

Lending Club Case Study

Understand driving factors behind loan default

Kokila Pawar
Prabhu Lingam

ANALYSIS

ROADMAP

- Data Understanding
- Data Cleaning
- Derived Columns
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Conclusion

DATA UNDERSTANDING

- No summary, header and footer rows are identified

Information provided by data :

- Customer Demographics : Annual income, Home Ownership, Employment Length, Debt to Income Ratio, Open Account, Zipcode and few other columns
- Loan Attributes : Loan amount, Term, Loan date, Purpose of loan, Interest rate, Verification status, Installment
- Major data issue : There are lots of columns with null values

Data Cleaning

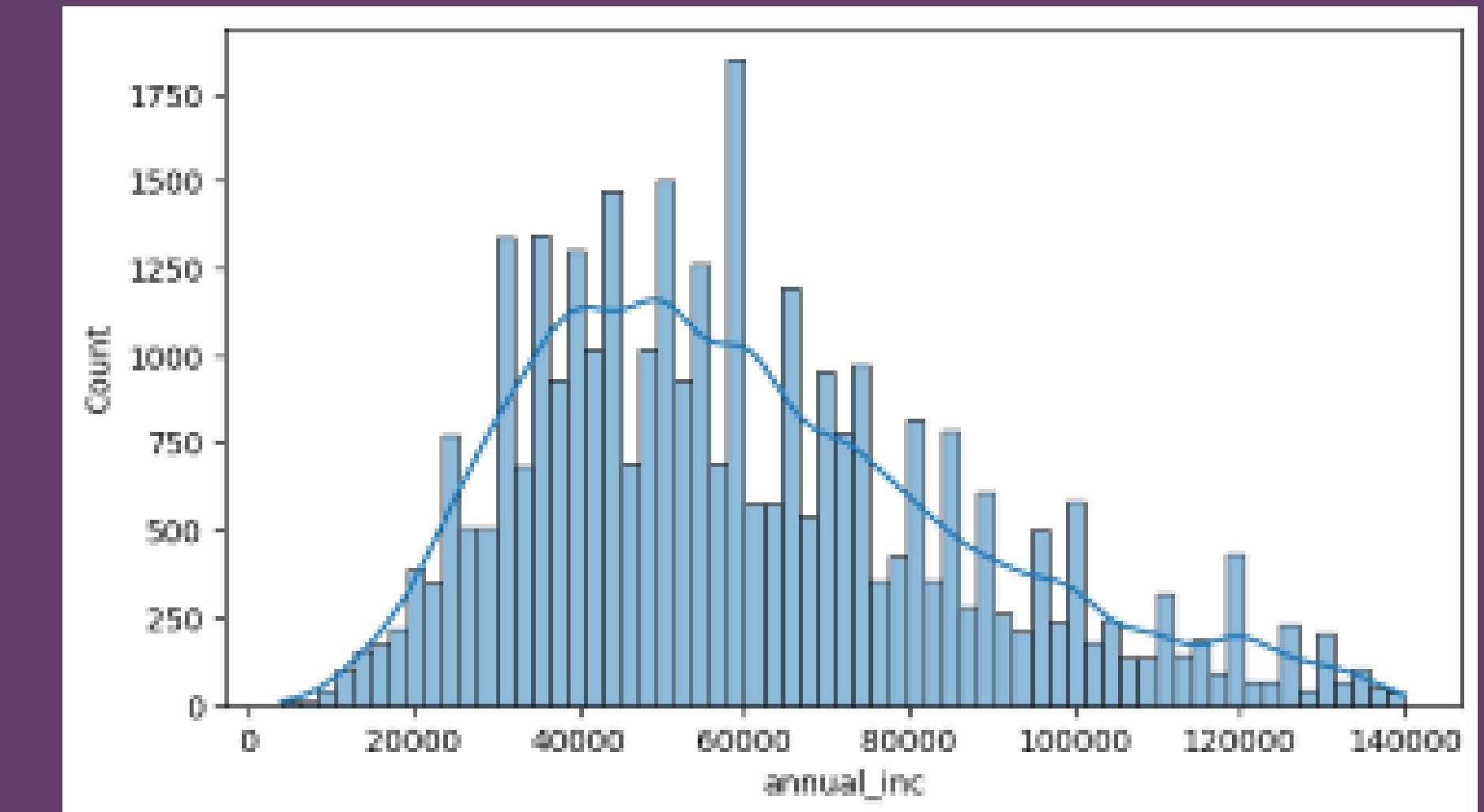
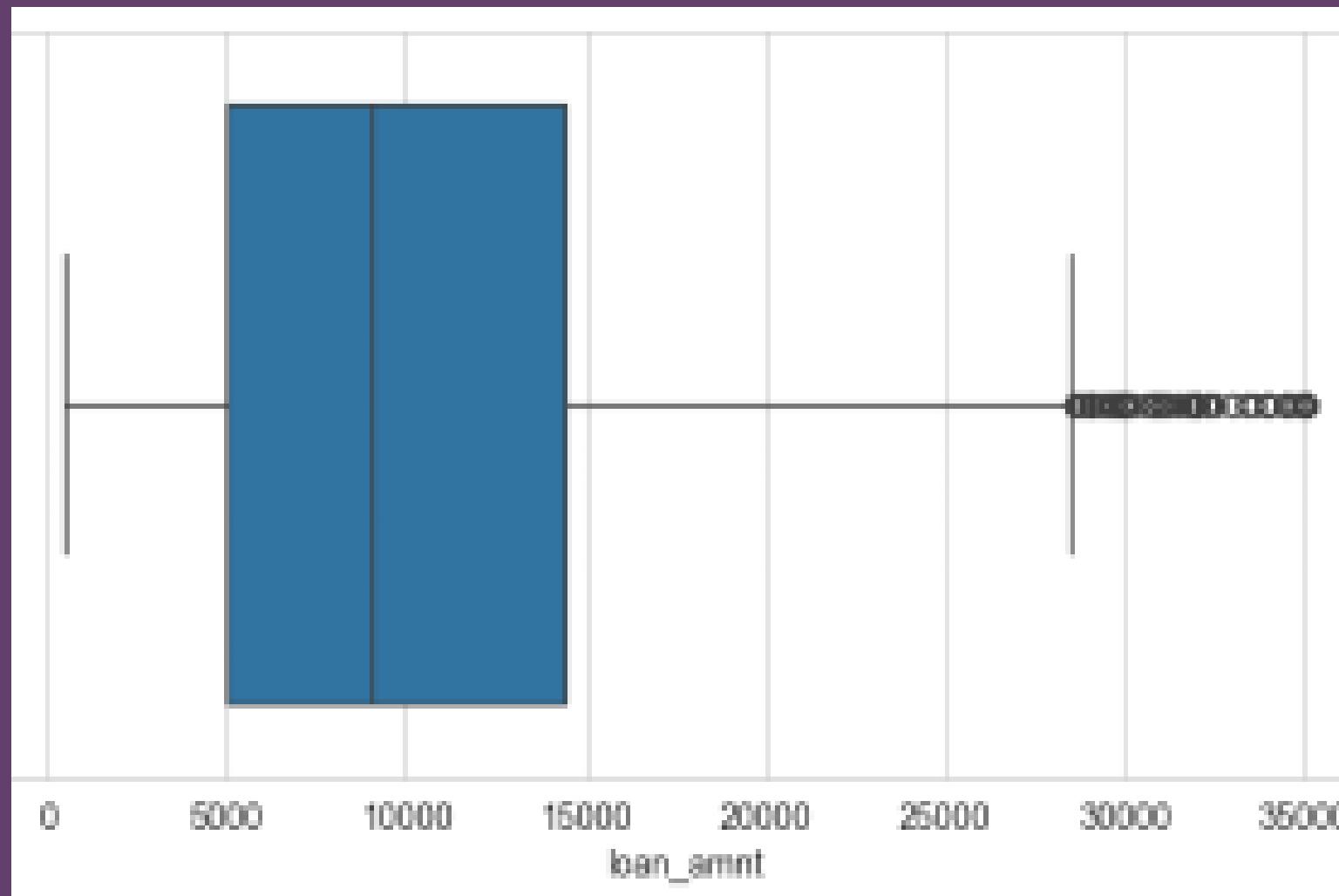
- Drop columns that contain more than 60% null values
- Drop columns that do not add value to analysis
- Subset historical data i.e Fully Paid and Charged Off columns
- Convert term, interest rate and revolving utility to suitable formats

Derived Columns

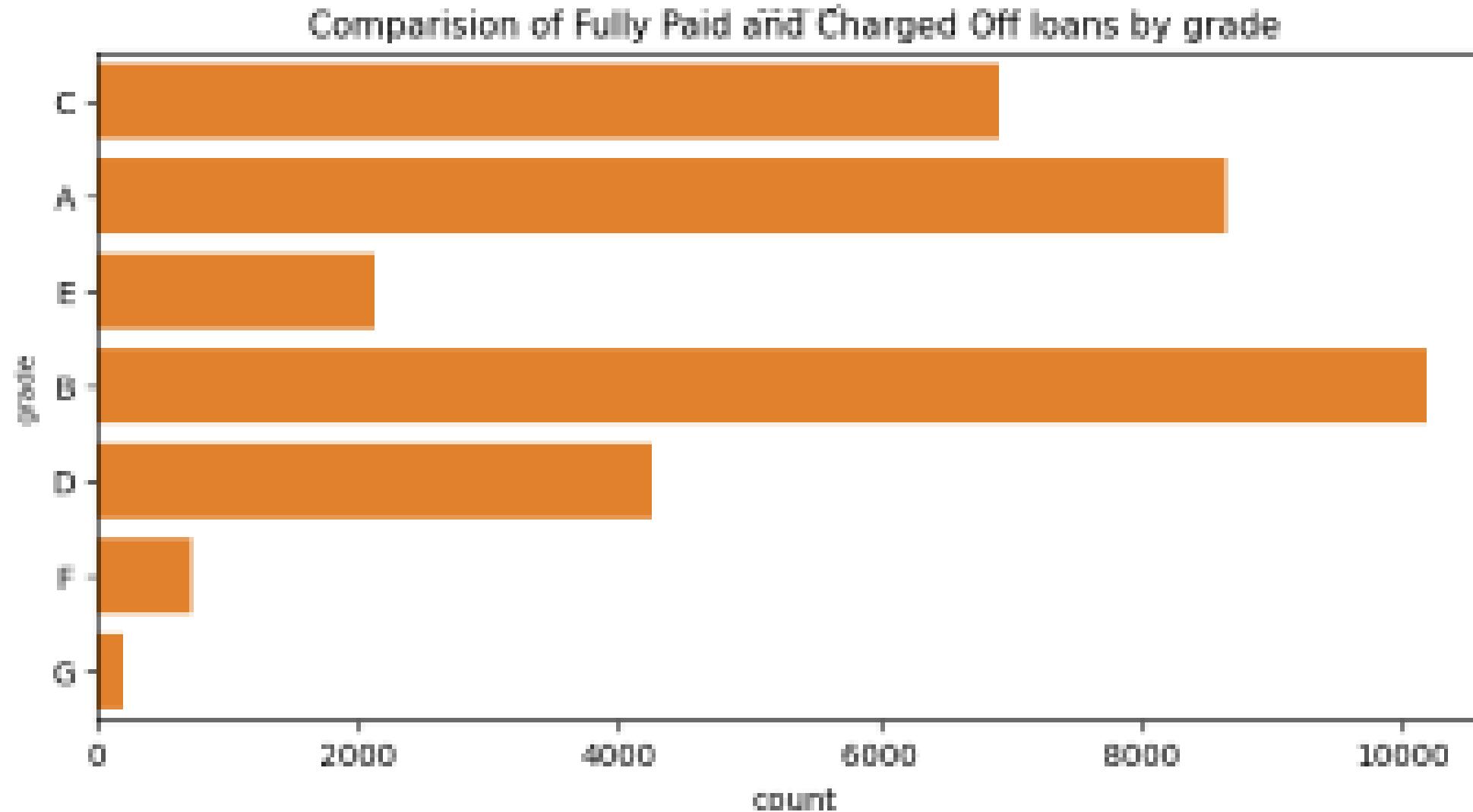
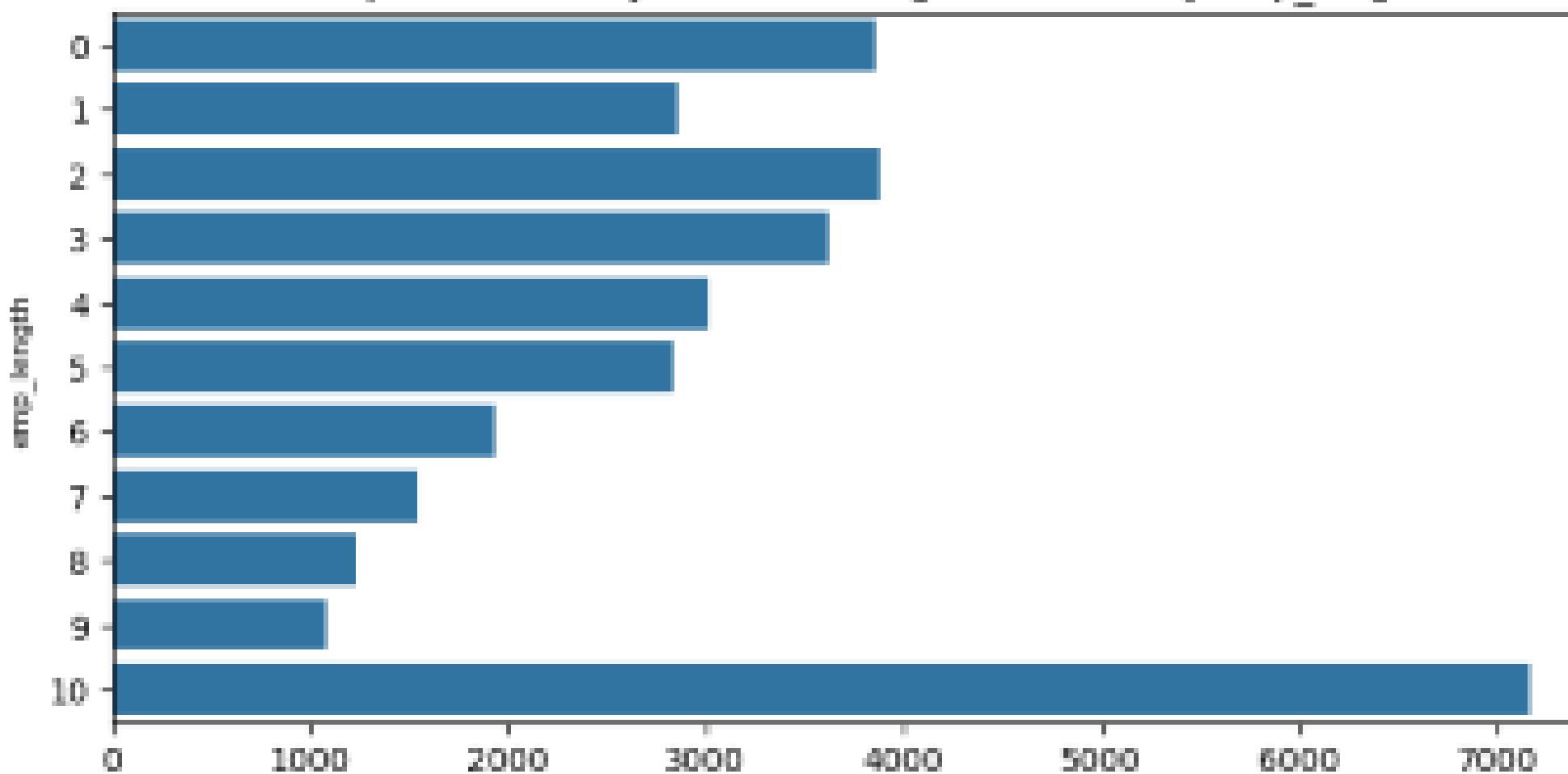
- Create month and year columns from issue date
- Create new columns for numerical data by binning them accordingly
- Variables installment, total account, annual income have outliers that skew analysis of data

Univariate Analysis

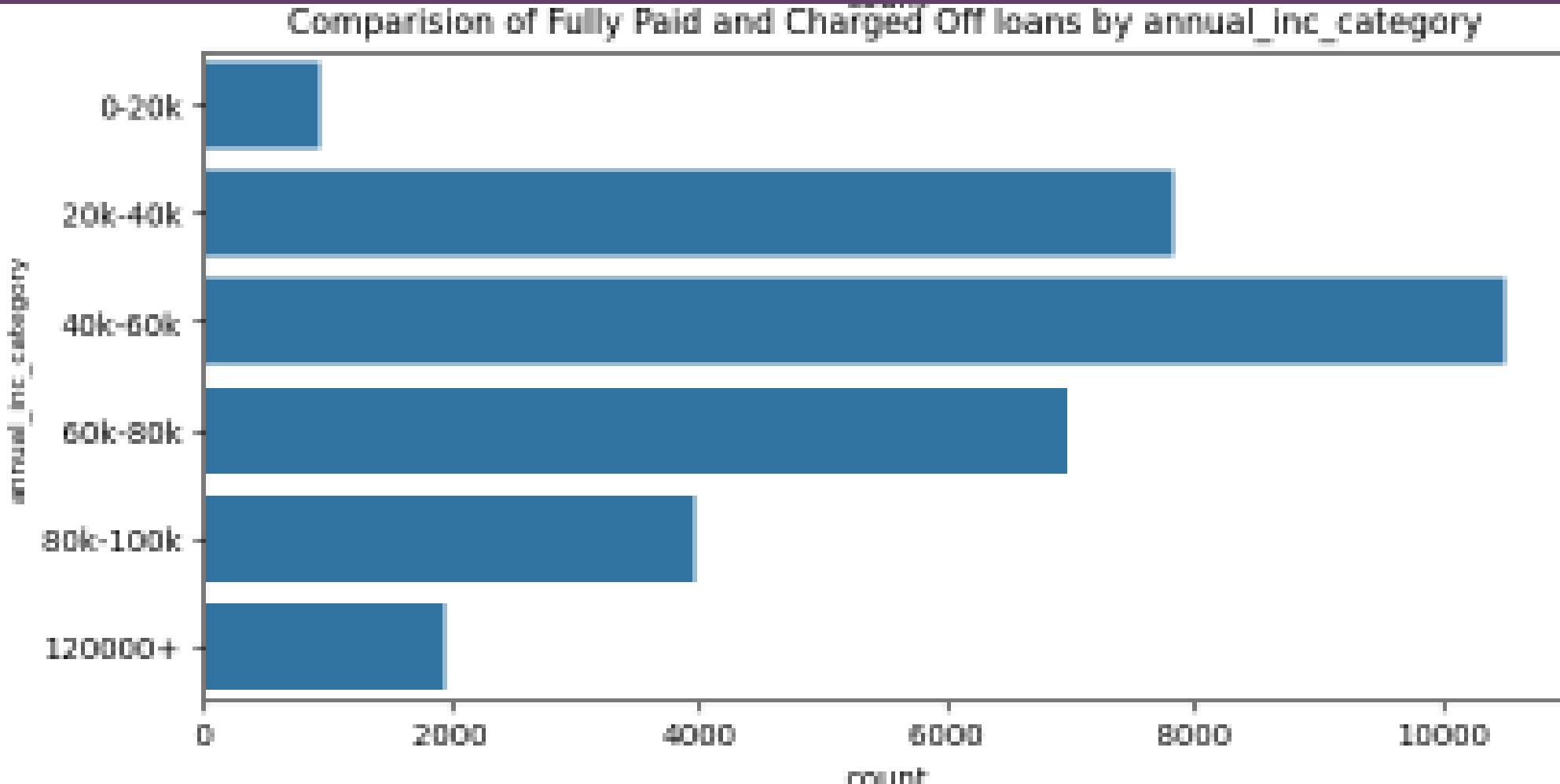
- Generate statistical data of quantitative variables using describe function
- Majority of the loan amount is in the range of 4k to 12k
- Majority of the annual income is in range of 40k to 75k



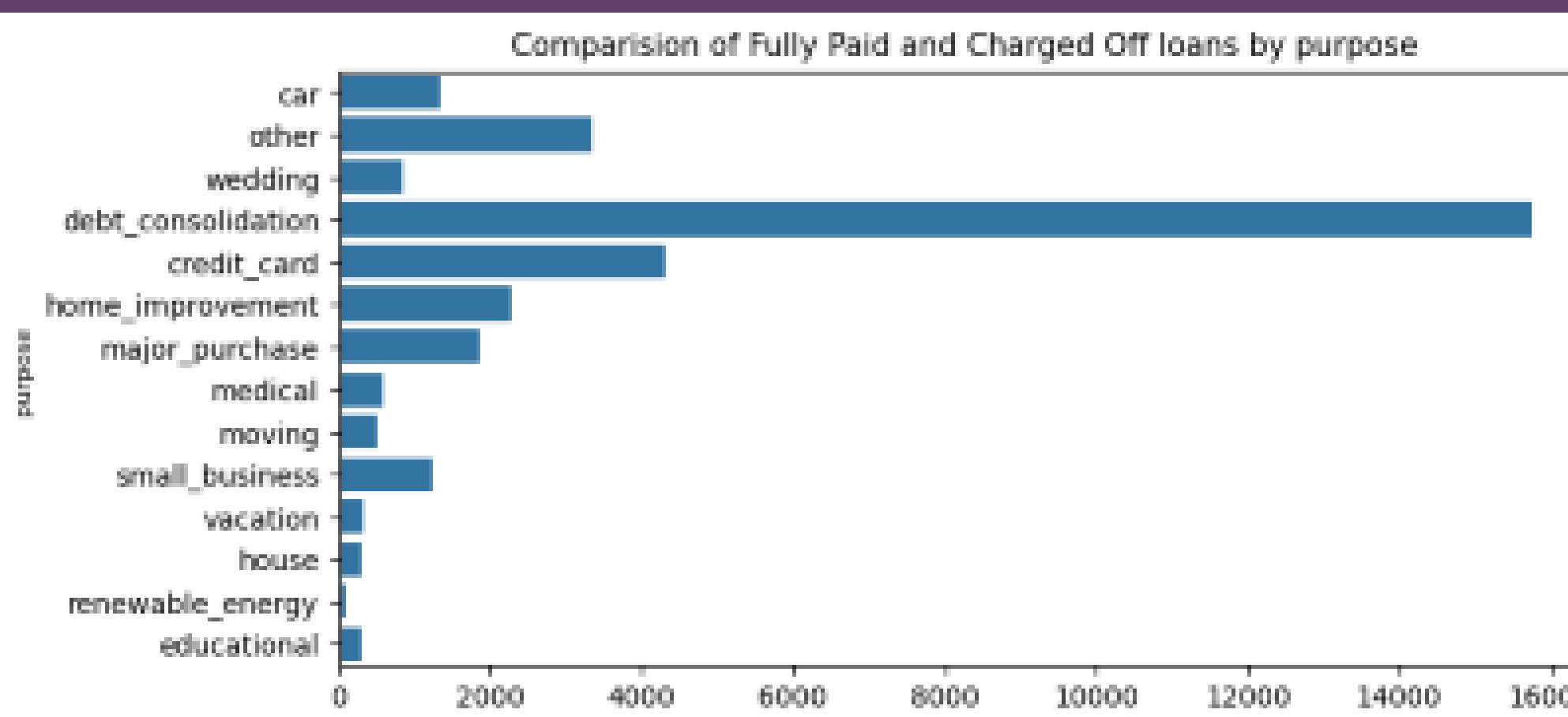
Comparison of Fully Paid and Charged Off loans by emp_length



- Most loan applications are from Grade B customers
- Max employee length is 10 years
- Maximum home ownership is Rent Category
- Most accounts are of not verified category
- Most of the Customers have opted for 36 month term



- Majority of the loan is taken for debt consolidation purpose



- zip code and address for majority of the customers is 945xx and CA respectively

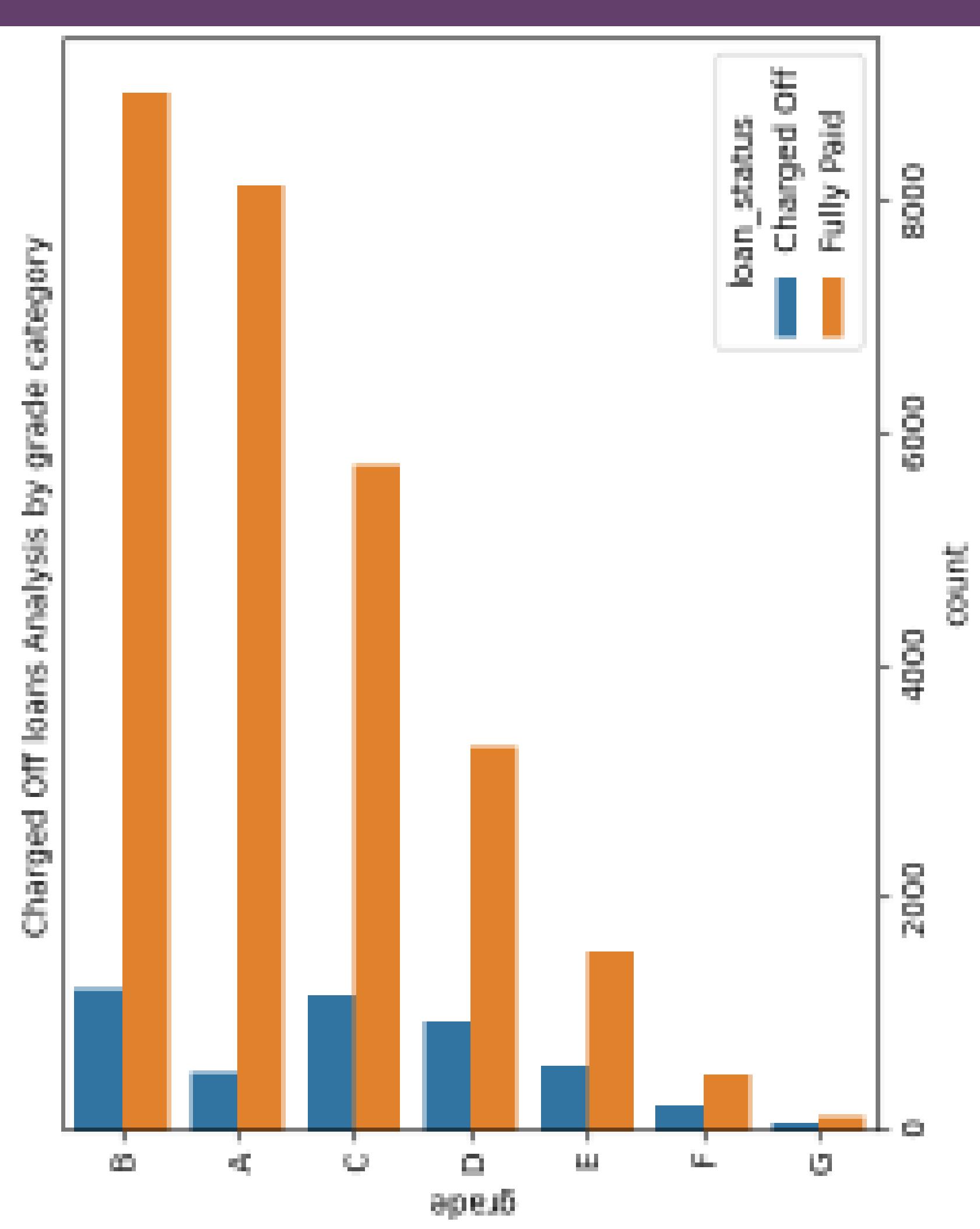
- Most of the customers either stay in rented or mortgage homes

Bivariate Analysis

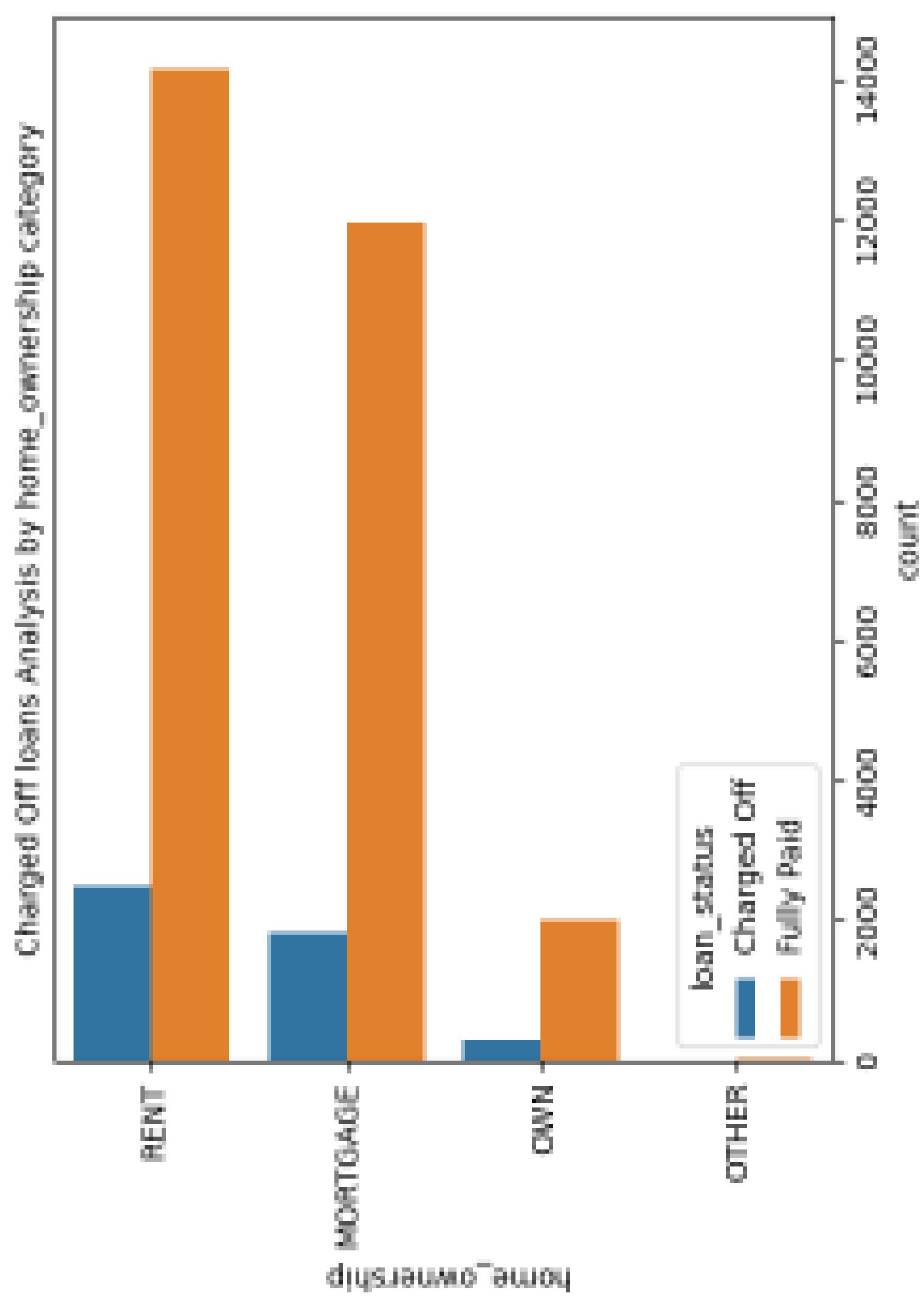
Default on loan is more likely when :

- When customer's home ownership is rent or mortgage
- When debt to income ratio is greater than 10
- When verification is not done

continue



- More customers tend to take loans after 10 years in job which has increased the number of defaulters, hence 10+ years employment length customers are likely to default more
- When revolving line utilization rate is more than 60
- When loan grades is B and C



- Customers from CA default the most followed by NY and TX
- Loan is taken for debt clearance and small business the most
- When interest rate on loan is more than 15%
- When loan term is 36 since most of the customers are taking loan in this term category, but default % is greater in 60 month term so 60 month term is more default likely

Multivariate Analysis

- More the revolving utility rate, higher the loan charged off
- Loan default tends to increase with increment of loan amount
- Increase in debt to income ratio increases the chances of default
- Loan default tends to increase with increase in interest rate
- For small business loan, interest rate is higher, hence more chances of default
- Term is co-related to loan amount
- Term and interest rate are co-related

Conclusion

Customers taking loans to consolidate debts

Customers taking loans at high interest rate are at higher risk of default

Most customers taking loans are employees of 10+ years of employment

Thank You