

# Lending CLUB Case Study

---

BY: VIKAS KUMAR GUPTA  
& VINAYAK SHARMA

# Business Overview & Objectives

---

## Loan Default Risk in Consumer Finance

The main challenge for lenders is predicting loan defaults, which can lead to significant financial losses. The goal is to identify high-risk applicants to minimize credit loss while avoiding missed business opportunities.

### Business Goals

- ✓ Minimize credit loss by identifying risky borrowers.
- ✓ Adjust loan terms or interest for high-risk applicants.
- ✓ Improve **risk assessment** to guide loan decisions.

### Loan Data Overview: Loan outcomes include:

- ✓ **Fully Paid:** Loan repaid in full.
- ✓ **Current:** Repayment in progress.
- ✓ **Charged-off:** Defaulted loan.

The focus is on identifying factors that contribute to **loan defaults**.

# EDA Approach & Risk Analytics

---

## EDA Approach

1. **Clean Data:** Handle missing values and outliers.
2. **Visualize Patterns:** Identify trends related to loan defaults.
3. **Correlation:** Find factors linked to defaults using Univariate, Bi-Variate and multivariate data analysis
4. **Risk Segmentation:** Group applicants into risk categories.

## Key Variables for Default Prediction

- ✓ **Financial Profile:** Income, credit score, debt-to-income ratio.
- ✓ **Loan Details:** Amount, interest rate, tenure.
- ✓ **Demographics:** Age, employment status.

**Risk Analytics in Banking :** Use data to predict defaults and adjust loan terms to reduce **financial loss**.

# Key highlights

---

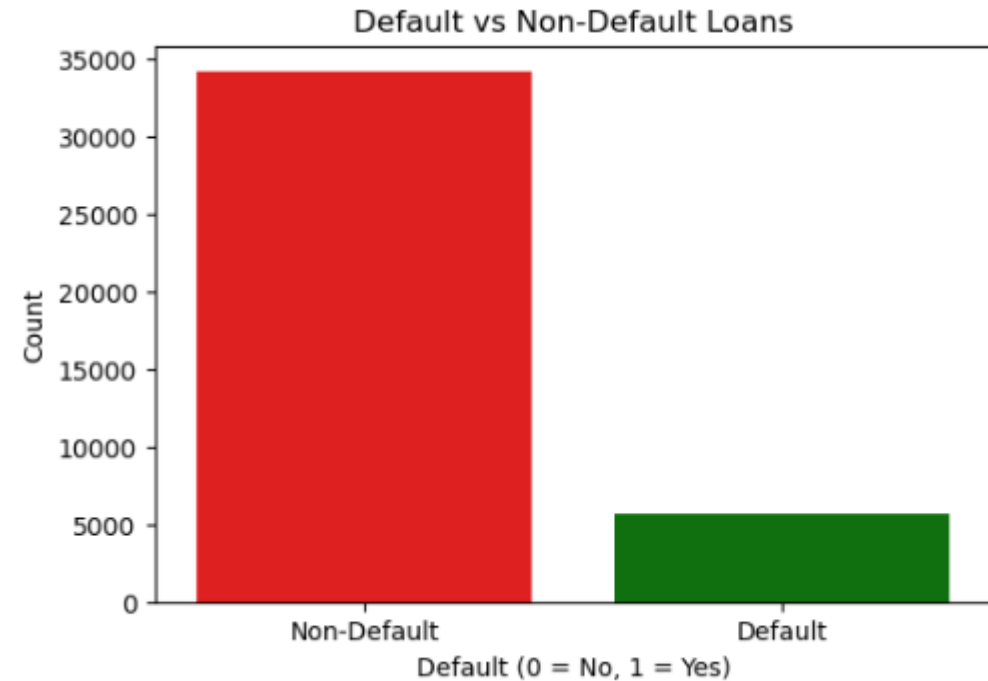
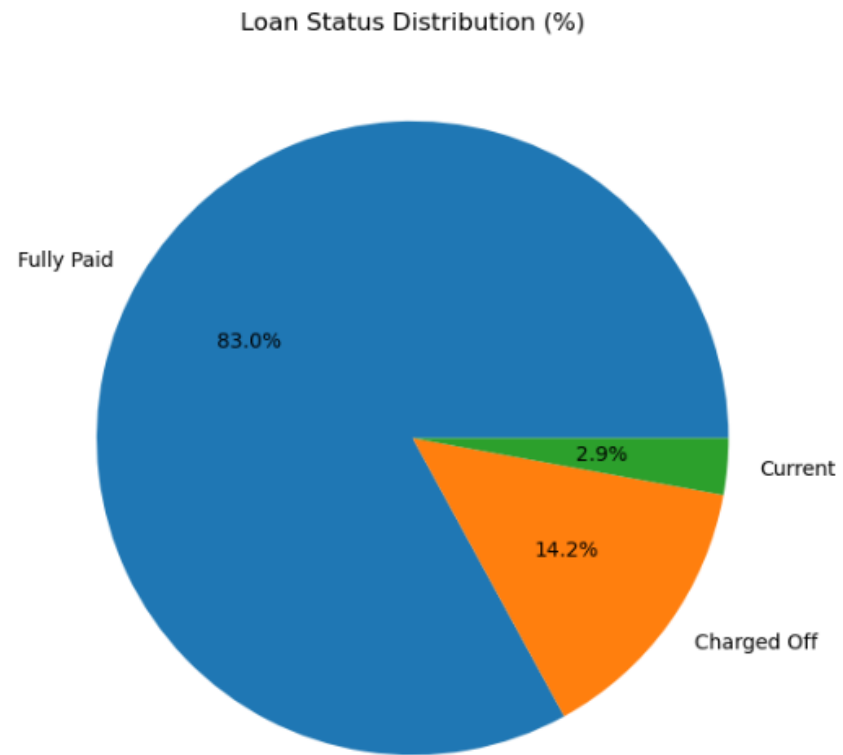
Data Cleaning: Data points dropped, due to insufficient/irrelevant information, it would be hard to drive any inference

- ✓ A lot of data points has only null values
- ✓ There were a few data points had more than 60% null values
- ✓ Dropping the data points having only distinct /single values like (ID, member id, pymnt\_plan)
- ✓ Data points having same min and max values

Data Transformation & Treatment:

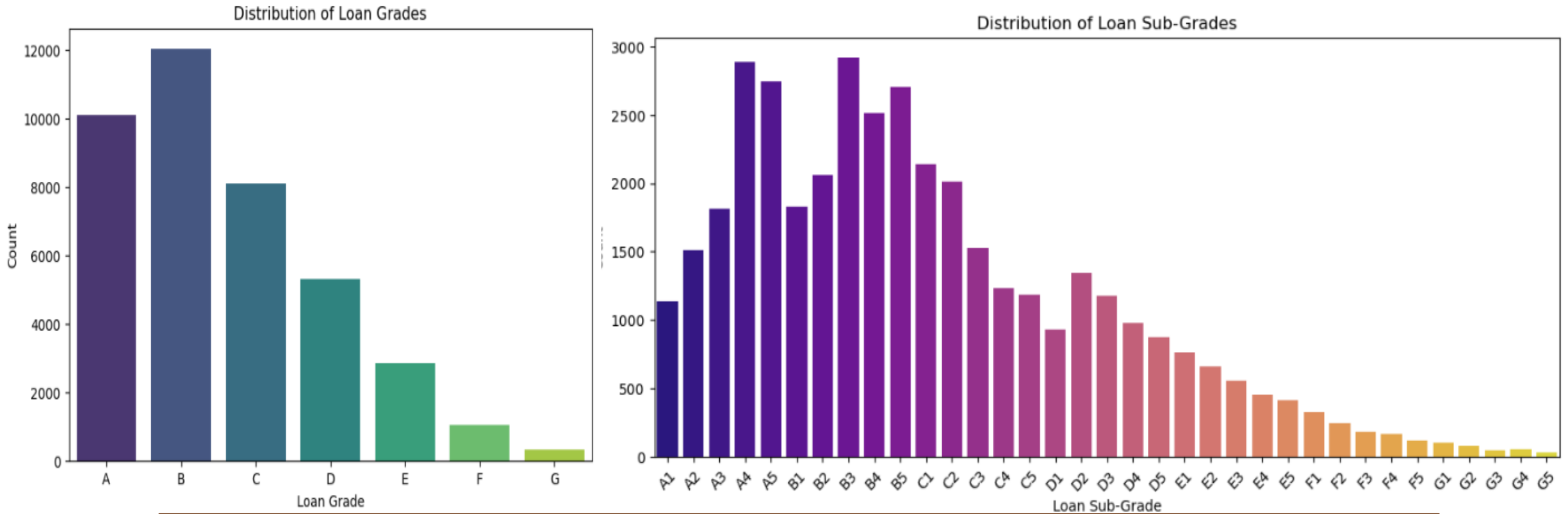
- ✓ Missing values in employee length has been replaced with “Business class” assuming the people how are employed did had already given the information about their experience range
- ✓ default\_loans: Created a derived column out of the loan status
- ✓ The analysis also includes many derived categorical variables from integer/float to bucket the big range into defined categories to simplify the analysis

# Univariate analysis: Loan status distribution



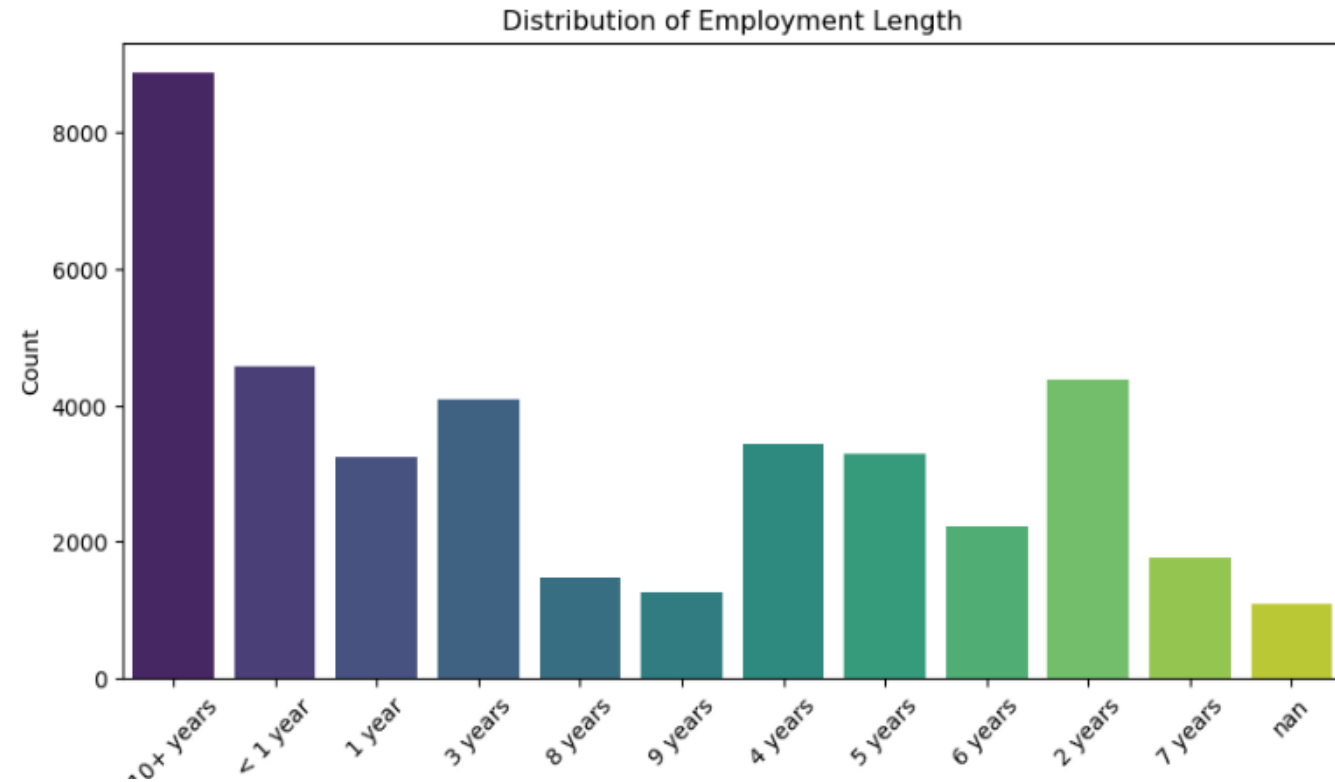
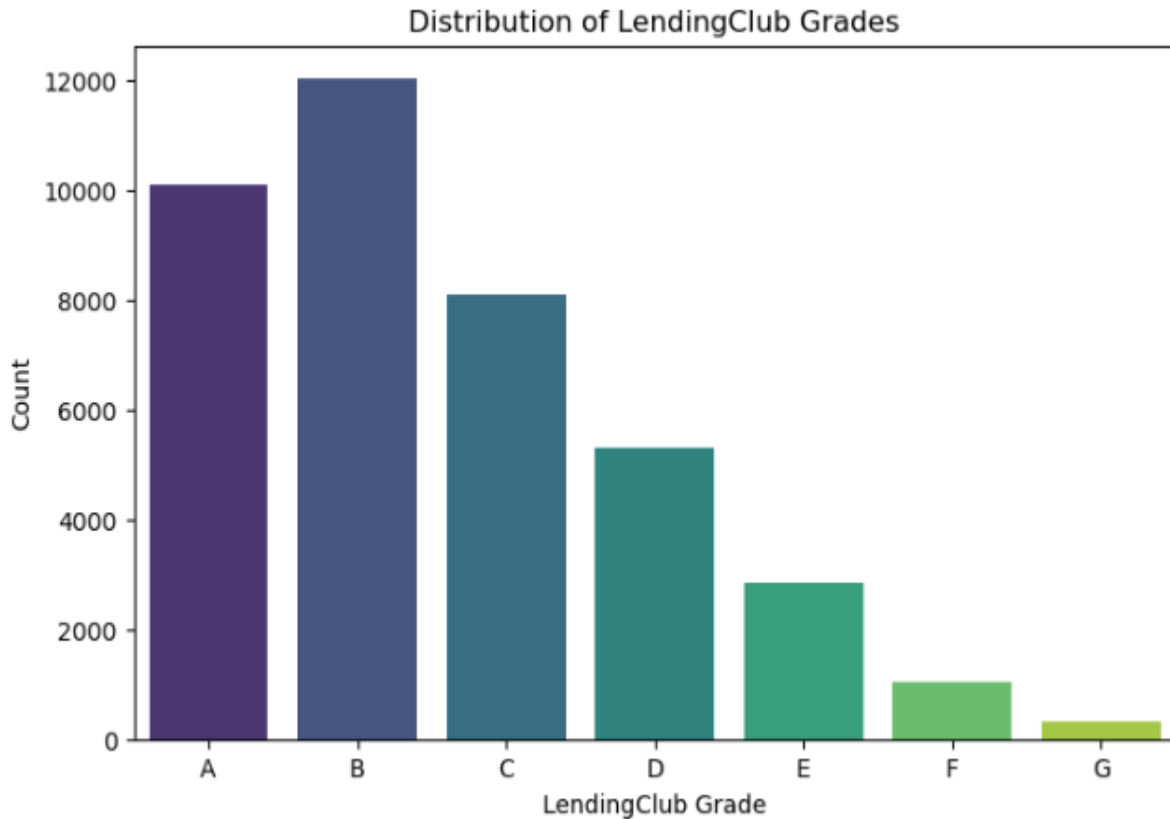
82.9% of applicants has paid loan, 14.1% of the applicants has not paid the loans

# Univariate analysis: Grade and sub-grade distribution



This provides insight that the loans within grade are more skewed towards lowered sub grades.

# Univariate analysis: Lending Club Rating & Experience

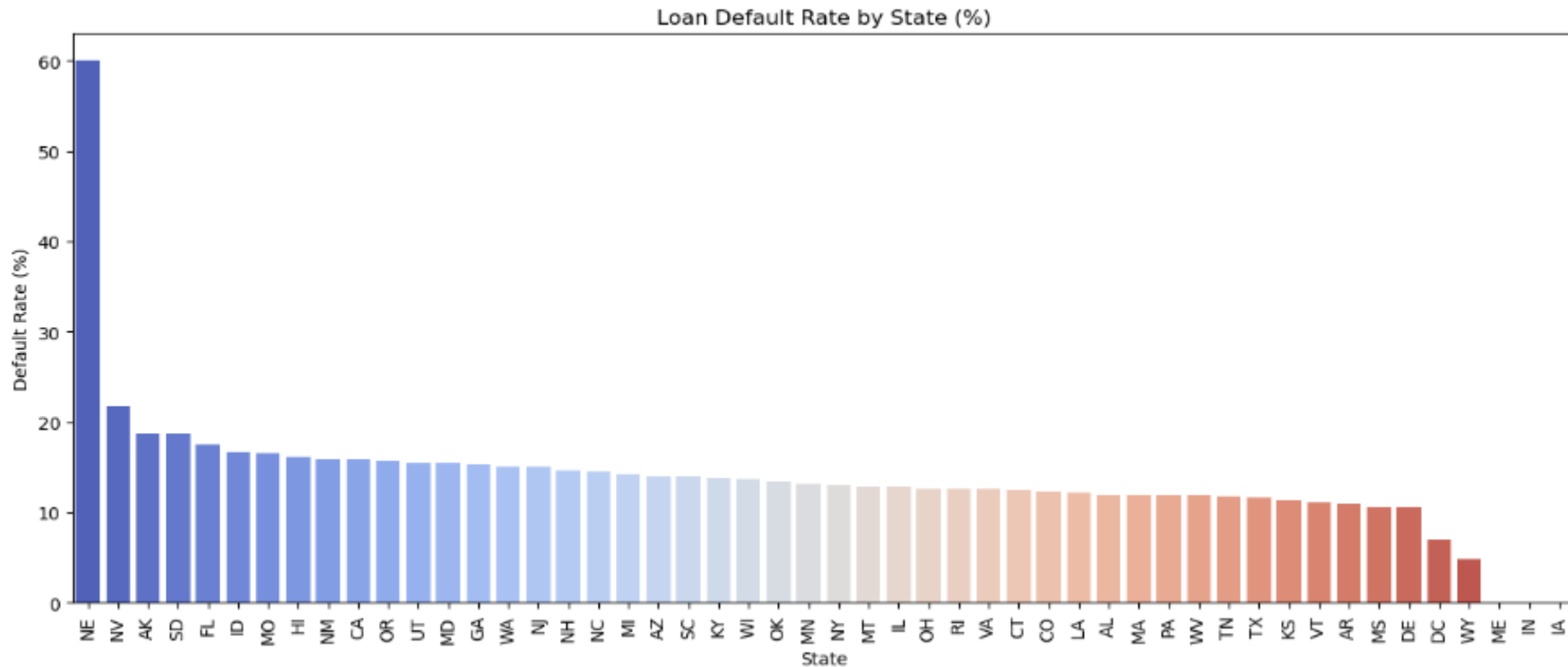


This provides insight that the fewer loans were given to the low Ratings, More loans were given to 10+ years of experience

# Bivariate Analysis: Zip Code and State vs Defaults

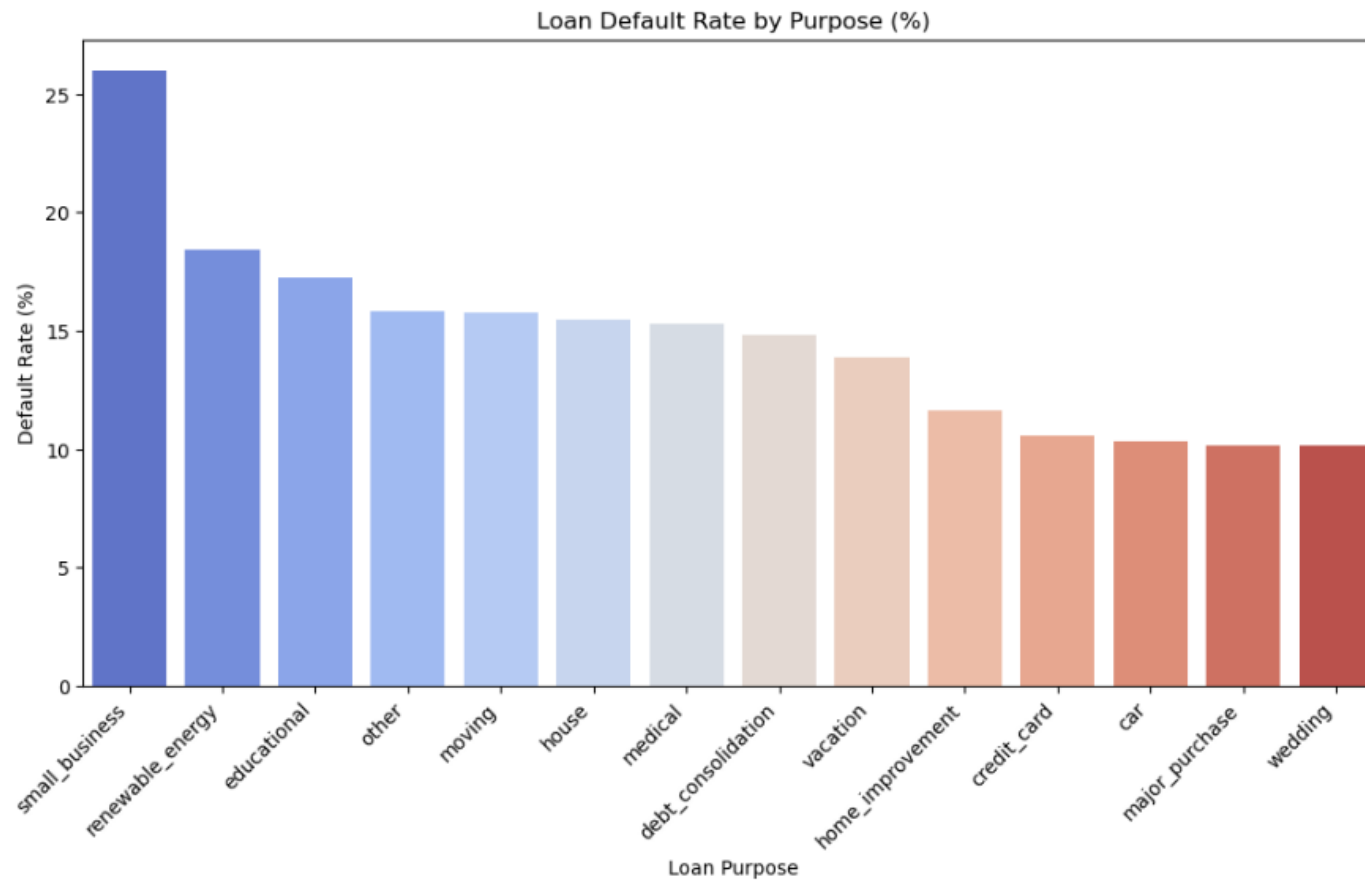
---

It is hard to conclude the defaults via zip code. However we can conclude that few states have higher default rates compare to others





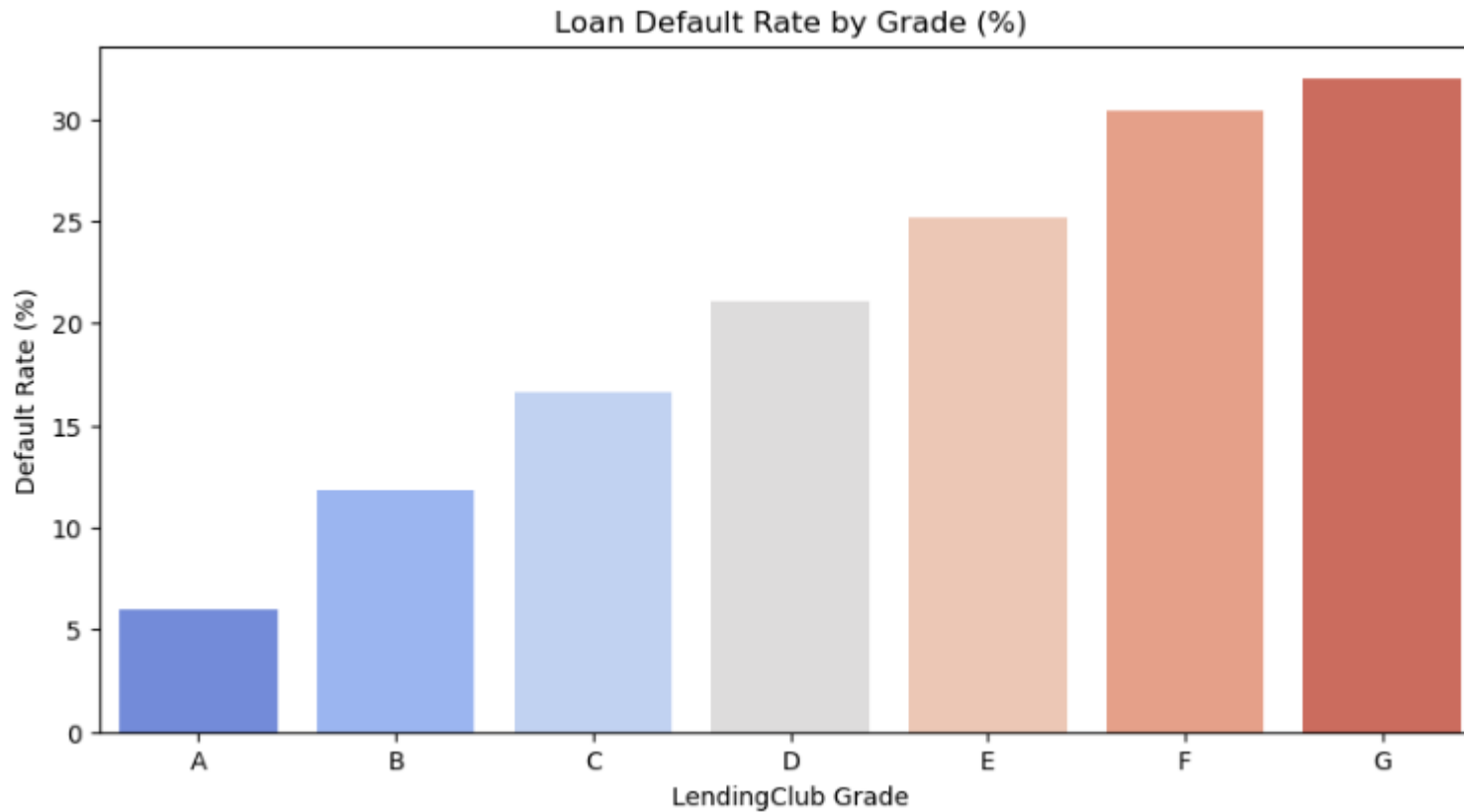
# Bivariate Analysis: Loan Purpose vs Defaults



You can see loan taken for business have highest default, lower being with car loan and wedding

# Bivariate Analysis: Grades vs Defaults

---

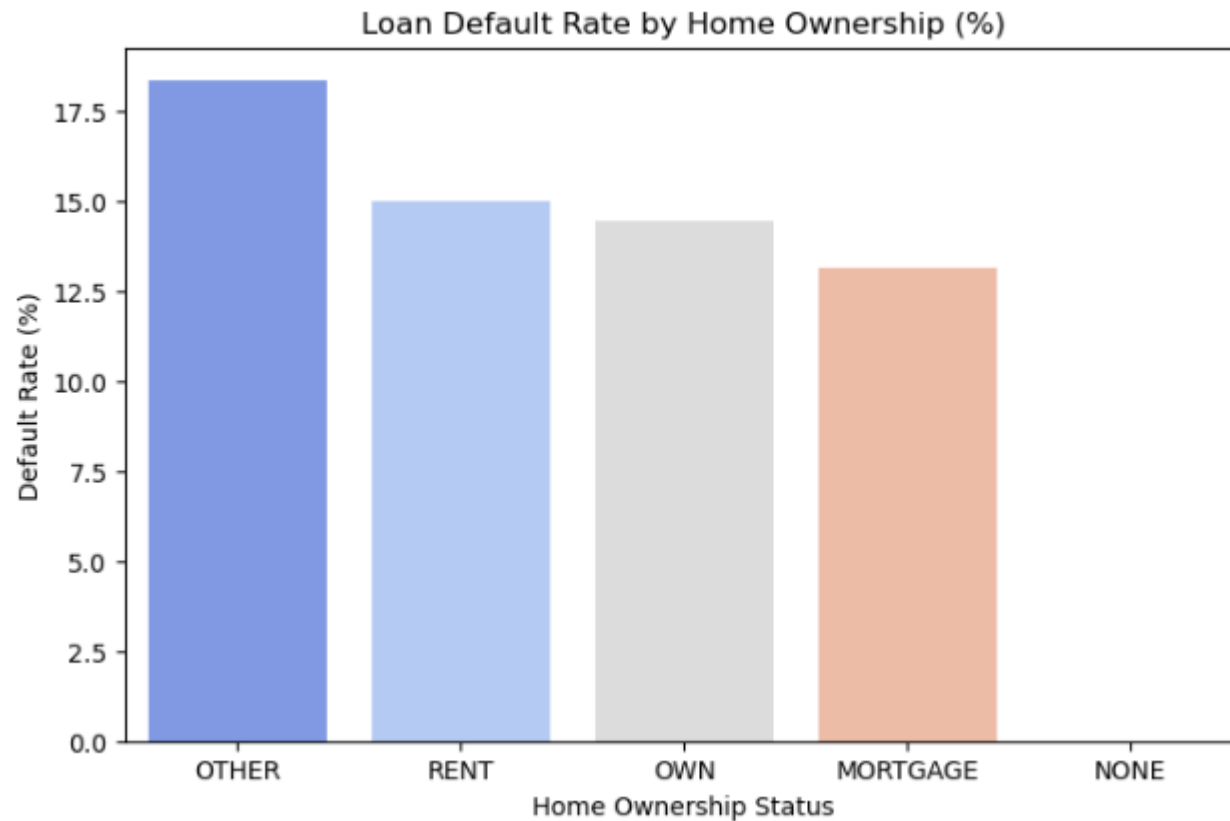


You can see that lending club grades have significant impact on loan default. Lender should take this score seriously.

Loans with lower grades have higher defaults.

# Bivariate Analysis: Home ownership vs Defaults

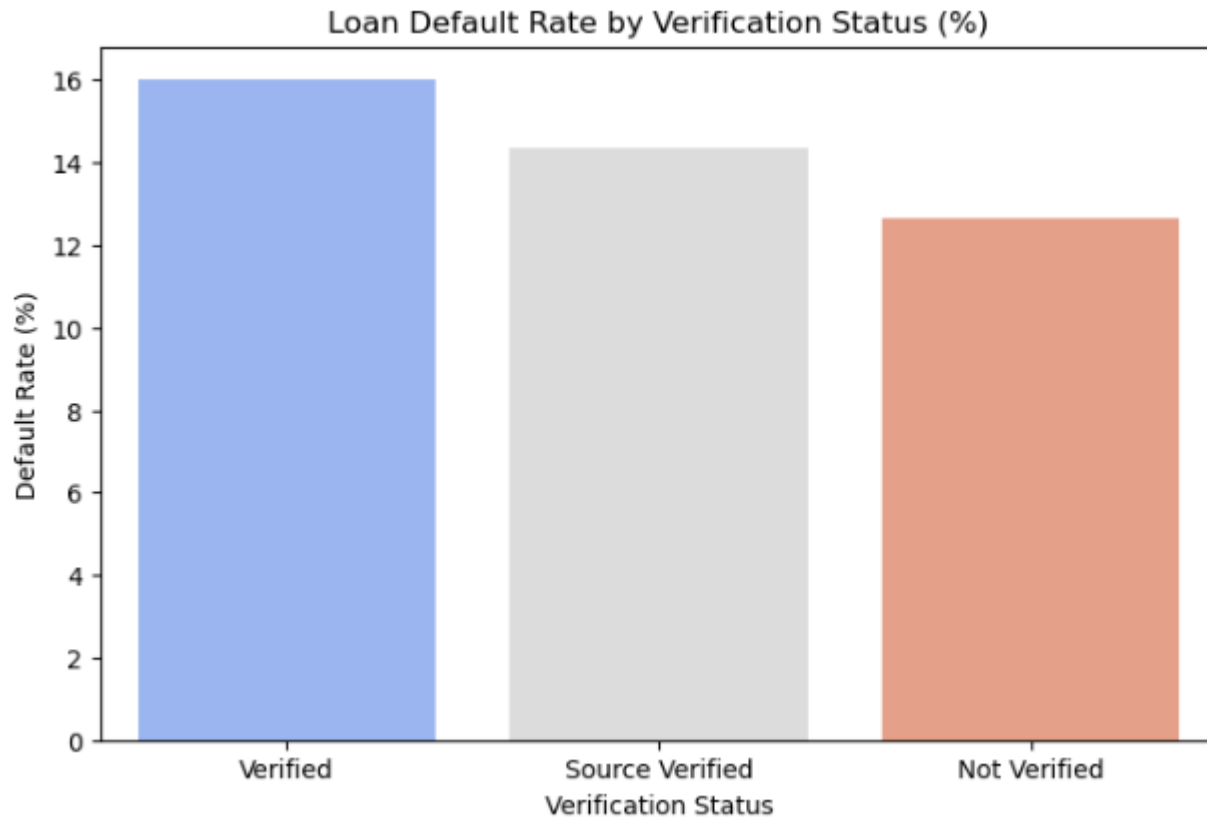
---



We can conclude that the home ownership does not have any impact non default rate

# Bivariate Analysis: Verification status vs Defaults

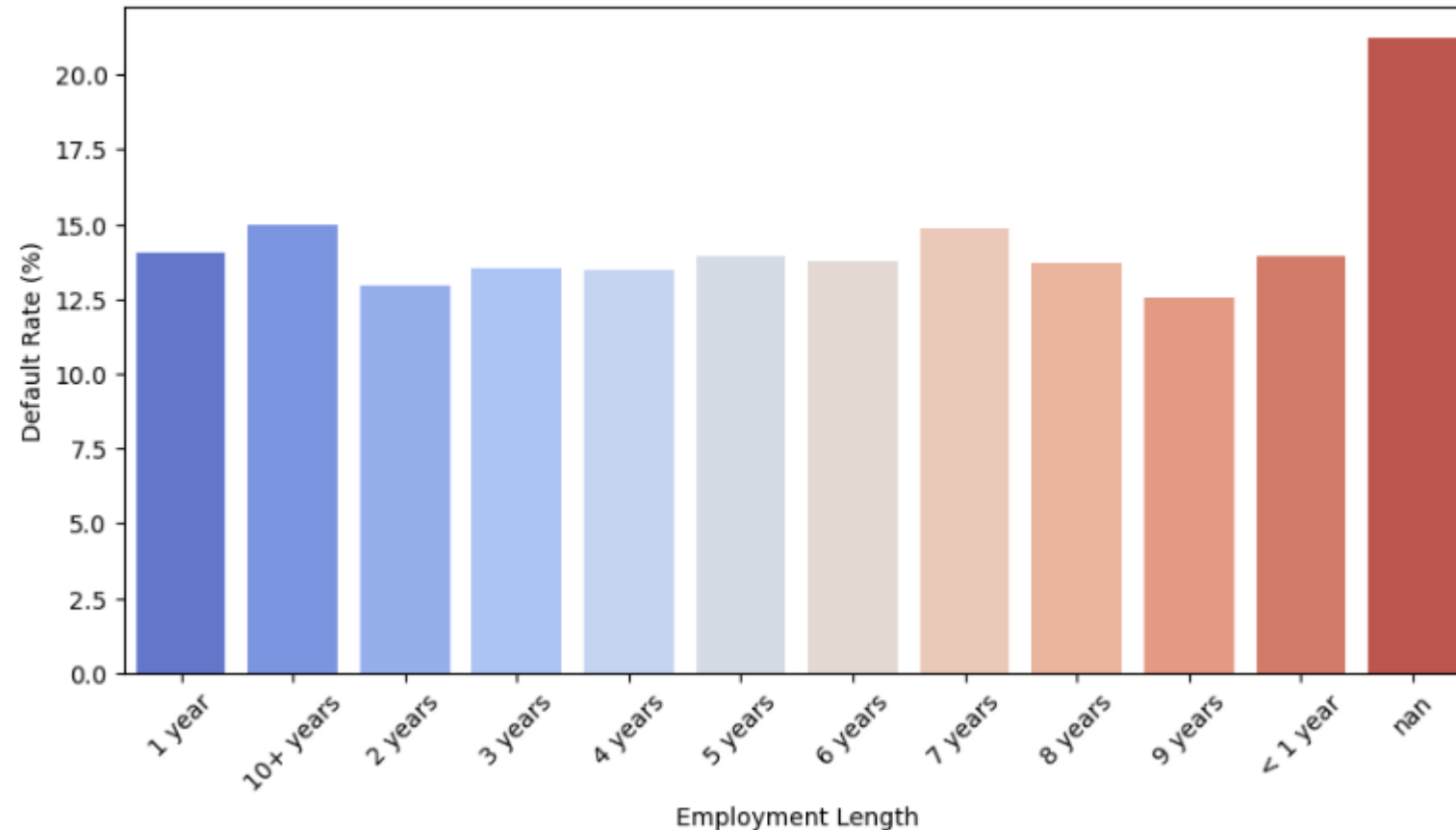
---



We can conclude that the verification status does not have any impact non default rate, it is higher on verified borrowers therefore we can conclude that the verification status can not be considered for our analysis. It does not signify that we should not do the verification

# Bivariate Analysis: Employment Length vs Defaults

Loan Default Rate by Employment Length (%)

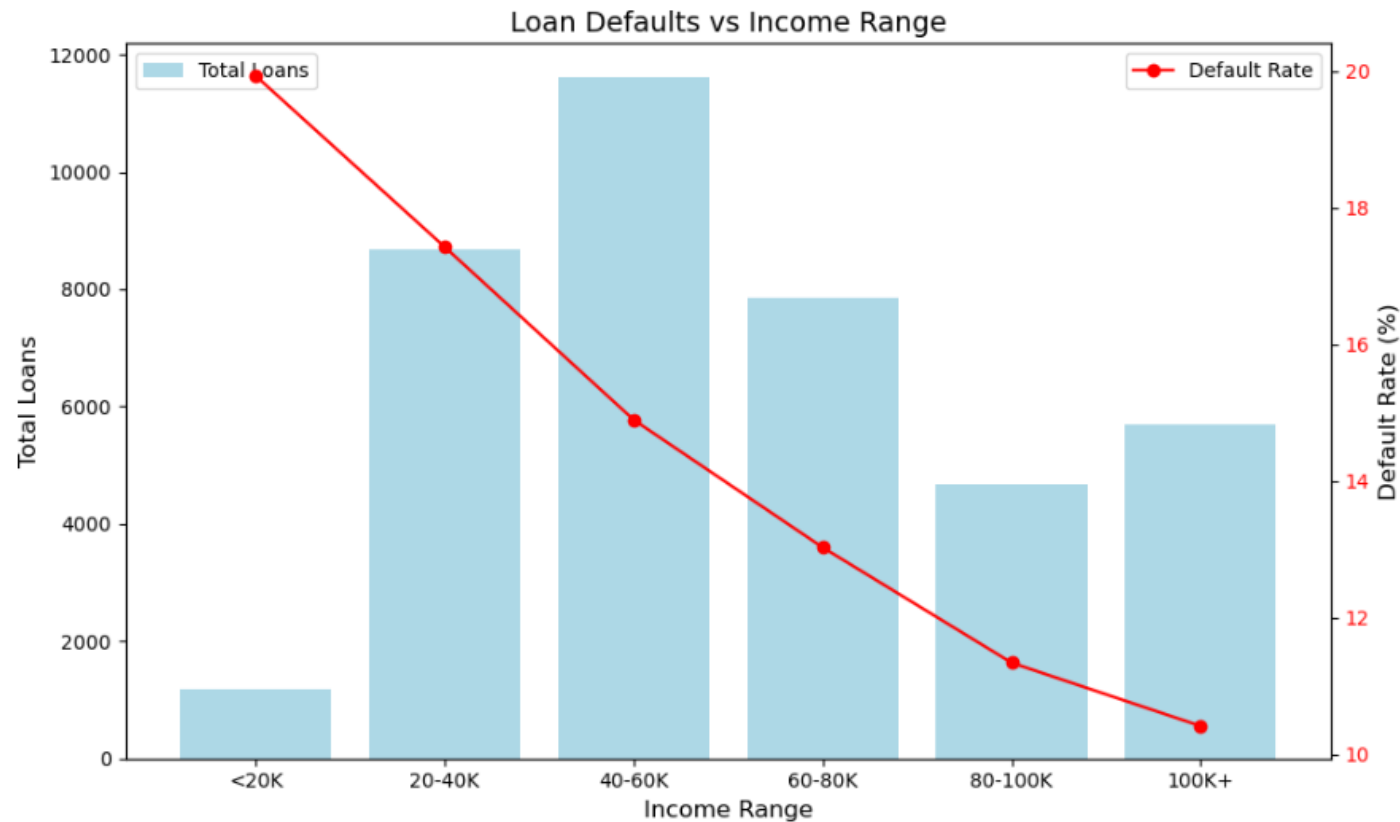


- All the employments length has the default

You can see the business class (nan) has highest loan default followed 10+ years.

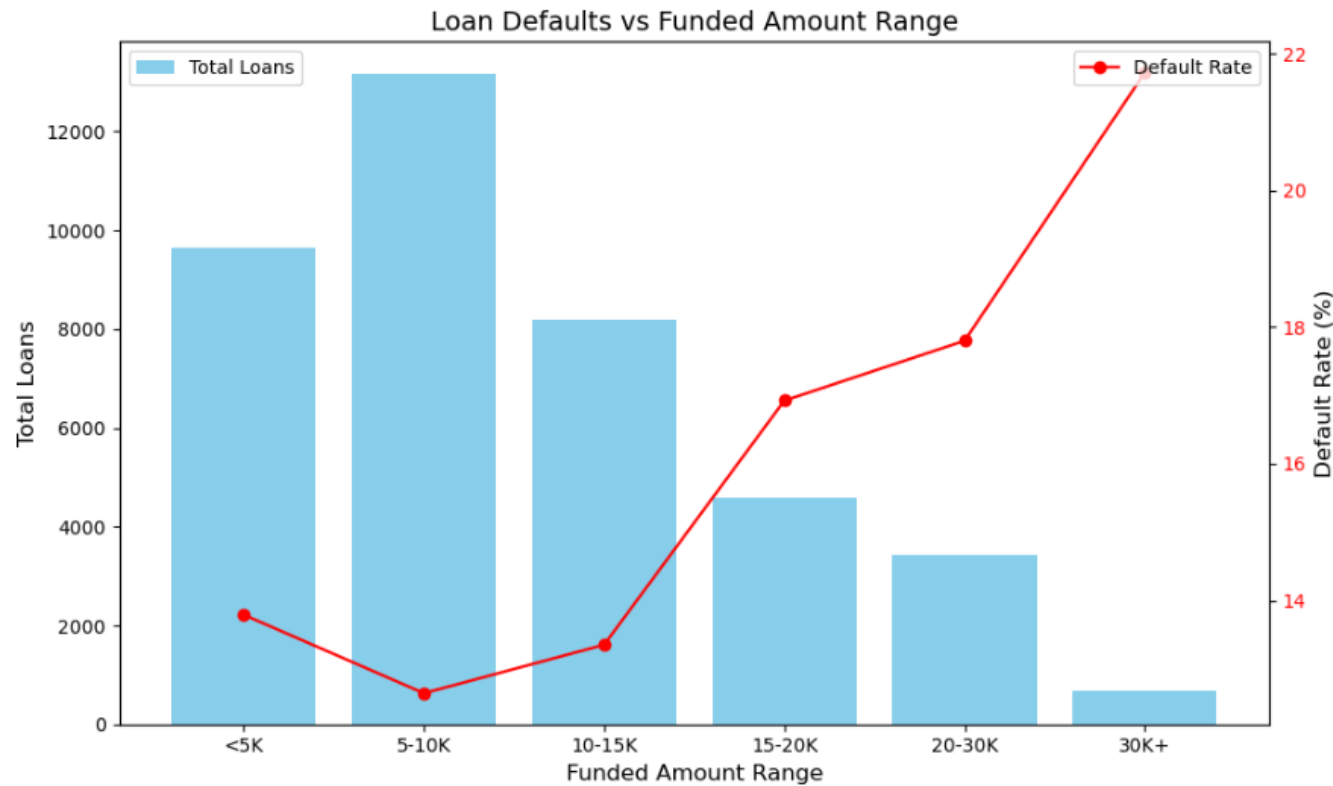
10+ years experience has got higher number of loans compare to the others

# Bivariate Analysis: Income Range vs Defaults



You can see low-income range has higher defaults and higher income range has less defaults

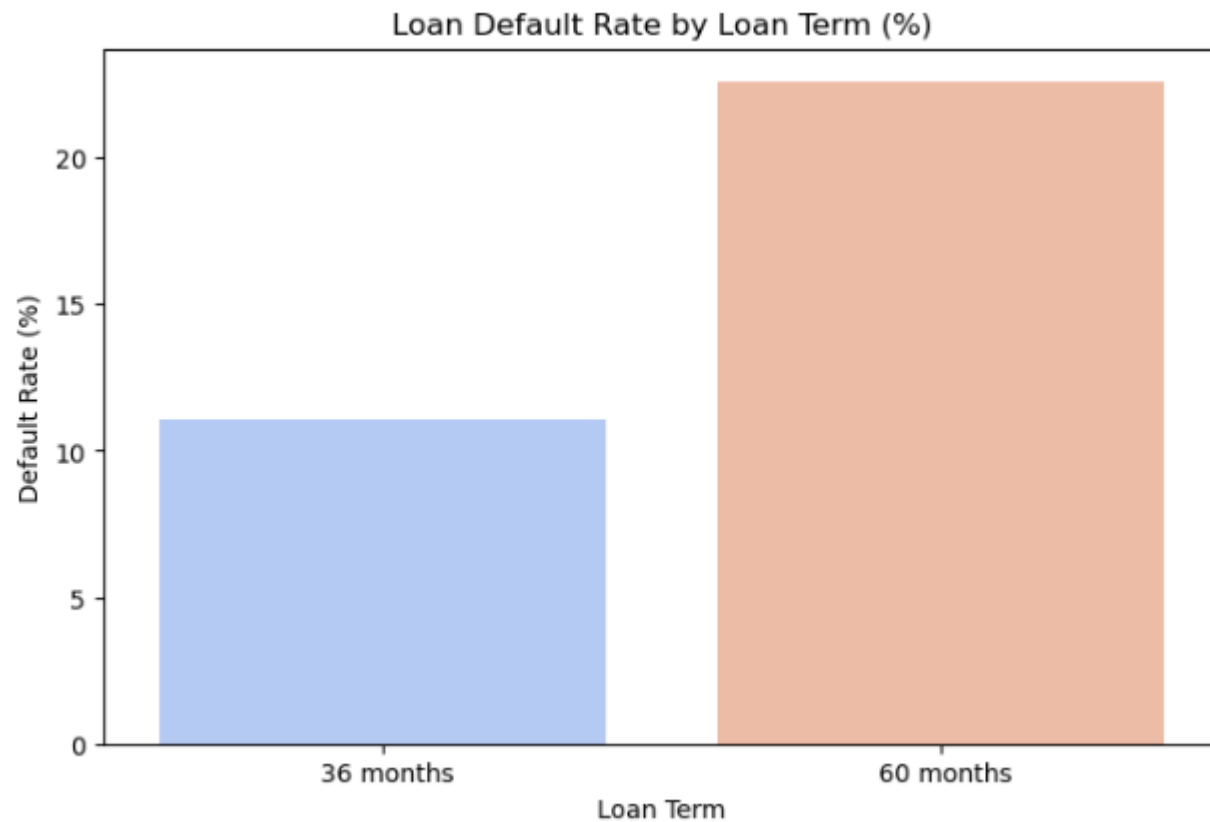
# Bivariate Analysis: Funded amount Range vs Defaults



You can see if higher funded amount has higher number of defaults

# Bivariate Analysis: Term vs Defaults

---

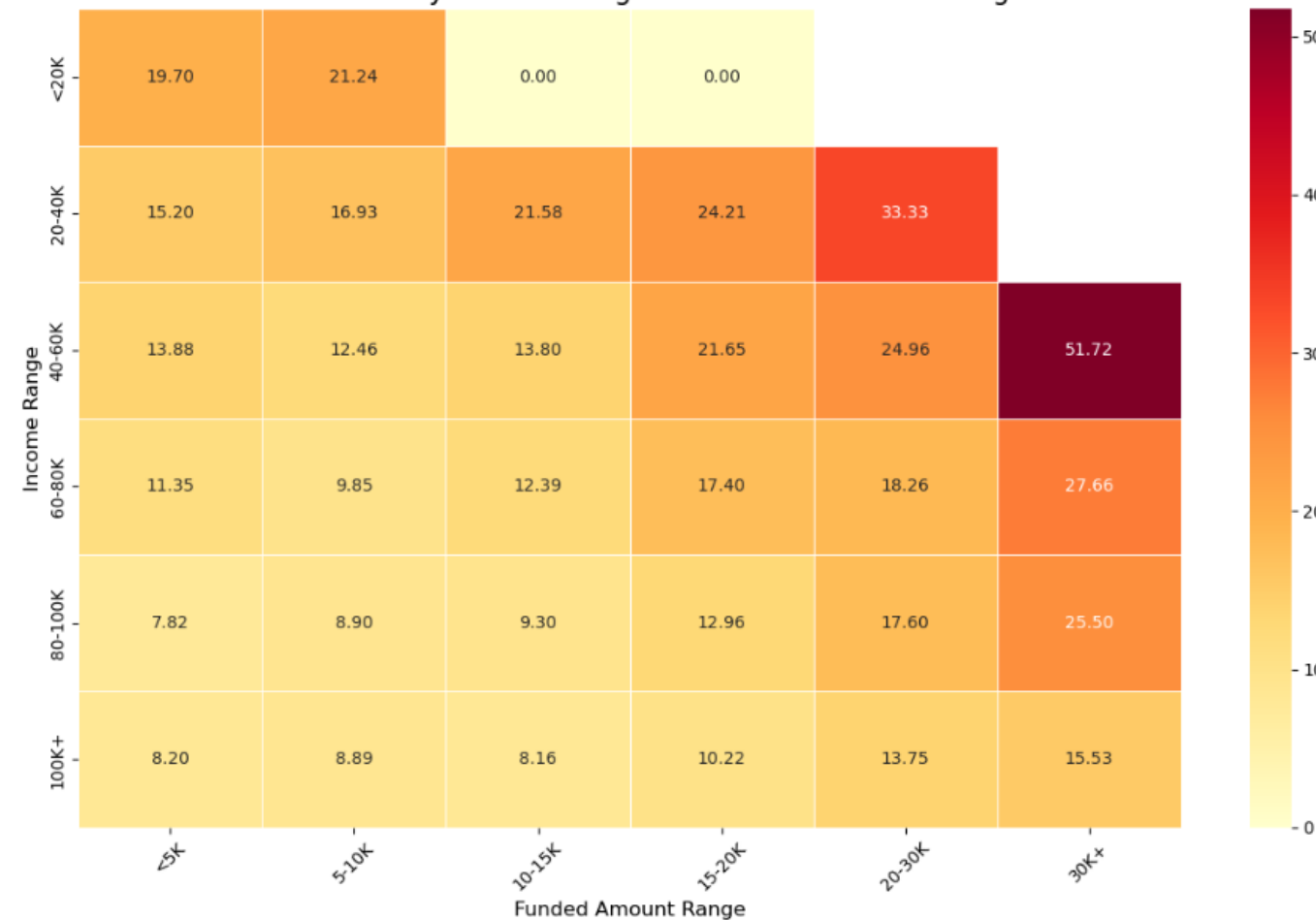


Short term has lower defaults



# Multivariate: Default Rate by Income Range and Funded Amount Range

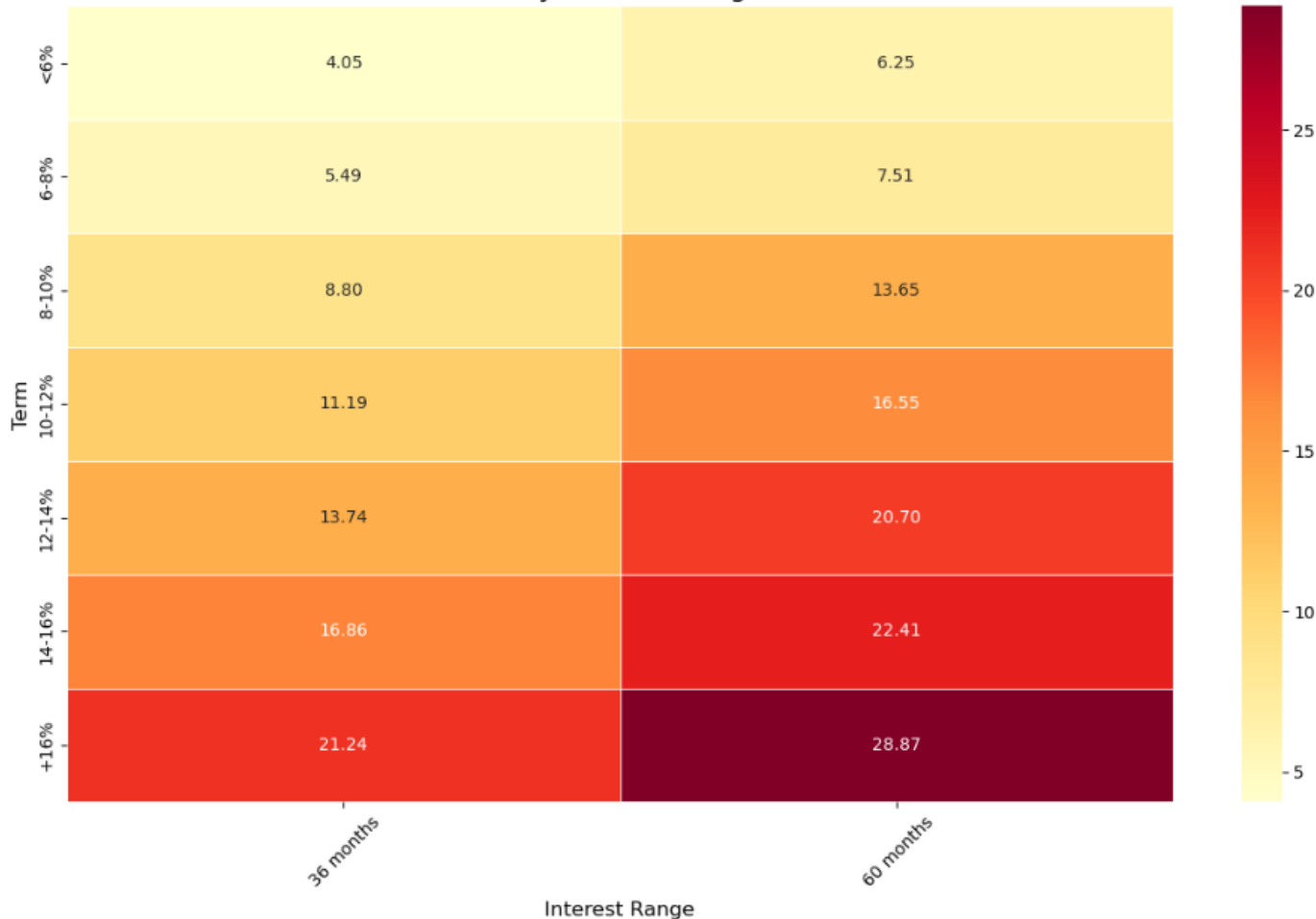
Default Rate by Income Range and Funded Amount Range



- Risk is higher for the low-income borrowers taking on large loans
- Higher income borrowers are likely to default regardless of loan amount, However the risk increases when the funded amount is higher
- As the funded amount increases the default rate increases
- From Bar chart, The income range 40-60 got higher number of loans, you can also see in heatmap they have the highest default rates for higher amount of loans

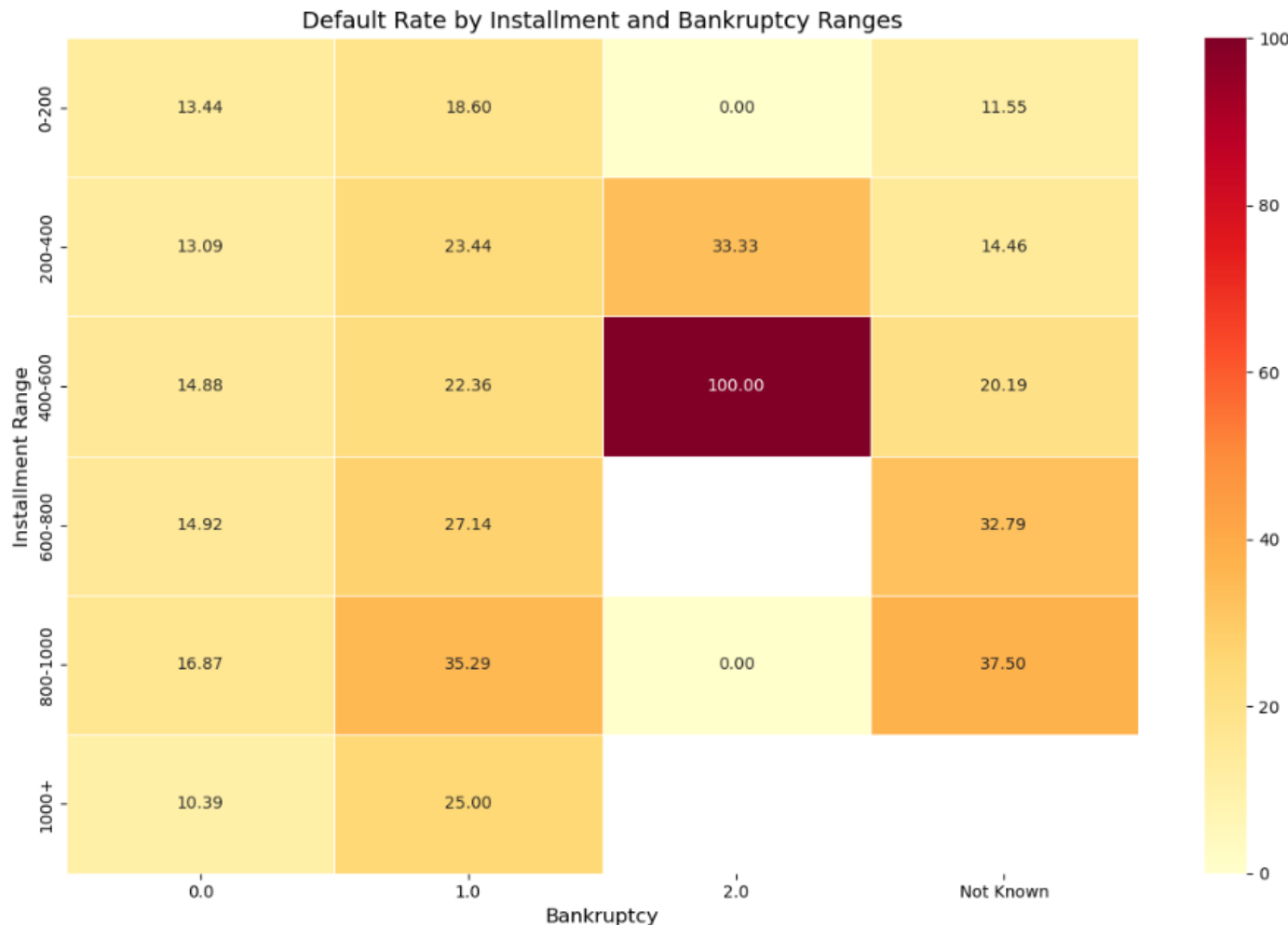
# Multivariate: Default Rate by Interest Range and term

Default Rate by Interest Range and term



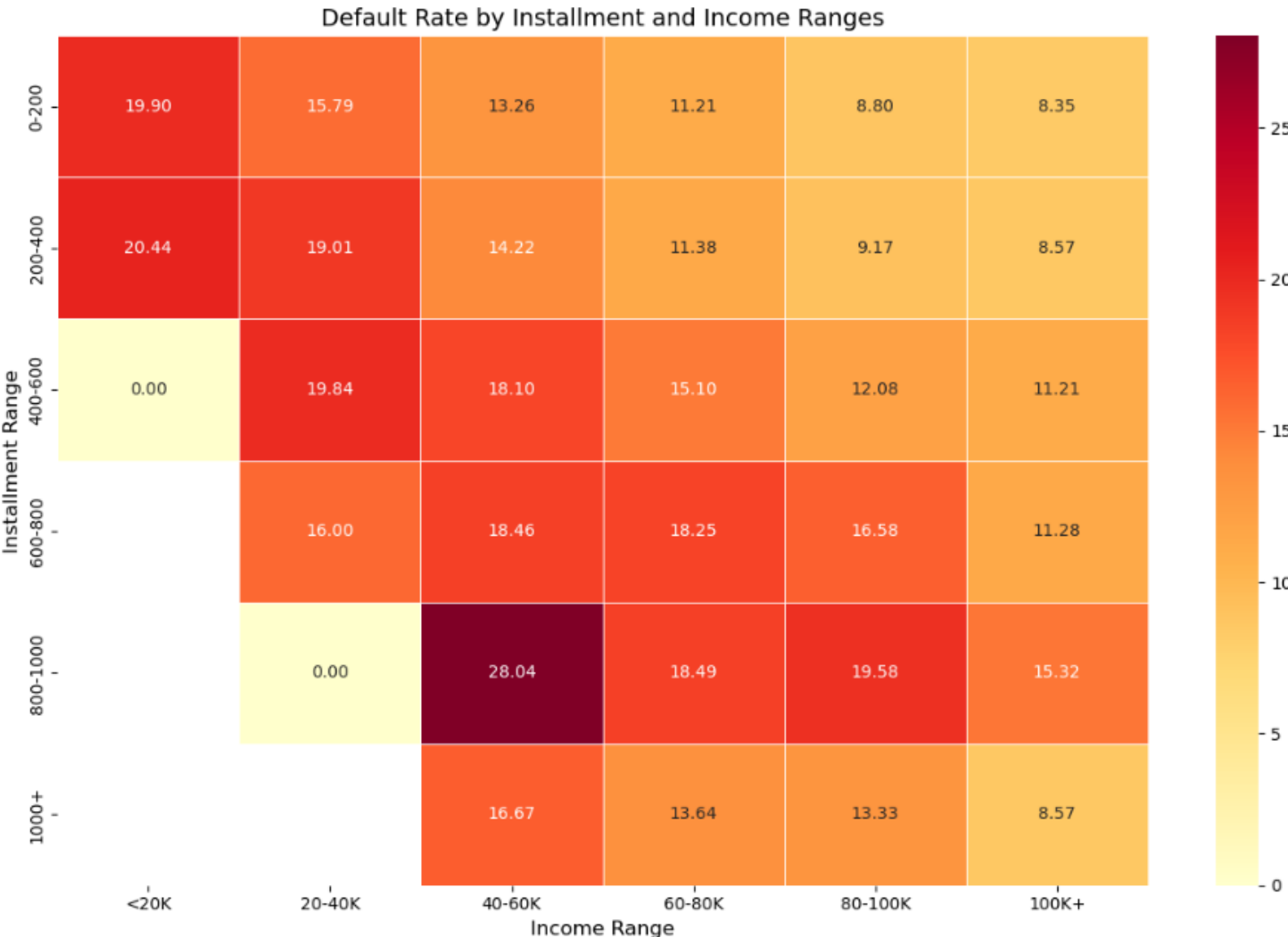
- Higher interest rates has higher default rate
- Lower interest rates has low default rate
- Higher term has higher chance of default

# Multivariate: Default Rate by Installment and Bankruptcy Range



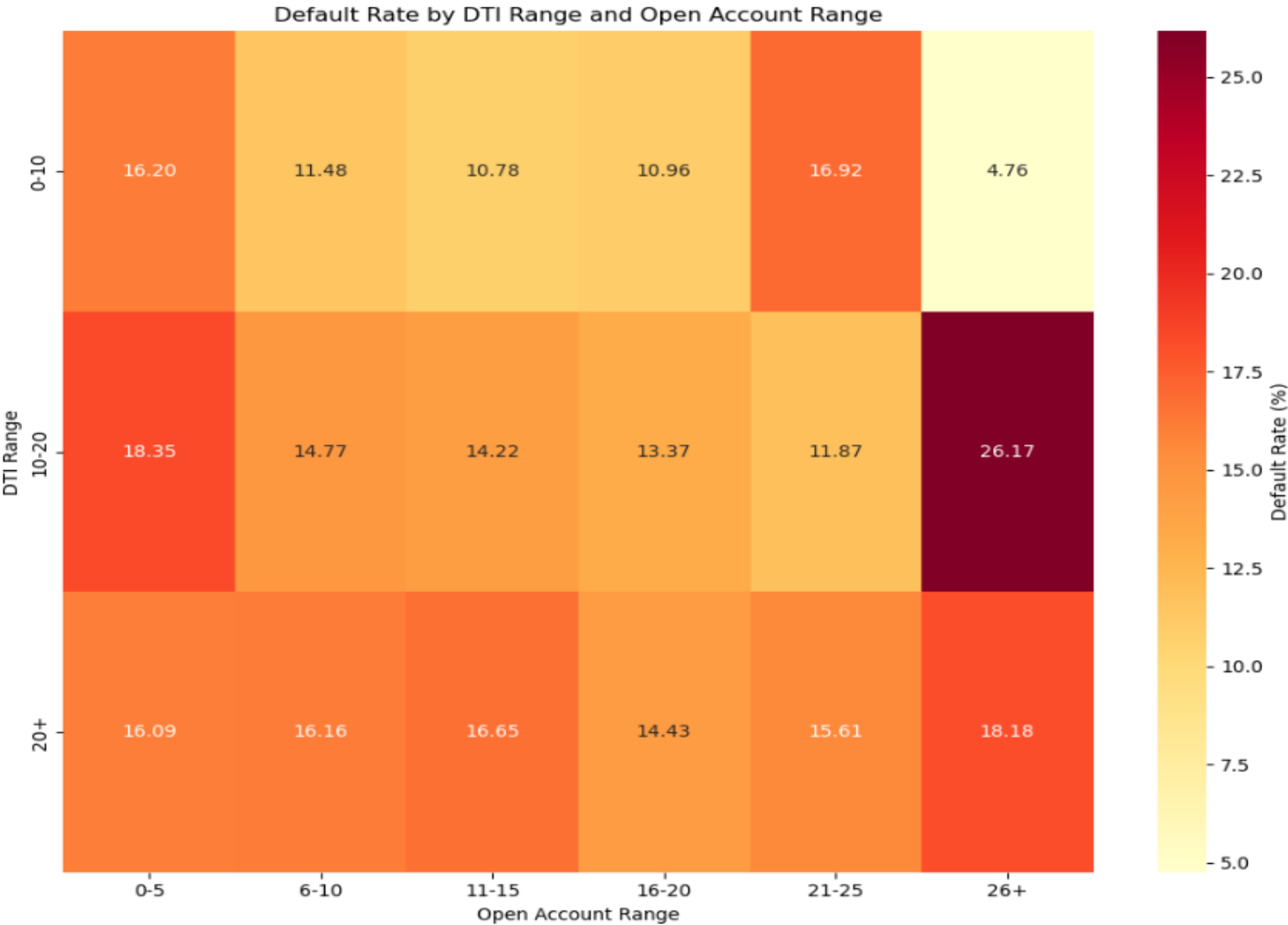
- Borrowers with a history of bankruptcy (especially 1 or more) are significantly more likely to default, especially in mid-to-high installment ranges.
- Higher installments correlate with higher default rates, but the trend flattens or decreases slightly beyond the 1500+ range, possibly due to a smaller, financially stronger borrower population.

# Multivariate: Default Rate by Installment and Income Range



- **Income Stability:** Higher income reduces default risk, even with higher installments.
- **Installment Threshold:** Installments between 800-1000 pose the highest default risk across income ranges, especially for low-to-middle-income groups.
- **Low-Income Borrowers:** Those earning less than 40K have a disproportionately higher risk of default across all installment brackets.

# Multivariate: Default Rate by DTI Range and Open Account Range



- **Higher DTI, Higher Default Rate:** As DTI increases, the default rate tends to rise, indicated by darker shades of red. This suggests that borrowers with higher debt burdens are more likely to default on their loans.
- **Open Accounts and Default Rate:** For lower DTI ranges (0-20), the default rate increases slightly with more open accounts. However, for the highest DTI range (20+), the default rate rises initially, then decreases slightly. This implies that borrowers with high DTI may manage a moderate number of open accounts better, reducing default risk.

# Conclusion

---

The analysis shows that 82.9% of applicants successfully paid their loans, while 14.1% defaulted. Key factors influencing loan defaults include:

1. **Loan Grade:** Loans with lower grades have higher default rates, highlighting the importance of grade in lending decisions.
2. **Income:** Borrowers earning less than \$40K exhibit higher default rates across all loan types. However, higher-income borrowers still default more frequently with larger loan amounts.
3. **Loan Amounts:** Larger funded amounts correlate with higher default rates, especially among low-income borrowers.
4. **DTI Ratio:** Higher DTI ratios are linked to higher default rates, with borrowers in the 20+ DTI range showing a complex pattern where a moderate number of open accounts reduces risk slightly.
5. **Interest Rates and Loan Term:** Higher interest rates and longer loan terms increase the likelihood of default, while lower interest rates and shorter terms reduce it.

These insights suggest that lenders should prioritize loan grade, income stability, and DTI ratio when evaluating default risk, while also considering loan amount and term structure.