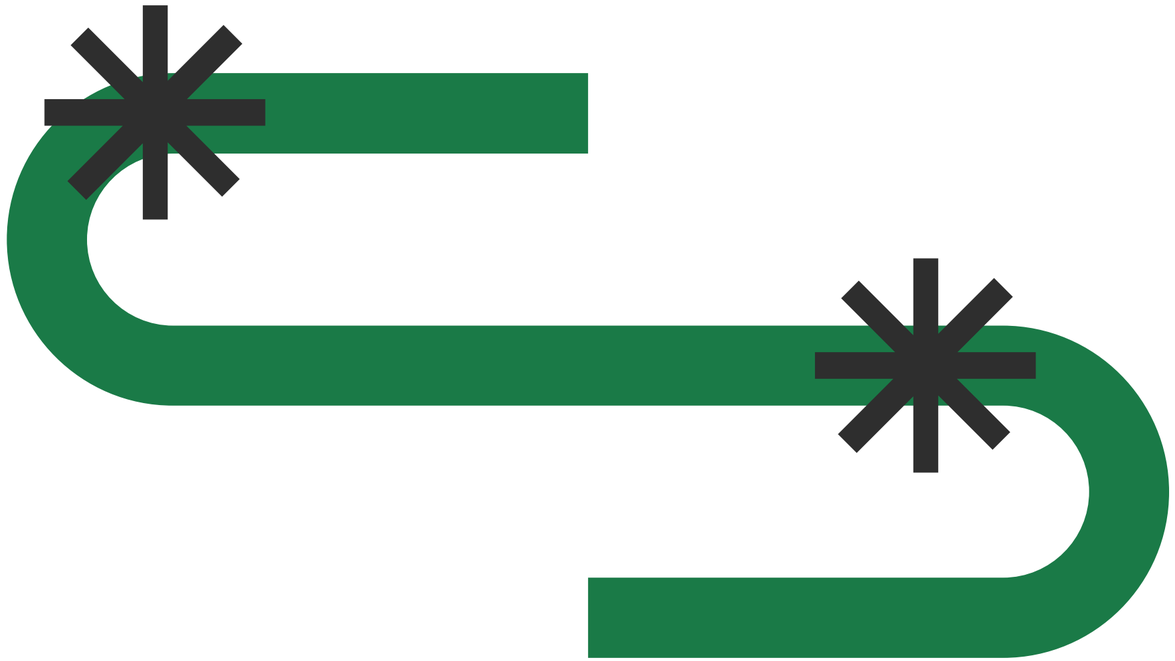
Big-sized businesses have seen an increase in defaults (%), but their median loan amount is similar to small-sized businesses. Mid-sized businesses apply for larger loans, but their default rate is the smallest as stated from the previous chart.



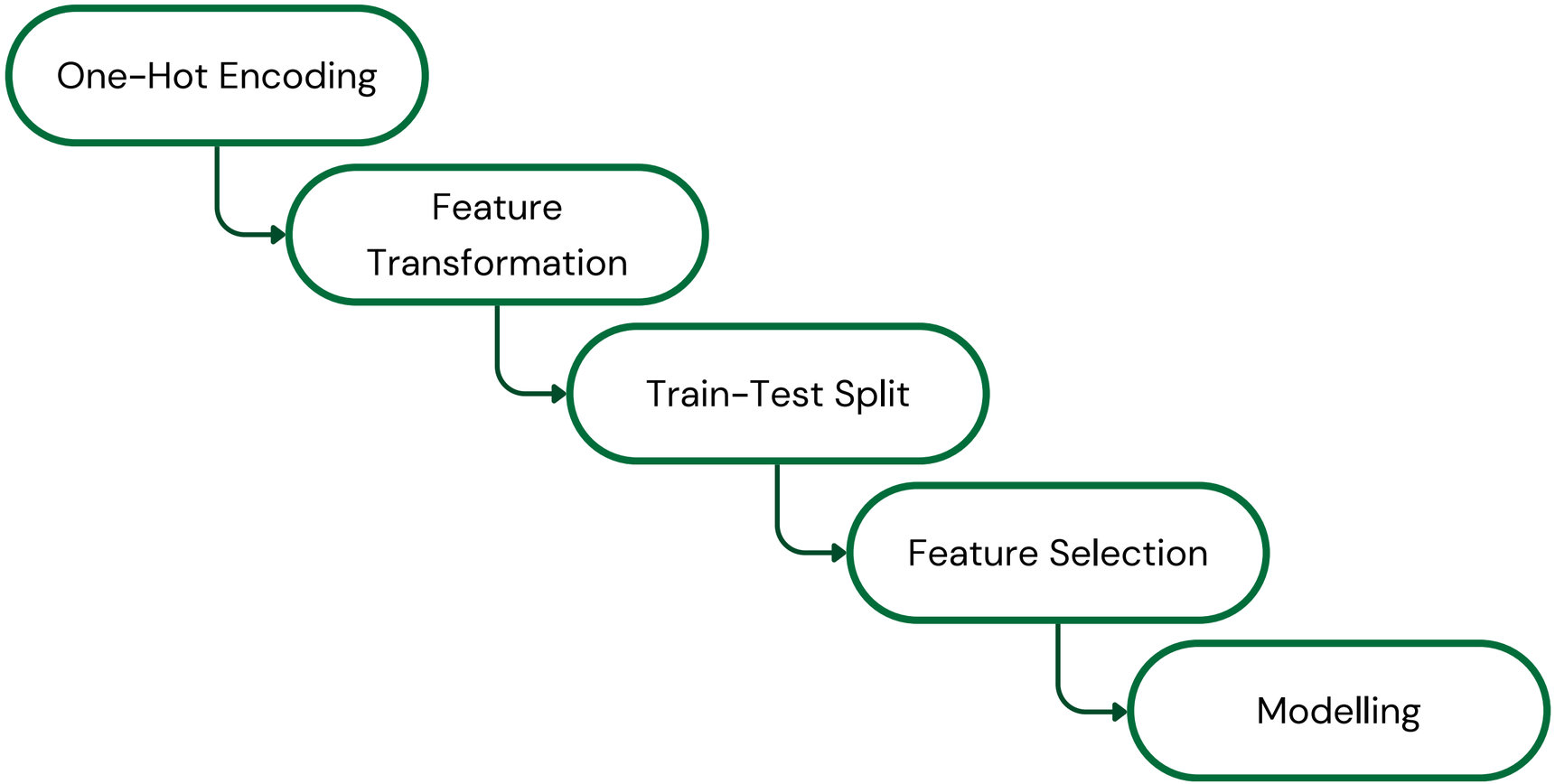



Section 03:

# Data Preprocessing

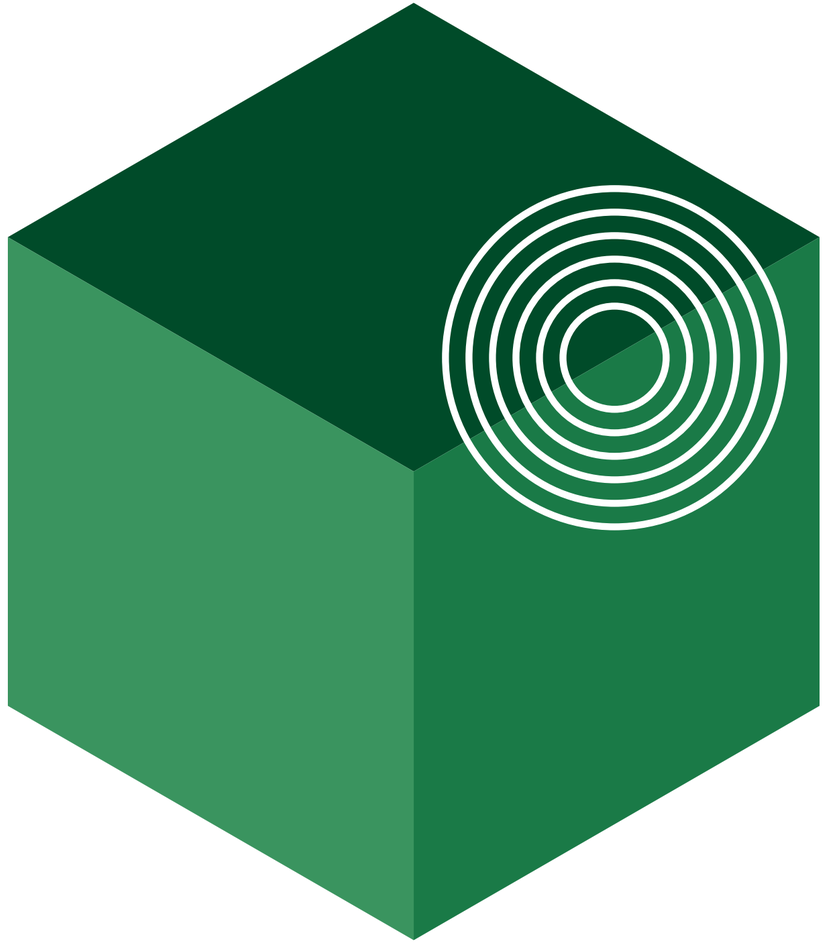


# Preprocessing Pipeline



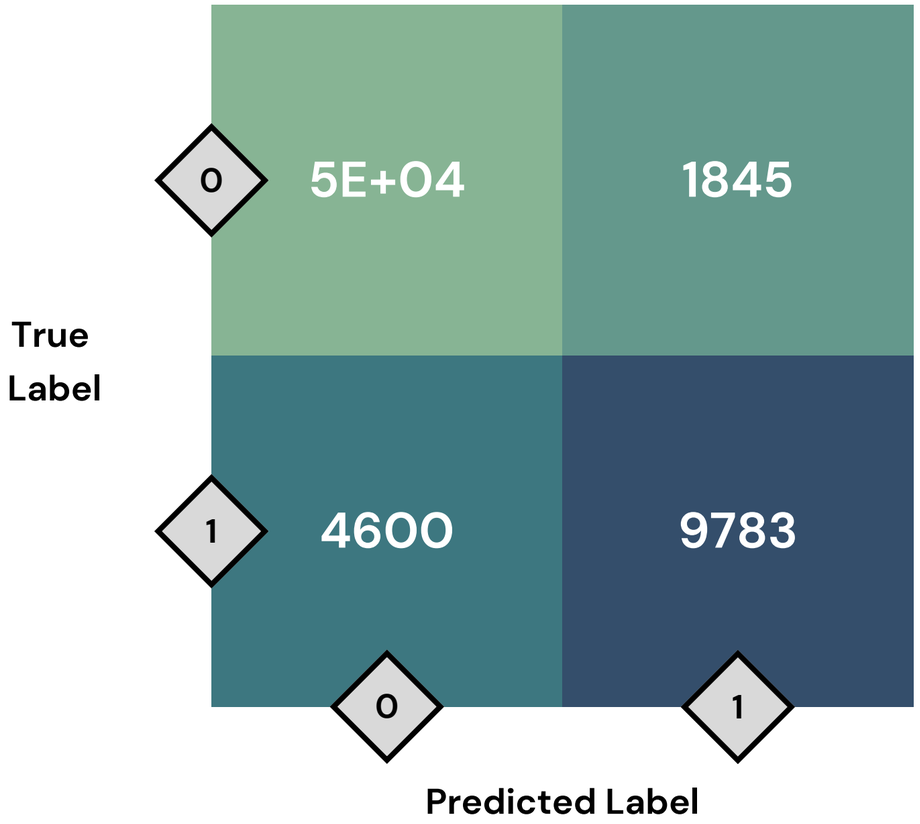
Section 04:

# Model & Evaluation



# XGBoost (tuned-1)

**Recall Score = 68%** **F1-Score = 75.2%**



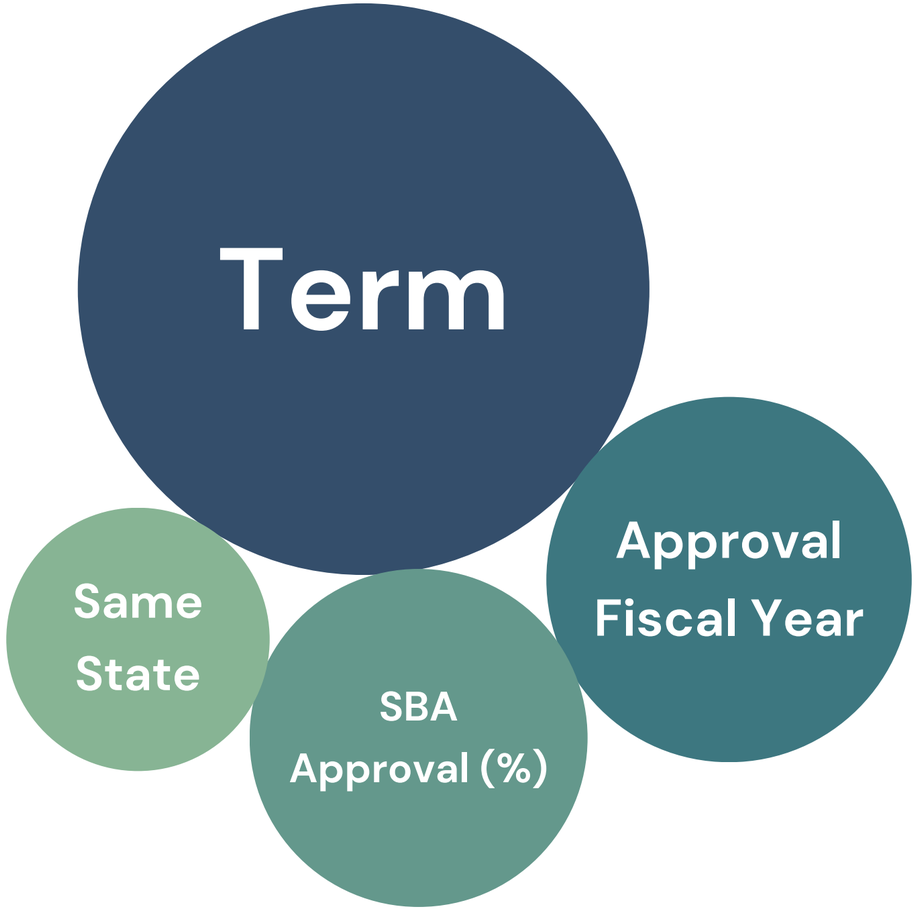

The **model was able to correctly identify 68%** of all business loans that actually defaulted.

The model demonstrates a **75.2% effectiveness in predicting loan defaults**. It can reliably distinguishes between which loans will default and which will not.

# Then what? (Model Interpretation)

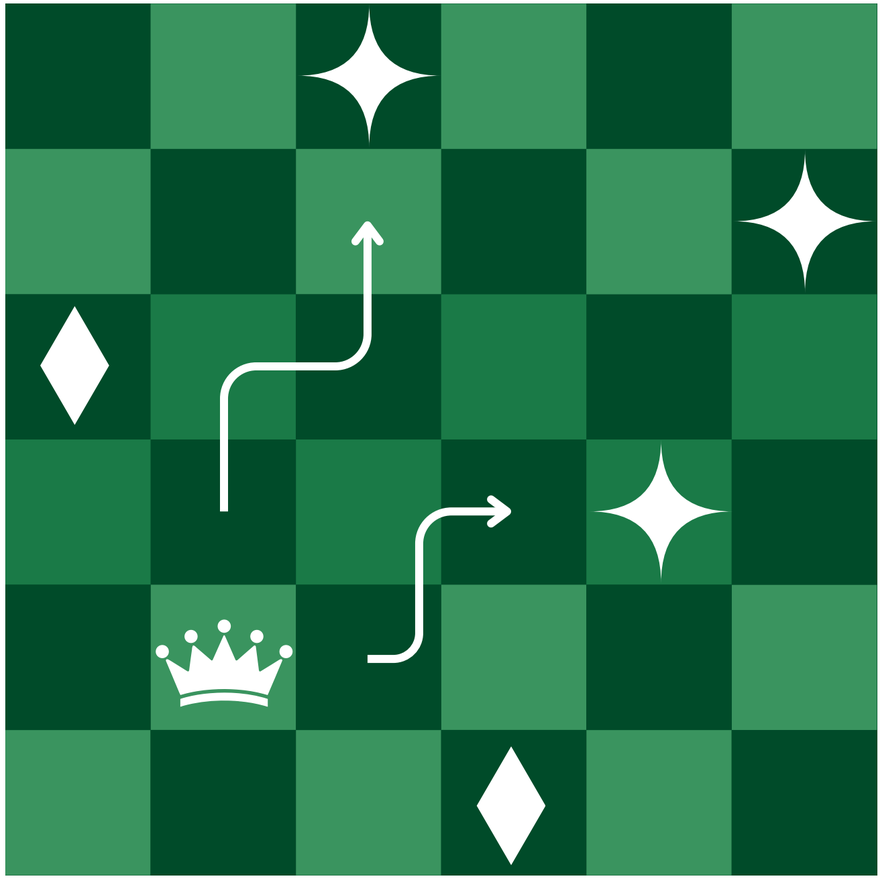
For example, if 100 businesses apply for a $100,000 loan. the bank will have to finance $10,000,000. If all these loans have the potential to default, this model can help prevent 68 defaults, potentially reducing costs by up to $6,800,000 gross.

# Key Features to keep a close eye on



Section 05:

# Business Recommendations



# Business Recommendations

To prevent more defaults, the SBA need to filter loan applications more thoroughly. Here are some recommendations to consider:

1. **Examine and strengthen the requirements for loan approvals**, particularly for high-risk sectors including construction, retail trade, and professional/scientific/technical services.
2. Use business information, such as **cash flow and transaction history**, to improve credit score.
3. **Keep a close eye on the borrowers' financial situation after payments** are made in order to spot any early warning indications of possible defaults.
4. To lessen the lender's liability, **demand collateral or personal guarantees** for loans with higher risk.
5. Banks and lenders with loose regulations can either **tighten their regulations or reduce their approval guarantee rate** from the SBA.

# Thank You!