Lending Club Case Study

Nachiket Shembekar
Punith Nagaraju

Lending Club Case Study - Agenda

1. Objective
2. Our Approach
3. Univariate & Bivariate Analysis - Observations
4. Recommendations
5. Detailed Results & Graphs

Objective

- As part of Lending Club, a consumer finance company, identify risky loan applicants who are likely to default in future.

- When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:
  - If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
  - If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

- This analysis is to be done using EDA over historical data available of all loans issued through the time period 2007 to 2011.

Our Approach

- Analyzing the dataset.
- Understanding different columns and data formats.
- Analyzing the overall volume to be processed.

- Univariate analysis for individual columns that are cleaned in previous step.
- Segmented Univariate analysis for individual columns, by categorizing values into different data ranges.

- Consolidate and present the analysis based on the observations recorded at each stage.
- Provide recommendations to the end user (in this case investors), for future purposes.

- Analysis of columns and elimination of redundant, irrelevant or less meaningful ones in perspective of the solution.
- Elimination of irrelevant rows.
- Data Type conversions if/as needed.
- Data standardization, replacement of missing values if/as needed.
- Defining derived columns to enhance the analysis.

- Analyzing behavior of 2 variables with respect to the criteria (i.e., loan applicants with Charged Off status.

Individual Factors (Univariate Analysis) – High Impact

Interest Rate

When the interest rate is in the range of 9% to 17%, there's significant rise in the defaulted applicants.

Annual Income

Applicants with low to medium income group (up to 58K) are at higher risk of defaulting. As income is the primary consideration of repayment, this is certainly a high impacting factor.

Loan Purpose

Applicants with Loan Purpose of Debt Consolidation have been found at most risk of defaulting, followed by credit card and small business.

Open Credit Lines

The number of open credit lines in the borrower's credit file have impact on their repayment trend. People with less than 10 open credit lines are more prone to defaulting than others.

Grade

As the grade changes from B to G, the possibility of defaulting goes on reducing significantly. Grades B, C & D have observed to have maximum possibility of defaulting.

Individual Factors (Univariate Analysis) – Medium/Low Impact

Home Ownership

Higher concentration of 'Charged Off' applicants who are either renting the home or have mortgage, which portrays impact of financial commitments on loan repayment.

Revolving line utilization rate

Number of defaulters is increases as the Revolving line utilization rate increases, until it reaches 80. Then the number stays nearly in the same range (~1200)

Debt to Income Ratio

Higher concentration of 'Charged Off' applicants for ‘dti’ in the range of 12 to 24 (follows normal distribution)

Earliest Credit Line

Highest number of defaulters are observed having credit line opening between 1992-2000.

Loan Funding Year

Highest number of defaulters are observed who had been funded between 2010-11. either credibility of applicants before 2010 was higher or possibly gaps in scrutinizing the applicants.

Couple Factors (Bivariate Analysis)

Annual Income & Loan Purpose

When considered individually, Loan purpose of 'Debt Consolidation' has highest impact. But when analyzes along with Annual income, applicants with small business purpose and lower income are the most prone for defaulting. Also, other purposes such as house or moving with lower income group applicants are prone to defaulting.

Annual Income & dti

Applicants belonging to high income groups are prone to defaulting if their debt-to-income ratio is in the range of 12 to 24. This would entail that they're heavily spending their income on debts and there's possibility of additional debts beyond current funding, thus risking the repayment to investors.

Annual Income & installment

As the installment figures increase, risk of defaulting increases. This applies to all the income groups, although the probability of higher installment is lesser for low-income groups. The loan amount funded (thus installment) to them is assumed to be based on their income and hence low-income group have lower installment figures.

Annual Income & Loan term

Loan term has an impact on defaulting for the lower income groups and for 36 Months. As the income increases, term impact reduces. For 60-month term, the defaulting possibility is reduced as the borrowers get more time to adjust their payments.

Annual Income & Revolving Utilization Rate

Possibility of defaulting increases with increased Revolving line utilization rate, across all the income groups. As the borrower uses more credit, his chances of not being able to repay will increase.

Annual Income & Interest Rate

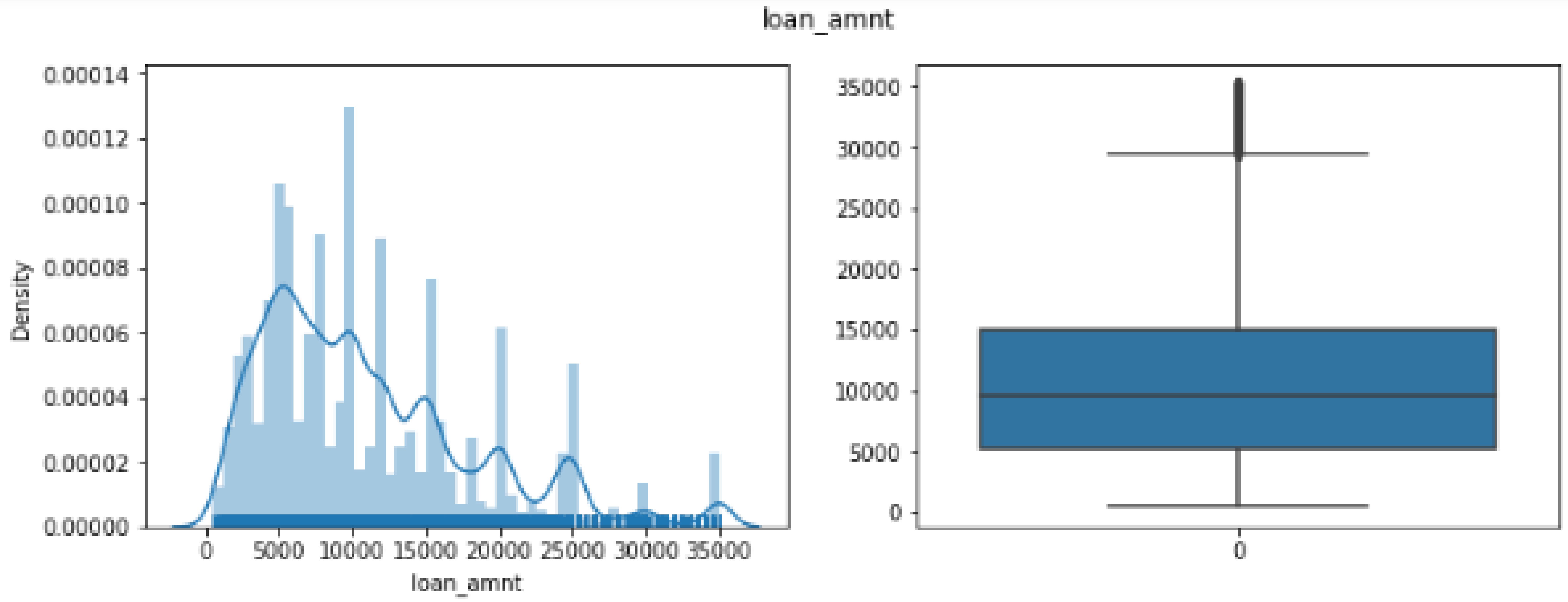
Significant impact of Interest rate. When the interest rate is between 9%-17%, possibility of defaulting increases irrespective of the income group.

Bivariate analysis is also done excluding 'Annual Income', e.g., purpose and home ownership. Although correlation is found to some extent, it didn't provide any conclusive observation for the impact on defaulting.

Recommendations for Lending Club

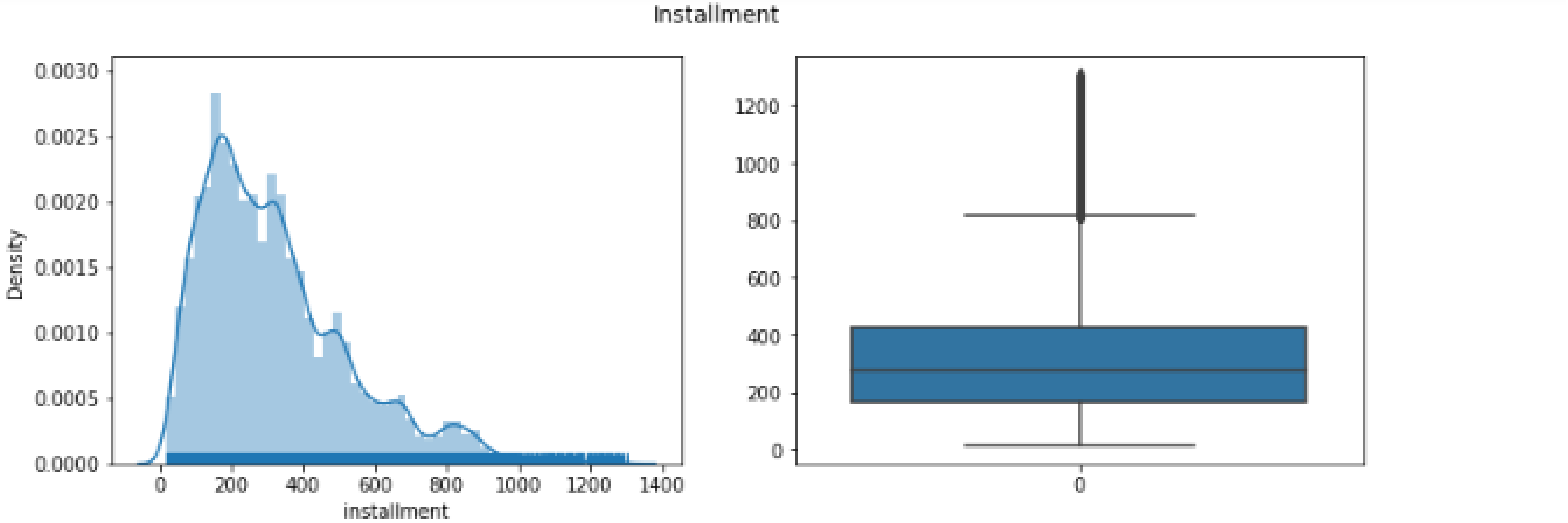
- ❑ Increase the level of scrutiny while funding people with lower income groups.
- ❑ For higher income groups, it's important to have low 'debt-to-income' ratio before funding is provided.
- ❑ Consider increasing Loan term to 60 months or also introduce a slab in between 36 to 60 months. This may reduce the defaulting probability.
- ❑ Avoid or reduce funding loans for purpose of 'debt consolidation' or 'small business'.
- ❑ Avoid or reduce funding loans for applicants having less than 10 open credit lines.
- ❑ Consider revising the Interest rates or try to be innovative in giving options to the borrowers. If any borrower is at the risk of defaulting, his interest rate may be reduced while increasing the tenure.
- ❑ Scrutinize the loans funded in 2010-11 if any gaps are identified. They need to be addressed to avoid any process or other issues.
- ❑ Scrutinize the Revolving Utilization Rate before funding the applicant.

Important Univariate Analysis – Results



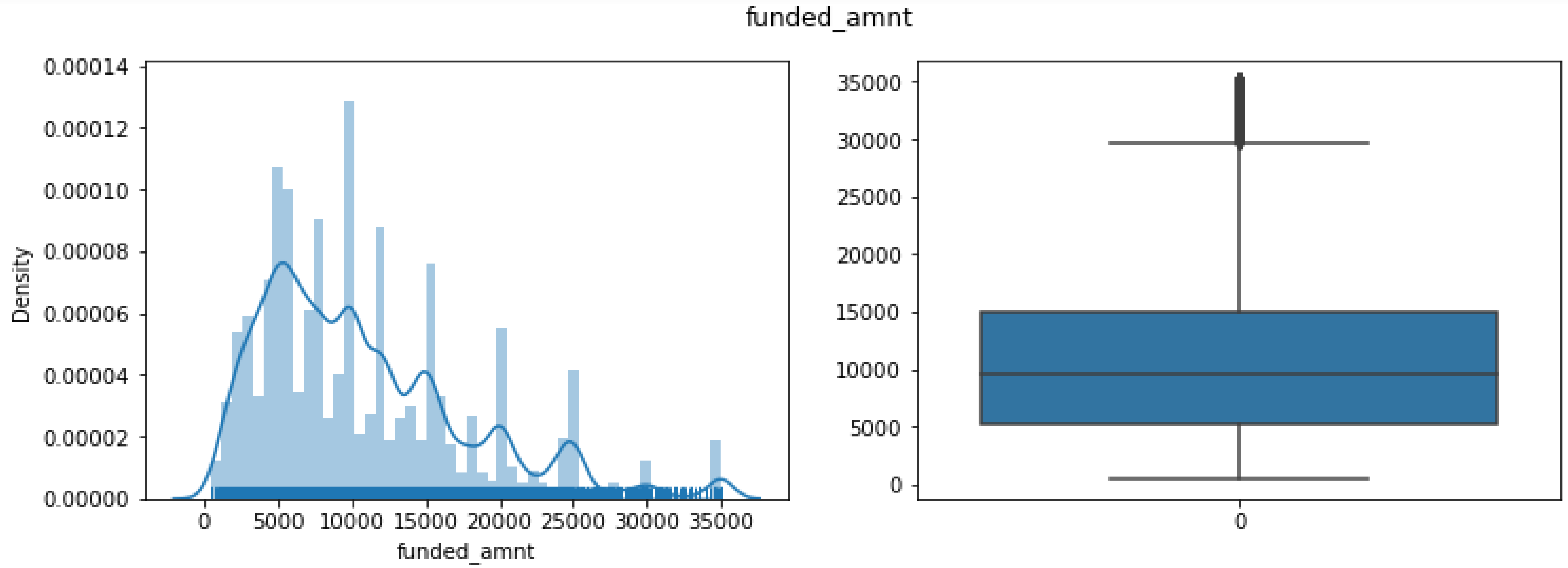
Observation : Highest number of applicants took loan around 10,000 which is between median and mean. Statistics show that most of the applicants have loan amount less than 15000.

Univariate Analysis – Results



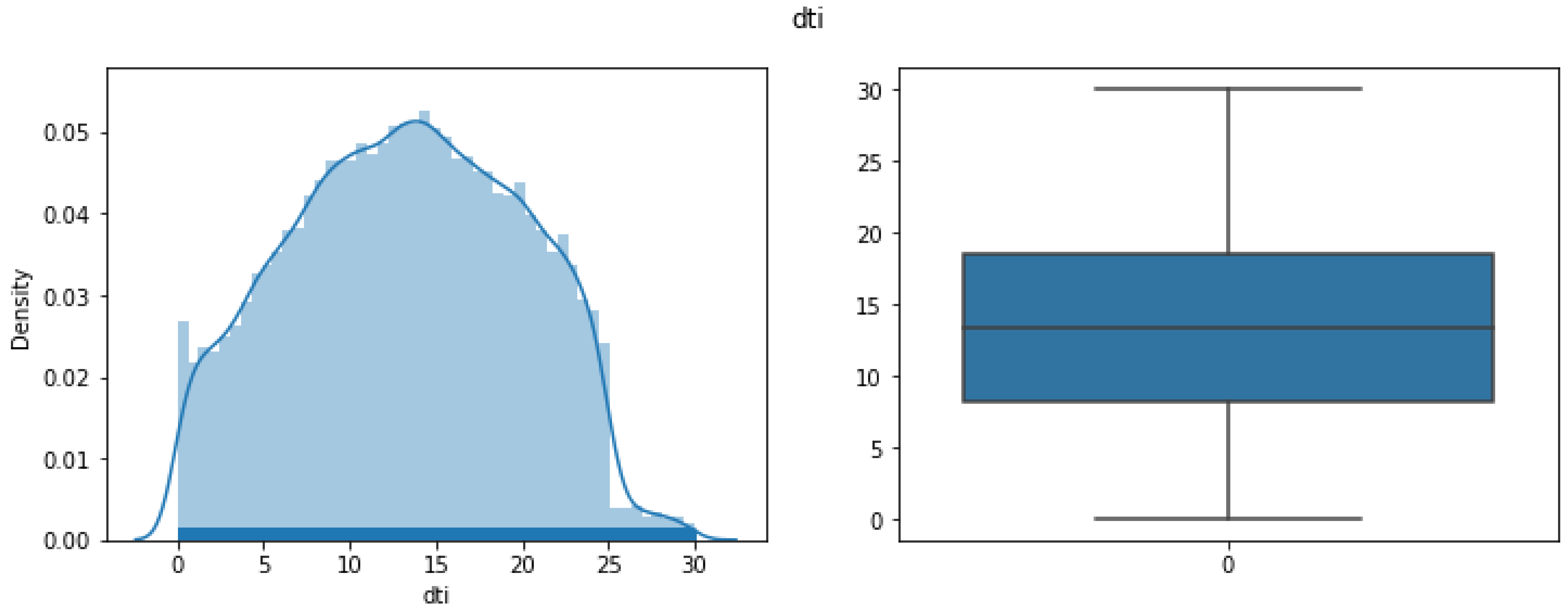
Observation : Majority of applicants have installment amount less than 500, which could be because of lower loan amount.

Univariate Analysis – Results



Observation : The funded amount statistics are similar to loan amount. This shows that almost in all the cases, complete amount is funded as requested (or approved by the intermediate organization before investers funded).

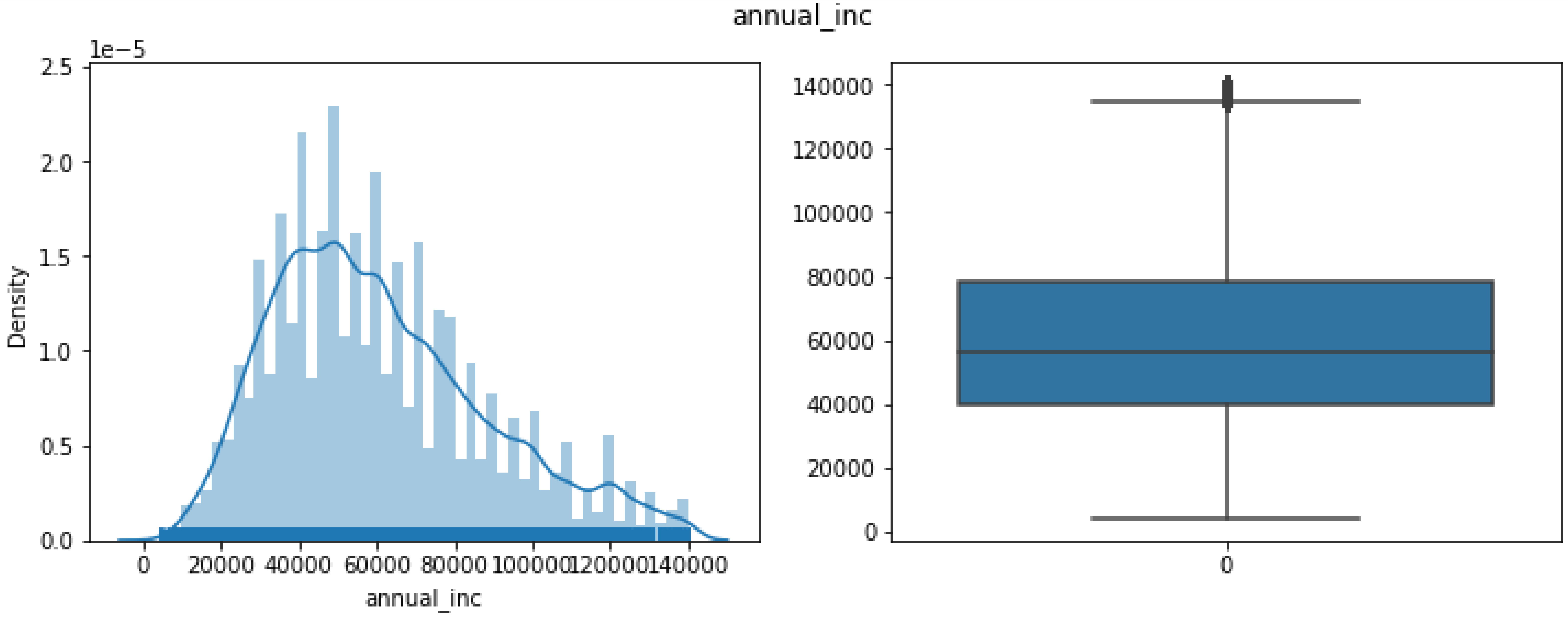
Univariate Analysis – Results




```
Median : 13.37
Mean : 13.272727013505502
```

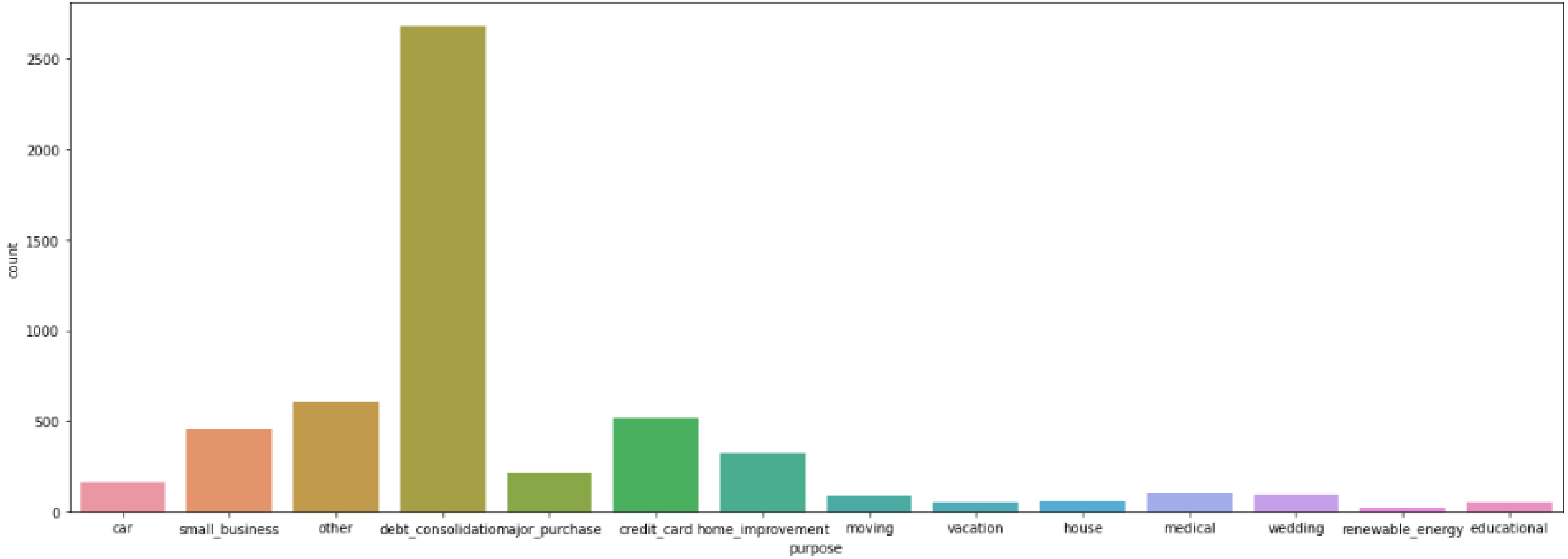
Observation : Loans are funded to applicants having mean/median DTI around 13. No applicants have dti greater than 30, and the graph follows nearly ideal normal distribution.

Univariate Analysis – Results



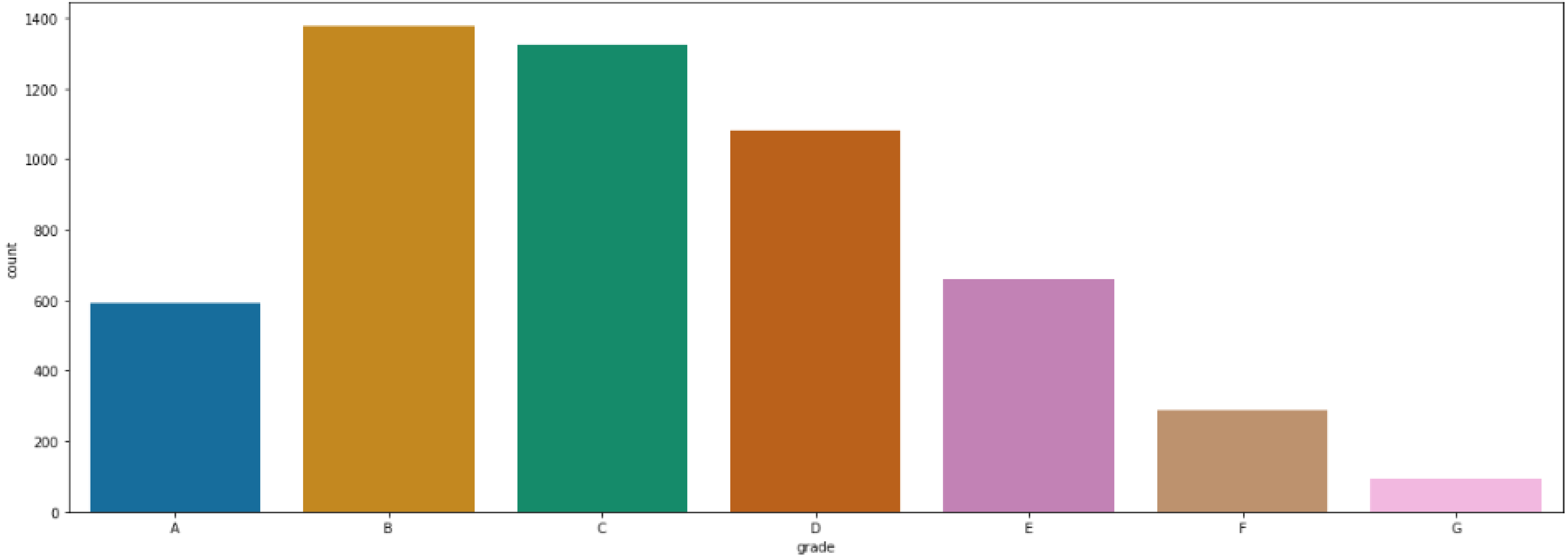
Observation : Annual incomes of more than 50% applicants is less than 60,000. This suggests people in lower income groups struggling to repay the loan in time.

Univariate Analysis – Results



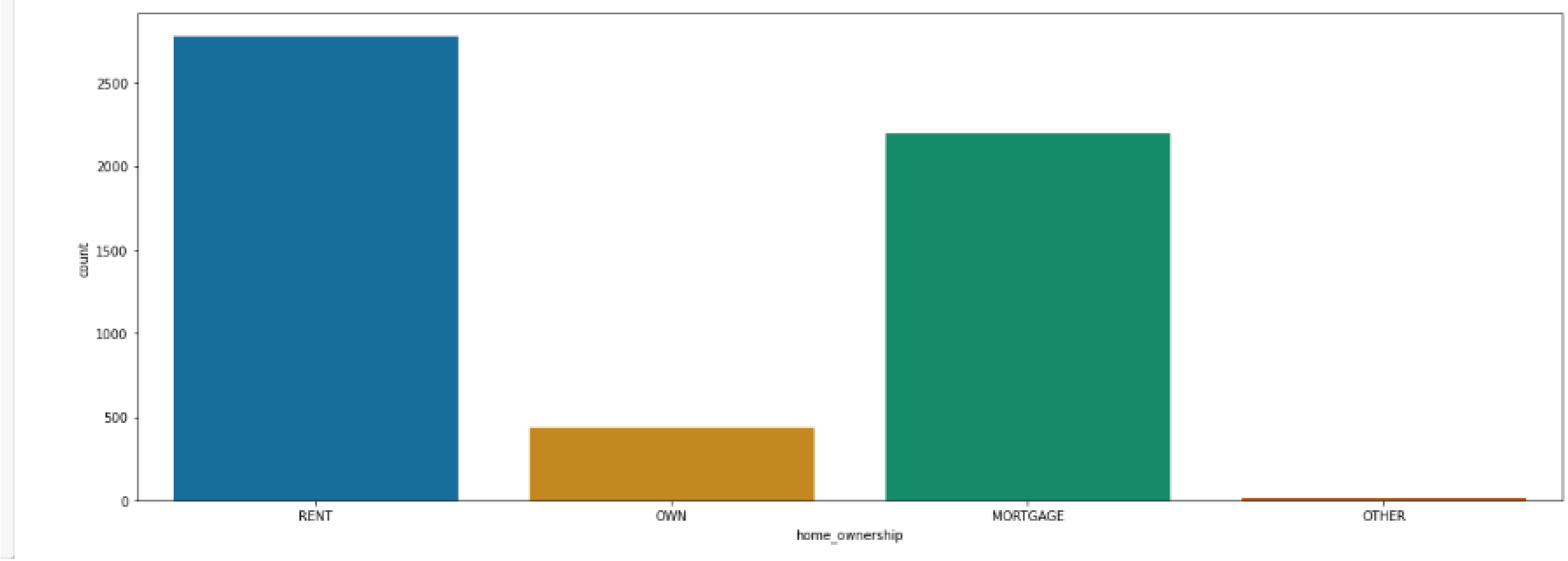
Observation : Highest number of 'Charged Off' applicants have purpose of 'Debt consolidation', which is an important factor. Other major purposes are credit_card, small_business and others.

Univariate Analysis – Results



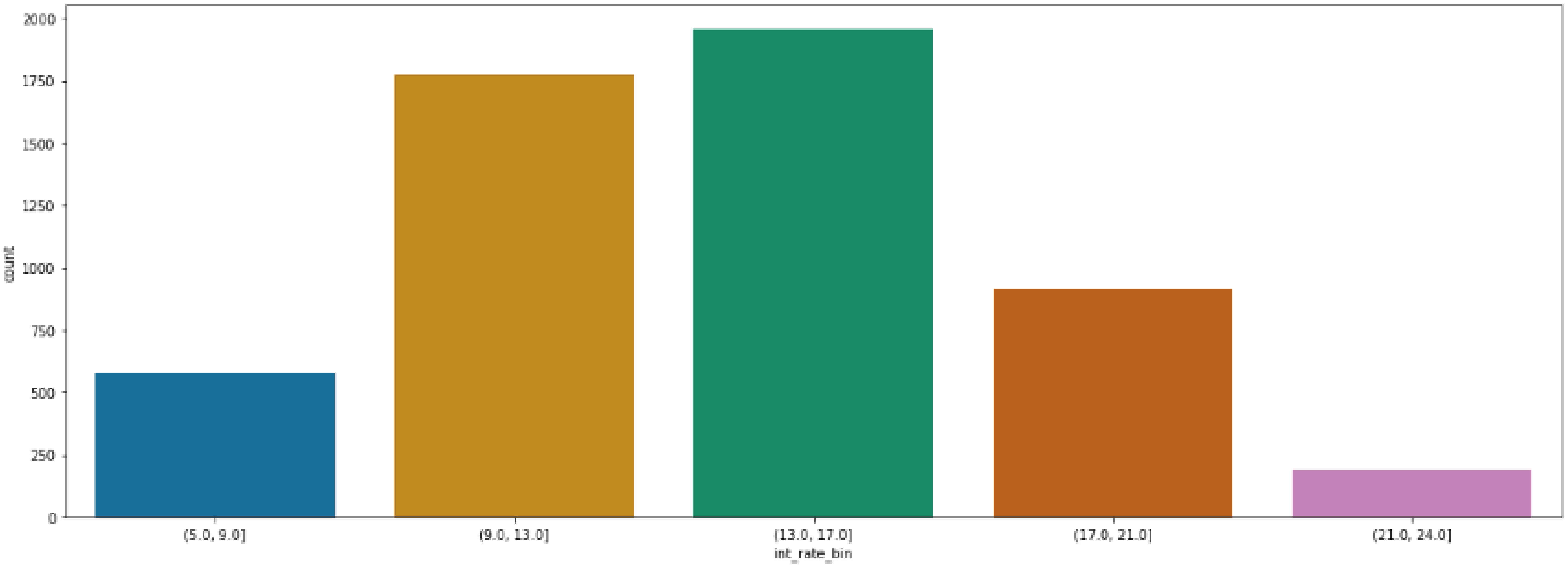
Observation : Higher concentration of 'Charged Off' applicants in the grades B, C, D.

Univariate Analysis – Results



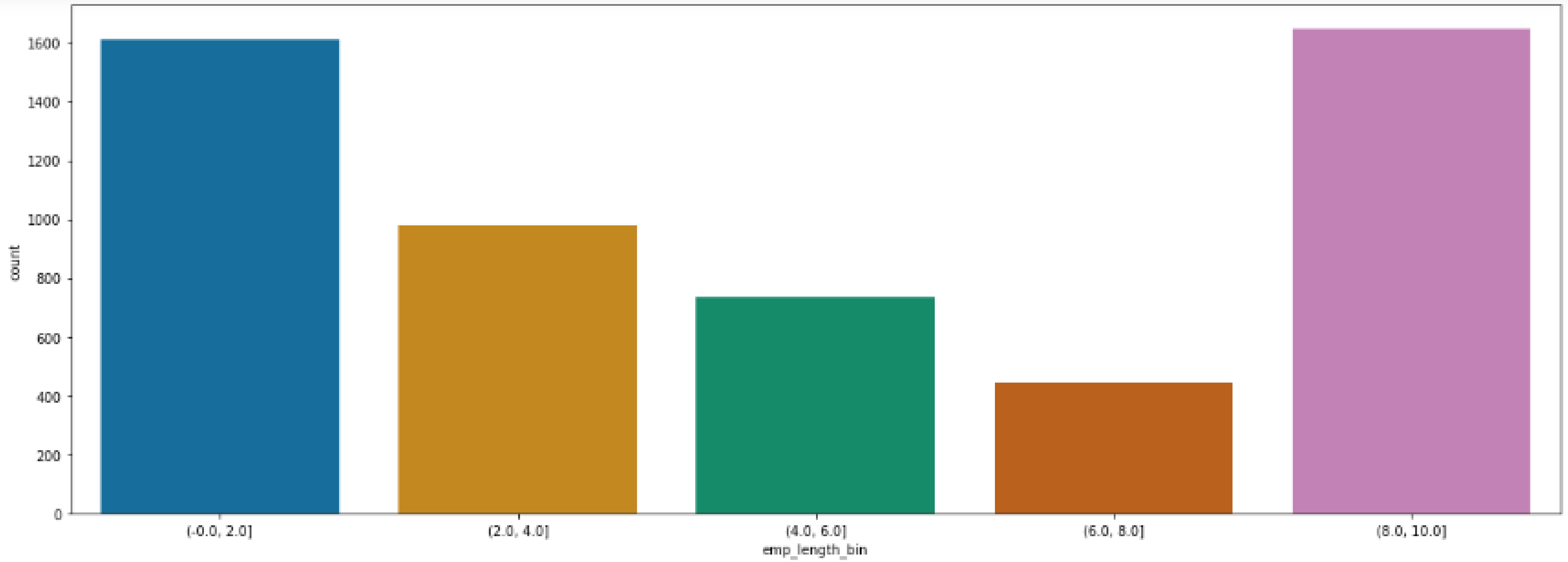

```
Observation : Higher concentration of 'Charged Off' applicants who are either renting the home or have mortgage. This is
logical as it shows important financial commitments that are impacting the loan repayment.
```

Univariate Analysis – Results



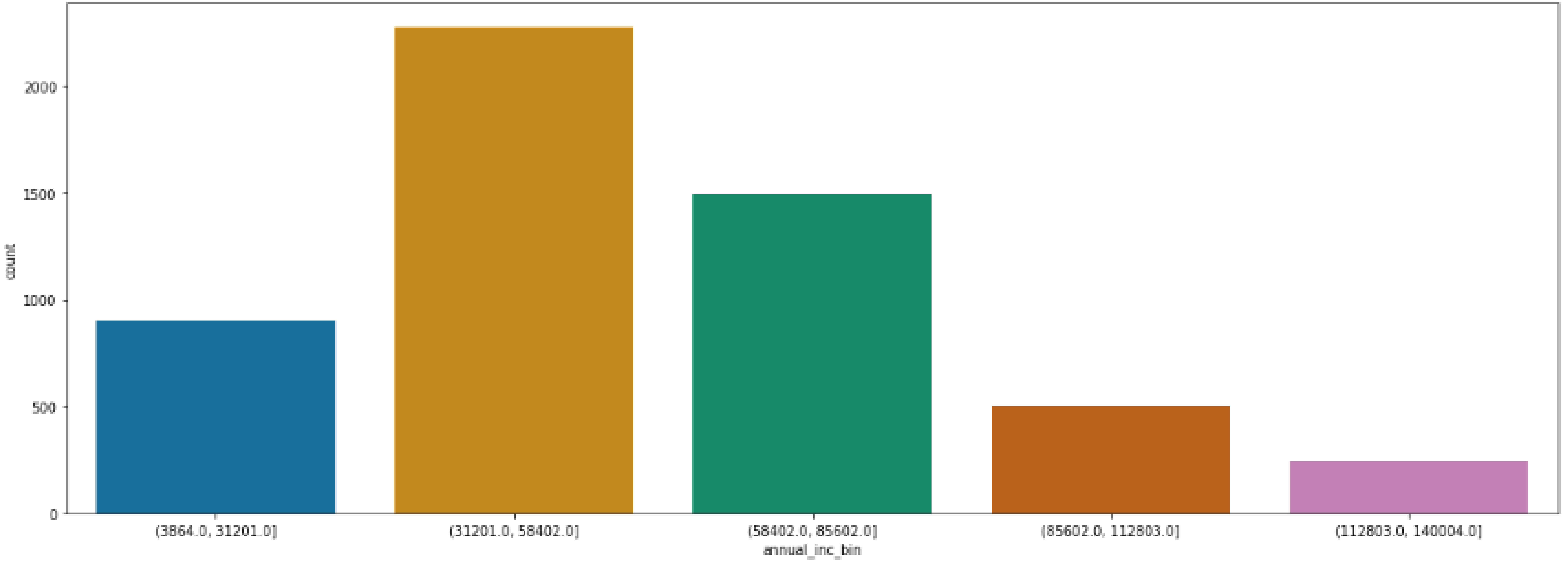
Observation : Higher concentration of 'Charged Off' applicants when the interest rate is in the range of 13%-17%. Lower concentration as the interest rate goes up, indicating either resistance from the borrower or hesitation from invester due to applicant's credibility.

Univariate Analysis – Results



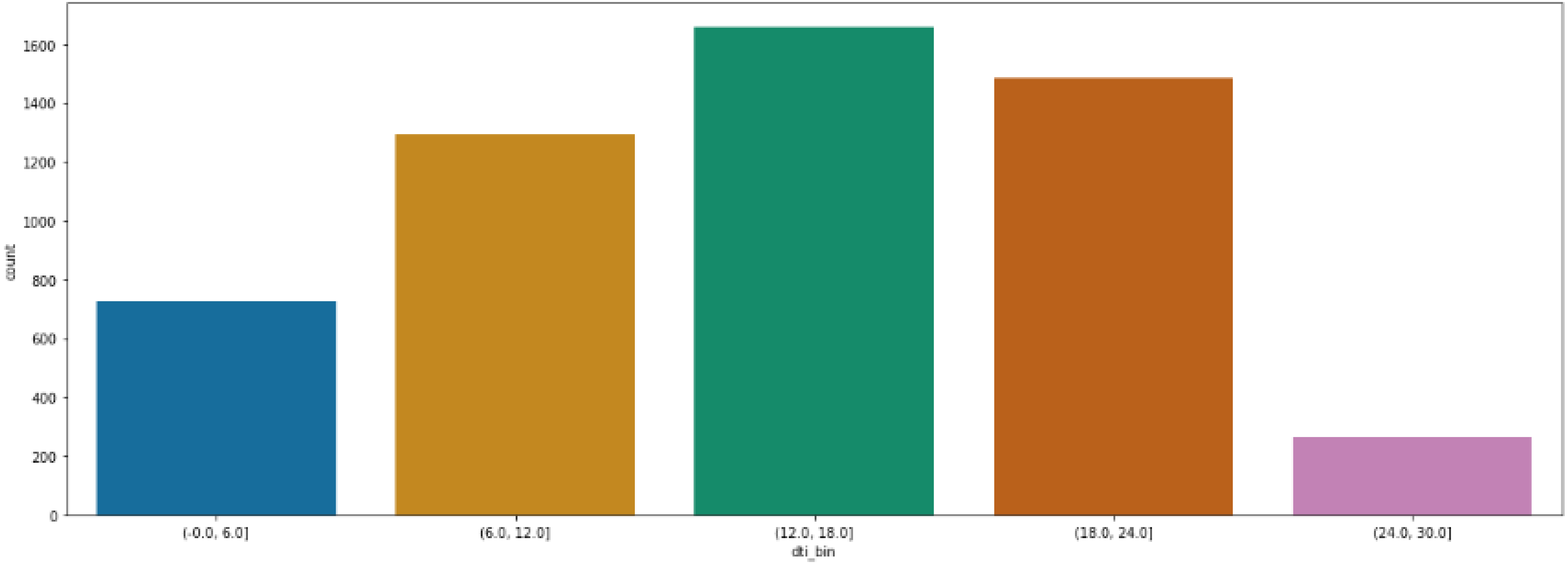

```
Observation : Higher concentration of 'Charged Off' applicants in the lower as well as higher emp_length. It's important to
note that the higher emp_lengh (10+ years) also contain the null rows as part of Data standardization, and this data may not
prove the impact of emp_length on the defaulting.
```

Univariate Analysis – Results



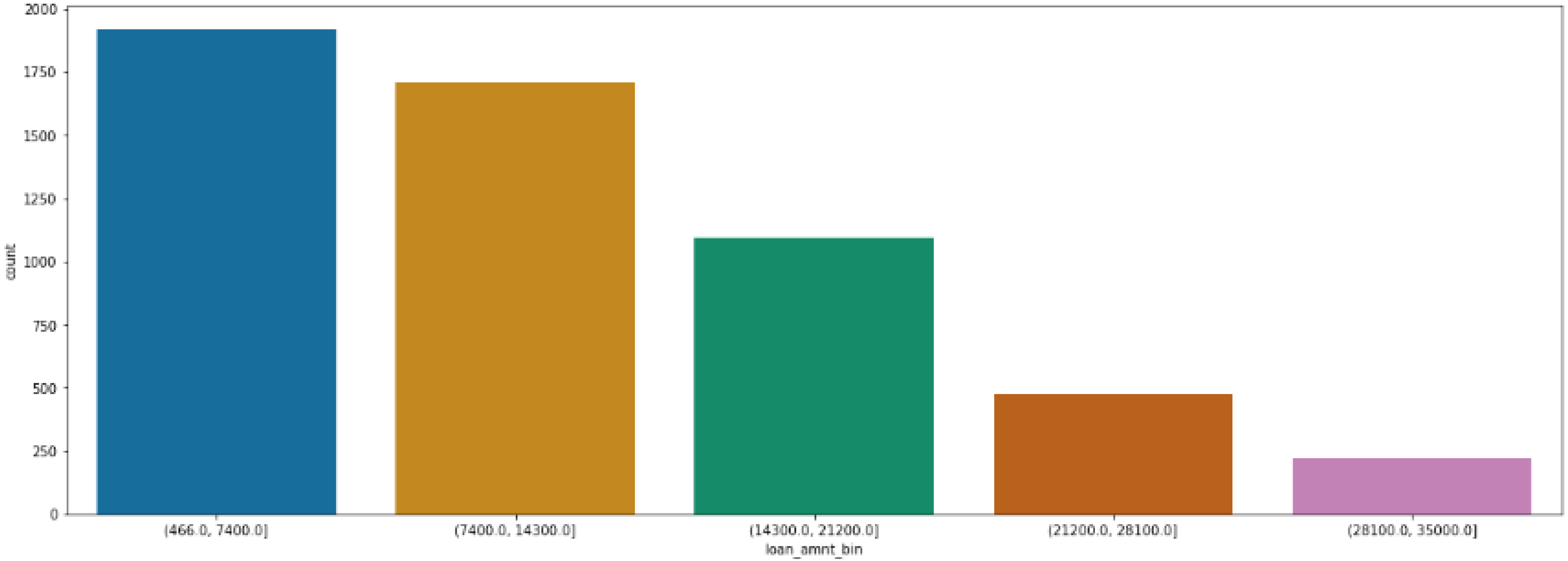

```
loan_ds['dti_bin'] = pd.cut(loan_ds['dti'], bins=5,precision =0)
Univar_Analysis_Count_Charged_Off(column_name='dti_bin')
```

Univariate Analysis – Results



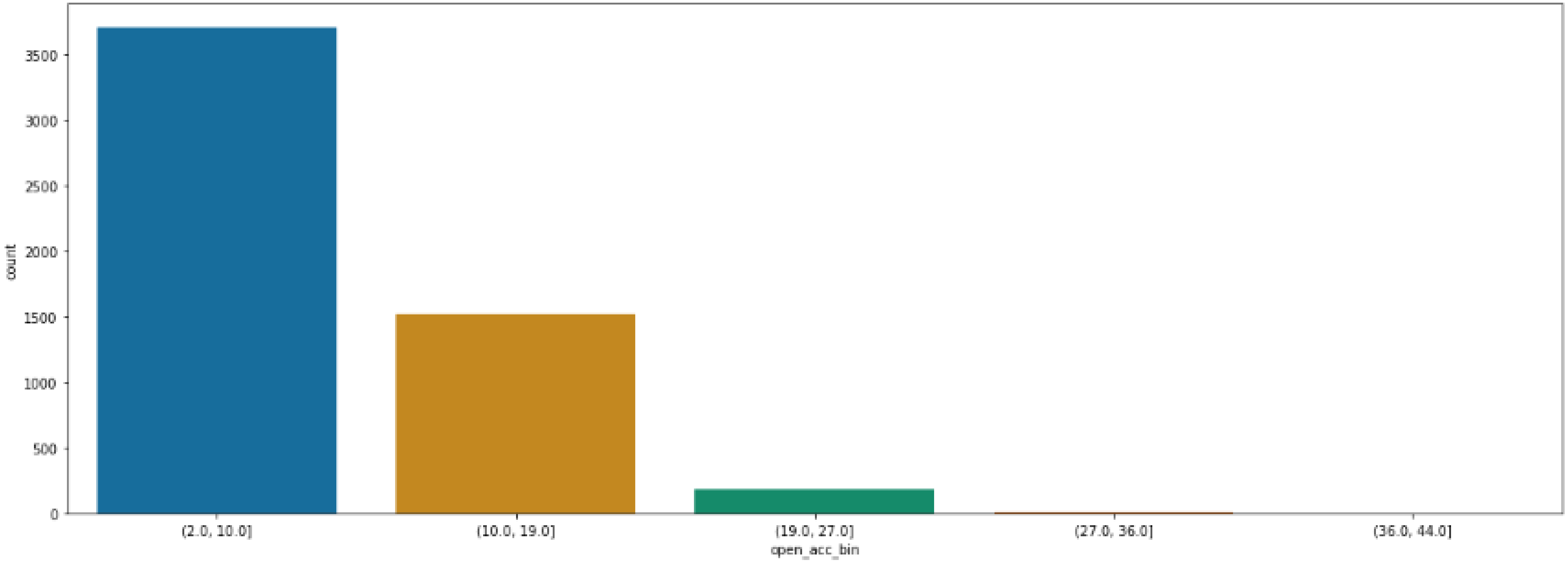
Observation : Higher concentration of 'Charged Off' applicants for dti in the range of 12 to 24 (as visible in normal distribution).

Univariate Analysis – Results



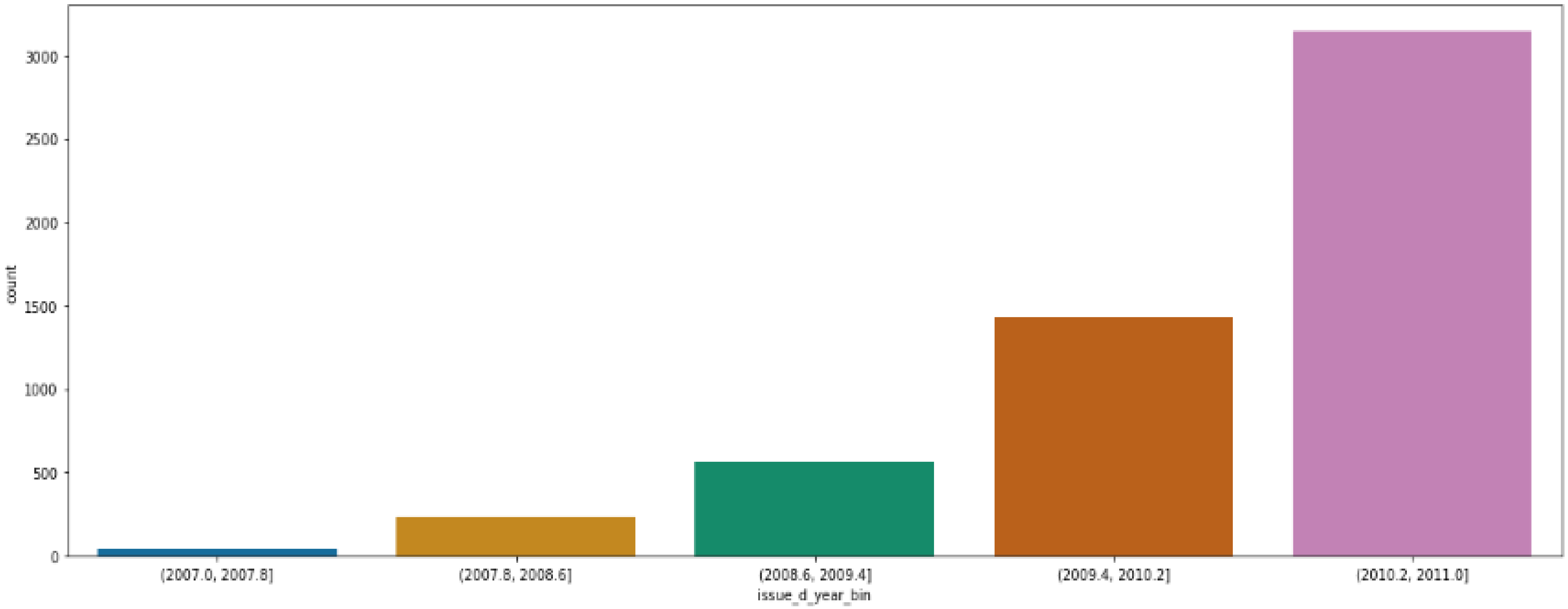
Observation : Higher concentration of 'Charged Off' applicants for lower loan amounts. The investers might have assumed low risk while funding smaller loan amounts, however it has resulted into financial losses overall.

Univariate Analysis – Results



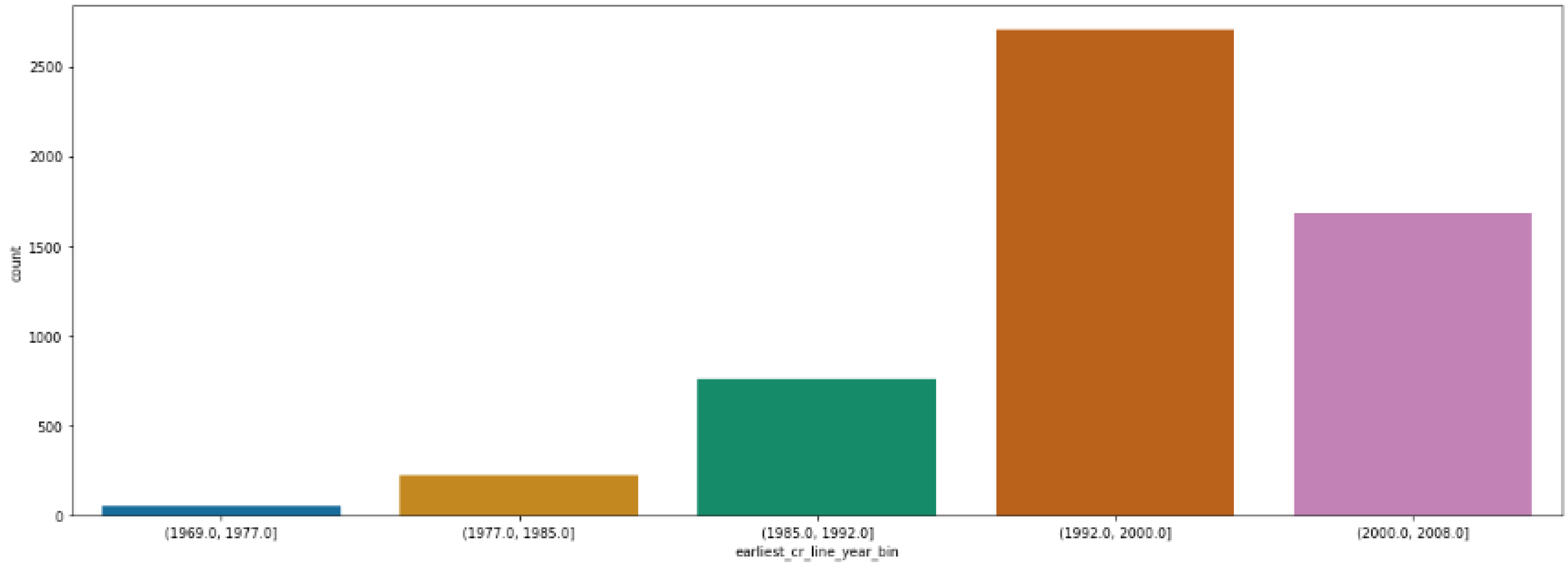
Observation : As the number of credit lines increase, defaulter count goes on reducing.

Univariate Analysis – Results



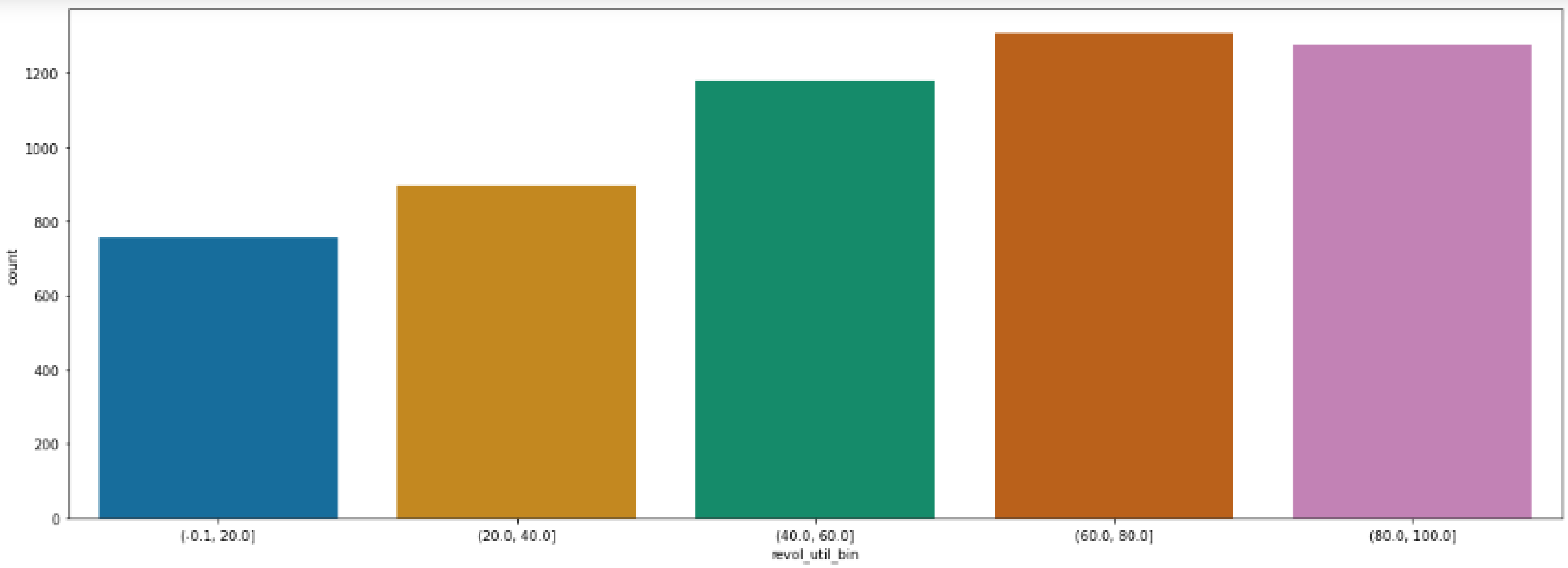
Observation : Highest number of defaulters are observed who had been funded between 2010-11. Credibility of applicants before 2010 was higher or there need to be improvements in scrutinizing the eligible applicants.

Univariate Analysis – Results



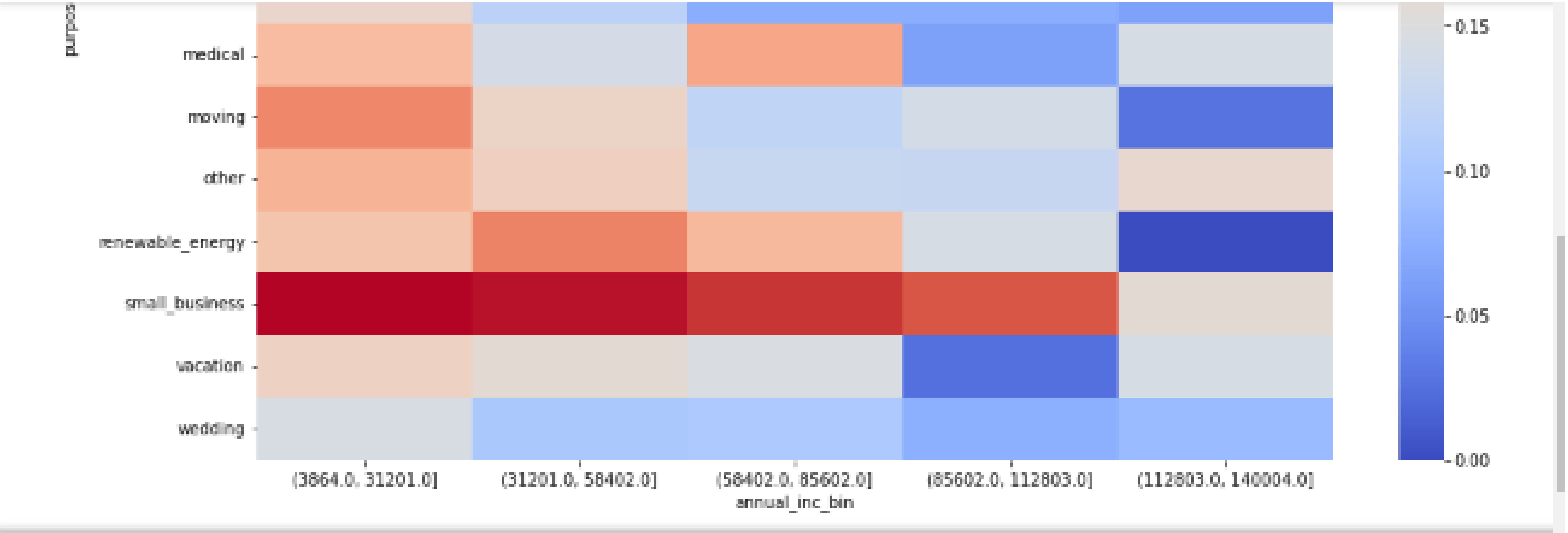
Observation : Highest number of defaulters are observed having credit line opening between 1992-2000.

Univariate Analysis – Results



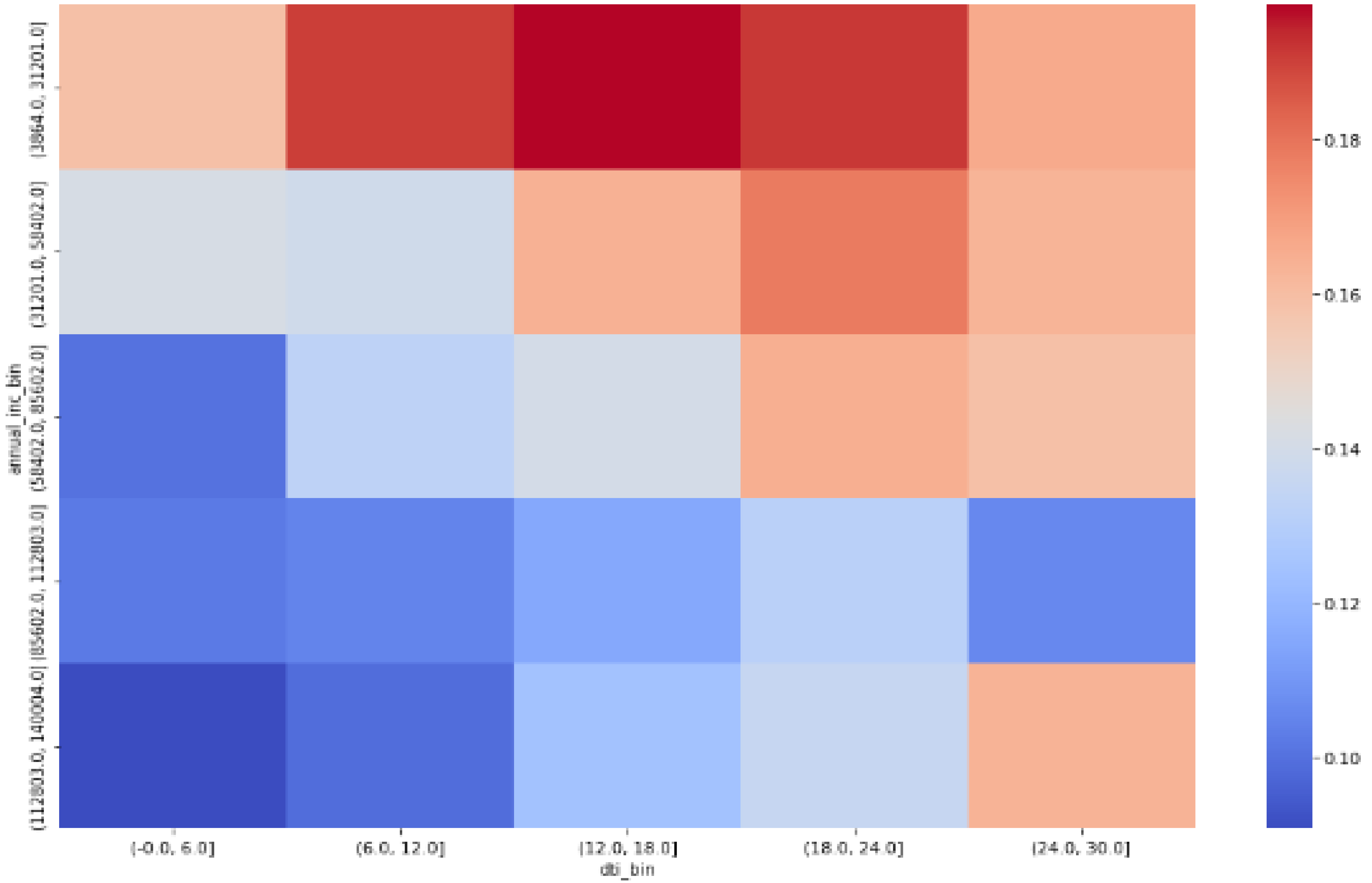
Observation : Number of defaulters is increases as the Revolving line utilization rate increases, until it reaches 80. Then the number stays nearly in the same range (~1200)

Bivariate Analysis – Results



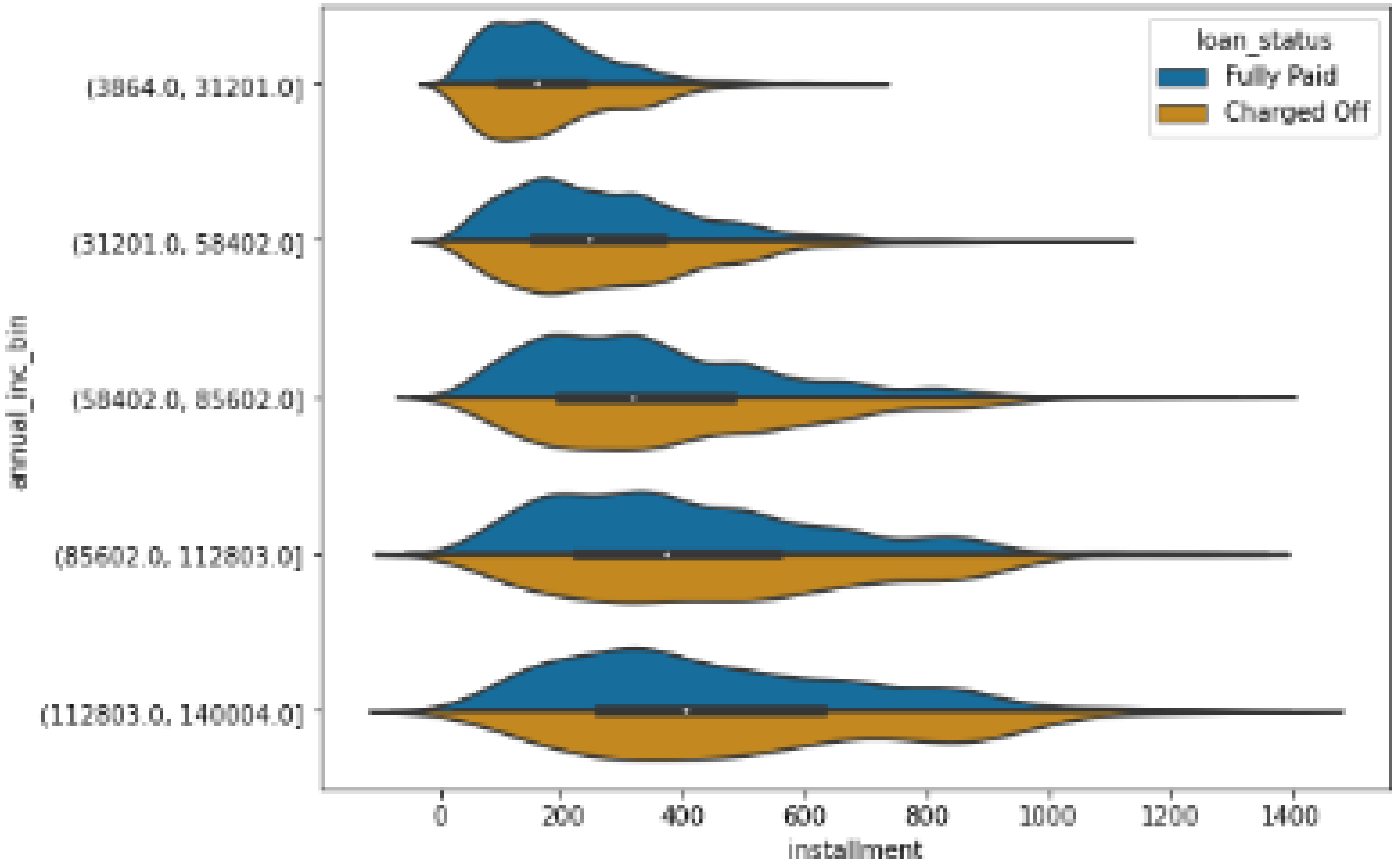
Observation : Applicants who have opted for loan for small business and lower income are the most prone for defaulting. Also, other purposes such as house or moving with lower income group applicants are prone to defaulting.

Bivariate Analysis – Results



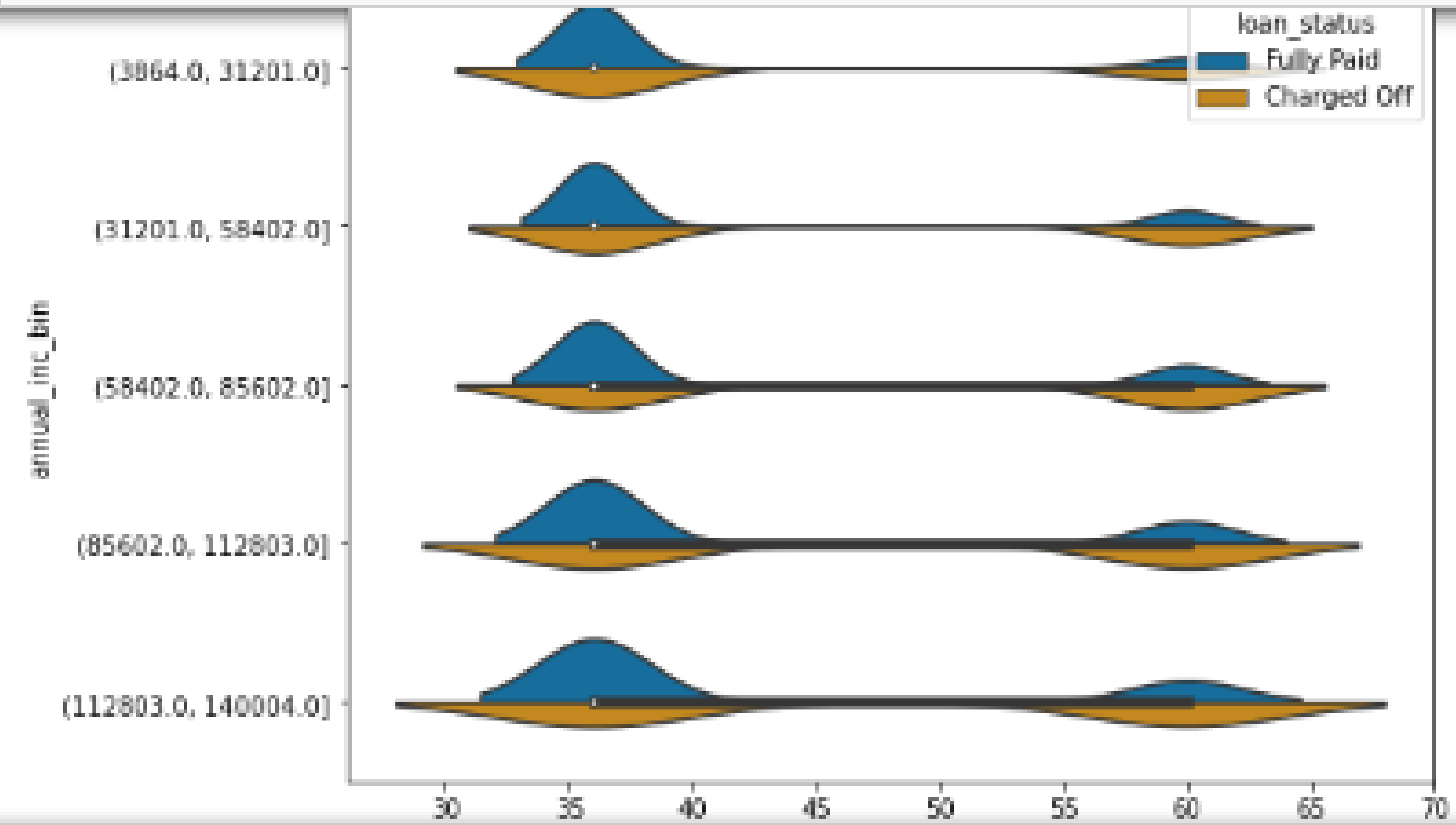
Observation : Applicants belonging to high income groups are prone to defaulting if their debt-to-income ratio is in the range of 12 to 24. This would entail that they're heavily spending their income on debts and there's possibility of additional debts beyond current funding, thus risking the repayment to investors.

Bivariate Analysis – Results



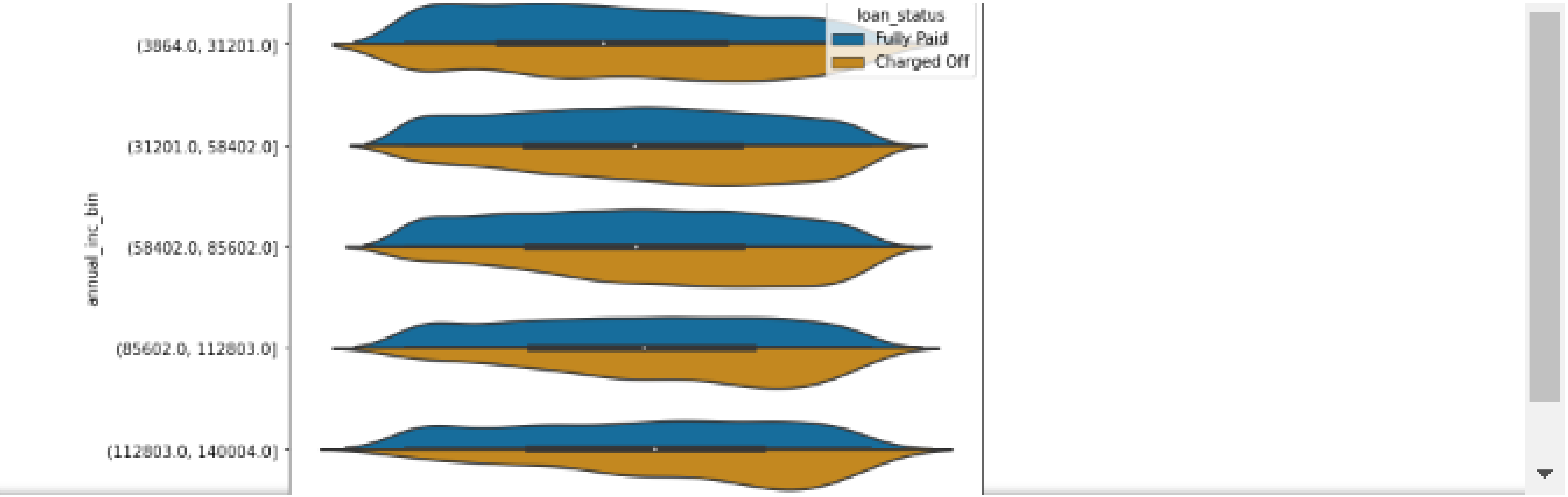

```
Observation : As the installment figures increase, risk of defaulting increases. This applies to all the income groups,
although the probability of higher installment is lesser for low-income groups. The loan amount funded (thus installment) to
them is assumed to be based on their income and hence low-income group have lower installment figures.
```

Bivariate Analysis – Results



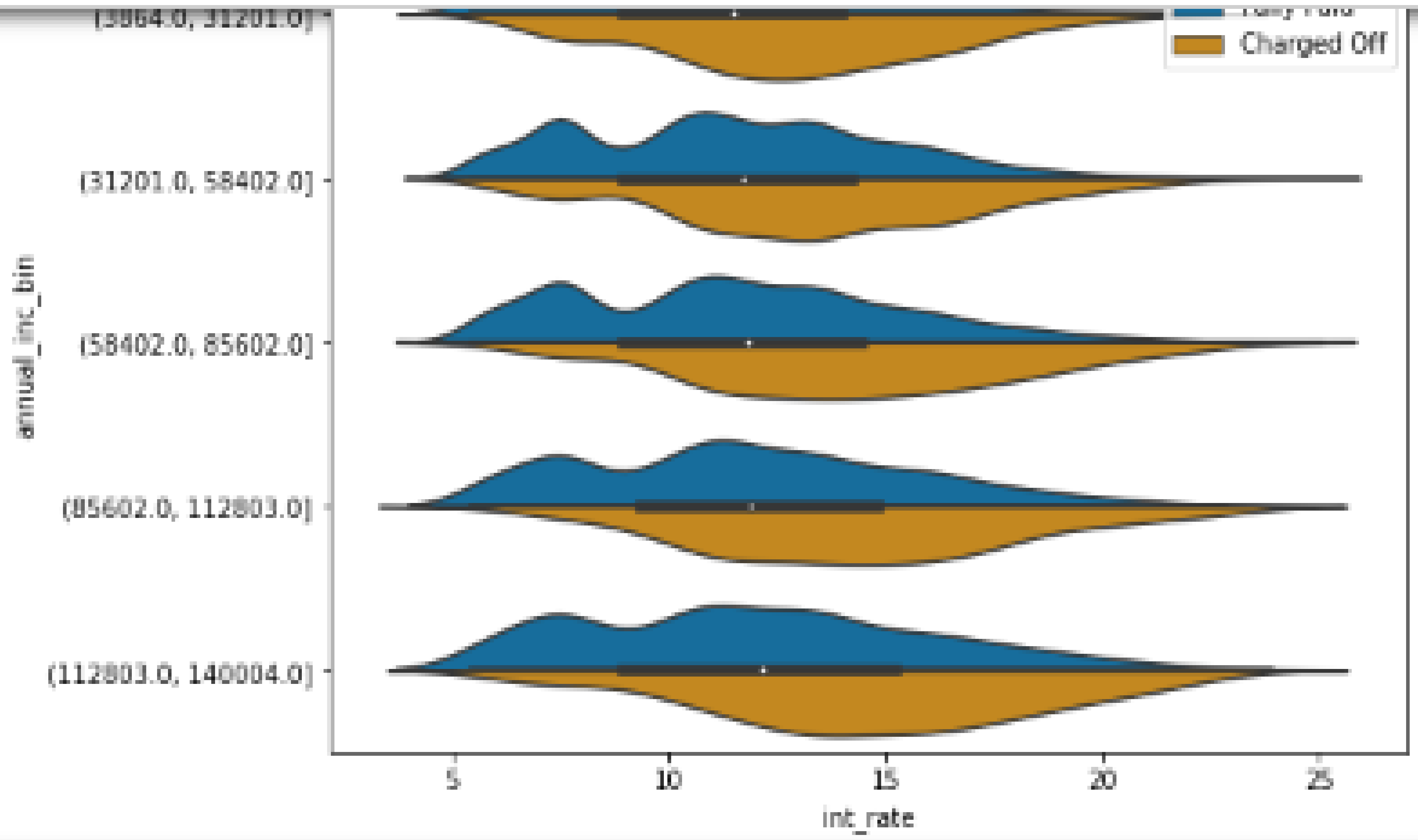

```
Observation : Loan term has an impact on defaulting for the lower income groups and for 36 Months. As the income increases,
term impact reduces. For 60-month term, the defaulting possibility is reduced as the borrowers get more time to adjust their
payments.
```

Bivariate Analysis – Results




```
Observation : Possibility of defaulting increases with increased Revolving line utilization rate, across all the income groups.
As the borrower uses more credit, his chances of not being able to repay will increase.
```

Bivariate Analysis – Results



Observation : Significant impact of Interest rate. When the interest rate is between 9%-17%, possibility of defaulting increases irrespective of the income group.

Thank You