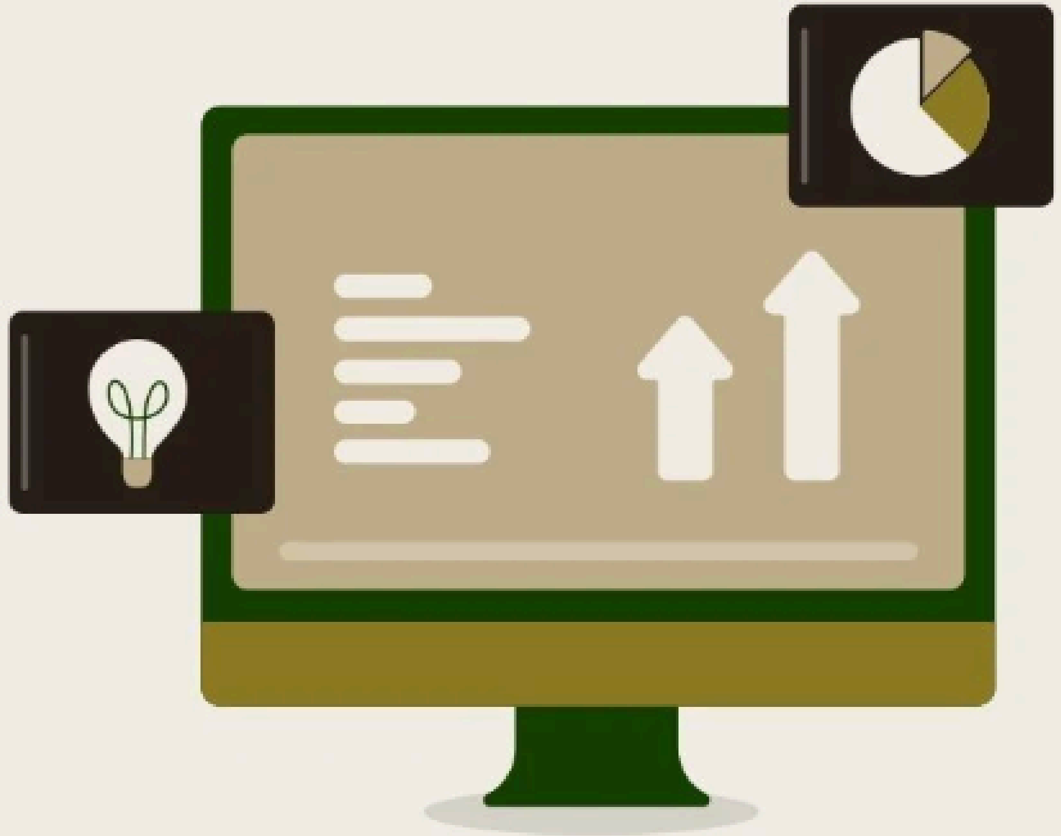
Lending Club Case Study

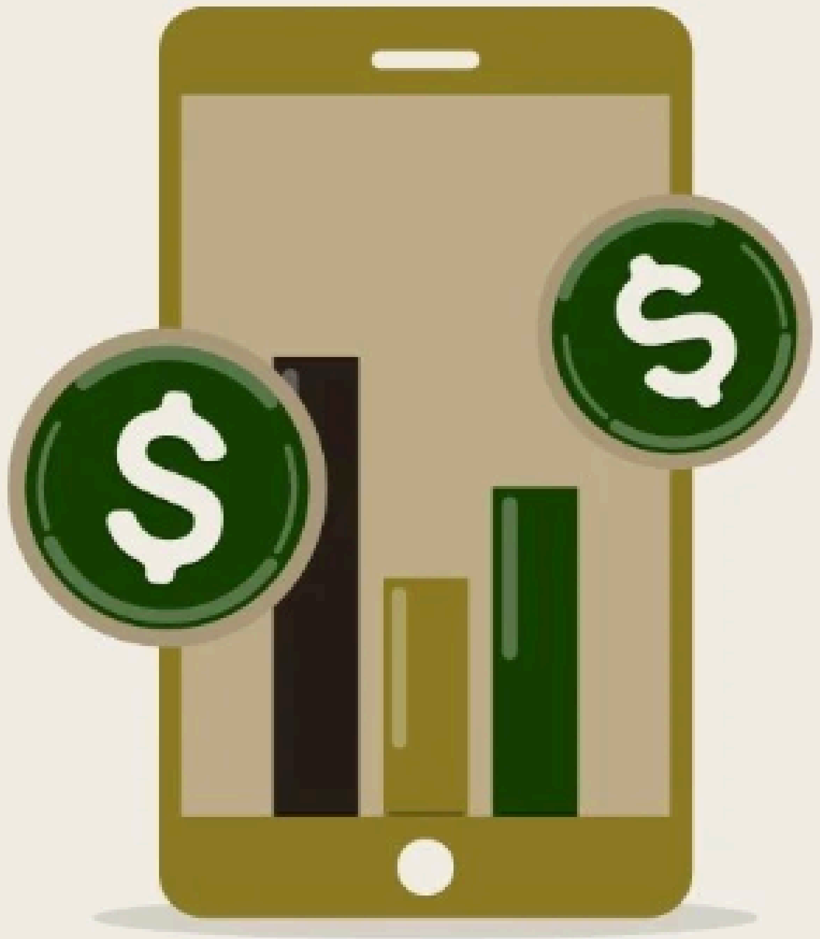
Ashwini Pujar
Anu Augustine

Table of Contents



Problem statement:

- The goal of this project is to identify key patterns and factors that predict the likelihood of a borrower defaulting on a loan.
- By leveraging consumer attributes (such as age, income, and credit history) and loan attributes (such as loan amount, interest rate, and loan term), the goal is to develop predictive insights into the behaviors and conditions that are most indicative of a borrower's propensity to default.
- These insights will be used to inform decision-making processes for financial institutions, enabling actions such as:
    - Rejecting loan applications from individuals with a high probability of default.
    - Offering reduced loan amounts or higher interest rates to mitigate risk.
    - Tailoring loan conditions to better manage the risk associated with specific borrower profiles.



Approach

DATA CLEANING
- DATA DUPLICATES
- NULL VALUE TREATMENT
- FEATURE ANALYSIS
- OUTLIER TREATMENT

DATA TRANSFORMATION
- DERIVED METRICS
- DATA TYPE CASTING
- DATA MANIPULATION

UNIVARIATE ANALYSIS
- BOX PLOTS
- DISTRIBUTION PLOT
- KDE PLOT
- COUNT PLOT

BIVARIATE ANALYSIS
- BAR CHART (HUE)
- BOX PLOT (HUE)
- COUNT PLOT (HUE)

MULTIVARIATE ANALYSIS
- HEAT MAP
- PAIR PLOT
- SWARM PLOT

BUSINESS CONCLUSION
- DATA INFERENCE
- RECOMMENDATION

Data Understanding

Dataset Overview

The dataset covers loans issued from 2007 to 2011, providing information on borrower attributes and loan characteristics for risk analysis.

Loan and Applicant Attribute

Attributes include loan amount, term, interest rate, applicant's income, employment details, and credit history, vital for risk assessment.

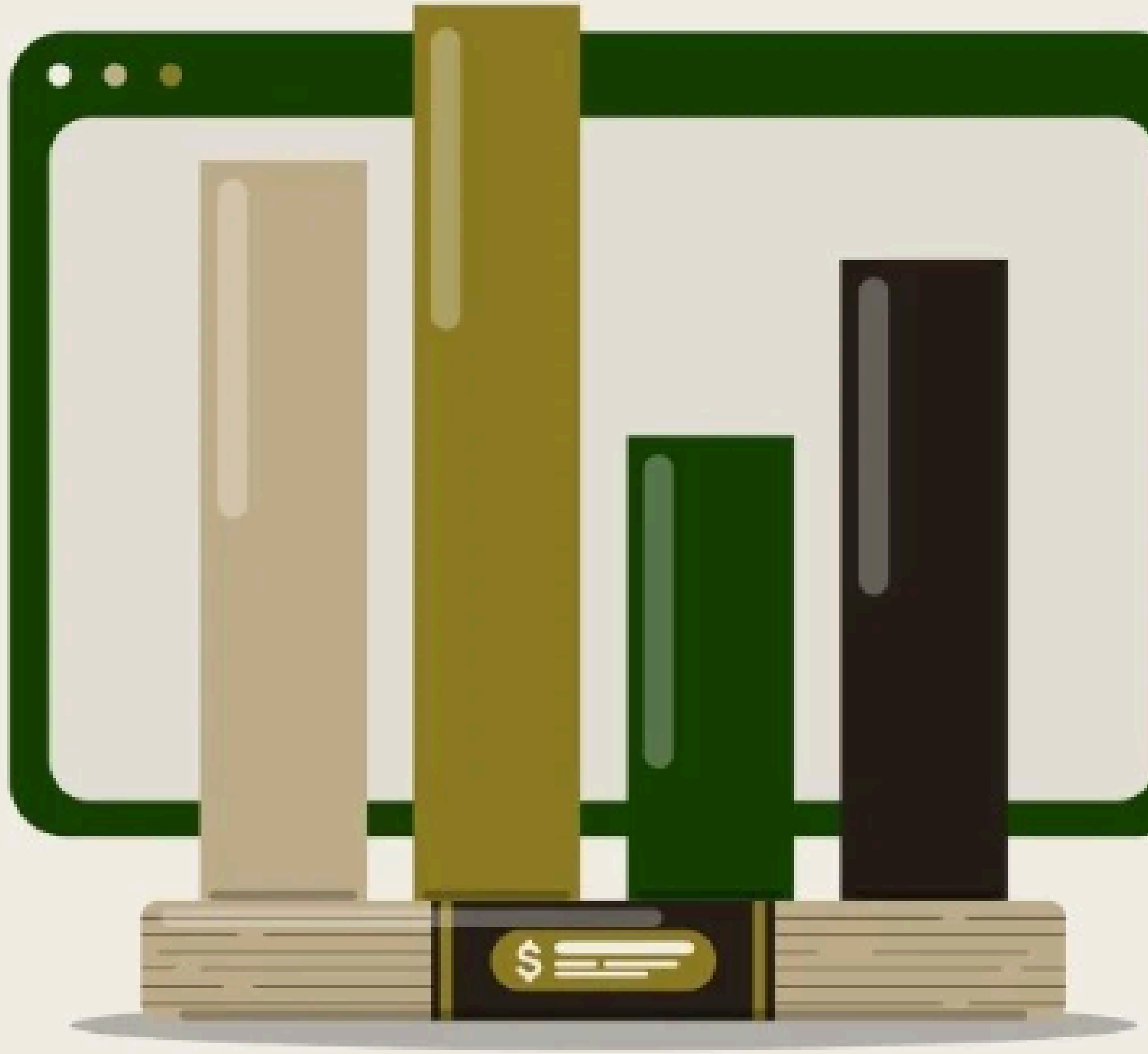
Significance of Risk Identification

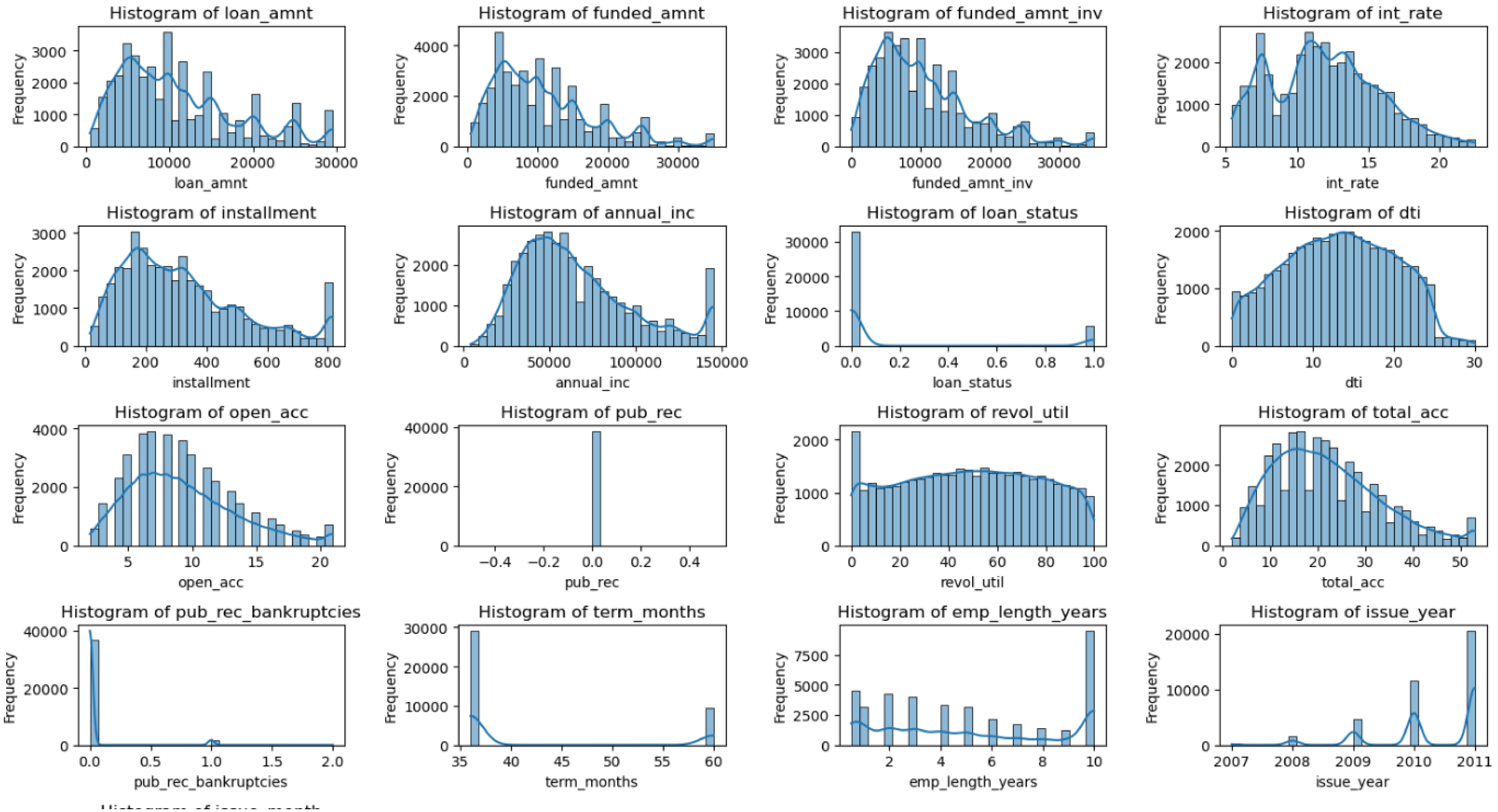
Ensures smart loan decisions, reducing defaults and financial losses, pivotal in enhancing overall portfolio performance.

Data Cleaning Steps

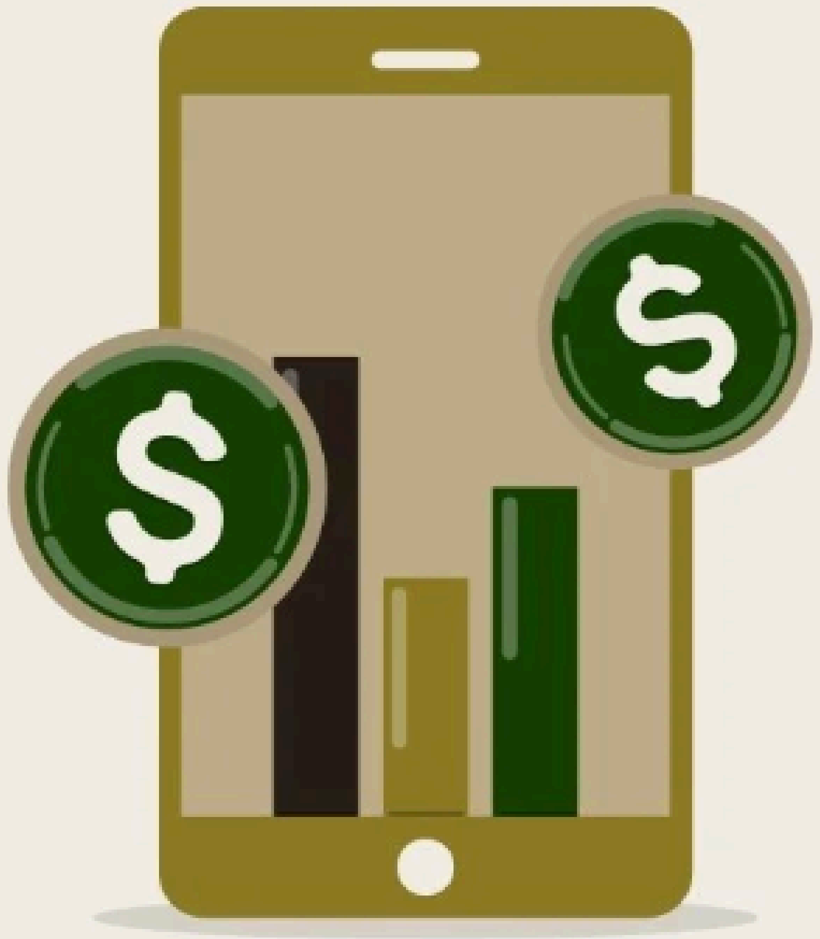
- Checked for empty columns and rows, deleted empty columns
- Deleted irrelevant columns
- Treated missing values with appropriate approach
- Manipulated strings and dates
- Data is converted to a suitable and convenient format
- Removed outliers

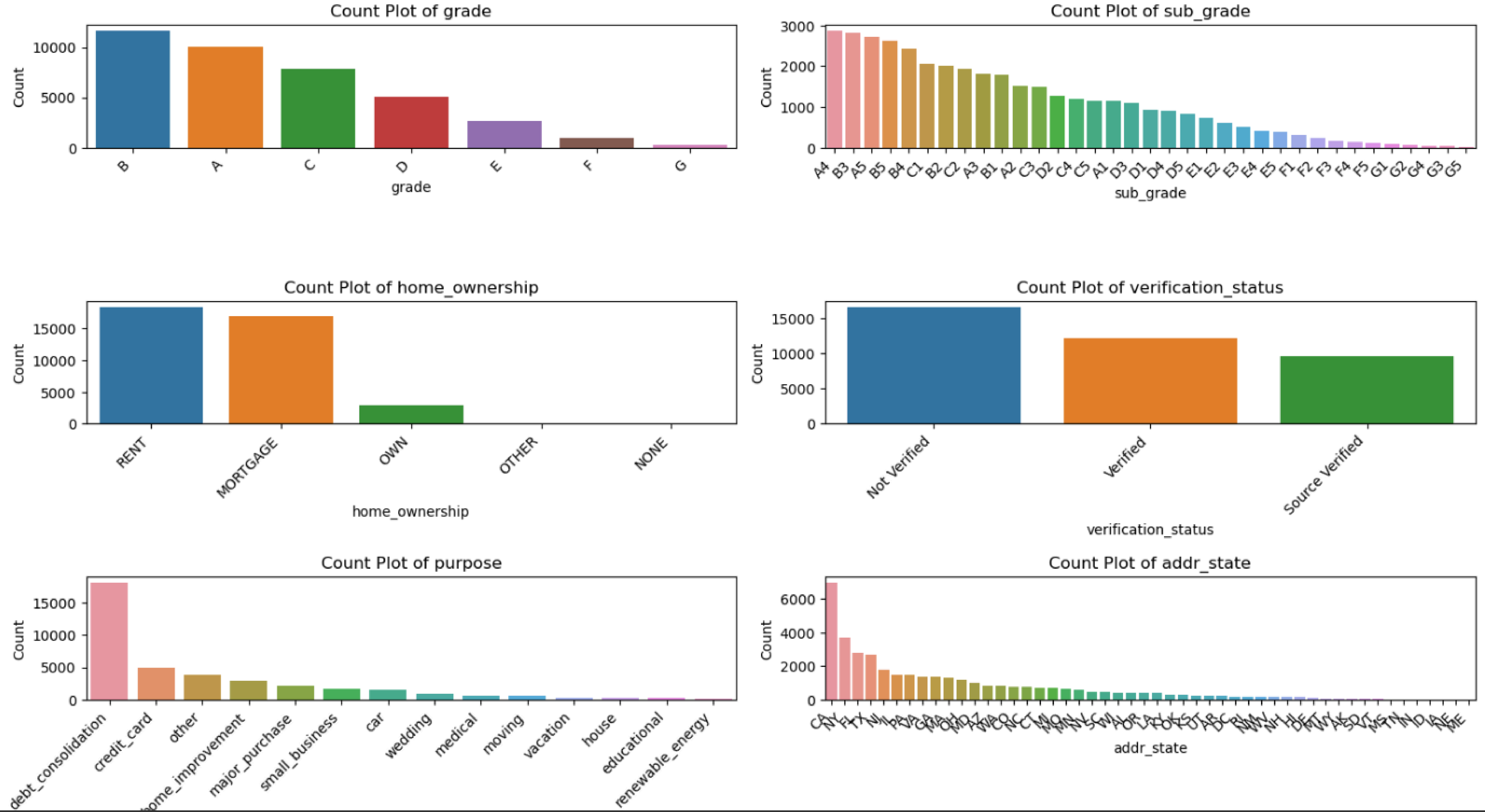
Univariate Analysis




Histogram of loan_amnt
Frequency
loan_amnt
Histogram of funded_amnt
Frequency
funded_amnt
Histogram of funded_amnt_inv
Frequency
funded_amnt_inv
Histogram of int_rate
Frequency
int_rate
Histogram of installment
Frequency
installment
Histogram of annual_inc
Frequency
annual_inc
Histogram of loan_status
Frequency
loan_status
Histogram of dti
Frequency
dti
Histogram of open_acc
Frequency
open_acc
Histogram of pub_rec
Frequency
pub_rec
Histogram of revol_util
Frequency
revol_util
Histogram of total_acc
Frequency
total_acc
Histogram of pub_rec_bankruptcies
Frequency
pub_rec_bankruptcies
Histogram of term_months
Frequency
term_months
Histogram of emp_length_years
Frequency
emp_length_years
Histogram of issue_year
Frequency
issue_year

- Loan Amounts: Most loans are between $10,000 and $15,000. Larger loans (over $20,000) pose higher repayment risk.
- Interest Rates: Most loans have interest rates of 10-15%. Higher rates (above 14%) indicate increased default risk.
- DTI: Borrowers with a DTI over 30% are more likely to default.
- Public Records: Bankruptcies increase default risk.
- Credit Utilization: Higher utilization (near 100%) indicates financial stress and higher default risk.
- Employment Length: Shorter job histories (0-5 years) suggest higher default risk.
- Loan Term: 60-month loans carry higher default risk than 36-month loans.
- Income: Lower incomes (below $50,000) increase default risk, especially with high loan amounts or DTI.



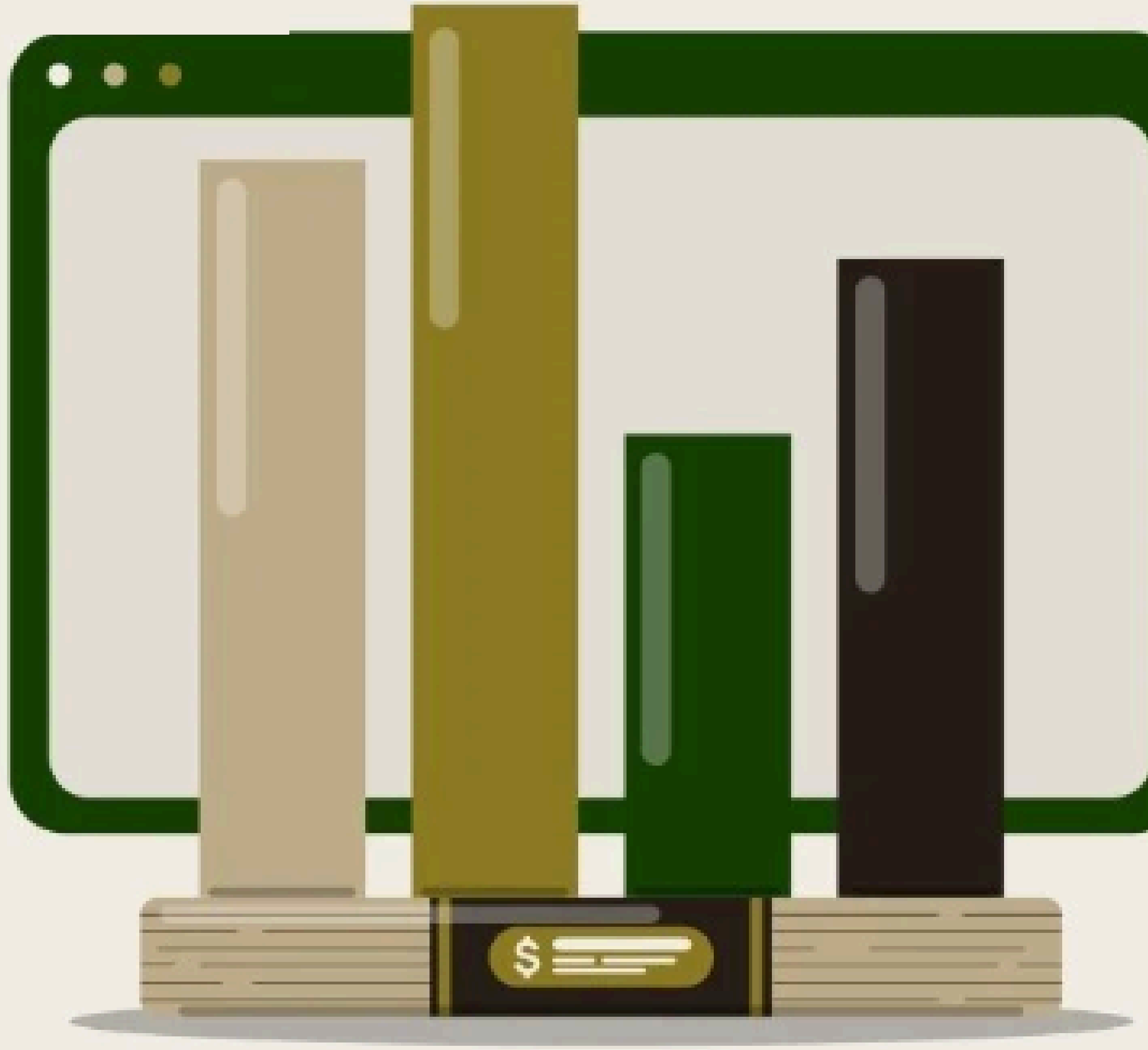


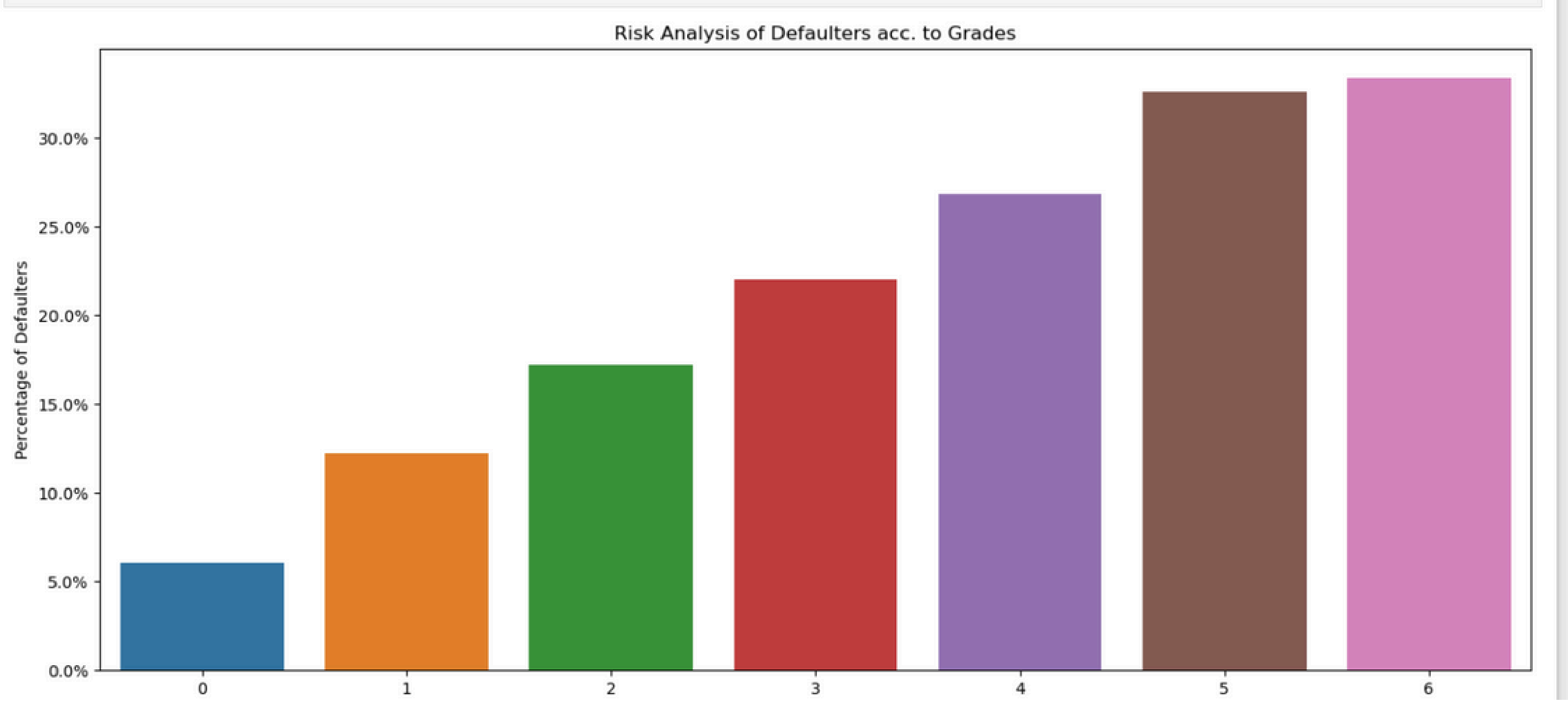
Count Plot of grade
Count
0
5000
10000
B
A
C
D
E
F
G
grade
Count Plot of sub_grade
Count
0
1000
2000
3000
A4 B3 A5 B5 B4 C1 B2 C2 A3 B1 A2 C3 D2 C4 C5 A1 D3 D1 D4 D5 E1 E2 E3 E4 E5 F1 F2 F3 F4 F5 G1 G2 G4 G3 G5
sub_grade
Count Plot of home_ownership
Count
0
5000
10000
15000
RENT
MORTGAGE
OWN
OTHER
NONE
home_ownership
Count Plot of verification_status
Count
0
5000
10000
15000
Not Verified
Verified
Source Verified
verification_status
Count Plot of purpose
Count
0
5000
10000
15000
debt_consolidation
credit_card
other
home_improvement
major_purchase
small_business
car
wedding
medical
moving
vacation
house
educational
renewable_energy
Count Plot of addr_state
Count
0
2000
4000
6000
addr_state

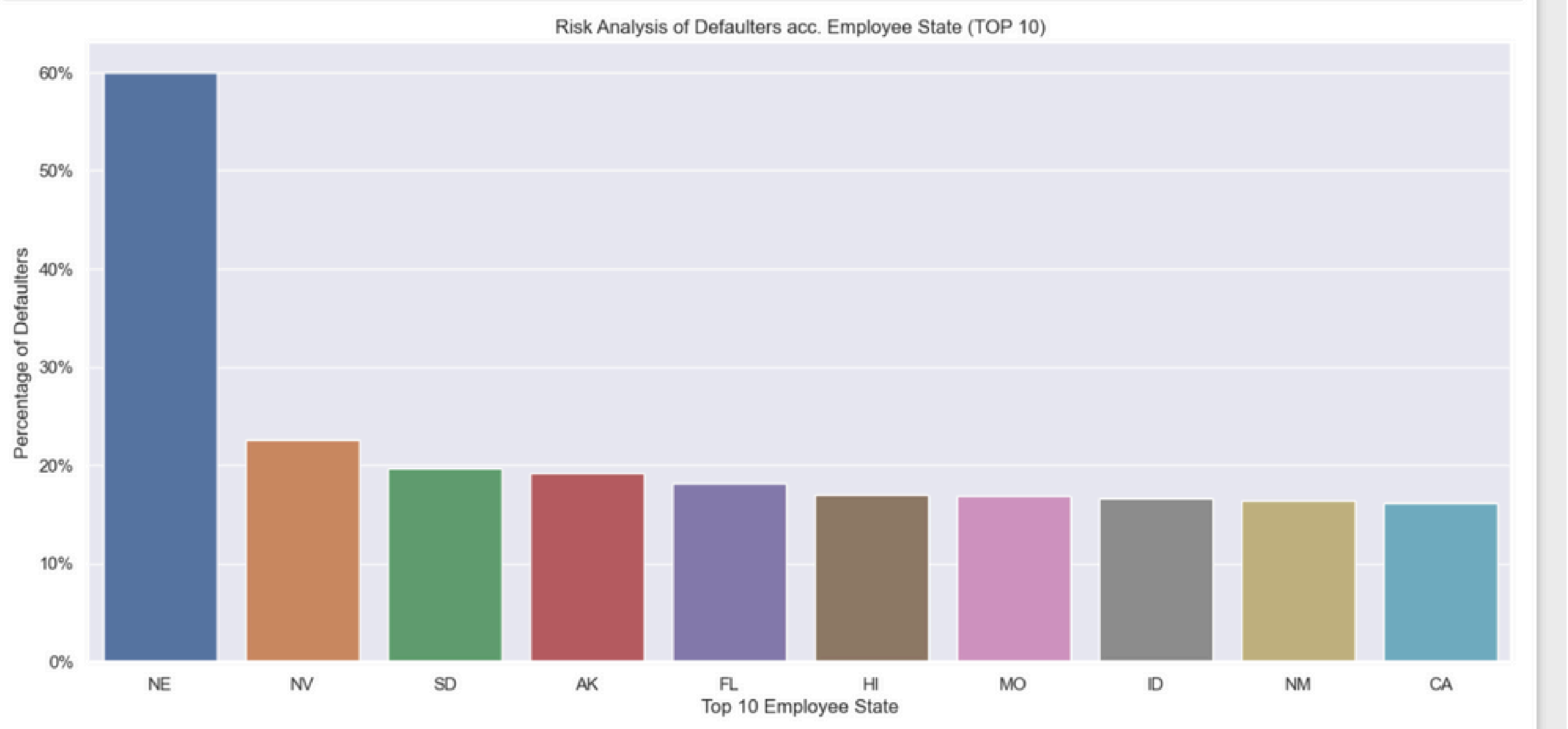
Inferences

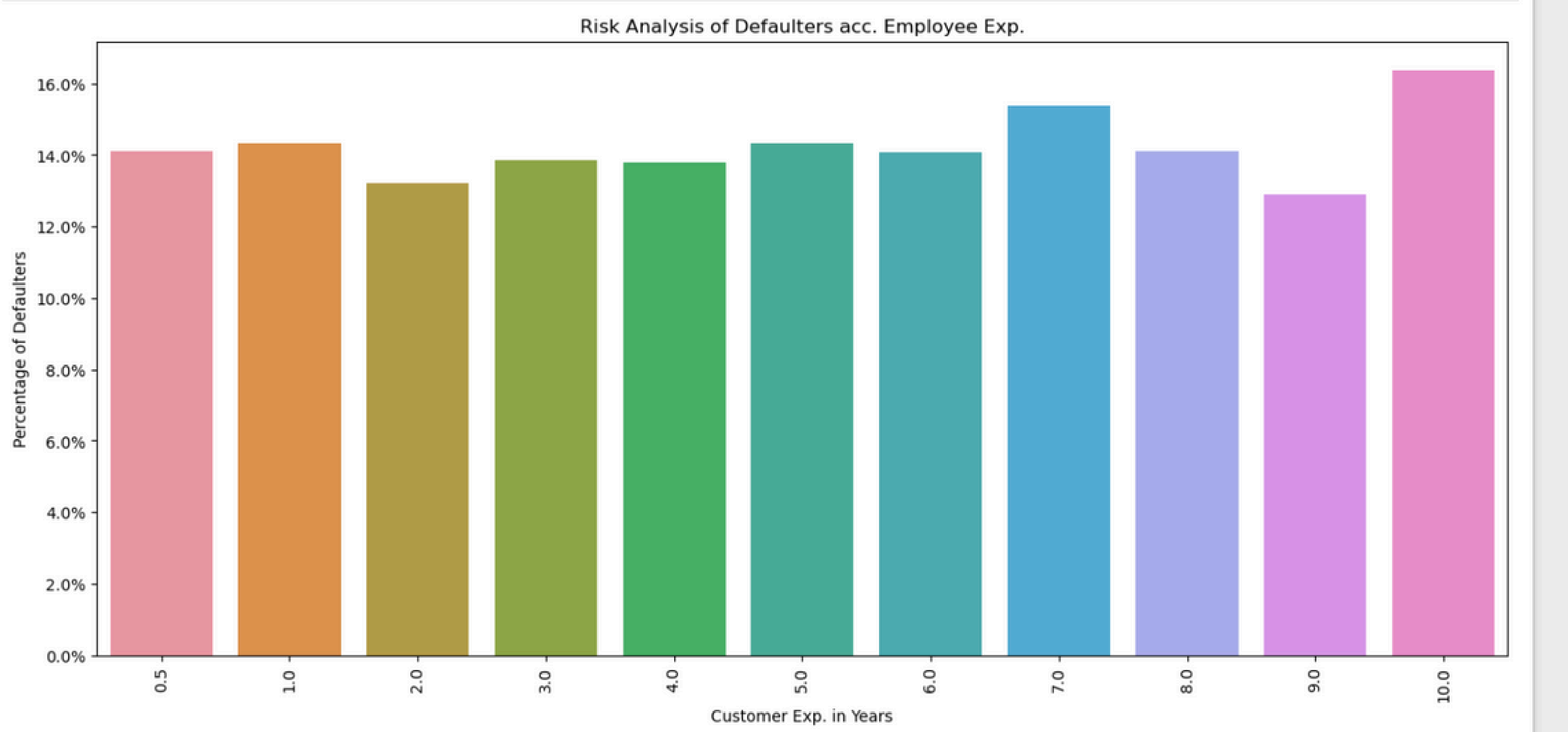
- **Credit Grade and Sub-Grade**: Borrowers with lower grades (C, D, F) are higher risk, potentially needing stricter loan terms or rejection.
- **Homeownership**: Renters may present more risk than homeowners, suggesting the need for tailored conditions.
- **Loan Purpose**: Debt consolidation loans, being most common, may involve higher risk.
- **Verification Status**: Unverified borrowers carry higher risk.
- **Geographic Factors**: Regional economic conditions can influence default rates.

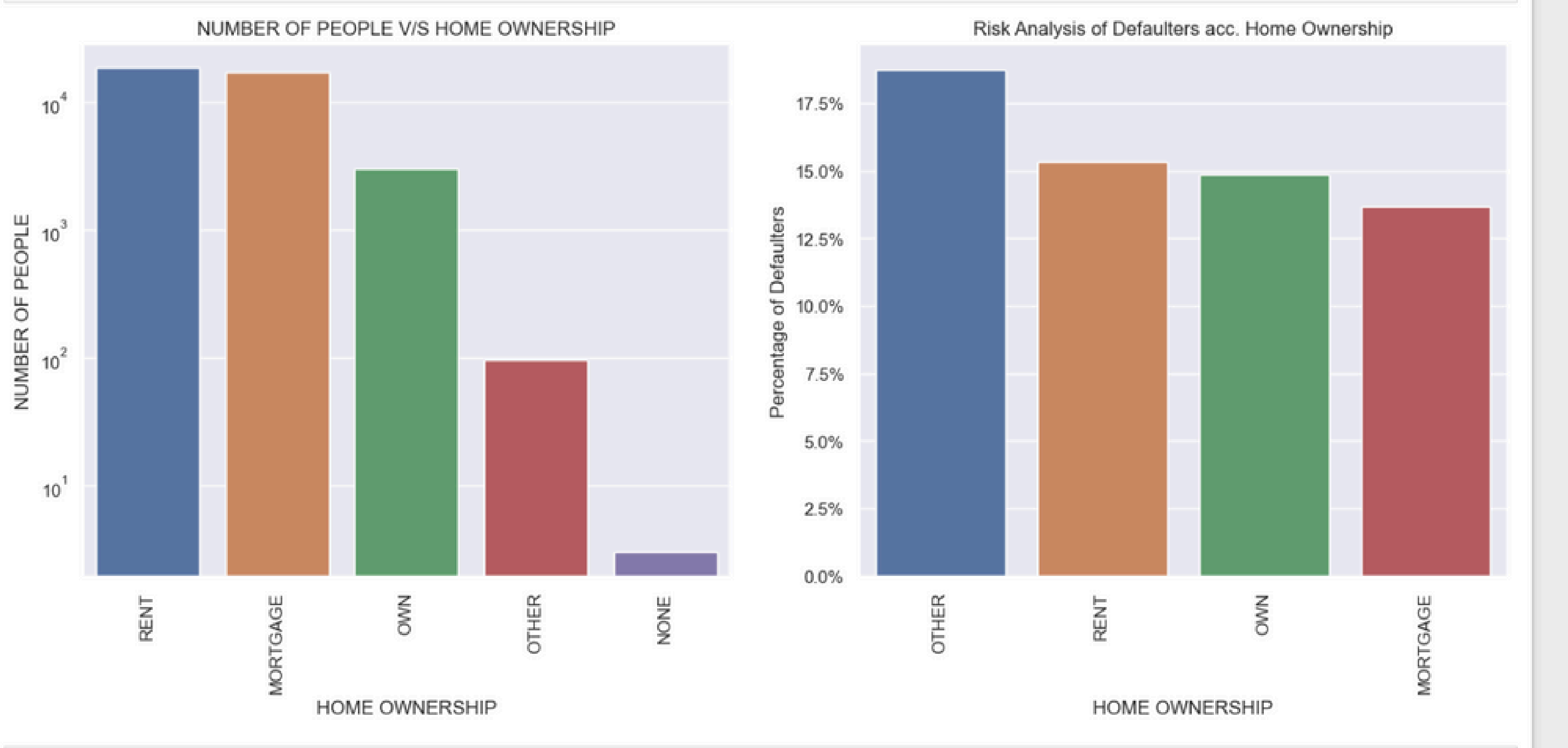
Bivariate Analysis

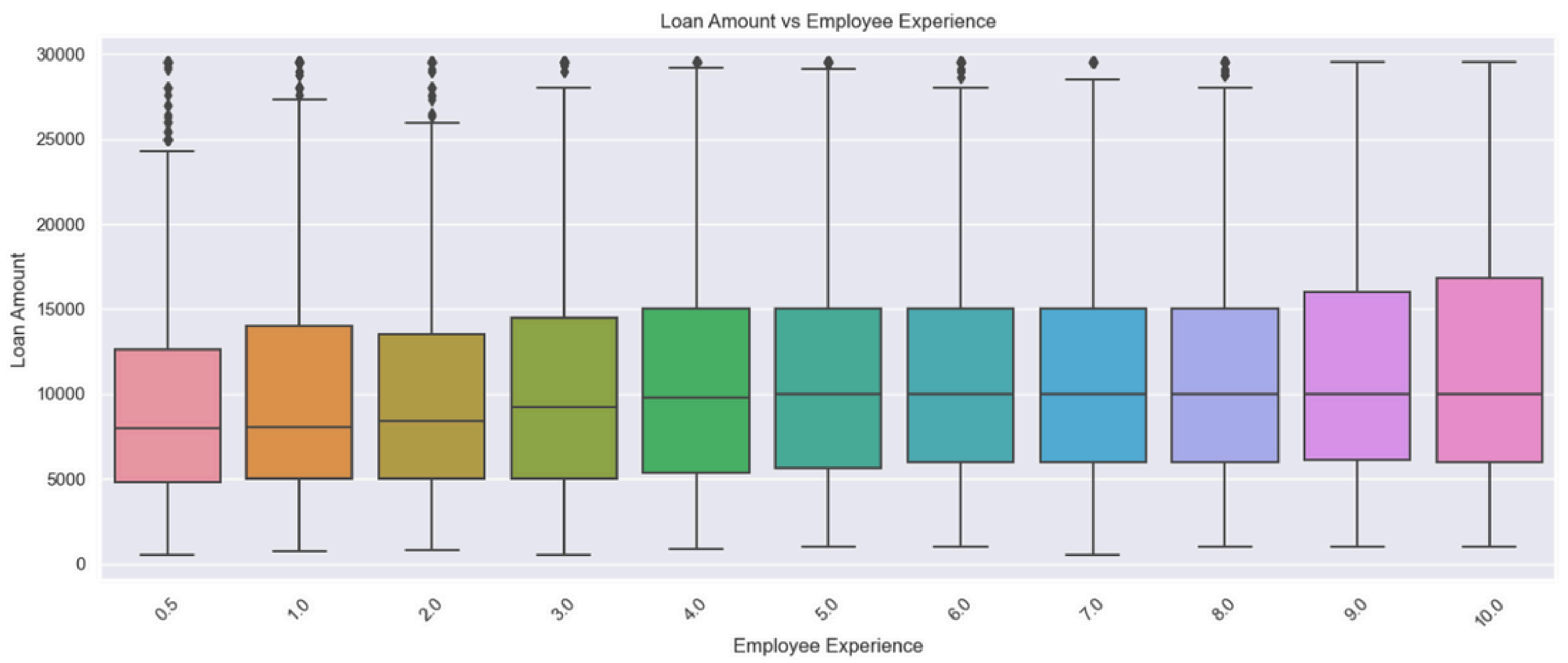


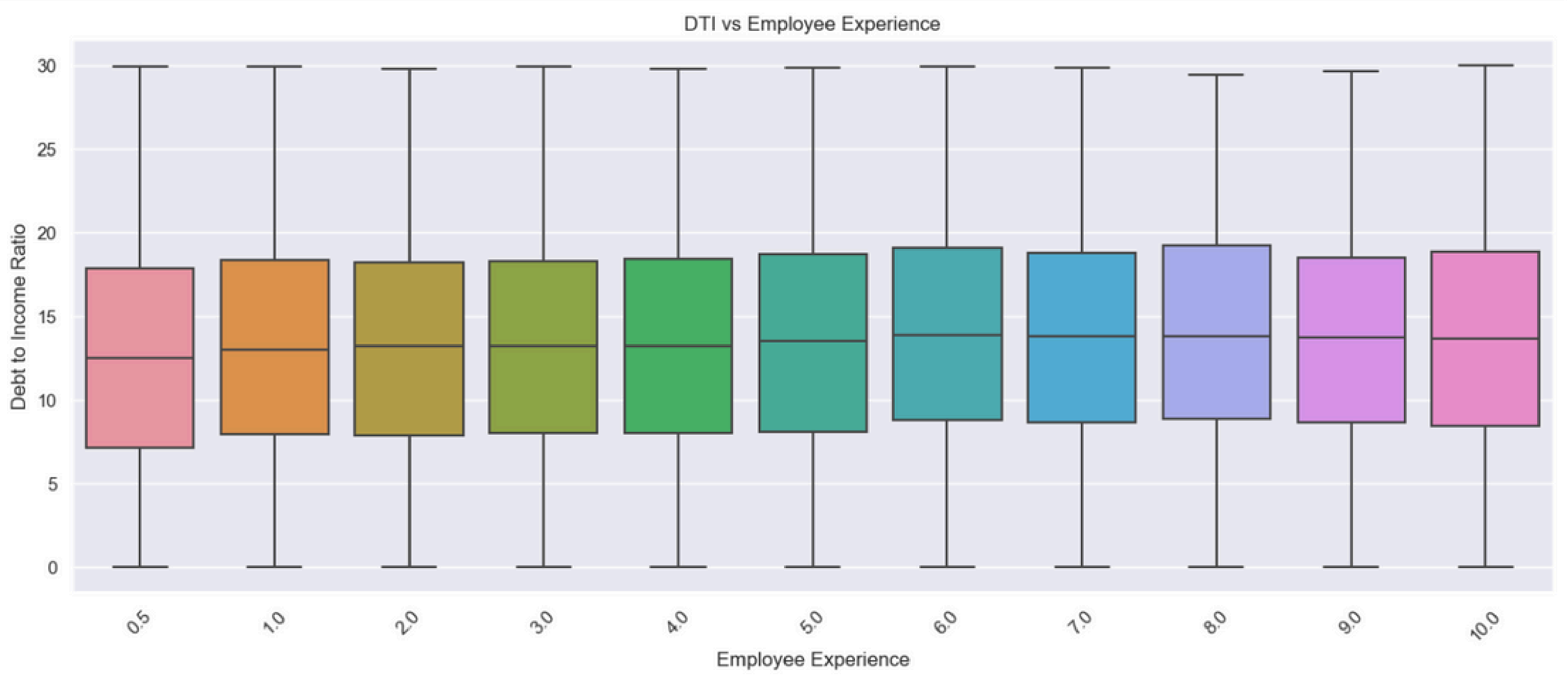

Risk Analysis of Defaulters acc. to Grades
Percentage of Defaulters
0.0%
5.0%
10.0%
15.0%
20.0%
25.0%
30.0%
0
1
2
3
4
5
6

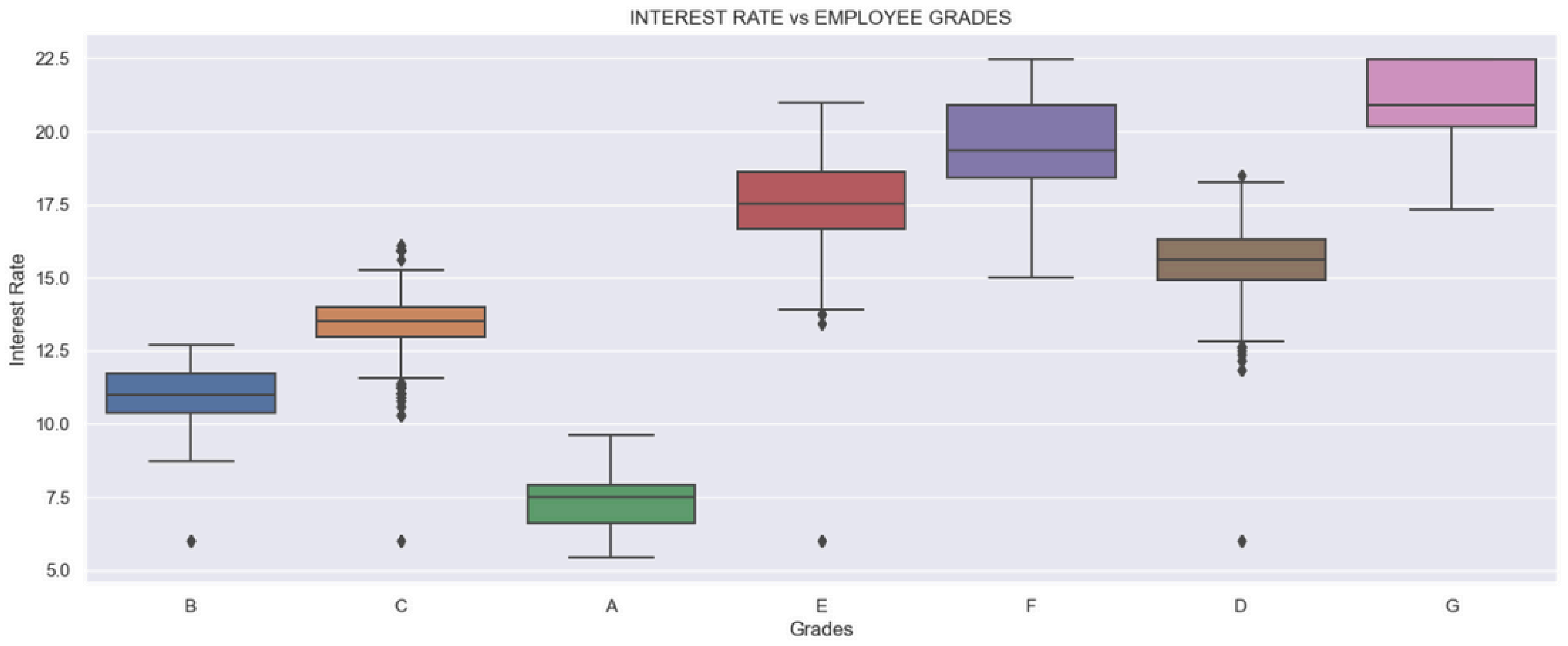

Risk Analysis of Defaulters acc. Employee State (TOP 10)
Percentage of Defaulters
0%
10%
20%
30%
40%
50%
60%
NE
NV
SD
AK
FL
HI
MO
ID
NM
CA
Top 10 Employee State

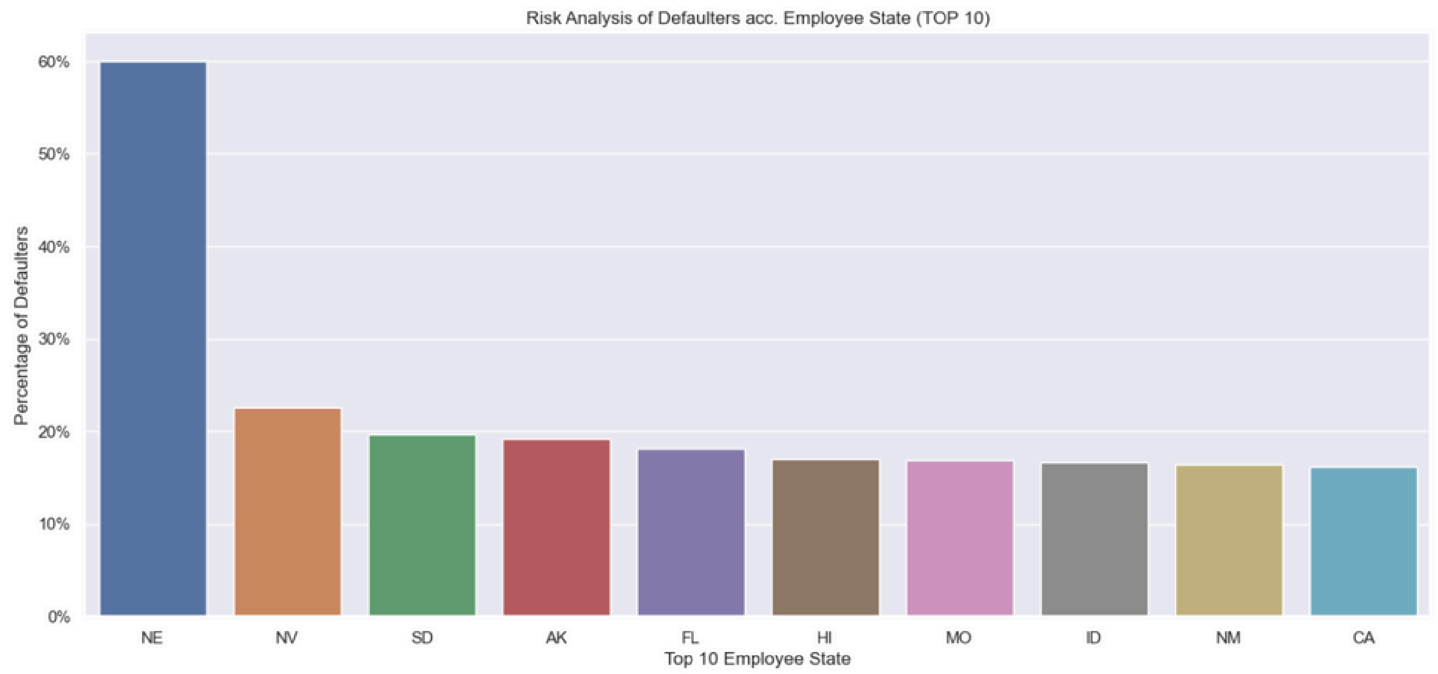

Risk Analysis of Defaulters acc. Employee Exp.
Percentage of Defaulters
0.0%
2.0%
4.0%
6.0%
8.0%
10.0%
12.0%
14.0%
16.0%
0.5
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
Customer Exp. in Years

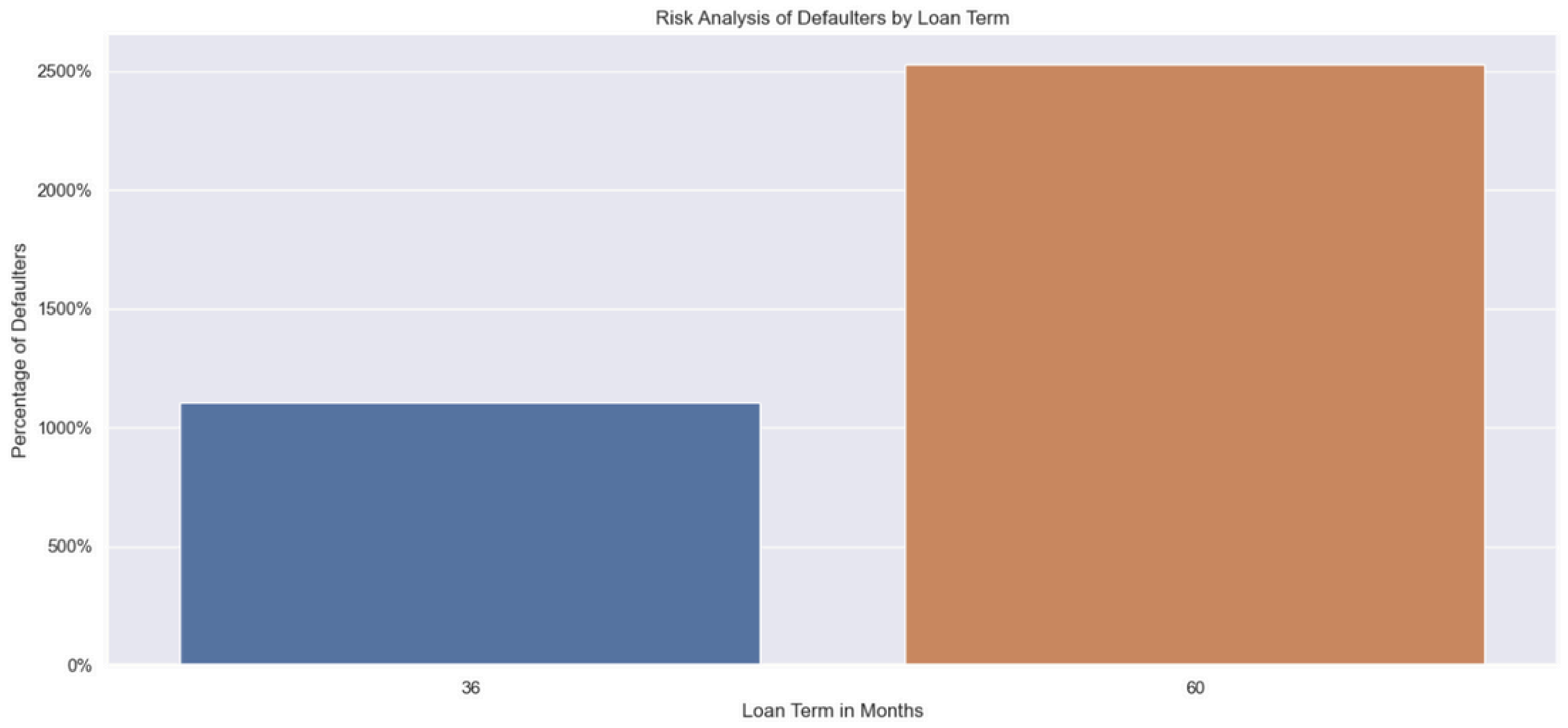

NUMBER OF PEOPLE V/S HOME OWNERSHIP
NUMBER OF PEOPLE
10^1
10^2
10^3
10^4
RENT
MORTGAGE
OWN
OTHER
NONE
HOME OWNERSHIP
Risk Analysis of Defaulters acc. Home Ownership
Percentage of Defaulters
0.0%
2.5%
5.0%
7.5%
10.0%
12.5%
15.0%
17.5%
OTHER
RENT
OWN
MORTGAGE
HOME OWNERSHIP

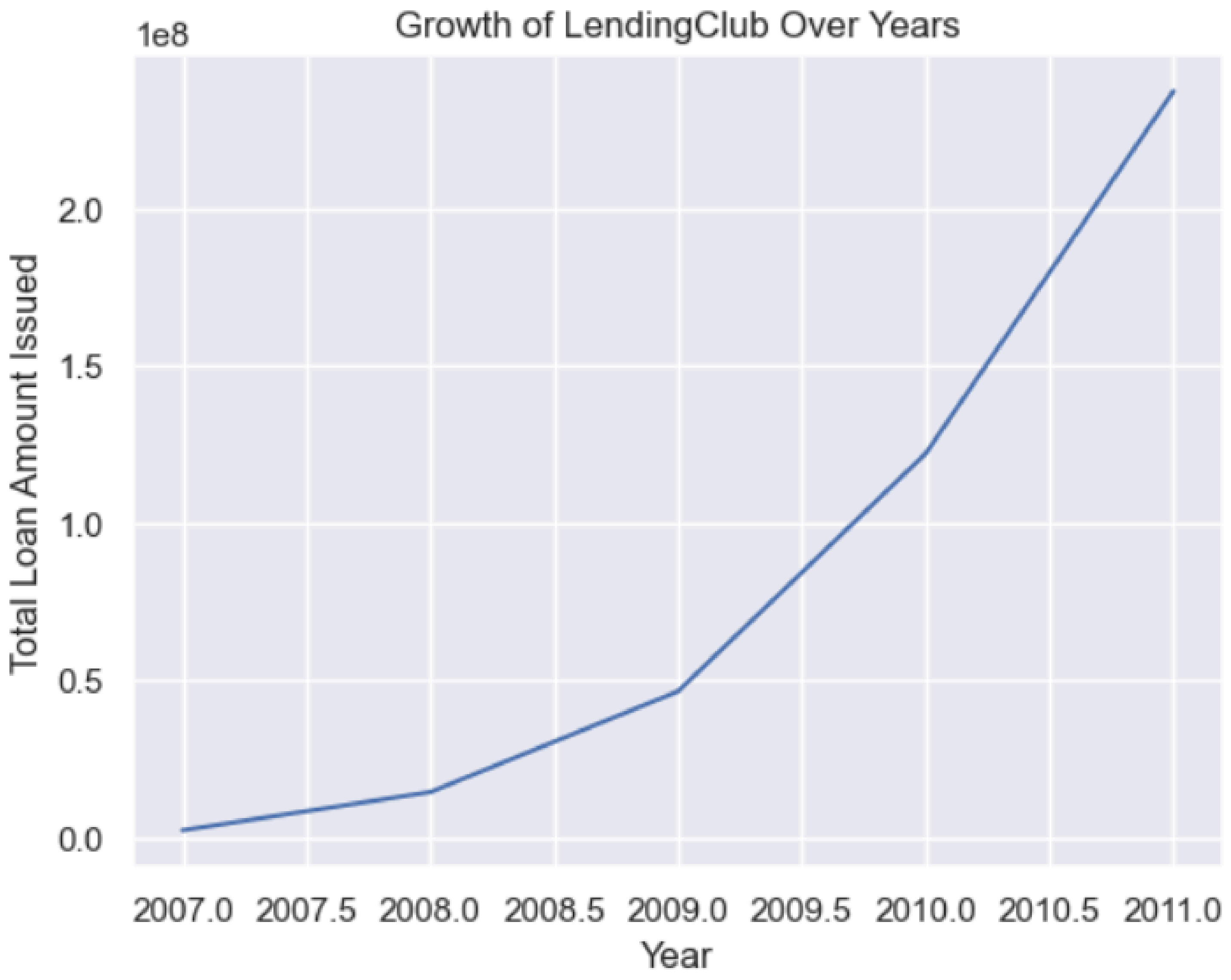

Loan Amount vs Employee Experience
Loan Amount
30000
25000
20000
15000
10000
5000
0
0.5
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
Employee Experience

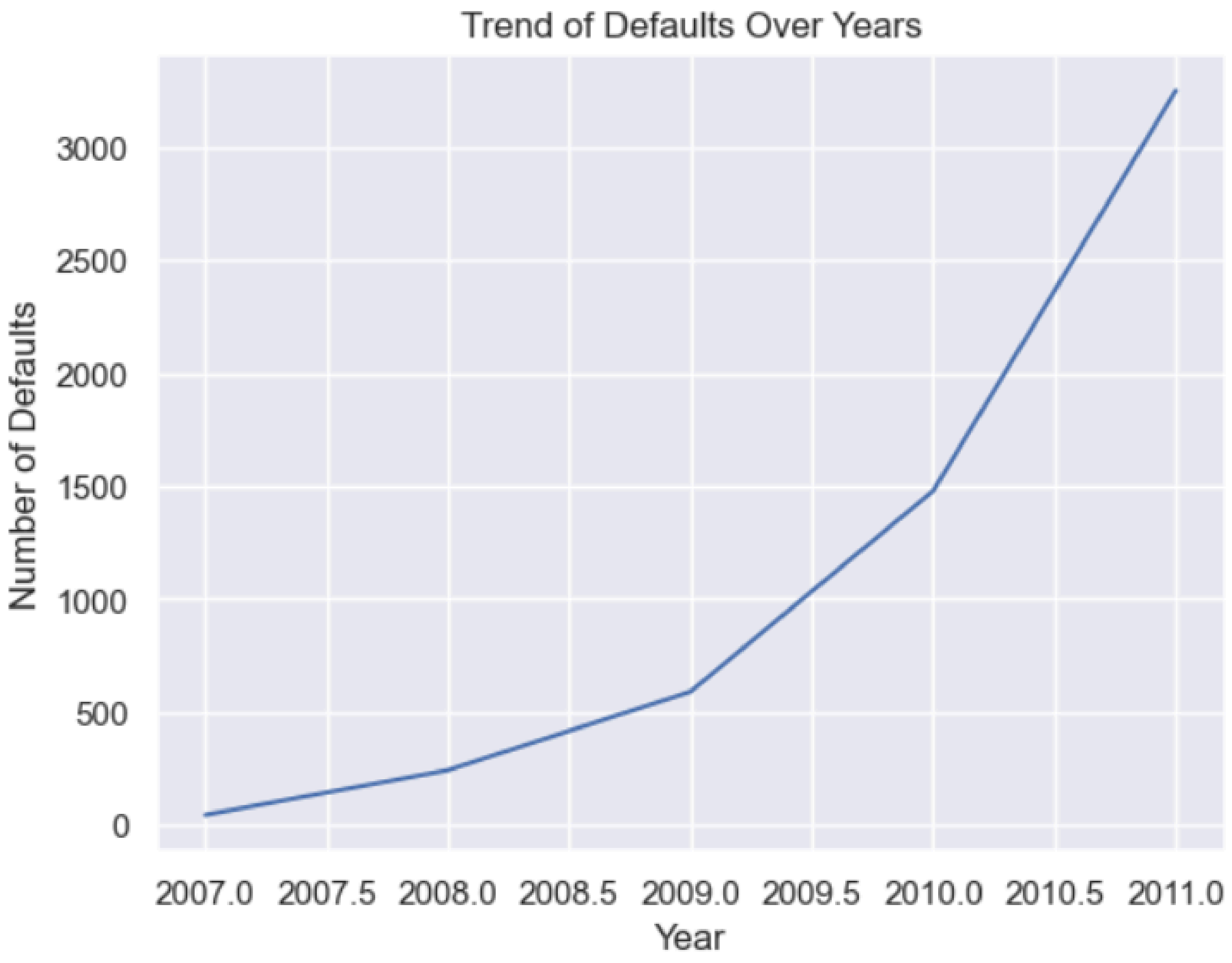

DTI vs Employee Experience
Debt to Income Ratio
30
25
20
15
10
5
0
0.5
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
Employee Experience


INTEREST RATE vs EMPLOYEE GRADES
Interest Rate
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
B
C
A
E
F
D
G
Grades


Risk Analysis of Defaulters acc. Employee State (TOP 10)
Percentage of Defaulters
60%
50%
40%
30%
20%
10%
0%
NE
NV
SD
AK
FL
HI
MO
ID
NM
CA
Top 10 Employee State


Risk Analysis of Defaulters by Loan Term
Percentage of Defaulters
2500%
2000%
1500%
1000%
500%
0%
36
60
Loan Term in Months


Growth of LendingClub Over Years
1e8
Total Loan Amount Issued
2.0
1.5
1.0
0.5
0.0
2007.0 2007.5 2008.0 2008.5 2009.0 2009.5 2010.0 2010.5 2011.0
Year


Trend of Defaults Over Years
Number of Defaults
3000
2500
2000
1500
1000
500
0
2007.0 2007.5 2008.0 2008.5 2009.0 2009.5 2010.0 2010.5 2011.0
Year

[ ]:

Inferences

- Higher Default for Lower Grades: Borrowers with lower grades (e.g., G) have higher default rates, requiring stricter approval criteria.
- Loan Purpose: Small business (27.10%) and renewable energy (18.63%) loans have higher default rates, while home improvement (11.99%) and vacation (14.17%) loans show lower risk.
- Employment Length: No clear pattern between years of experience and default risk, though employees with 10 years show the highest defaults (16%).
- Home Ownership: Renters have higher default risk compared to mortgage holders.
- DTI and Experience: Less experienced employees with high DTI and loan amounts are more likely to default.
- State Risk: NE, NV, SD, and AK have the highest default rates.
- Loan Term: 60-month loans have a higher default risk than 36-month loans.

Conclusions:

These insights support improved risk assessment and lending strategies:

- Enhanced Risk Models: Adjust criteria based on loan size, interest rate, DTI, and credit grade for better borrower evaluation.
- Targeted Interventions: Focus on high-risk borrowers (e.g., high DTI) with tailored loan products or education to reduce defaults.
- Strategic Adjustments: Use regional default data for targeted marketing and outreach.
- Policy Development: Inform regulations to encourage responsible lending and protect vulnerable borrowers.

Thank You