# Lending Club Case Study

SUBMITTED BY:

VEENUGOPAL T

LAVANYA KUMAR V

# Problem Statement and Objective

Lending Club is a consumer finance company that specialises in providing various types of loans to urban customers. One of the critical challenges we face is making accurate loan approval decisions.

The challenges are two ways:

- Loss of Business: If they reject a loan application from an applicant who is likely to repay, they miss out on potential business opportunities. This not only affects our revenue but also their market competitiveness.
- Financial Loss: On the other hand, if they approve a loan for an applicant who is likely to default, they risk incurring massive financial losses. Defaults can lead to a loss of the principal amount and the interest, which affects their overall financial health.

The main objective is to identify patterns that indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. by using the (Exploratory Data Analysis )EDA on the dataset provided.

# Analysis Approach

1. Understanding the Dataset.
2. Data cleaning.
3. Data Analysis
    I. Derived Variables
    II. Univariate Analysis
    III. Multivariate analysis
    IV. Correlation Analysis
4. Conclusion
5. Recommendations

# Understanding the Dataset

The data set contains information about past loan applicants and whether they ‘defaulted’ or not. The aim is to identify patterns that indicate| if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of the loan, lending (to risky applicants) at a higher interest rate, etc.

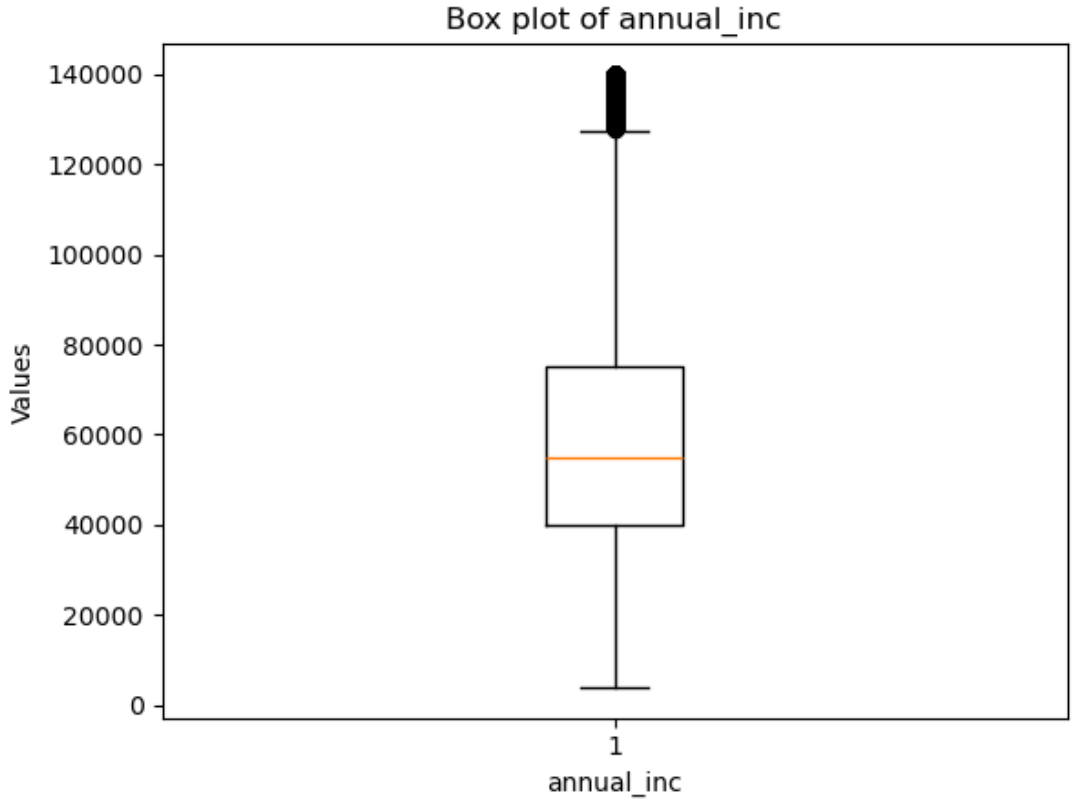
When a person applies for a loan, two types of decisions could be taken by the company:

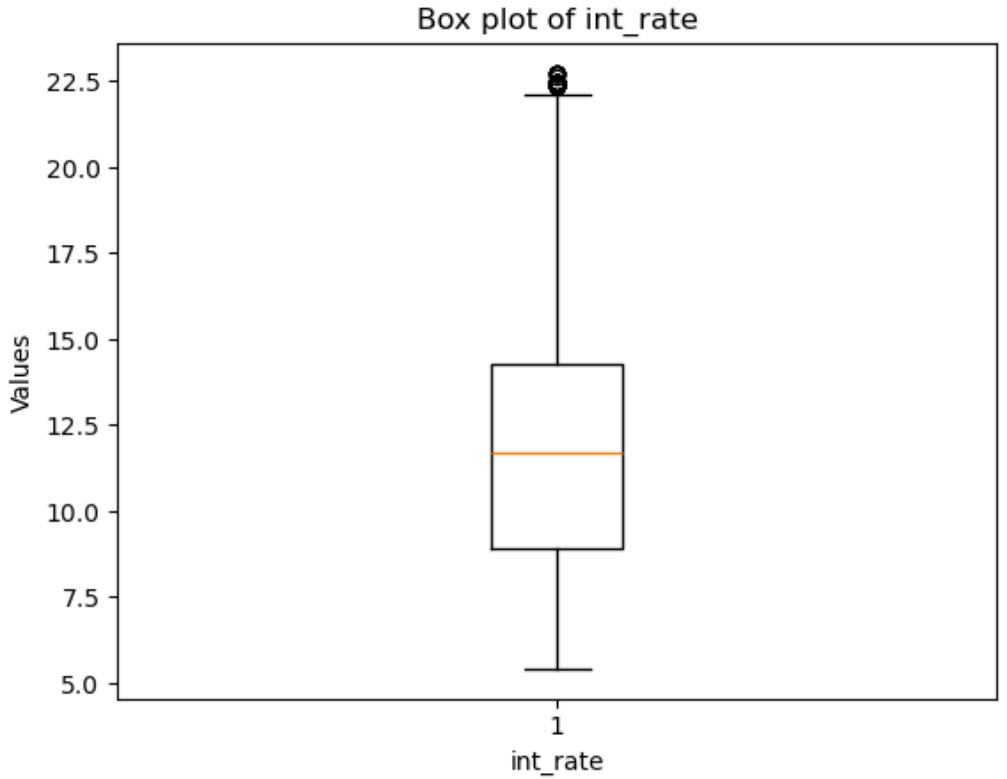
1. Loan Accepted: If the company approves the loan, there are three possible scenarios:
    1. Fully Paid: The applicant has fully paid the loan (the principal and the interest rate).
    2. Current: The applicant is in the process of paying the instalments, i.e., the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted’.
    3. Charged-off: The applicant has not paid the instalments in due time for a long period, i.e., he/she has defaulted on the loan.
2. Loan Rejected: The company rejects the loan because the candidate does not meet their requirements. Since the loan was rejected, there is no transactional history of those applicants with the company, and thus this data is not available in the dataset.

# Data Cleaning

- Dropped 54 columns  from the dataset, which are filled with null values as they don`t add any value to the analysis
- Dropped the columns which are considerable number of null values.
- Deleted the rows having value “current” for the Column - loan_status.
- Stripped the string 'months' from the values under the column 'term' and converting it into an integer.
- Stripped the percentage(%) symbol from the values under the column 'int_rate' and converting it into a float.
- Altered the value from 'Source Verified' to 'Verified' in the column - verification_status to Normalise the values.
- Rounded of the values of columns with float datatype to two decimal places.
- Reconstructed the emp_length column values to integer values and filled the null values with Zero.
- Removed the outliners for the columns: 'int_rate’(Interest Rate), 'dti’(Debt-to-Income Ratio), 'installment', 'funded_amnt_inv’(Fund Amount Investors) , 'annual_inc’(Annual Income), 'loan_amnt’(Loan Amount) by plotting the Box plots.

# Data Cleaning – Outliner Removal



# Data Analysis

## Derived Variables:

Converted the column – “issue_d” into Date format. Splitting the values into new columns issue_month and issue_year.

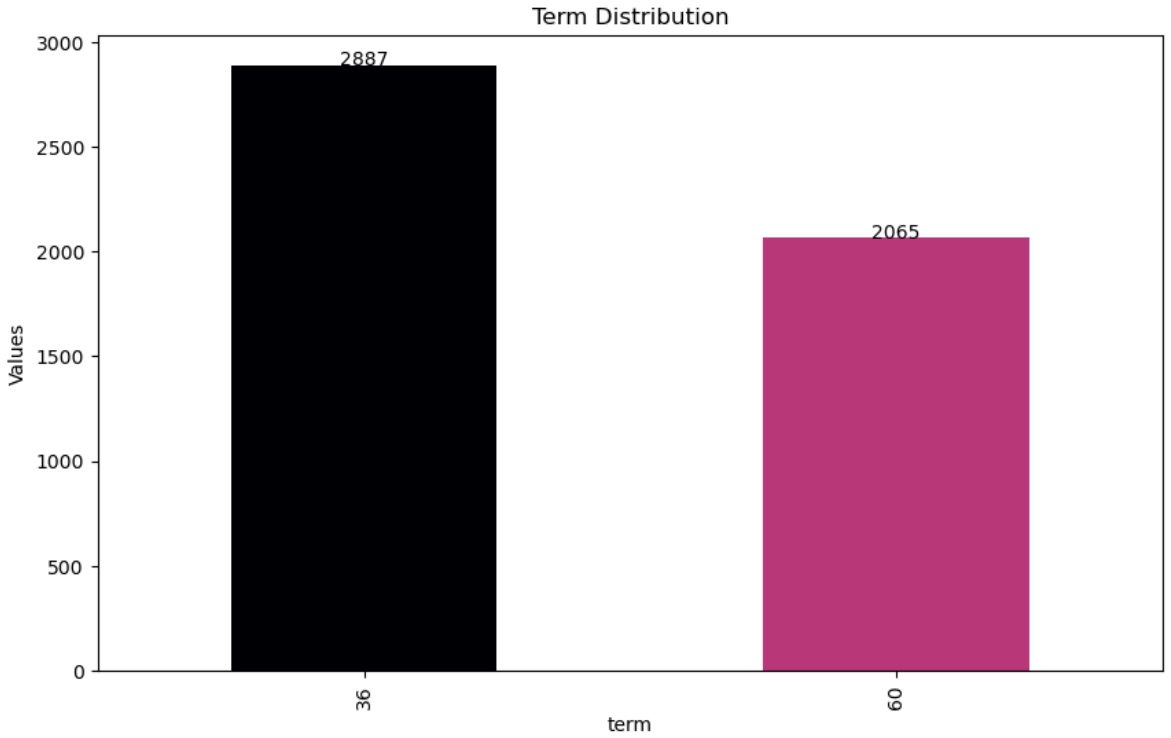
Created bin ranges for the following columns based on their minimum and maximum values:

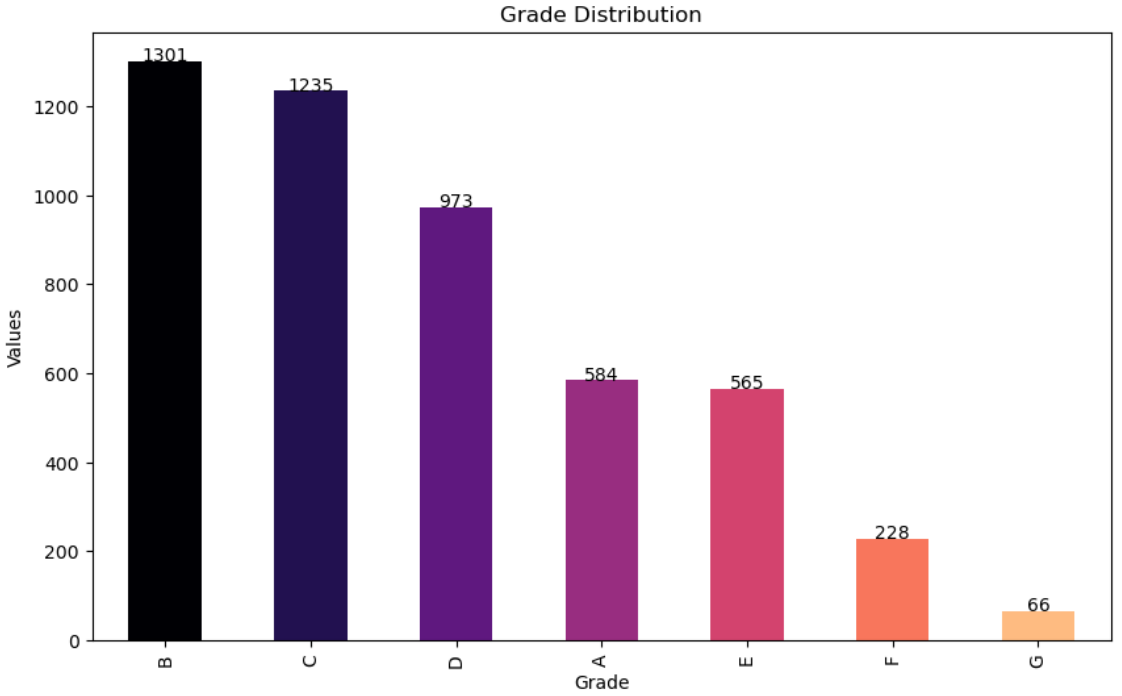
- For annual_inc(Annual Income) we have created annual_inc_bin(Annual Income Range).
- For loan_amnt(Loan Amount) we have created loan_amnt_bin(Loan Amount Range).
- For int_rate(Interest Rate) we have created int_rate_bin(Interest Rate Range).
- For dti(Debt-to-Income Ratio) we have created dti_bin( Debt to Income Ratio Range).

# Key Findings (Univariate Analysis)

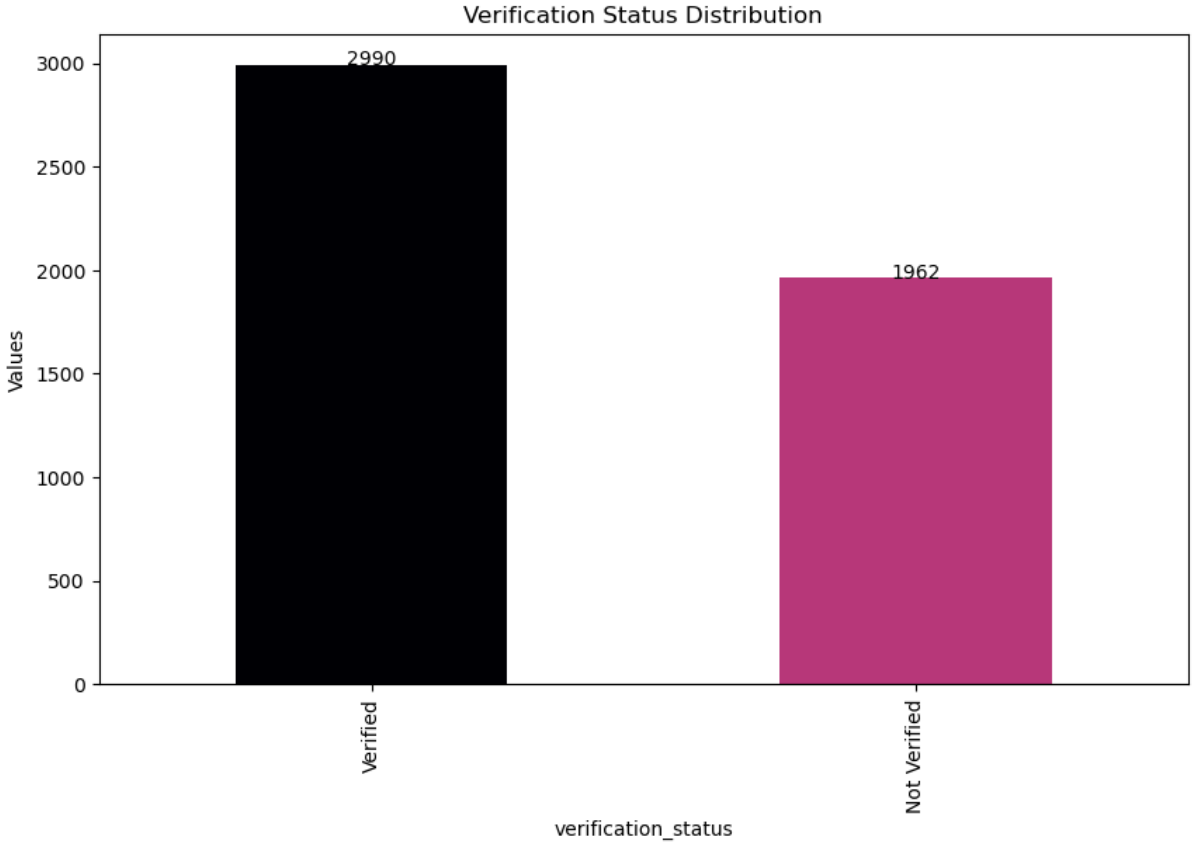
| Ordered Categorical Variable | unordered Categorical Variable | Quantitative Variable |
|---|---|---|
| emp_length(Employment Length) | addr_state(State Address) | annual_inc(Annual Income) |
| grade | home_ownership | Dti(Debt to Income Ratio) |
| issue_month | loan_status | funded_amnt(Funded Amount) |
| issue_year | purpose | funded_amnt_inv(Funded Amount from Investoers) |
| pub_rec_bankruptcies (Public Record Bankruptcies) | verification_status | installment |
| sub_grade | | int_rate(Interest Rate) |
| term | | loan_amnt(Loan Amount) |
| | | annual_inc_bin(Annual Income Range) |
| | | loan_amnt_bin(Loan Amount Range) |
| | | int_rate_bin(Interest Rate Range) |
| | | dti_bin( Debt to Income Ratio Range) |

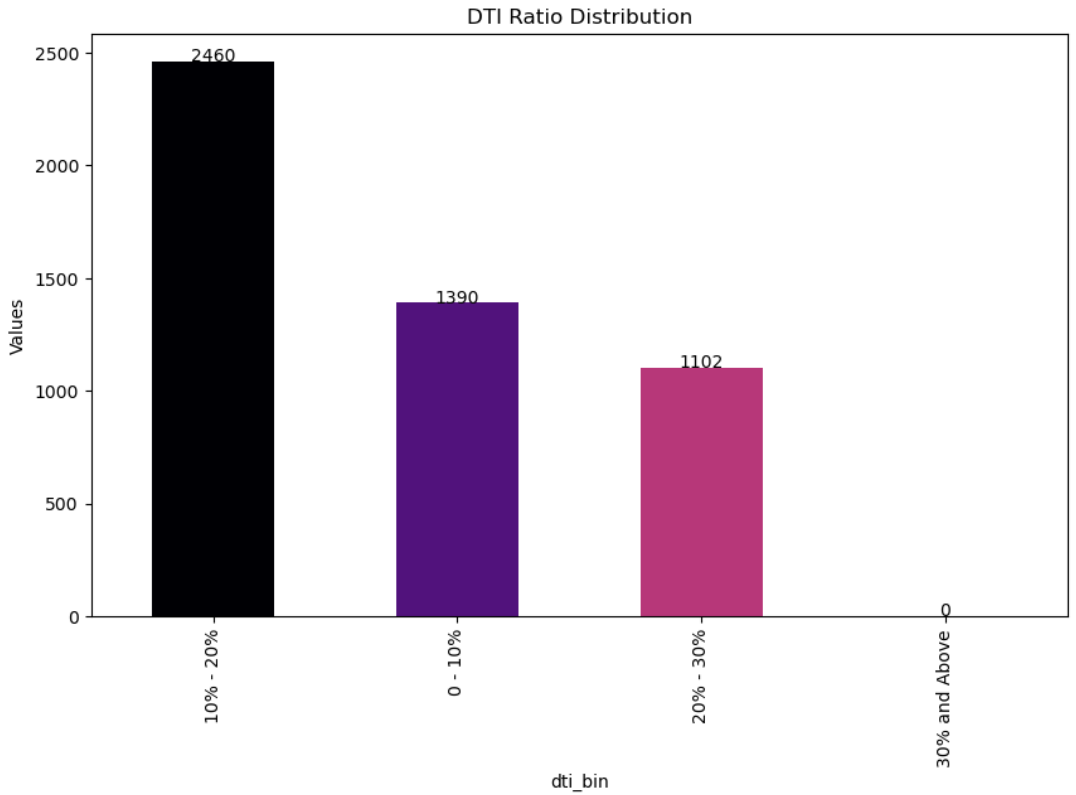
# Univariate Analysis Graphs



# Univariate Analysis Graphs



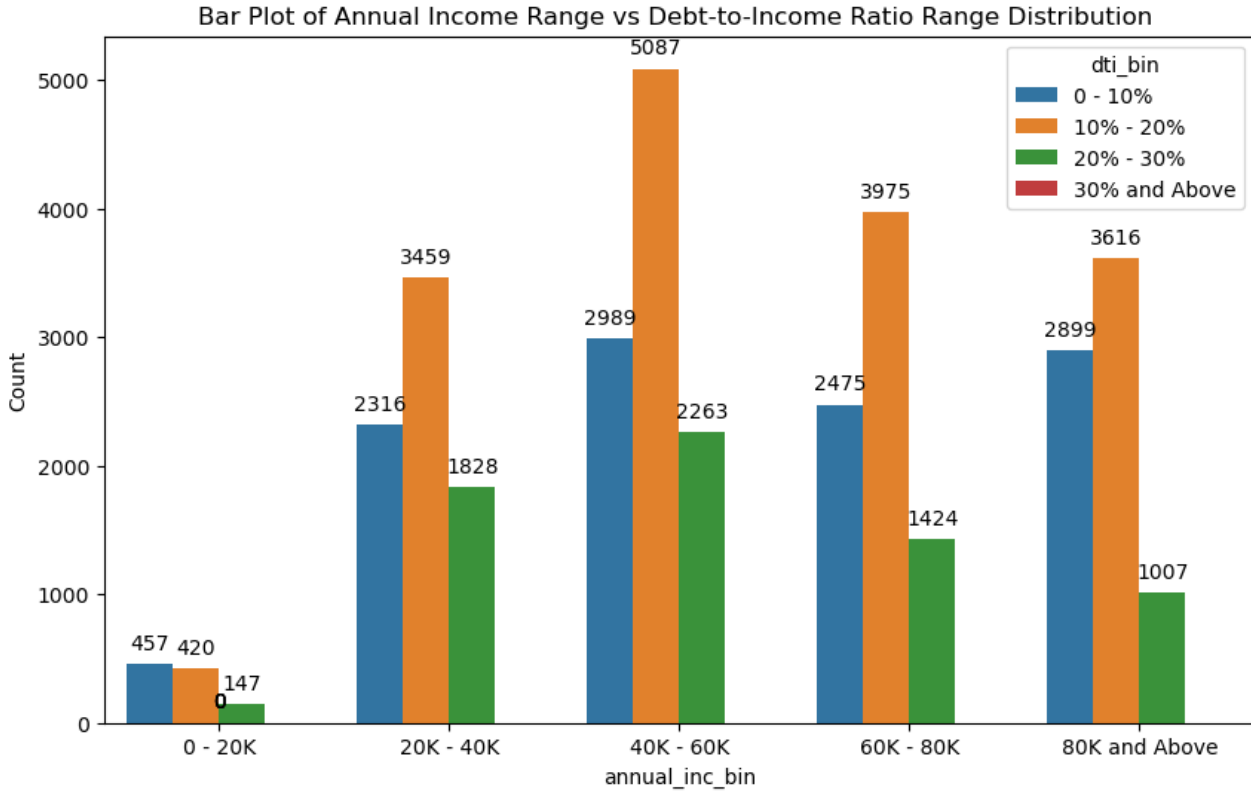
# Key Findings (Univariate Analysis)

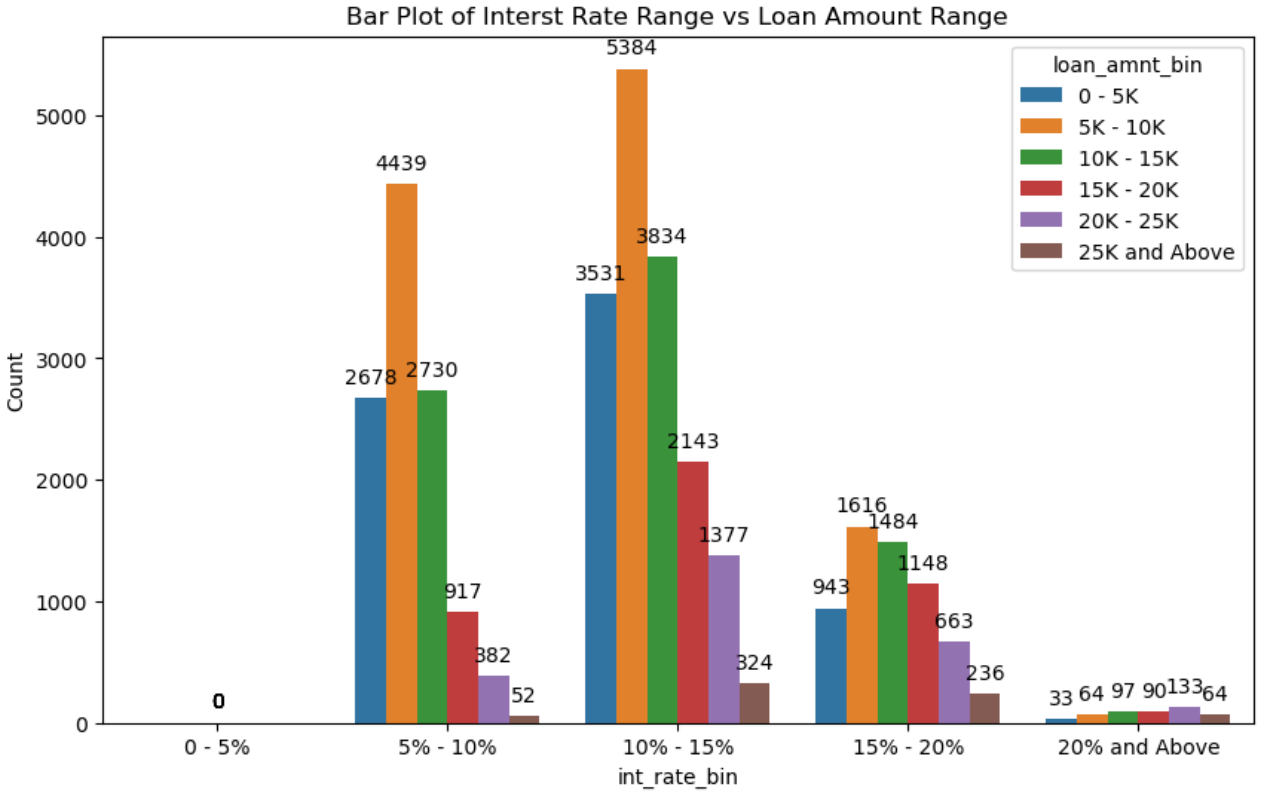
1. Term - There are more defaulters for 36 months compared to 60 months.
2. Grade – There are significant number of defaulter whose applications are graded as B,C and D followed by A,E and F. The lowest defaulters are for grade G.
3. Sub-Grade: There are significant number of defaulter whose applications are sub-graded as B5, B3, B4, C1, C2, C3.
4. Issue_month - - There are more defaulters for applications approved in the month of December and November. The lowest defaulters are applications approved in the month of February.
5. Issue_year – The highest defaulters are the applications approved in 2011. while the year 2009 is in the mid range of defaulters and lowest defaulters are recorded in the year 2007.
6. Emp_length(Employement length) – The highest defaulters are the ones with 10 years of experience and followed by zero year`s of experience. The lowest defaulters are form the employees with 9 years of experience
7. home_ownership – The highest defaulters are the ones who are in Rented or Mortgage houses.

# Key Findings (Univariate Analysis)

1. Verification_status – The highest defaulters are from the application which are verified compared to non verified applications
2. Purpose – The highest defaulters are the ones who took the loan for the Debt consolidation purpose followed by other category. Whereas the ones took loan for renewable_energy purpose are the least defaulters.
3. Addr_state(Address State) – The highest defaulters are from the state of California then followed by Florida and Newyork. Whereas applicants from MS & TN are with least defaulters.
4. Annual_inc_bin(Annual Income Range): The highest defaulters are from the annual income ranging between $40K to $60K. whereas least defaulters are from the income ranging below $20K.
5. Loan_amnt_bin(Loan Amount Range): The highest defaulters are from the loan amount ranging between $5K to $10K followed by $10K to $15K . whereas least defaulters are from the loan amount ranging from $25K and above.
6. int_rate_bin(Interest Rate Range): The highest defaulters are from the interest rate ranging between 10%-15% followed by 15%-20% . whereas least defaulters are from the interest rate ranging between 0-5% .
7. Dti_bin(Debt to Income Ratio Range): The highest defaulters are from the Debt-to-Income Ratio Range ranging between 10%-20% followed by 0-10%. whereas least defaulters are from the Debt-to-Income Ratio Range ranging from 30% and above.

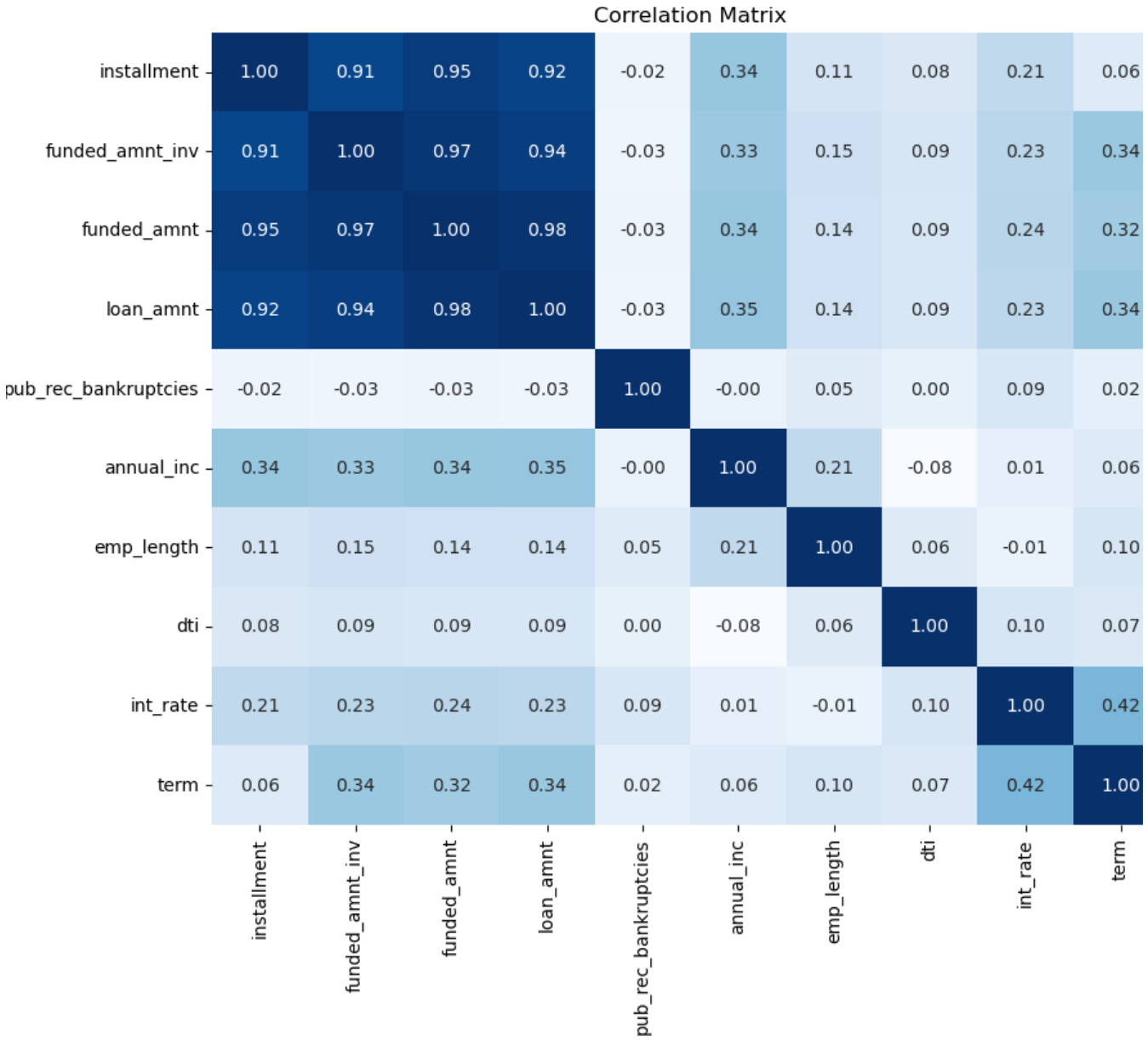
# Key Findings (Bivariate Analysis)



# Key Findings (Bivariate Analysis)

1. The highest defaulter are the applicants with the income range $40K – $60k  and the Debt-to-income ratio range is 10% to 20%
2. The highest defaulter are the applicants with the Interest Rate range 10%-15% and the Loan Amount range is 5k – 10K
3. The highest defaulter are the applicants with the Debt-to-income ratio range is 10% to 20% and the Loan Amount range is 5k – 10K




# Correlation Analysis

# Correlation Analysis (Cont.)

**Strong Correlation:**

1. installment and funded_amnt(Funded Amount): 0.95
2. funded_amnt_inv and loan_amnt(Loan Amount): 0.97
3. funded_amnt(Funded Amount) and loan_amnt(Loan Amount): 0.98
4. funded_amnt_inv(Funded Amount from Investors) and funded_amnt(Loan Amount): 0.97
5. installment and loan_amnt(Loan Amount): 0.92

**Weak Correlation:**

1. installment and term: 0.06
2. Dti(Debt to Income Ratio) and annual_inc(Annual Income): 0.08
3. emp_length(Employee Experience) and pub_rec_bankruptcies(Public Records for Bankruptcies) : 0.05

**Negative Correlation:**

1. pub_rec_bankruptcies (Public Records for Bankruptcies) and all others: ranges from -0.03 to -0.00.

# Recommendations

1. Consider offering more long-term loans, as they have fewer defaulters compared to 36-month loans
2. Be cautious with loans graded B, C, and D, as they have higher default rates. Strengthen the evaluation process for these grades.
3. Be cautious with renters and those with mortgages, as they default more often.
4. Ensure thorough verification, as verified applications have higher default rates.
5. Be cautious with loans for debt consolidation, as they have the highest default rates. Use stricter criteria for these loans.
6. Focus on applicants with debt-to-income ratios, as they default more. Use stricter criteria for these ratios.
7. Be cautious with applicants earning $40K-$60K, as they default more. Do extra financial checks for this income range.

# Conclusion

The analysis highlights the main reasons for loan defaults and offers practical suggestions.

By following these insights and recommendations, you can lower default rates and improve risk management.