



LENDING CLUB EDA


UNDERSTANDING THE DRIVERS BEHIND LOAN DEFAULT

PROBLEM STATEMENT

- Lending club is a consumer finance company which wants to identify the driving factors behind consumer defaulting.
- There are two risks associated with banks 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

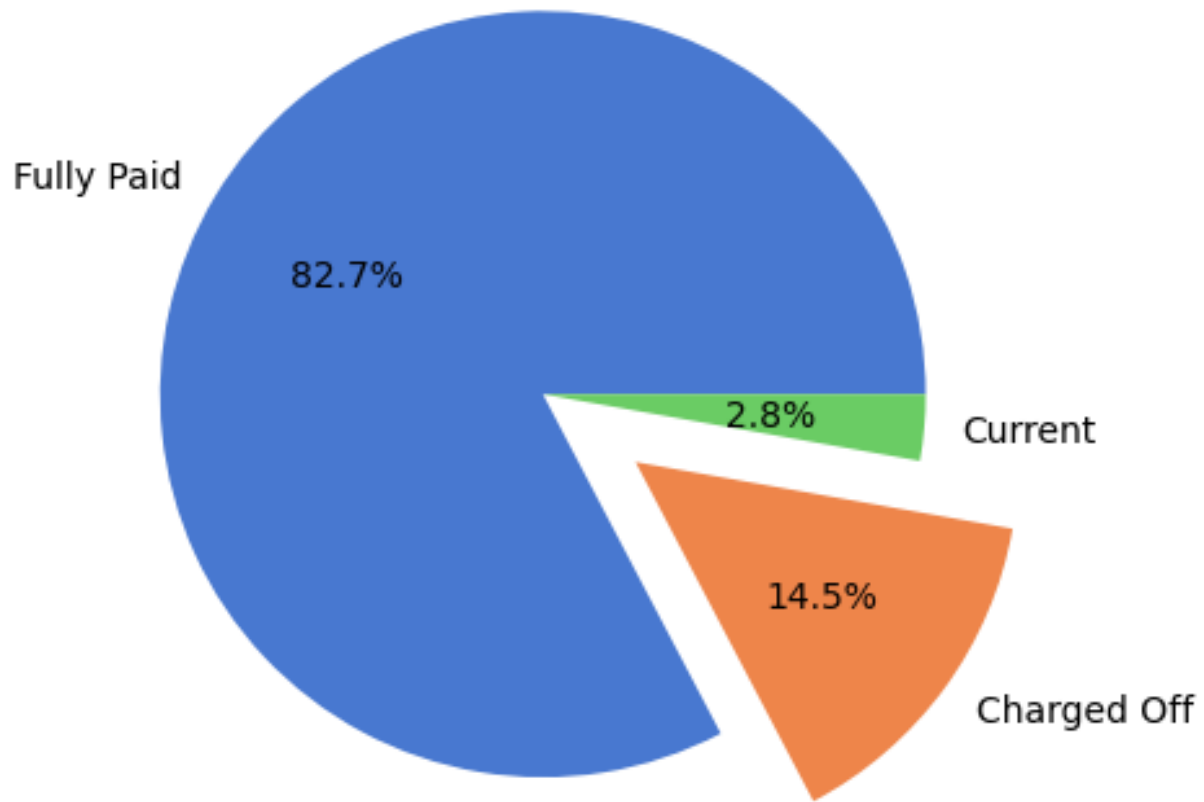
DATA CLEANING

The background features a series of flowing, translucent ribbons in vibrant colors: red, orange, yellow, green, and blue. These ribbons curve and swirl across a solid black background, creating a sense of dynamic movement and depth. The lighting on the ribbons gives them a three-dimensional appearance, with highlights and shadows that emphasize their fluid, ribbon-like texture.

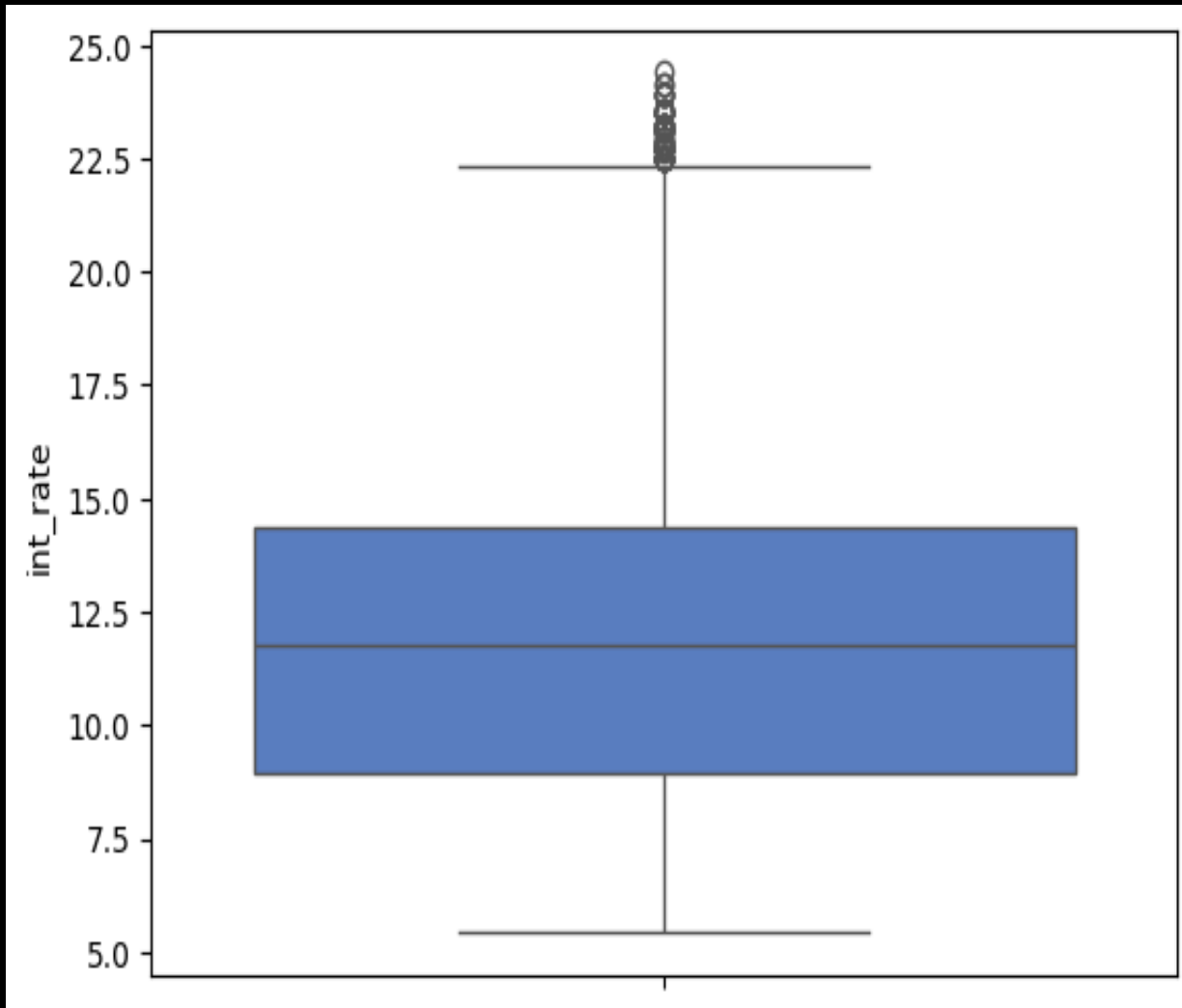
- 
- FIRST PART WAS DATA UNDERSTANDING AND CLEANING. IN THIS WE IDENTIFIED SOME DATA WHERE THE MISSING NUMBER OF ROWS IN A PARTICULAR COLUMN WERE $> 30\%$ WE STRATIGHAWAY DROPPED THOSE COLUMNS.
 - FOR EMP_LENGTH WE CLEANED AND CONVERTED THOSE VALUES TO NUMERIC FORMAT. WE REMOVED '%' SYMBOLS IN REVOL_UTIL AND INT_RATES USING FUNCTIONS SO THAT WE DON'T DO REPETITIVE TASK. FOR PUB_REC_BANKRUPTCIES WE IMPUTED '0' INPLACE FOR MISSING VALUES.
 - FOR ANNUAL INCOME WE IDENTIFIED SOME OUTLIERS BECAUSE THE MEAN AND THE MEDIAN WERE HAVING A HIGHER DIFFERENCE IN VALUES AND ALSO THE ANNUAL INCOME WAS DEPENDANT ON EXTERNAL FACTORS.
 - AFTER ALL THESE PROCESSES WE WERE LEFT WITH 35692 ROWS AND 40 COLUMNS ONTO WHICH WE DID OUR EDA.

UNIVARIATE ANALYSIS

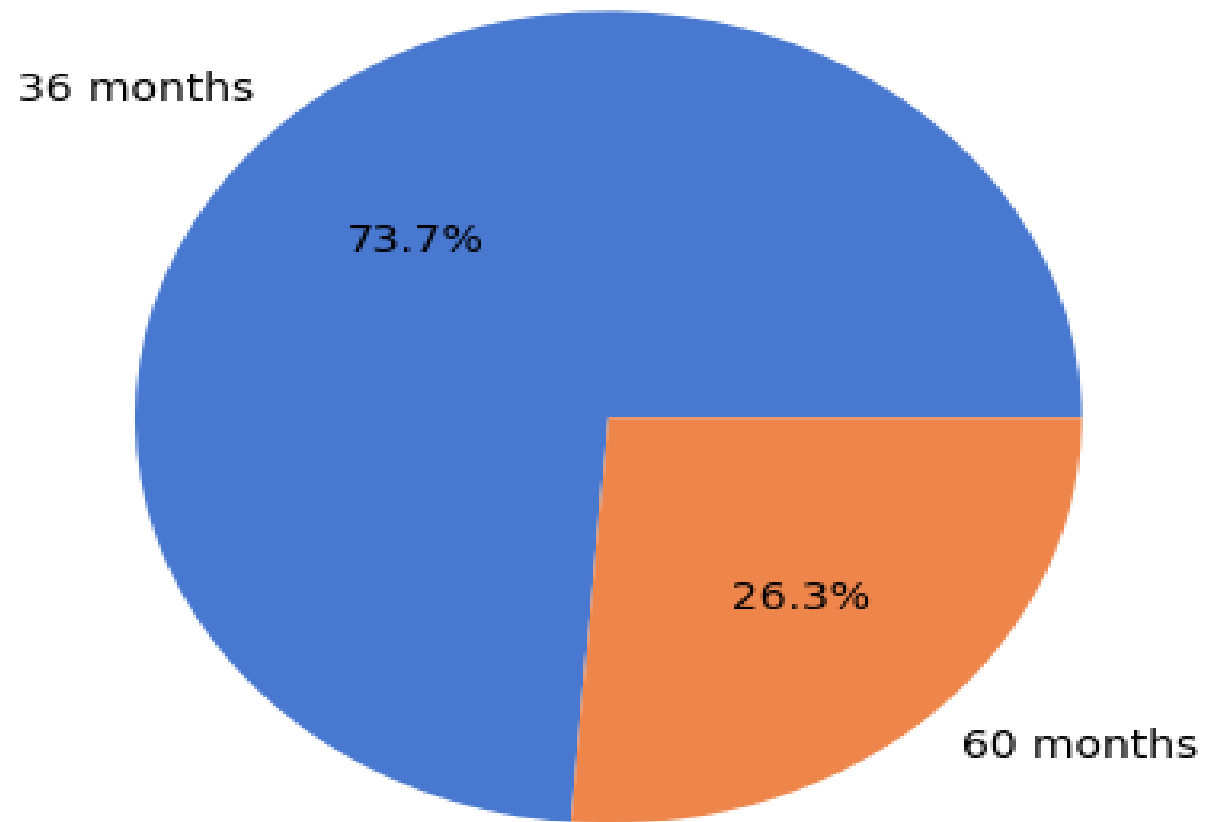
The background features a series of flowing, translucent ribbons in vibrant colors including red, orange, yellow, and blue. These ribbons curve and swirl across a solid black background, creating a dynamic and modern aesthetic. The lighting on the ribbons gives them a three-dimensional appearance.



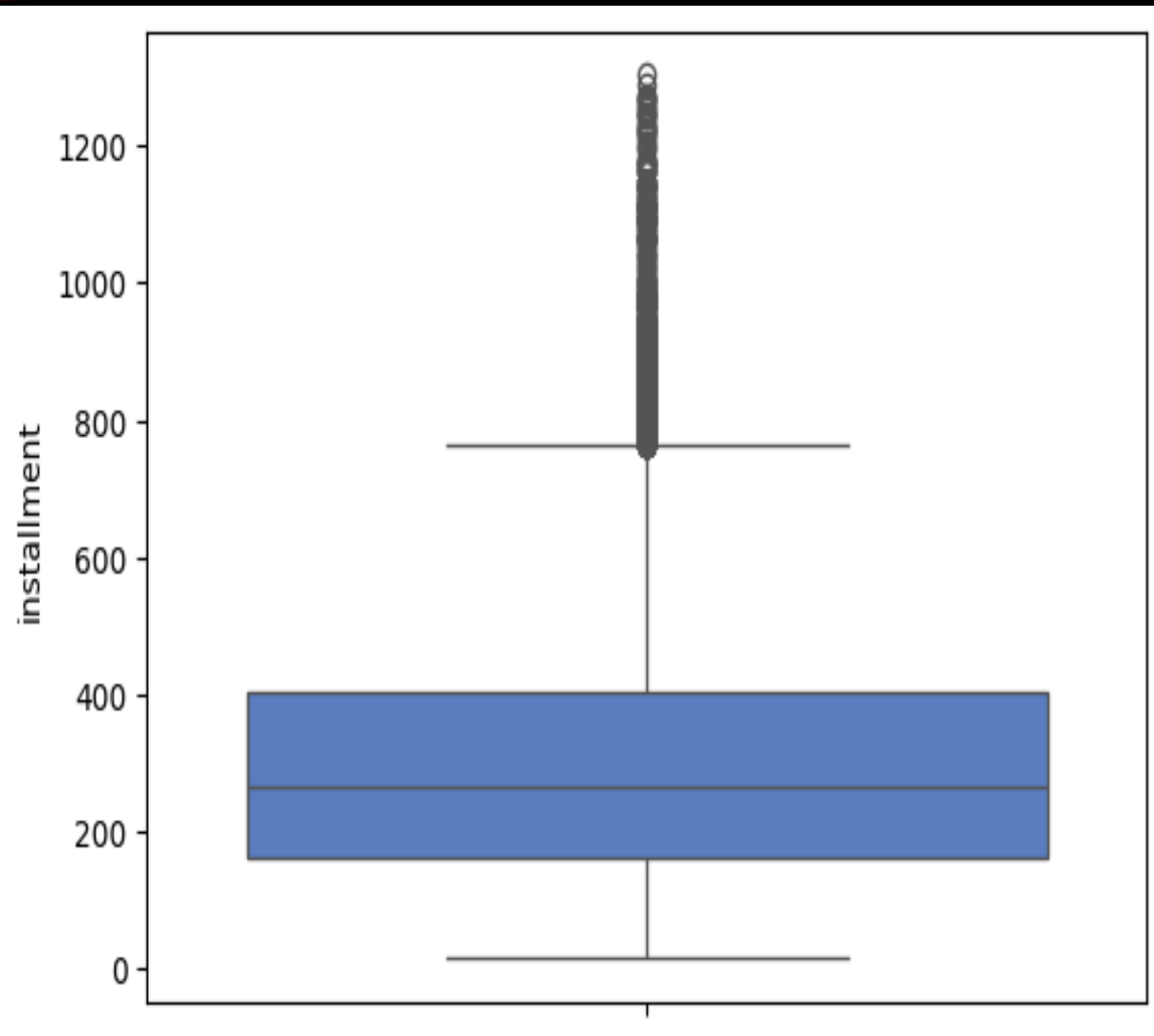
- Here we can see there are 14.5% of people as defaulters.
- Meaning out of every 100 customers we can get 14 defaulters.



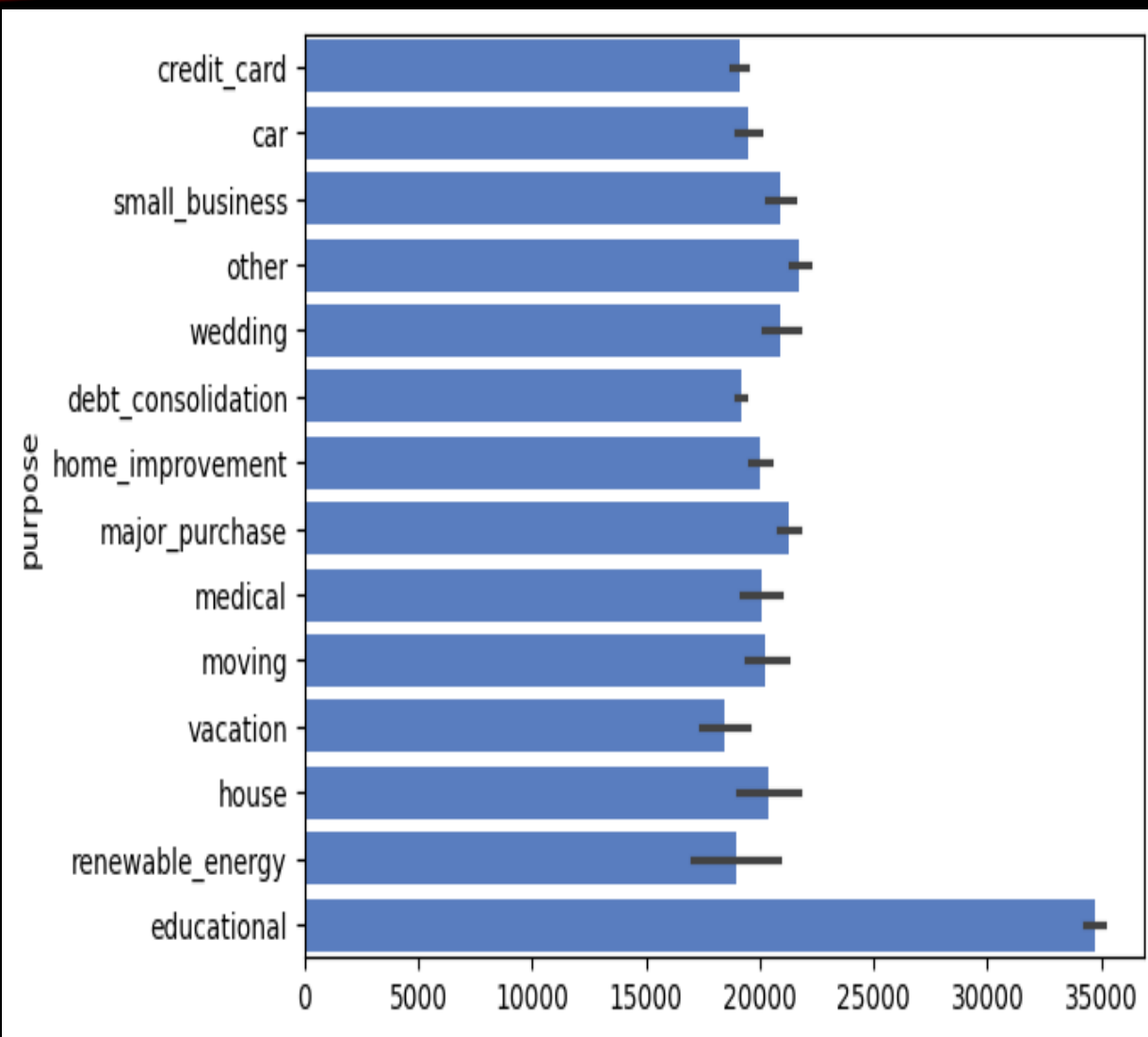
- Interest rates stand at a range of mostly around 9% to 15%.
- Average interest rate stands at around 12%.
- Lowest interest rate stands at around 5.42%.
- Highest interest rate stands at around 24%.



- The most number of loans are for 36 months.
- So most people go for lower tenure meaning they have higher instalment amounts as compared to 60 month term.



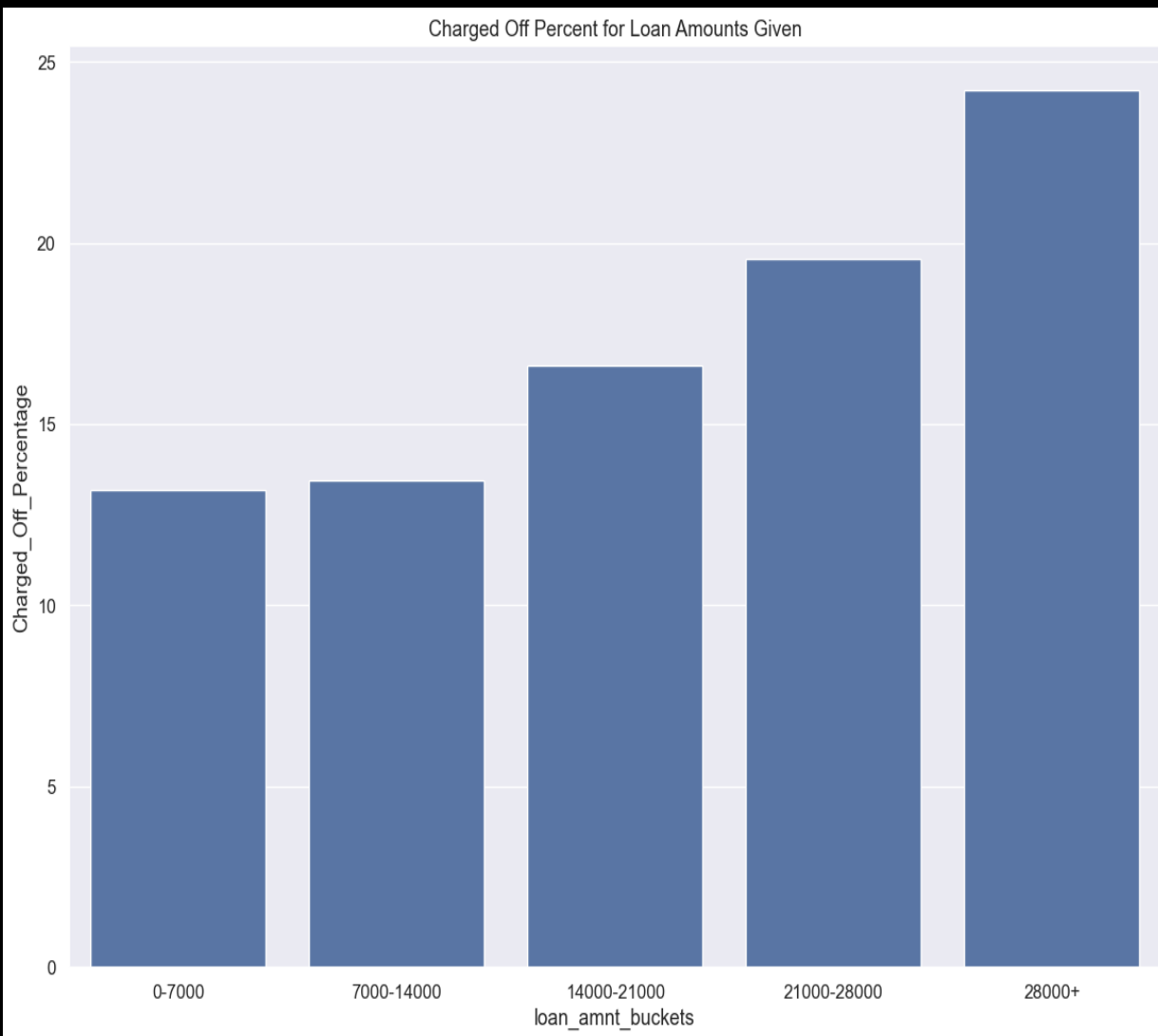
- Here we see that majority of instalments are at a range of 200 to 400.
- The average instalments stand at around 300.
- The minimum instalment is around 16.
- The maximum instalment is around 1300.



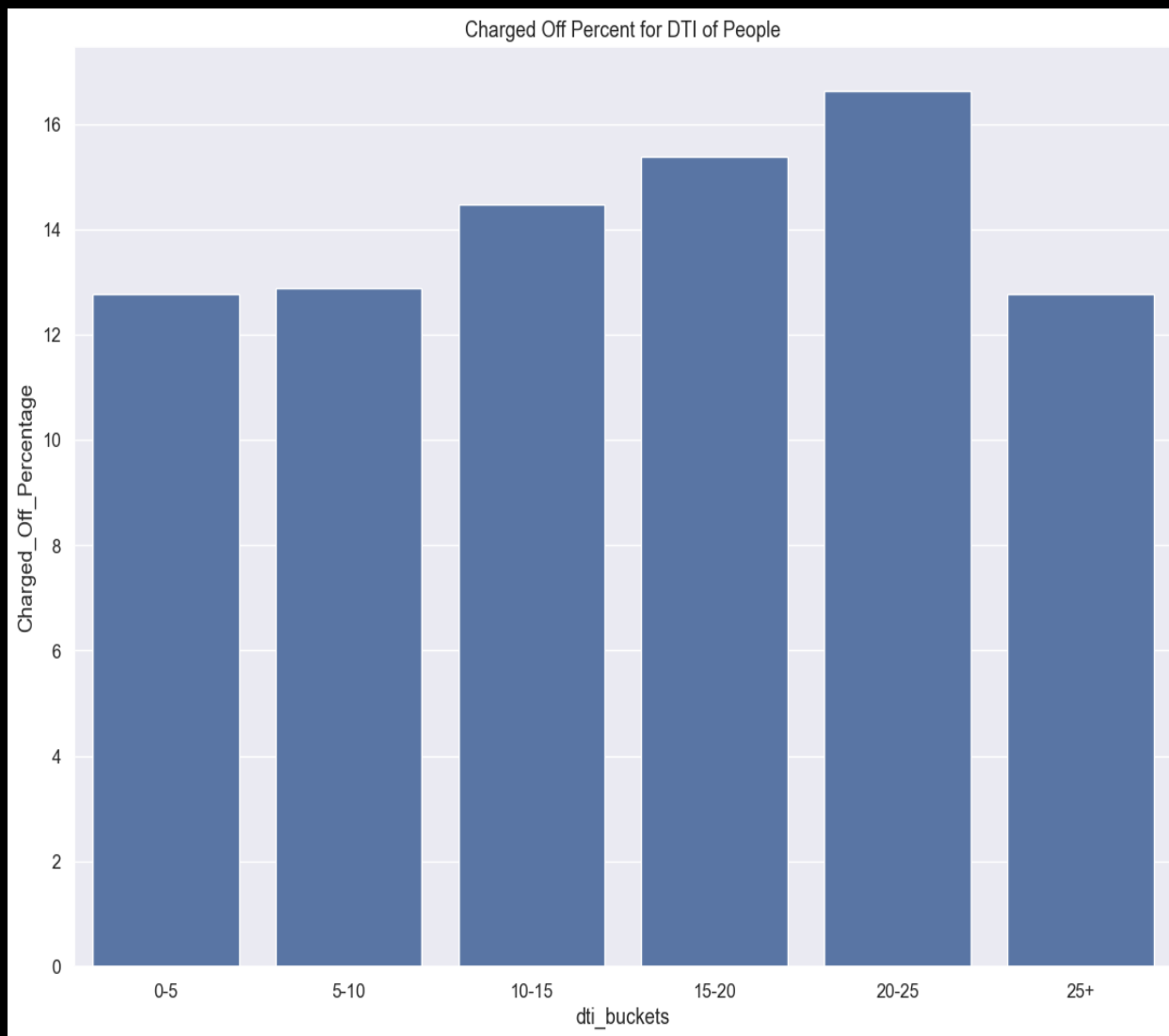
- The most number of loans taken are for education.
- Also we see some loans also taken for credit card, debt consolidation etc.



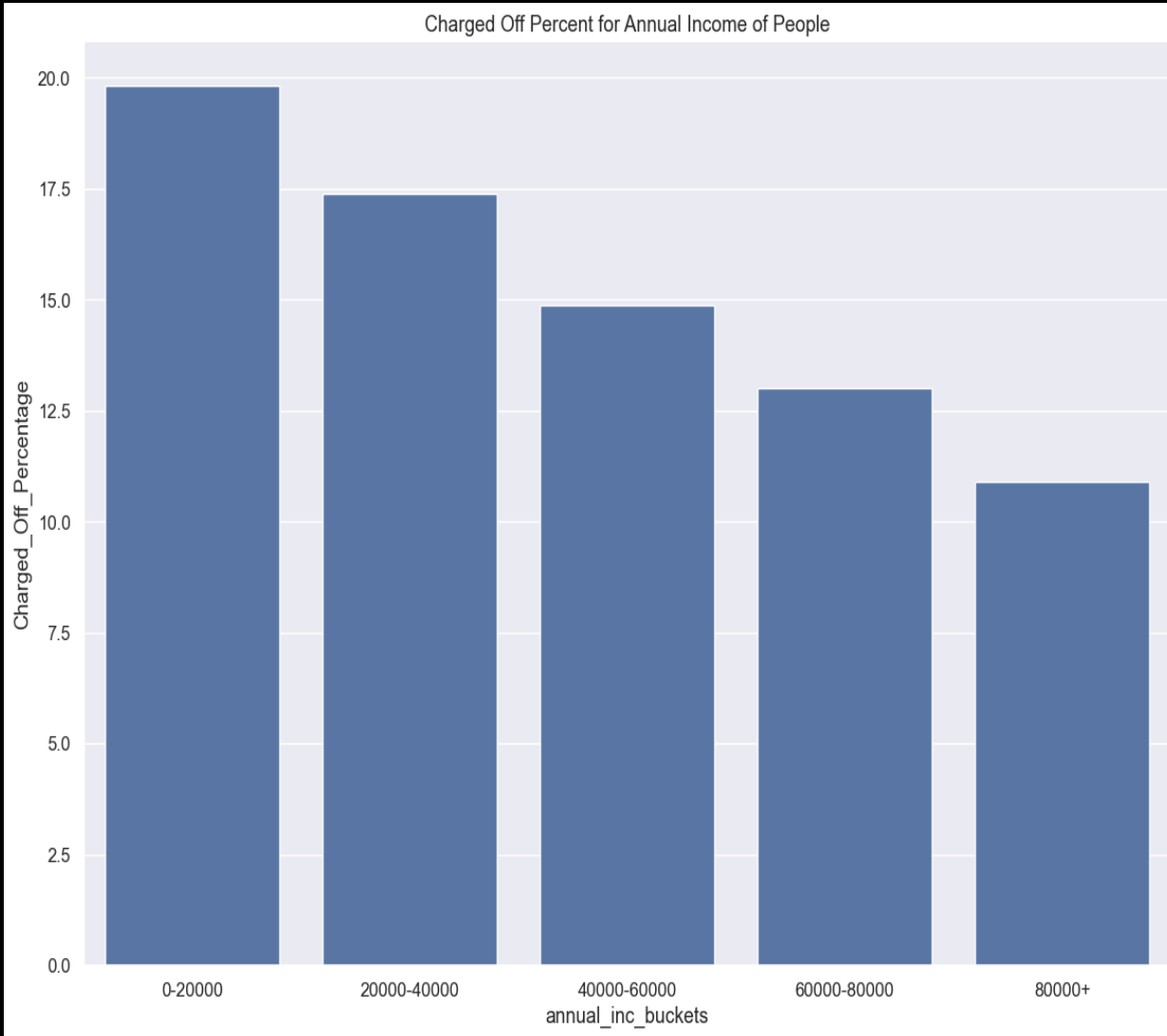
BIVARIATE ANALYSIS



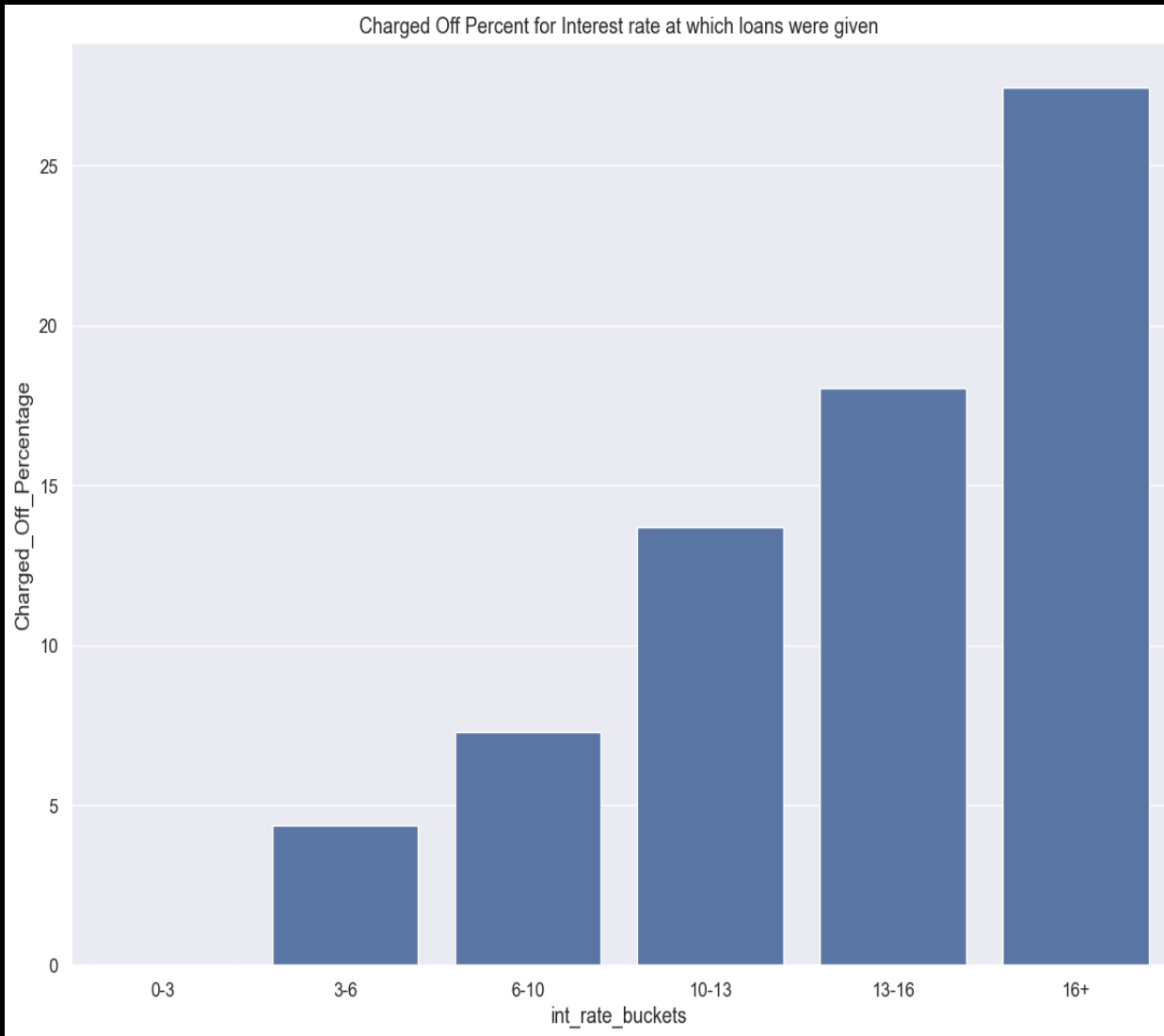
- We can clearly see that the loans given for amounts of 28000+ are a huge problem for the business.
- Business has to be careful with high ticket loan amounts.



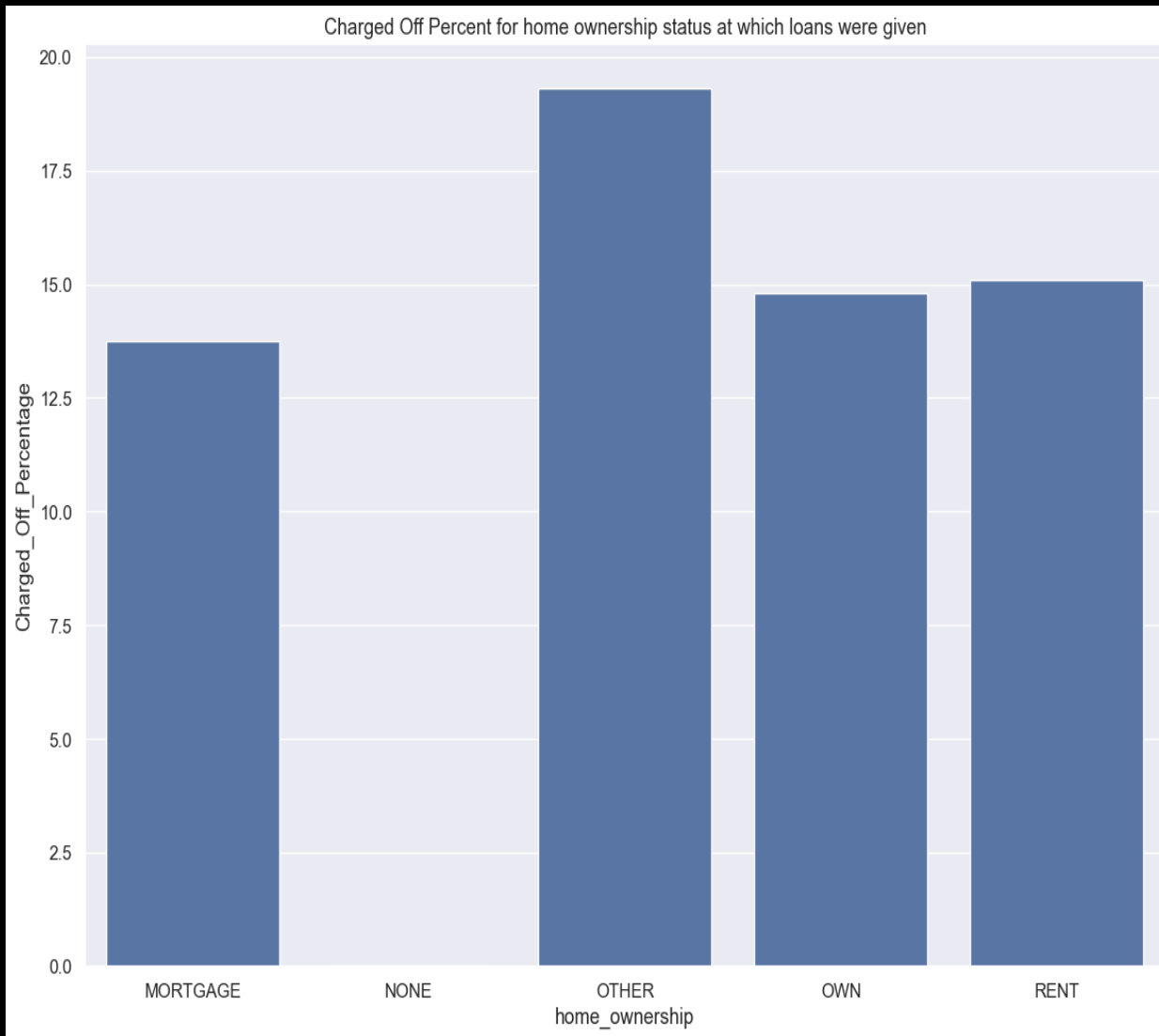
- The DTI (Debt To Income) ratio tells us that the people in the 15 to 25 range are the most probable to default in general.



- Here is where we can clearly see people with low income which is 0 to 20000 are the most important factors contributing to the defaulters category.



- Here we can see that the people with higher interest rates are a good and easy way to make money but are also very high risk customers.
- Hence we can very clearly see that the people who were given loans with interest rates $> 16\%$ were defaulting.
- This might be due to the bad DTI and also low income where the annual income of the people was also bad.



- Here we can say that the people who generally tend to hide their home ownership status are defaulting.
- Meaning that the chart shows people who don't tell you that if they are living on rent or its their own house or if it is mortgaged are the people with probably some bad intent and tend to default.

CONCLUSION

- Loan amount, dti, annual income, interest rates, home ownership, verification status, state, zip codes and some months are some of the factors where we found major number driver variables.
- The business has to first keep a close eye on loan amount, dti, annual income and instalment amounts.
- Also not blindly trust verification status as we saw verified customers are the ones contributing the most for loan defaults.