Lending Case Study

By,

Lakshmi Rao

Manik Kapoor

Problem statement

A consumer finance company which specializes in lending various types of loans to urban customers has to make a decision for loan approval based on the applicant's profile.

Two types of risks that are associated with the bank's decision are:

- 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

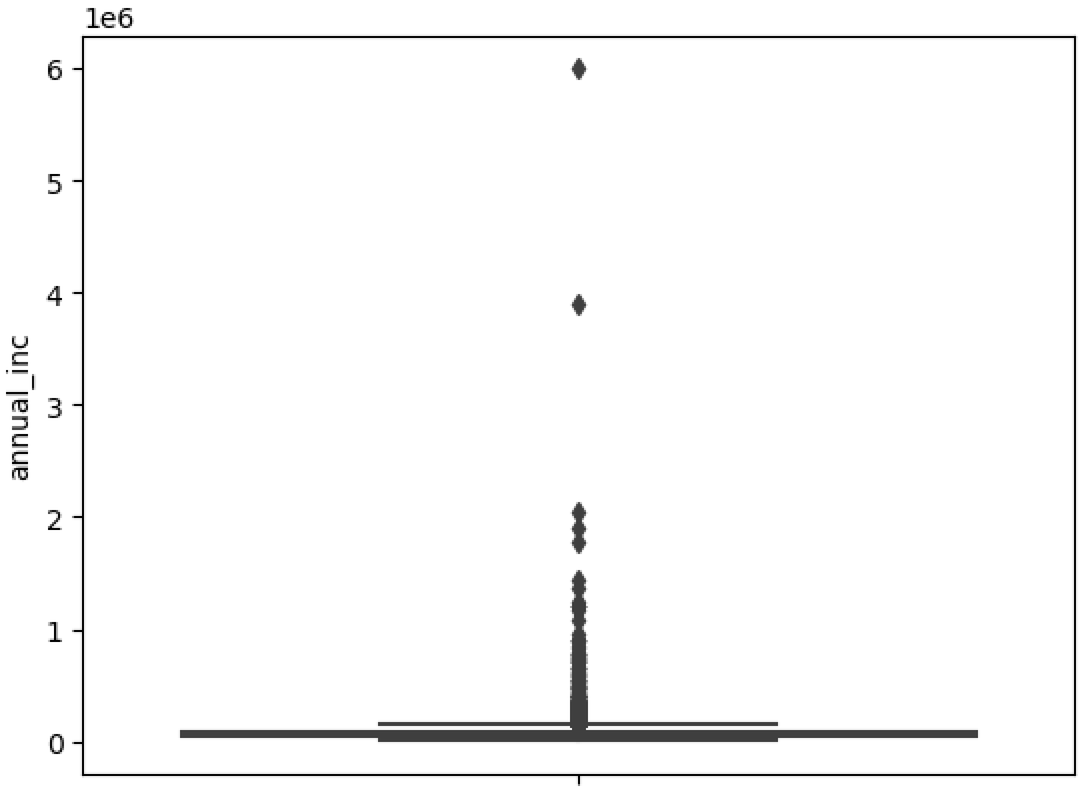
The lending case study leverages Exploratory Data Analysis (EDA) to uncover patterns in past loan applications. By examining consumer and loan attributes, we aim to develop insights that will guide decisions on loan approvals, optimizing our risk management strategy while maximizing business potential.

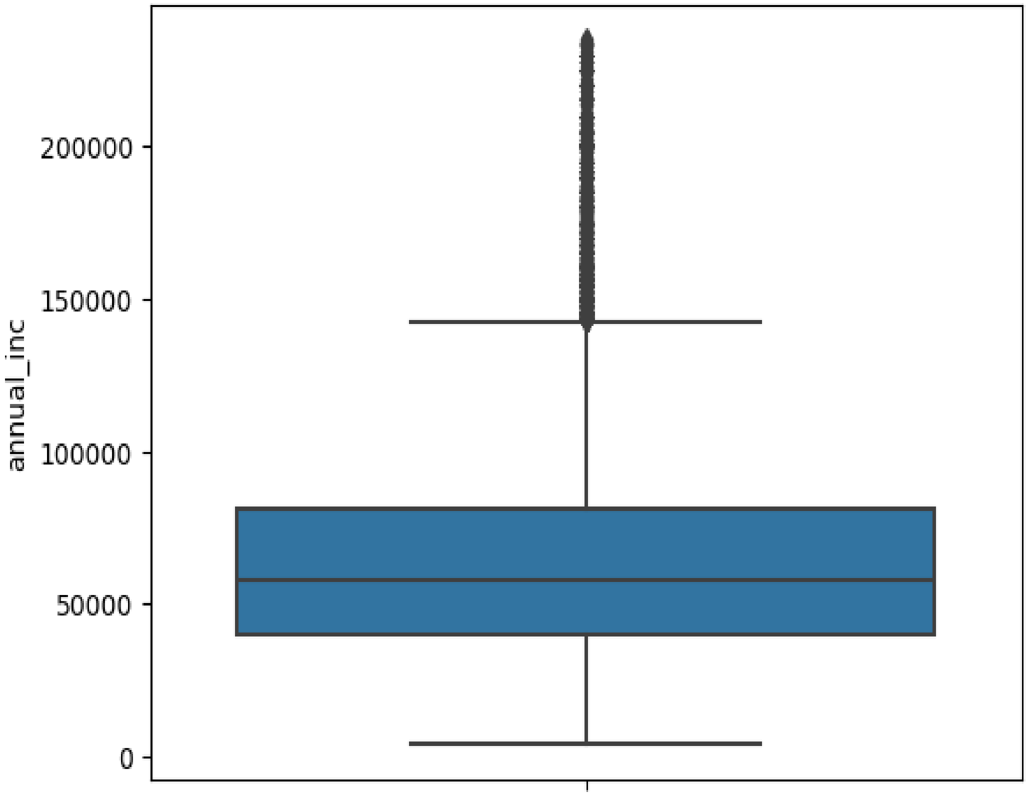
First step : Data cleaning

After importing the csv file, and reading it as a dataframe,

- The first step was to check for missing/null values, duplicate values and clear the dataframe of columns with all values as NULL.
- The data that we were provided, consisted of dirty values and missing values. We assess this data for any missing values and implement appropriate methods to address them.
- A few of the columns all has their values as NULL which were omitted from the analysis.
- Duplicate records were checked for and removed.
- A few of the rows that contained any values that were considered very high as outliers were removed.

Visualization of outliers





The visual shows the outlier lying on the 100th percentile to be around 600000, hence to neutralize the value, we considered and dropped any value for annual income greater than the 99th percentile.

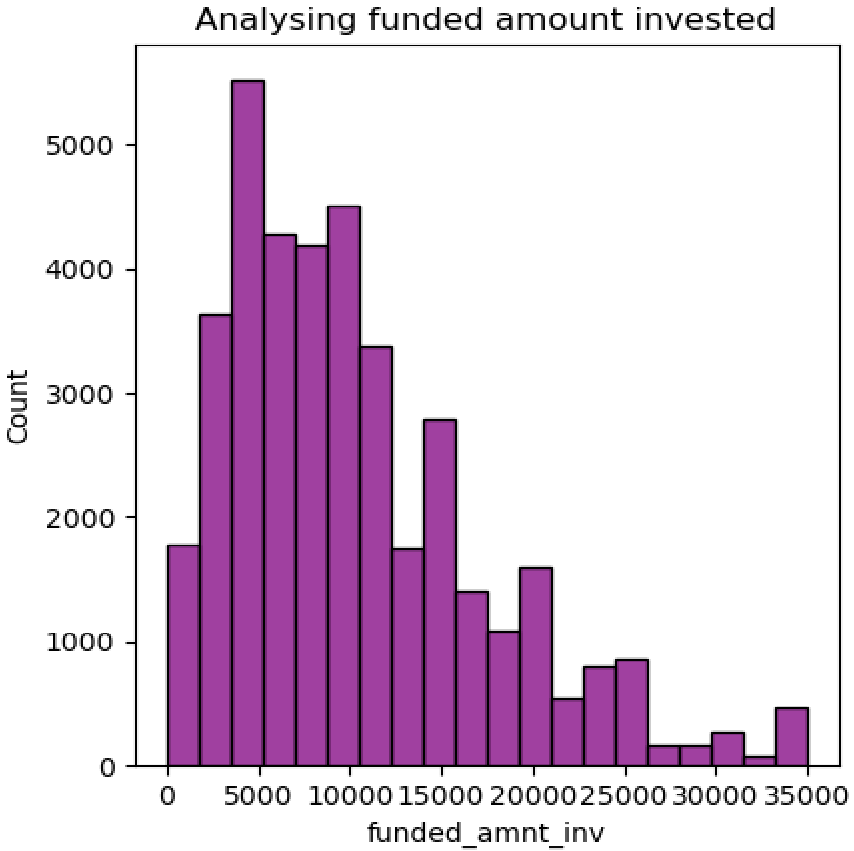
Insights after the data cleaning process

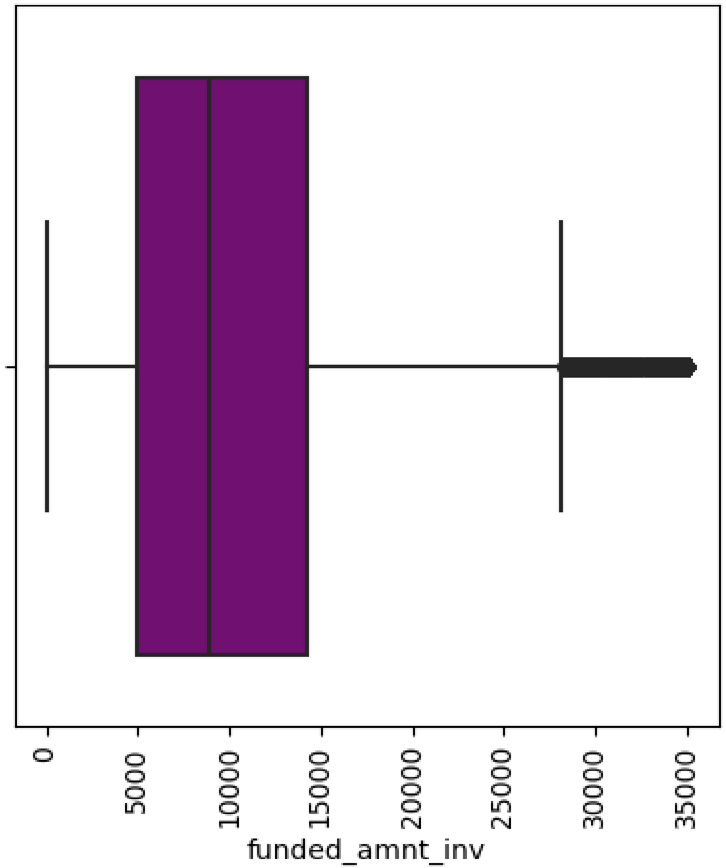
- The verification status are categorised as not verified, verified and source verified.
- The average interest rate on loans is 12%, with a minimum of 5.42% and a maximum of 24.59%.
- The average annual income of borrowers is 65524.21, with a minimum of 4000.00 and a maximum of 234996.00.
- The average debt-to-income ratio (DTI) of borrowers is 13.37, with a minimum of 0 and a maximum of 29.99.
- Most loans are for a term of 36 months and 60 months
- The status of loans are split into 3 categories as fully paid, charged off and current.

Univariate analysis

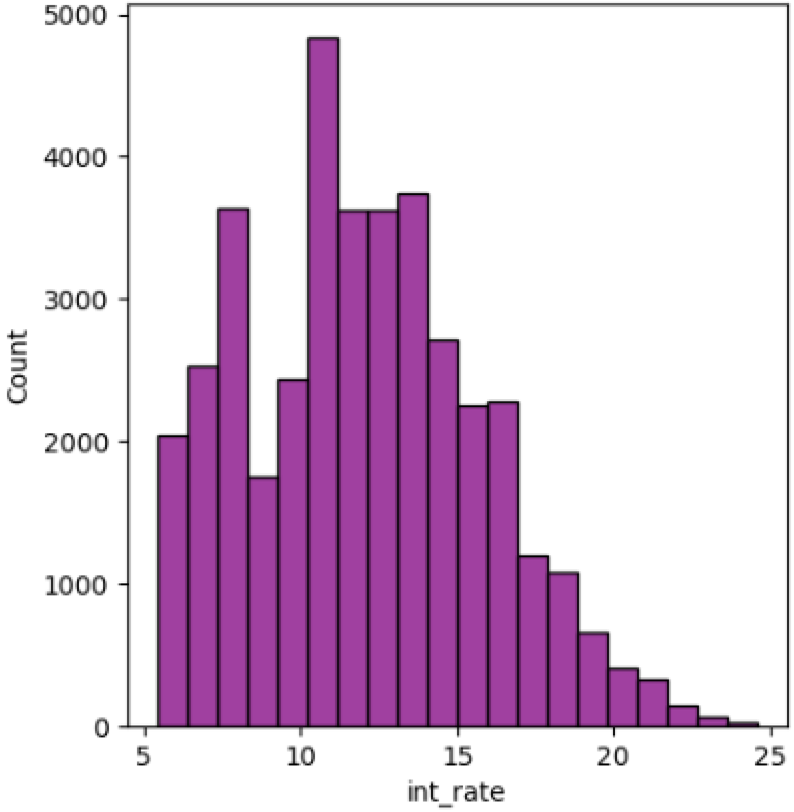
The below images(graphs) show that the values are correct for the analysis, all the outliers have been removed while cleaning the data.

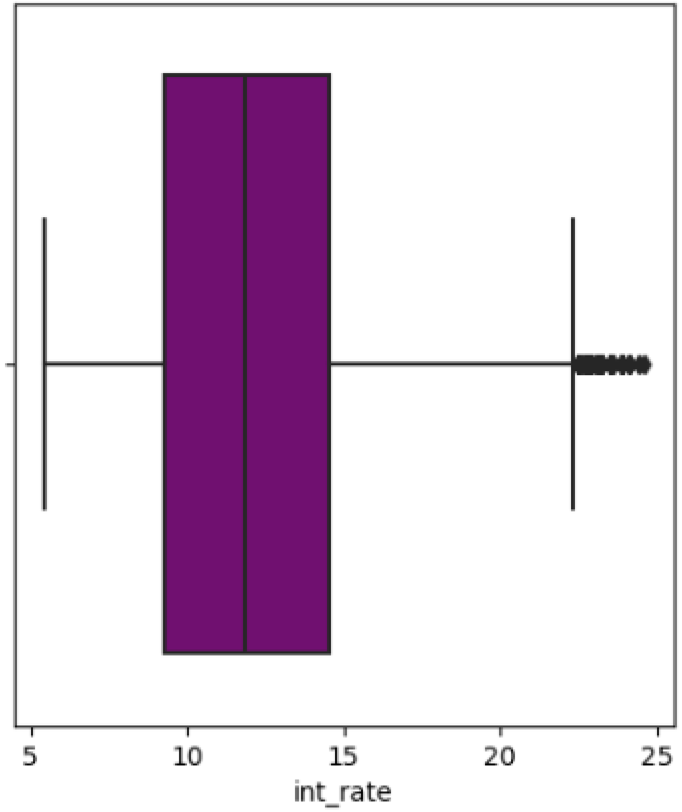
The number of funded amount invested by the bank is below 5000 is more than 5000 and for 35,000 is below 1000 counts.
As per the boxplot, the funded amount is 5000 for the 25th percentile and 8900 for the 50th percentile.



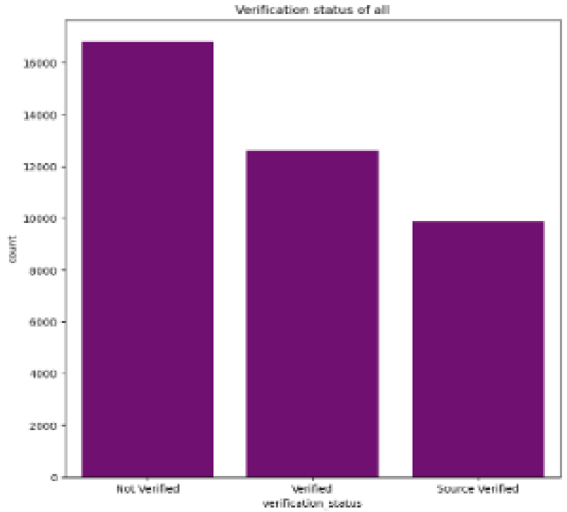


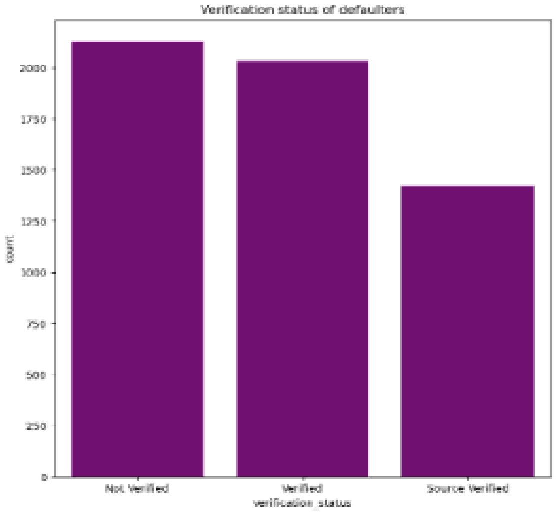
The number of funded amount invested by the bank is below 5000 is more than 5000 and for 35,000 is below 1000 counts.
As per the boxplot, the funded amount is 5000 for the 25th percentile and 8900 for the 50th percentile.

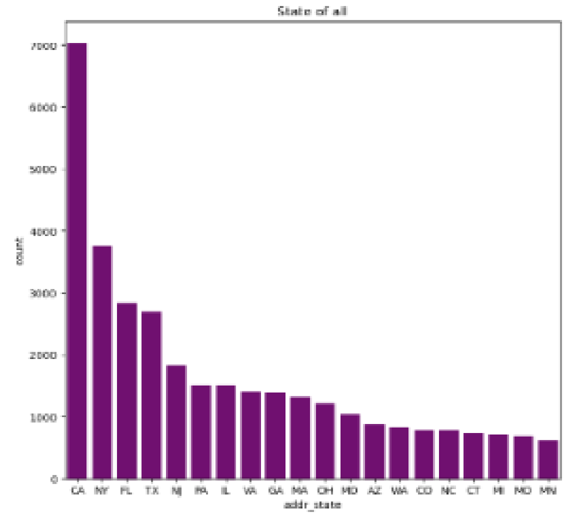


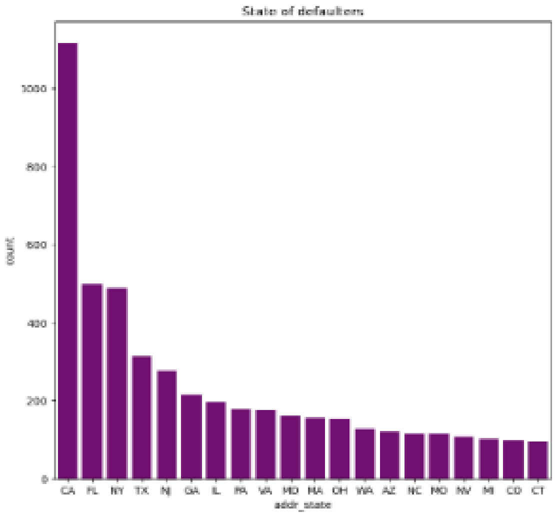


The analysis was done between all the loan holders and the defaulters to show a clear visual on how the prediction would sway.







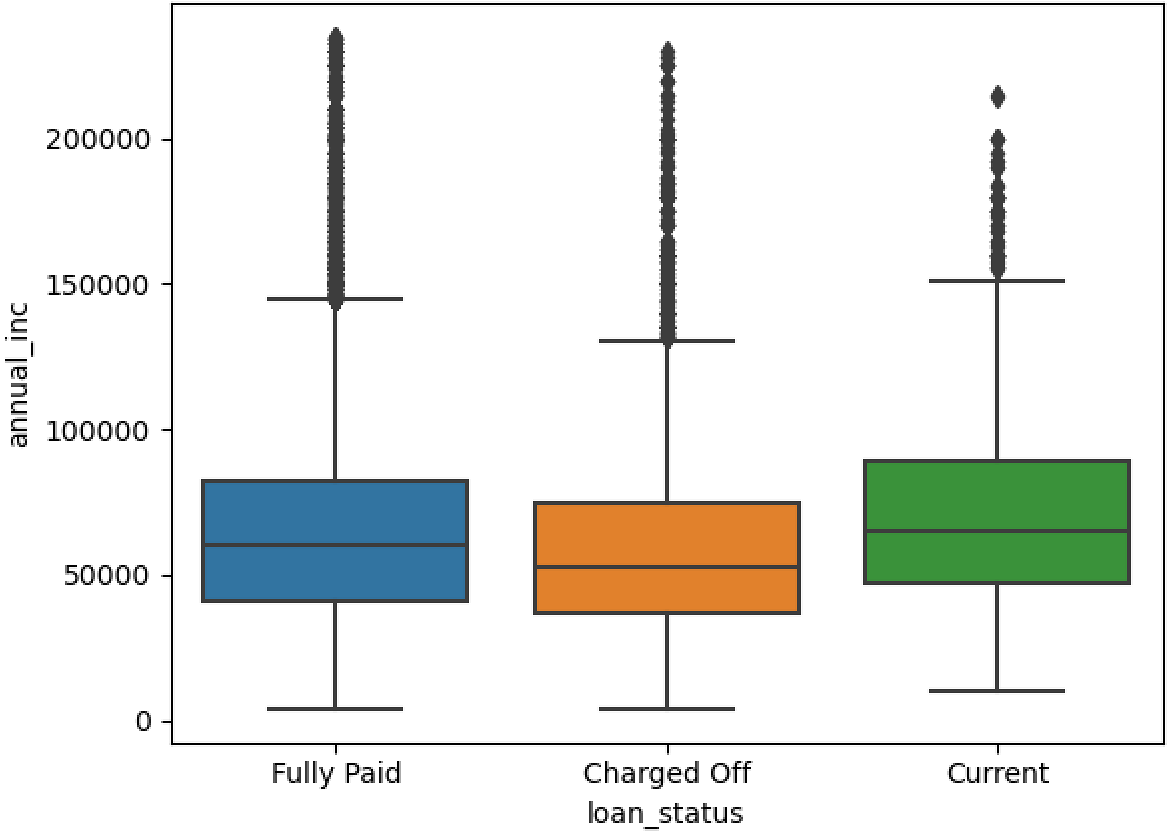


From the graph, the analysis can be such as

1. There is a high number of people who have not been verified before giving the loan
2. This also reflects under the defaulters that the large number of defualters were also not verified.
3. The highest number of people applied and approved for a loan were from state of CA, which in turn also shows that the highest number of people of defaulters are also from the same state.

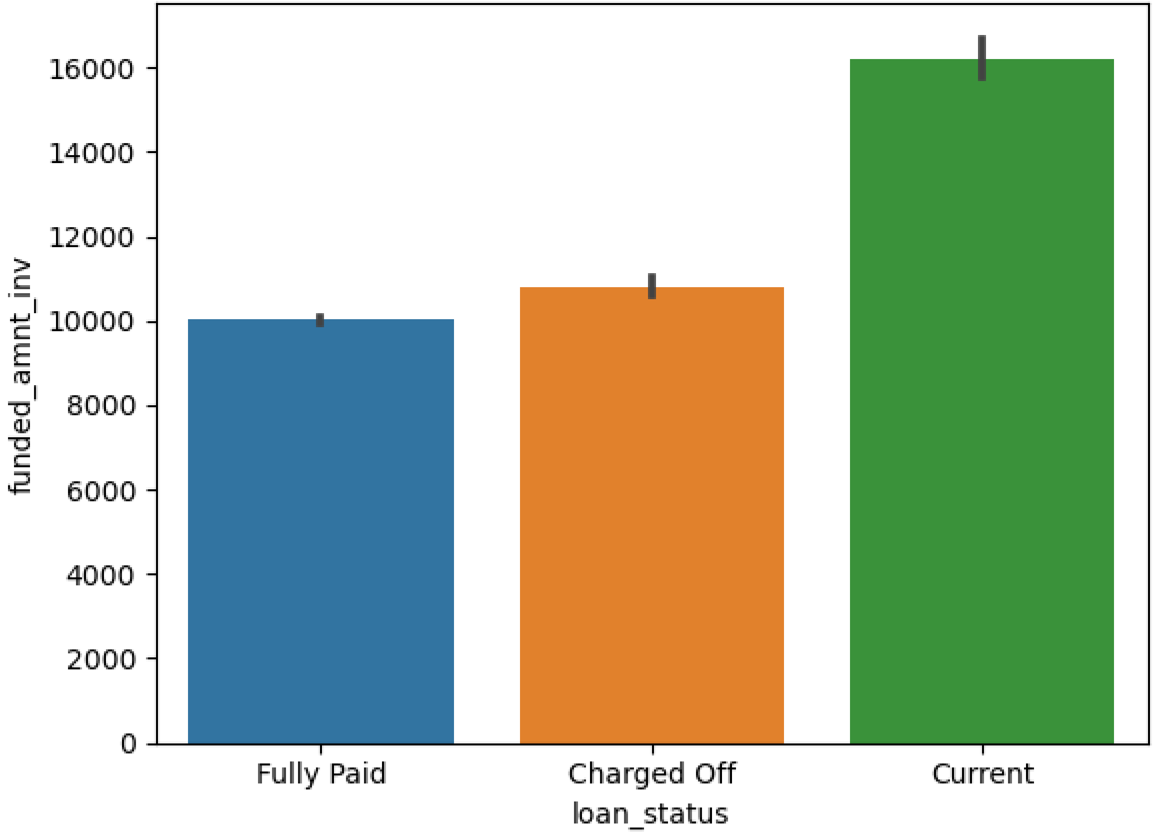
Bivariate analysis

The graph below shows that the defaulters were from the annual income below 5000.



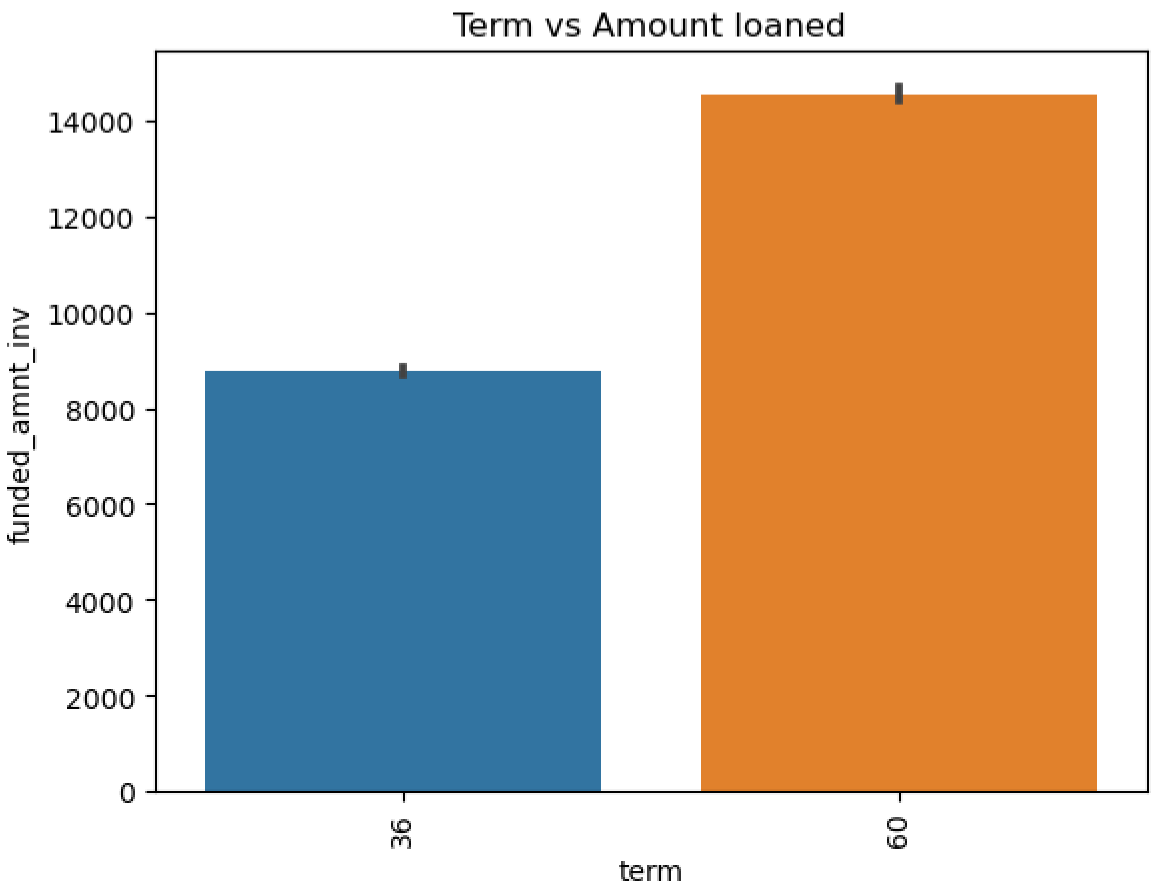
- The median annual income (the middle line in each box) appears to be slightly higher for the "Current" status compared to "Fully Paid" and "Charged Off."
- "Charged Off" loans seem to have the lowest median income.
- The interquartile range (IQR), represented by the height of the boxes, is relatively similar across all three categories.
- "Fully Paid" and "Current" statuses have slightly wider boxes, indicating a broader range of income in the middle 50% of the data.

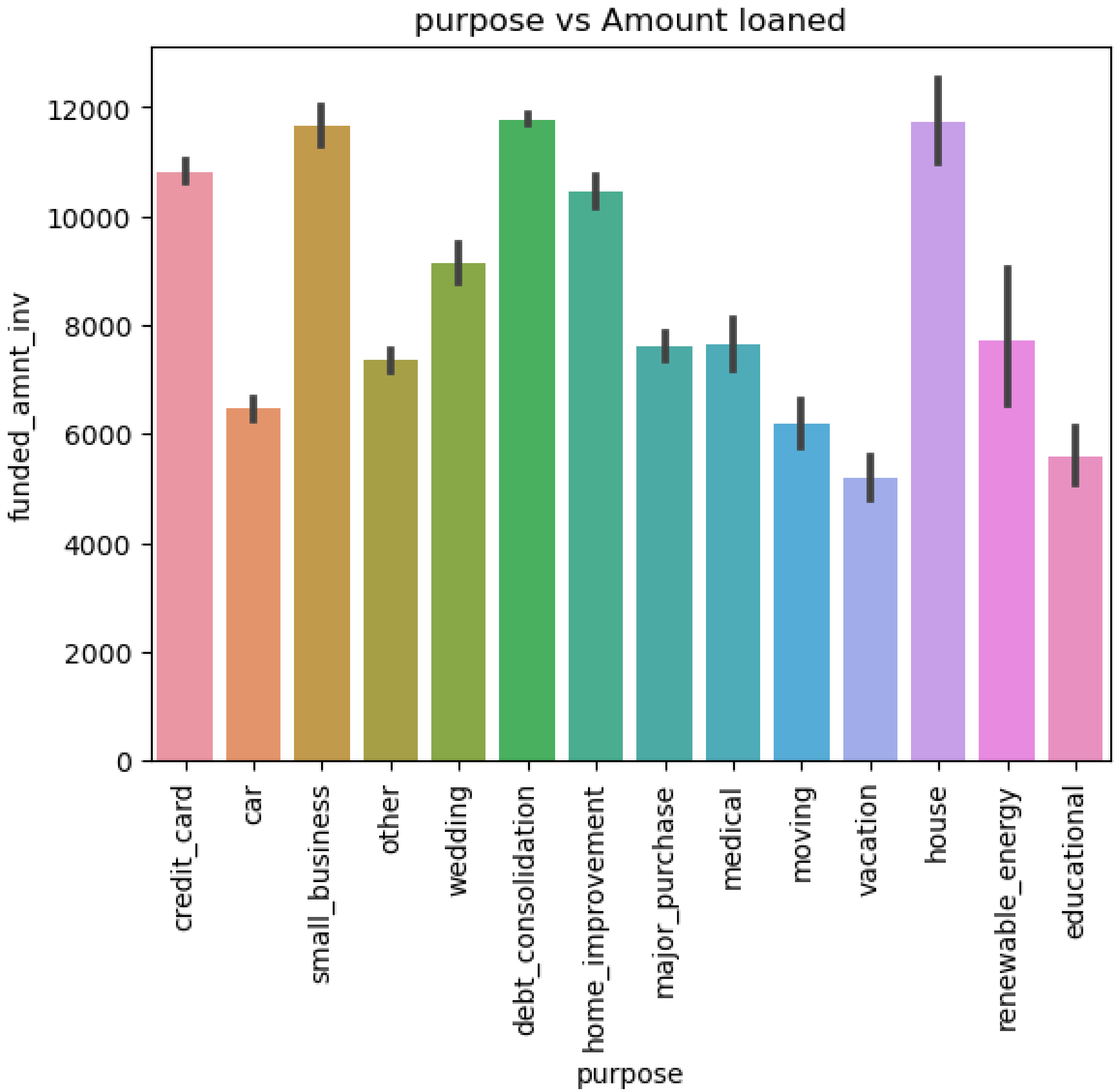
The image is a bar chart that shows the average funded amount invested (funded_amnt_inv) across different loan statuses (loan_status). The x-axis represents the loan statuses, which include "Fully Paid," "Charged Off," and "Current." The y-axis represents the average funded amount invested.



- The average funded amount for loans that have been charged off is slightly higher than that of fully paid loans, but still lower than current loans, with an average of around 11,000.

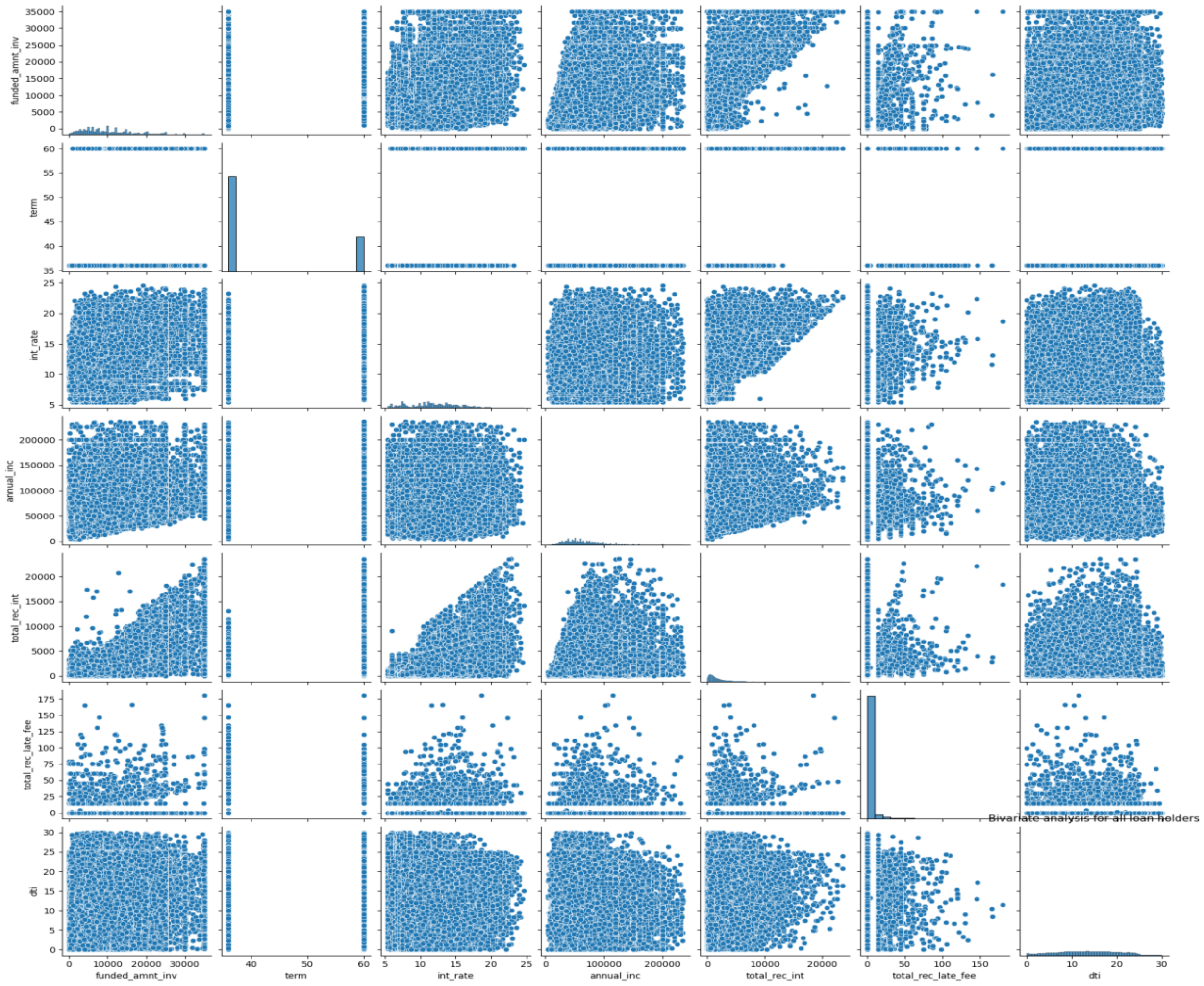
The image is a bar chart that shows the average funded amount invested (funded_amnt_inv) across different term. The terms are of 36 and 60. The highest funded amount investment can be seen for the longer duration of the loan for 60 .



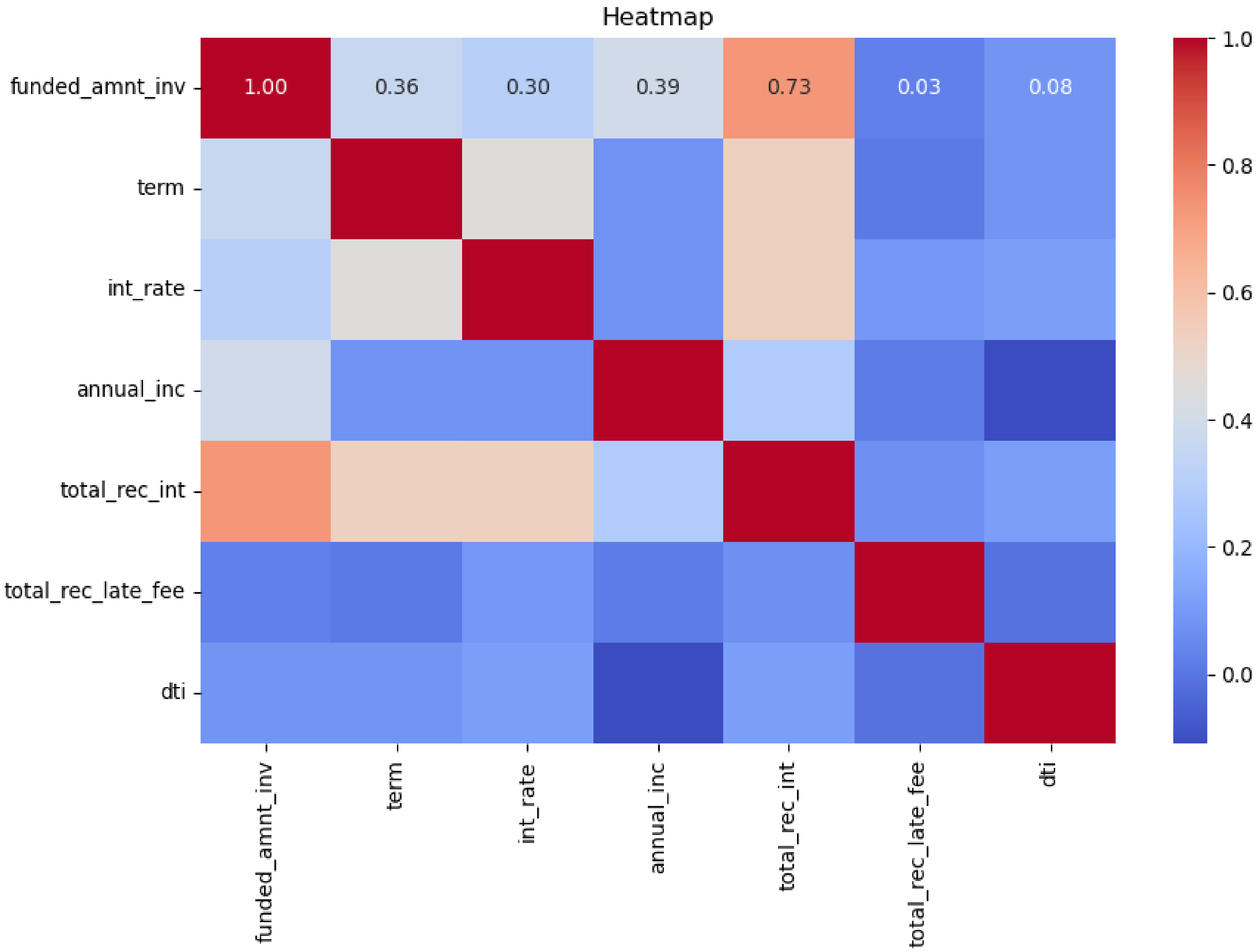


Debt consolidation, credit cards, small businesses and house improvement all have an average loan of more than 10000

The pair plot provides a visual summary of how these financial variables relate to each other, highlighting both strong and weak correlations



- There is a clear positive linear relationship between funded_amnt_inv and total_rec_int. This makes sense because higher loan amounts generally result in higher total interest received.
- A positive relationship is also seen between funded_amnt_inv and annual_inc, suggesting that higher-income individuals tend to borrow more.
- int_rate and dti (Debt-to-Income Ratio) may also show some correlation, which could suggest that higher debt levels might be associated with higher interest rates.



- Interest rate and term have a significant correlation.
- Loan amount by the bank and received interest total are also correlated.

Conclusion

After analyzing the data of the past and current loan holders,

- Individuals with lower incomes are more likely to have their loans charged off, as indicated by the lower median income in the "Charged Off" category.
- The distribution of income among those who are current on their loans or have fully paid them off is somewhat similar, though the median is slightly higher for those who are current, potentially indicating better financial stability.
- The income may have some correlation with loan outcomes, with those having higher incomes being slightly more likely to keep their loans current or fully paid off.
- The fact that charged-off loans have a slightly higher average funded amount than fully paid loans might imply that higher amounts may correlate with a higher risk of default, although the difference is not large.
- Overall suggestion, is that larger loans are more likely to still be active (current), while smaller loans are more likely to be fully paid off. The risk of loans being charged off appears to be slightly higher for loans with a moderate amount of funding.