Lending Club Casestudy

- Akash Singhal
- Amirthagadeswaran Gowrissankar

Submission Date:
21-Aug-2024

Business Objective

- The objective of this case study is to understand Lending Club loan data, use our skills learnt on EDA, apply them to produce a business usable output

- We use our understanding to cleanse the given data, eliminate data that does not provide us any insight

- Understanding on risk underwriting methodologies for loan acquisition

- Use Univariate and Bivariate analysis through different charts to plot and compare the data

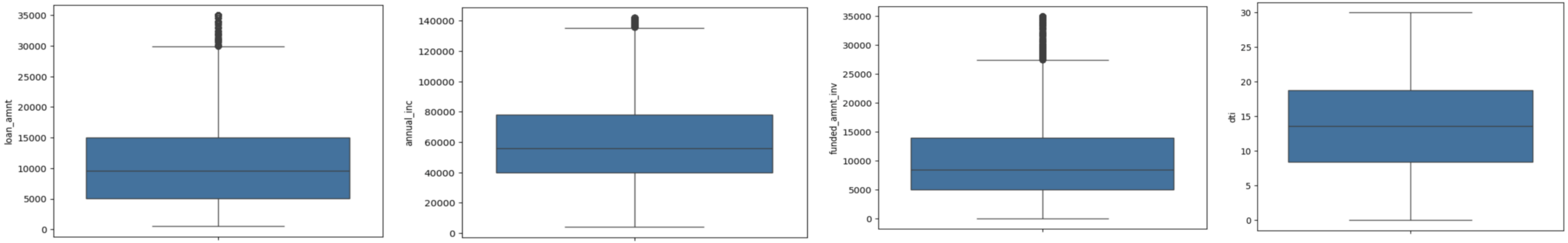
- Derive a conclusion based on the analysis

Note: We have focussed our analysis extensively on charged off customers

Data Understanding

Data

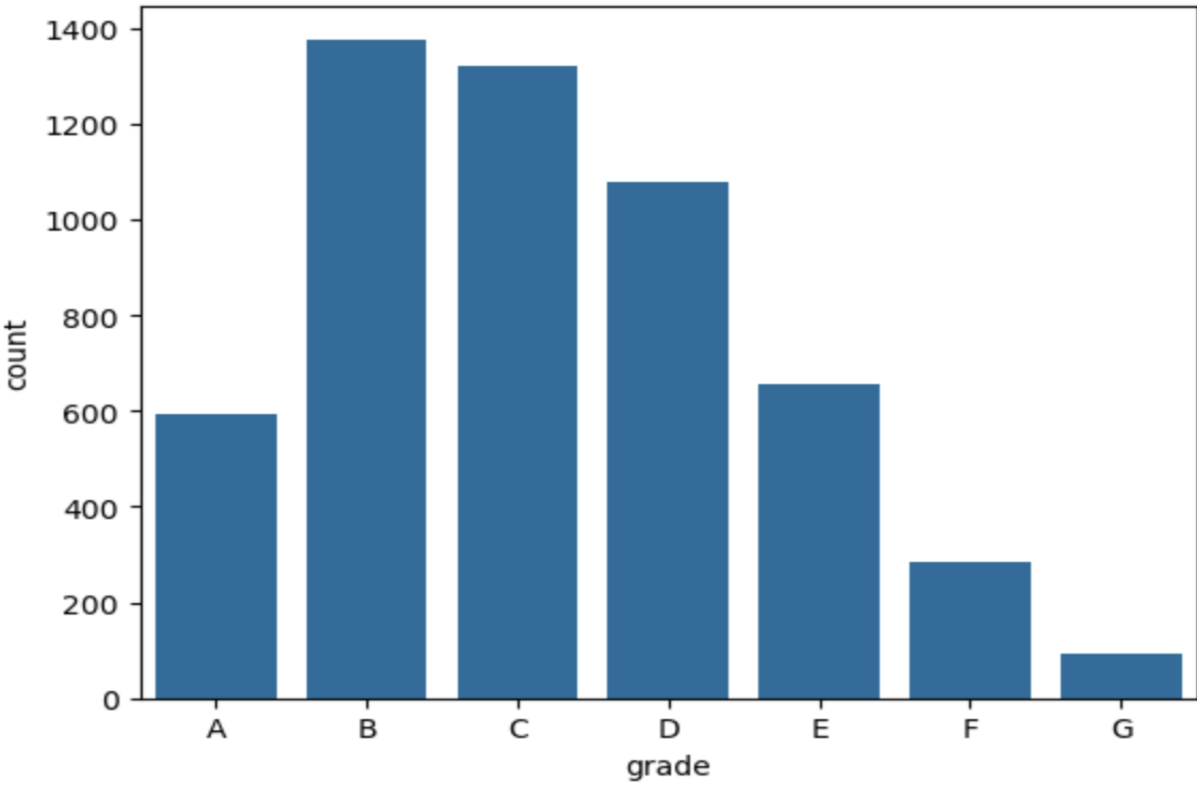
- Data contains details of loans, status of the loan and information on the borrower.
- There are multiple columns that has null values. For effective analysis we have to remove them.
- Some of the columns have only one value for all records. Using them cannot provide conclusive output. Hence we are dropping them
- Some of the variable such as url, title etc. are not quantitative and hence we dropping them
- Handling Missing Values – We replace missing values with mode
- Converting data types for certain columns to be used for analysis
- Use box plot to identify outliers and remove them appropriately
- Filter data to contain data for customers with status “Charge Off” for our analysis

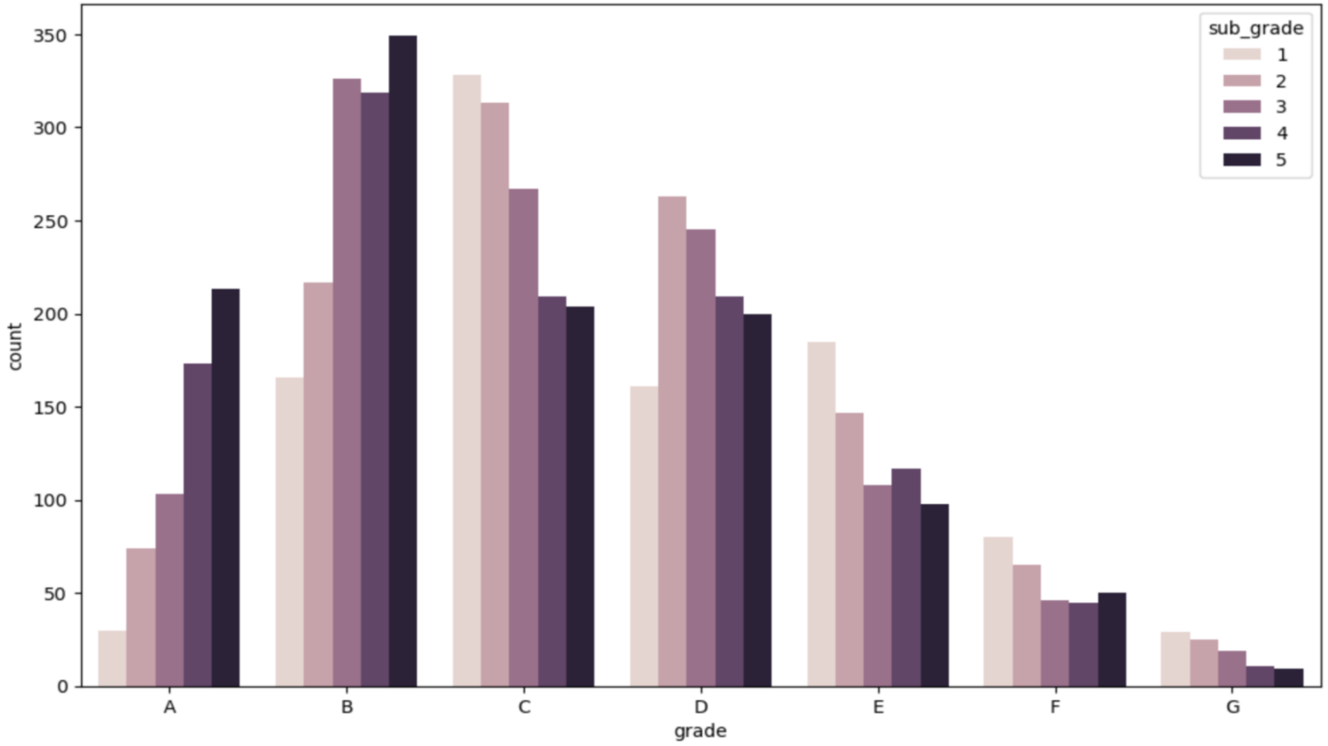


Univariate Analysis

Riskiest Grade – B5

- For the charged off account, the most riskiest in the portfolio is Grade B and sub grade 5



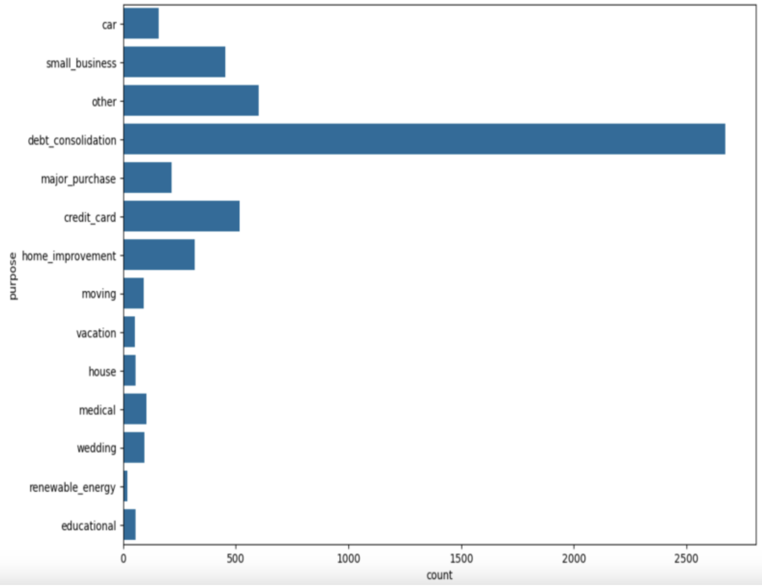


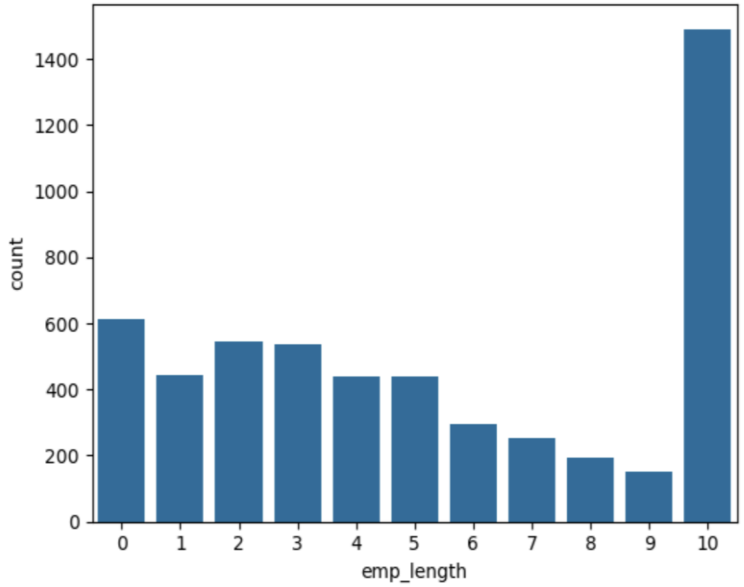
Risk Profiles by

- **Home Ownership**
- Customers Renting homes instead of owning or paying mortgages are most likely to be charged off

- **Purpose**
- The level of default risk is high for customer who apply for loan for debt consolidation

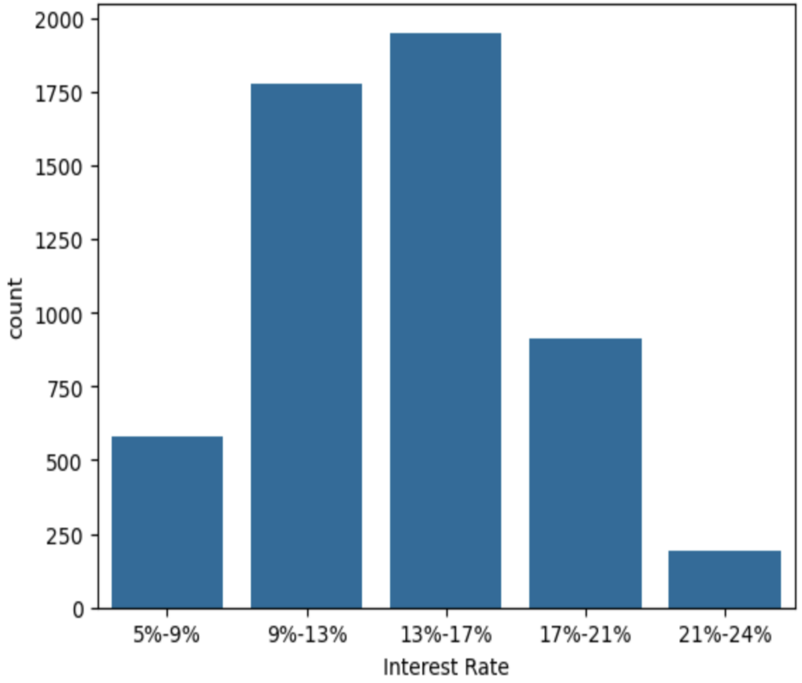
- **Employment Length**
- Its surprising to see that people who are employed for longer are at high risk of default



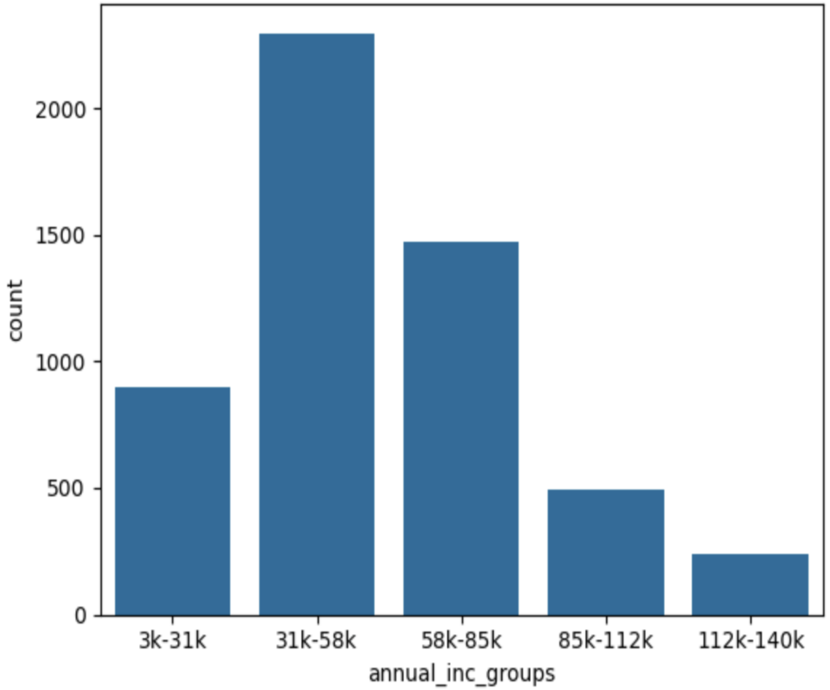


Risk Profiles by

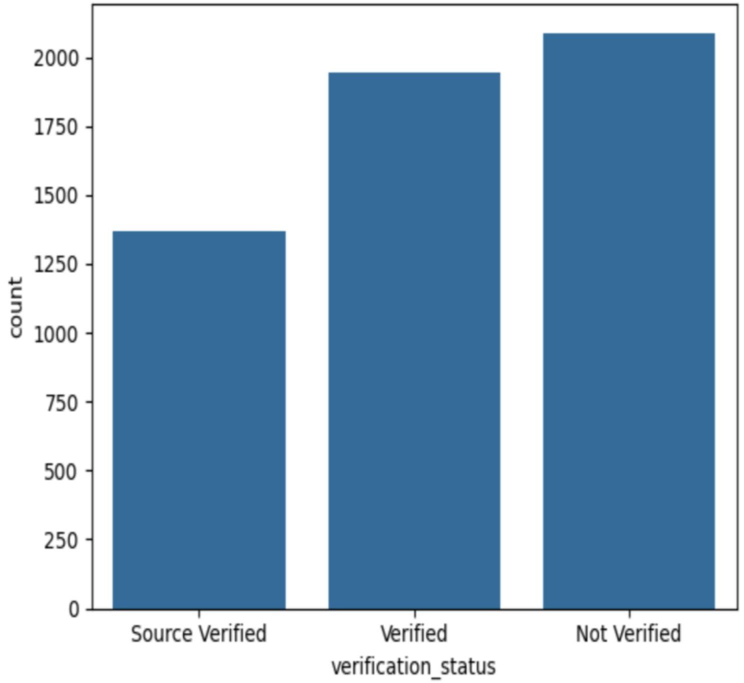
- **Interest Rate**
- High number of customers on the interest rate category 9-17% tending to be riskier



- **Annual Income**
- Customers with income < 31K or > 85k are performing better than customers in-between

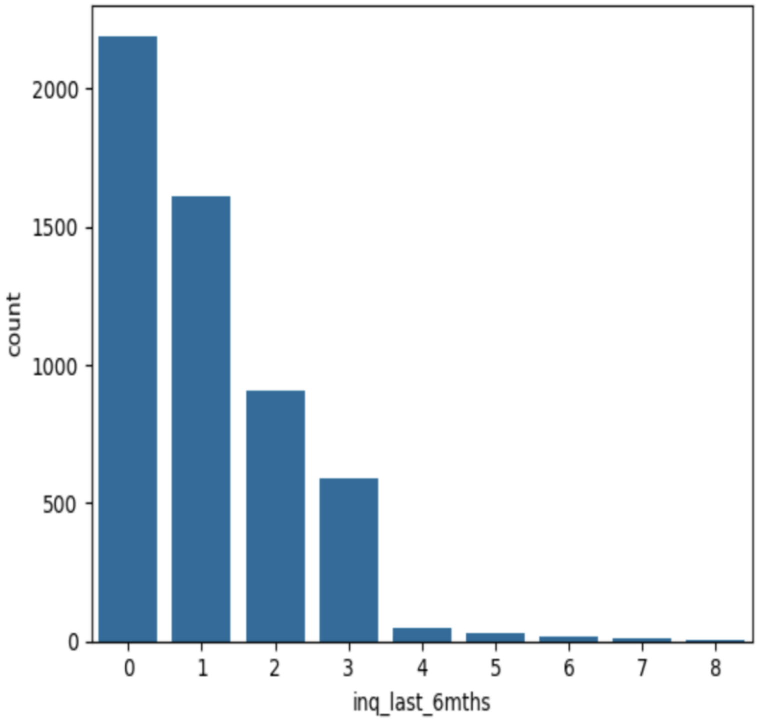


- **Verification Status**
- Income source verification should be mandated to reduce risk of borrowing

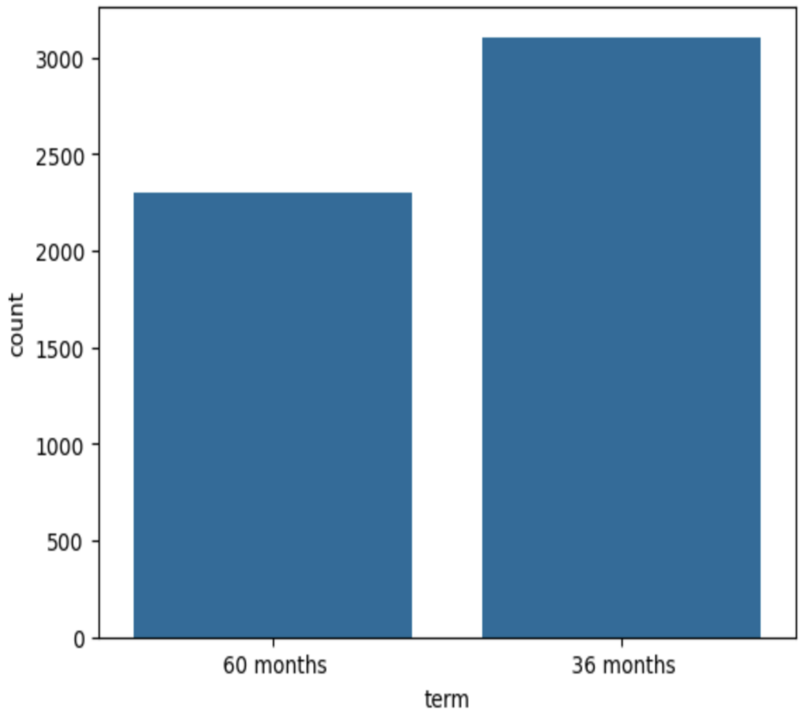


Risk Profiles by

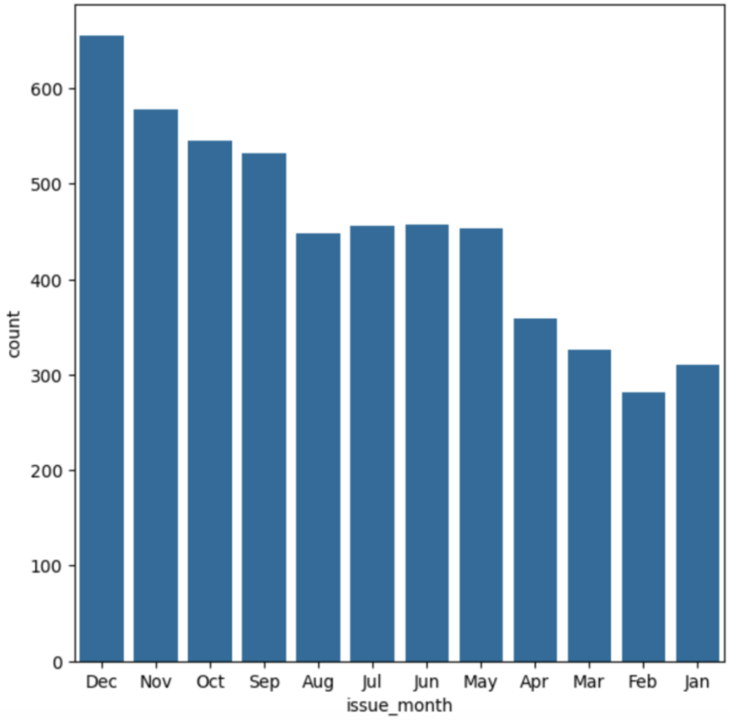
- **Inquiries in Last 6 Months**
- Customers who have not inquired for loan products (excluding auto or mortgage) are more likely to default



- **Term**
- Customers who opt for long term are at less risk.
- This may be due to less installment owing to longer term



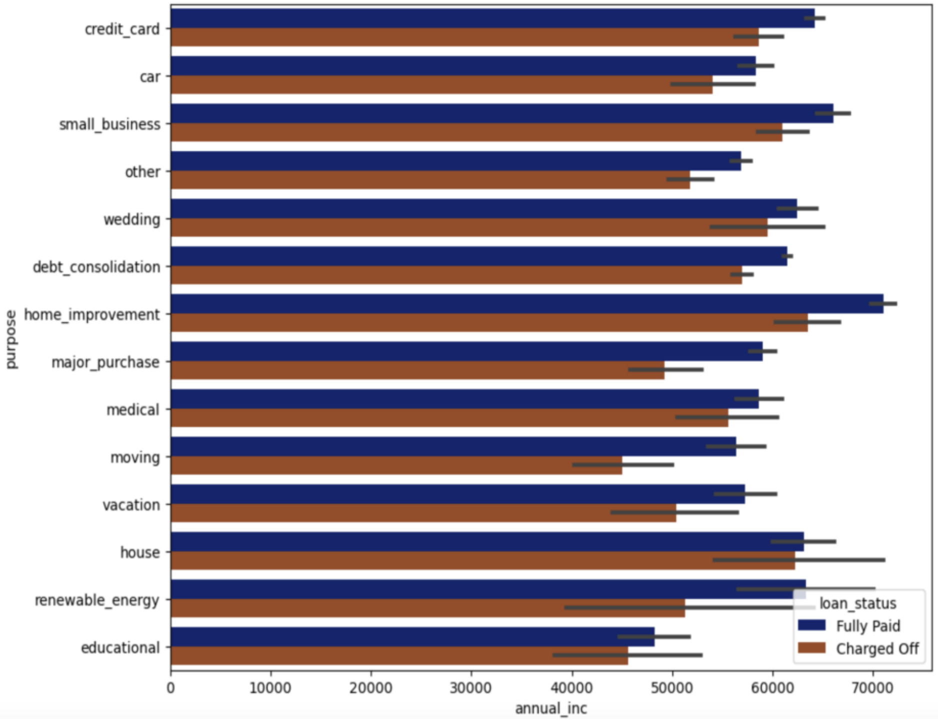
- **Borrowing Month**
- We can infer that customers borrowing during end of the year are at high risk
- They may use the borrowed money for holiday spending



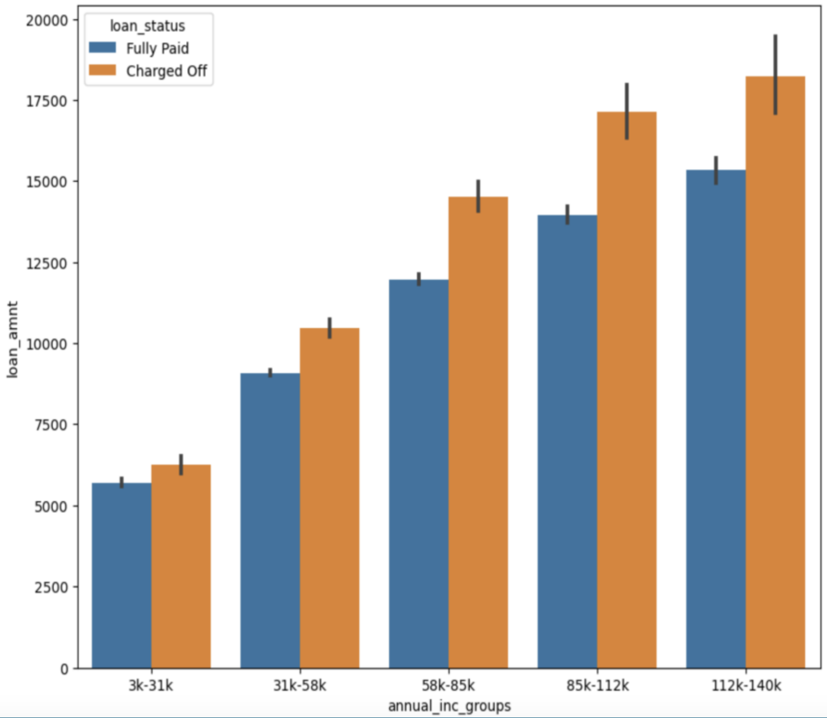
Bivariate Analysis

Annual Income vs

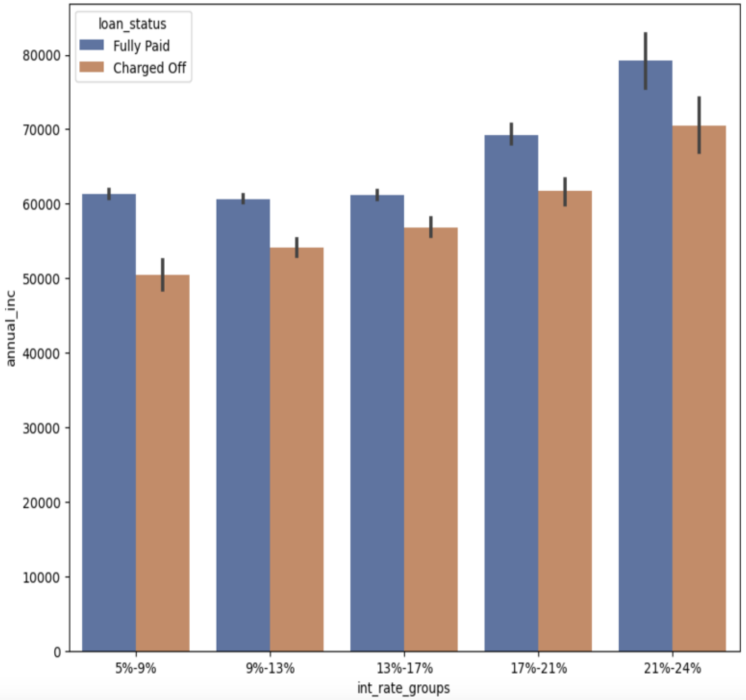
- **Purpose**
- Customers with higher income tend to go for loans for home improvement or house related expenses.



- **Loan Amount**
- Higher income category, get higher loan amount.
- This may be due to their dti availability

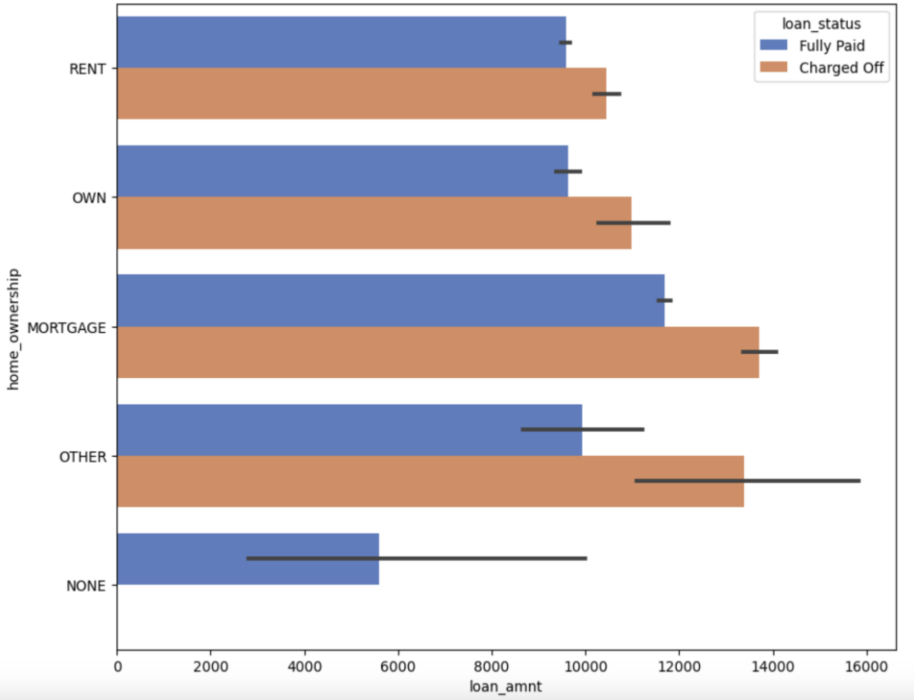


- **Interest Rate**
- Customers in the lower income group get better interest rates compared to higher income groups

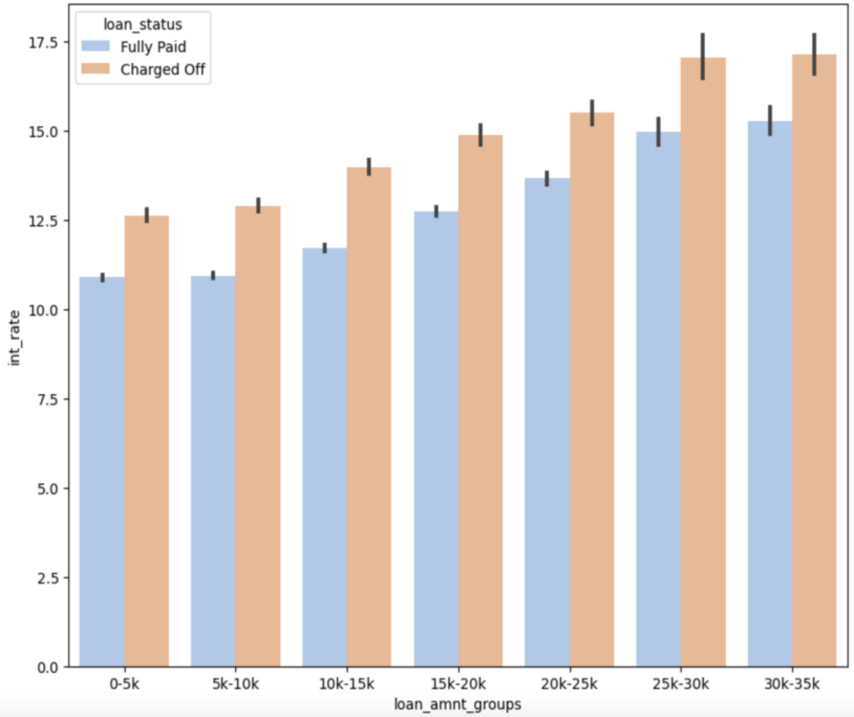


Loan Amount vs

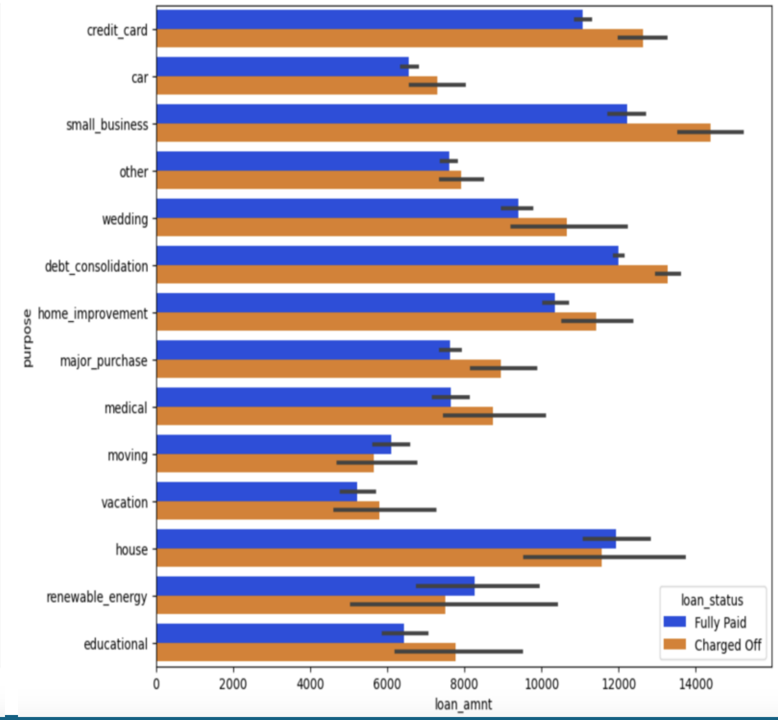
- **Home Ownership**
- Customers who reside in rental properties have lesser loan amount compared to people who reside in mortgaged houses



- **Interest Rate**
- Higher interest rates are observed for higher loan amounts



- **Purpose**
- Higher loan amounts are taken by customers for purpose of small business

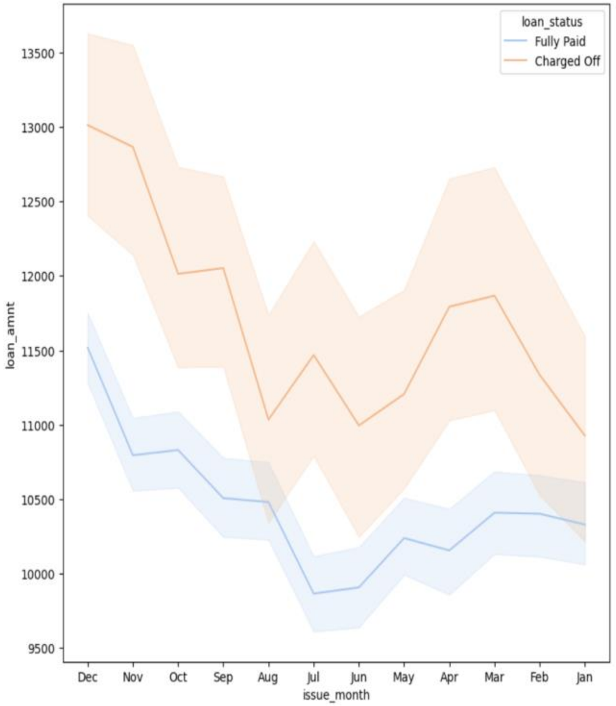


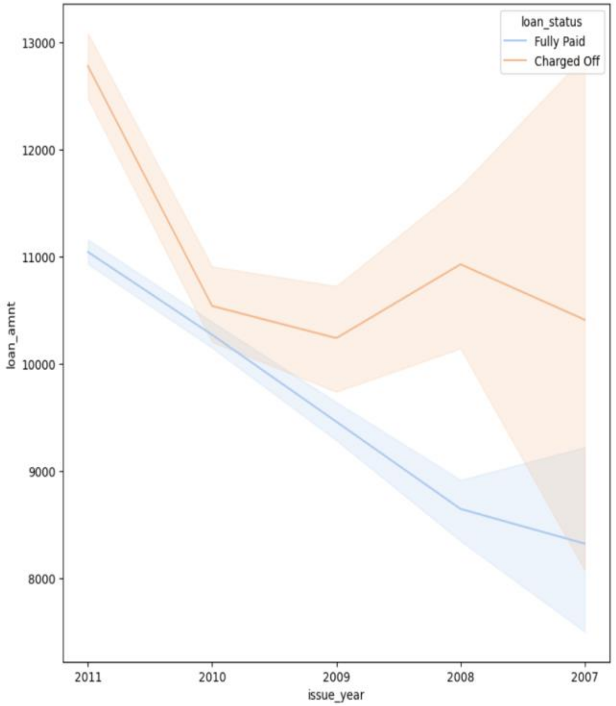
Loan Amount vs

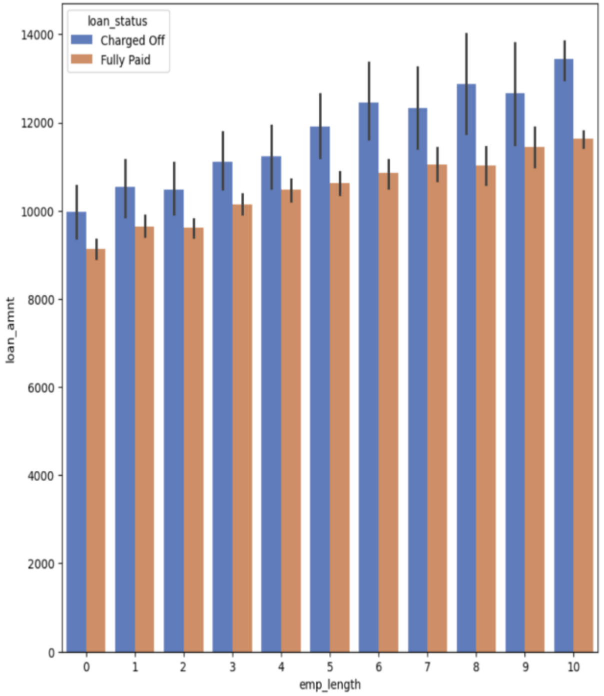
- **Month and Year**
- December month loan sanction amounts are higher compared to January
- 2011 has seen the highest loan amount allocation

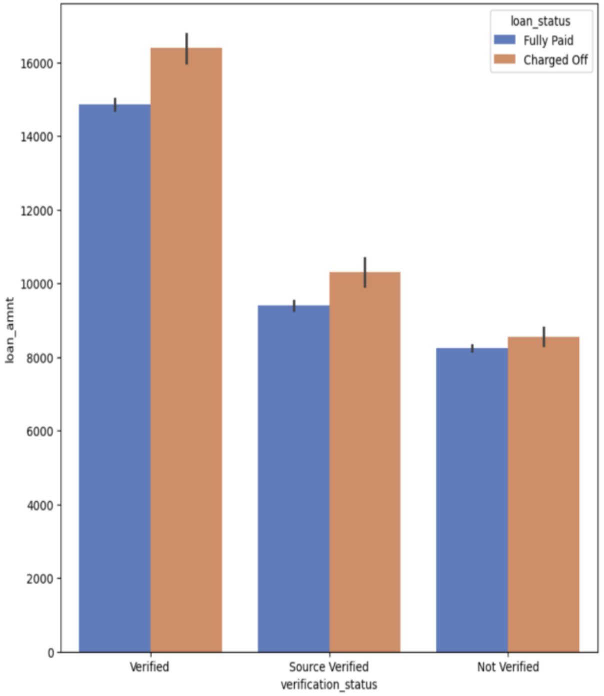
- **Employment Length**
- Customers with higher length of service get higher loan amounts

- **Verification Status**
- LC income verified customers get higher loan amount
- In univariate analysis, we have seen that Source verified customers perform better







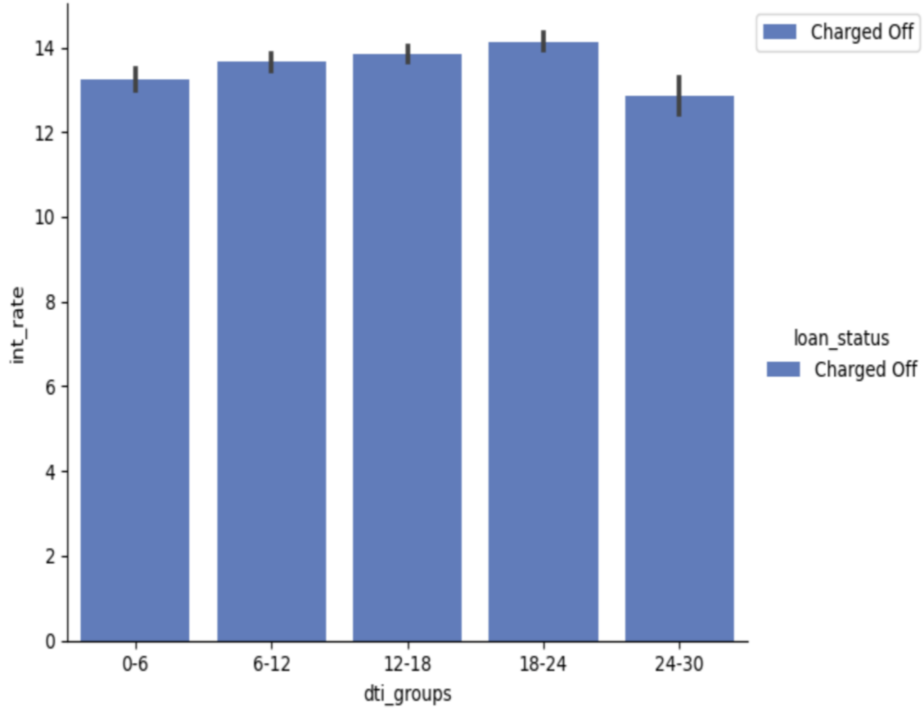


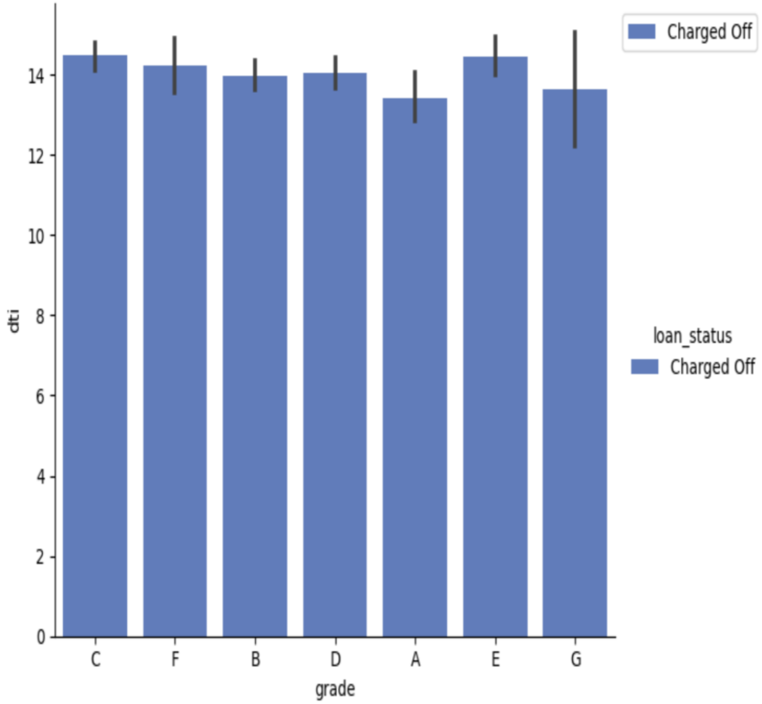
DTI vs

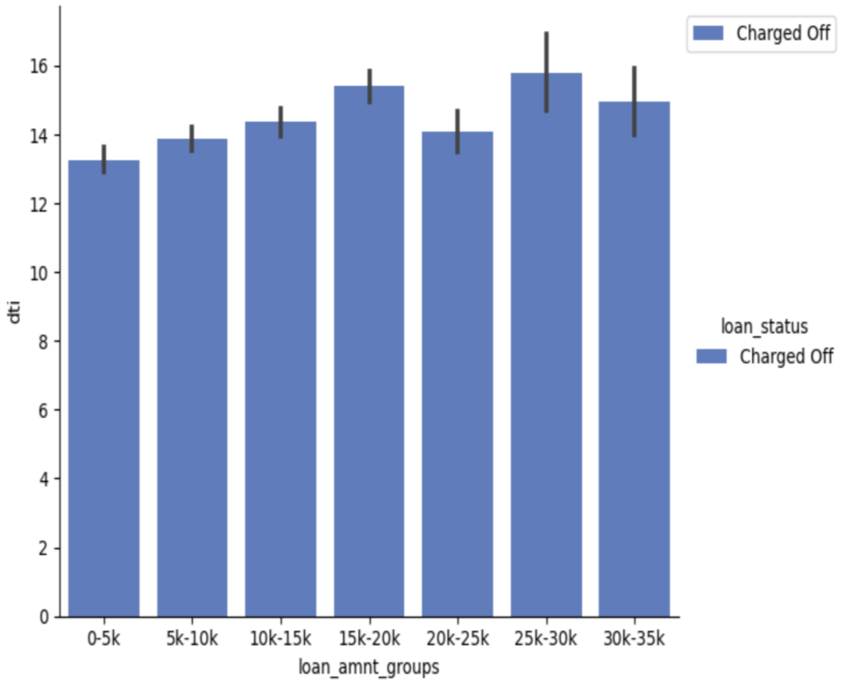
- **Interest Rate**
- Customers with higher DTI gets lower interest rates

- **Grade**
- Though we have observed that B grade B was worst performing, they have lower DTI compared to Grade E

- **Loan Amount**
- There is a trend observed where DTI increases with loan amount till 20k.
- Loan amounts > 20K does not follow any trend with respect to DTI







Conclusion

- Grade B5 is the riskiest of all customers and likely to be charged off
- Count of customers with home ownership ‘Rent’ is the highest
- Employees with longer emp_length got the loan approved for a higher amount.
- When the loan is verified, loan amount is above 16k.
- For grade G and interest rate above 20% -Verified loan applications tend to have higher loan amount. Which indicates that the firms are first verifying the loans with higher values
- The interest rate for charged off loans is higher than that of fully paid loans in all the loan_amount groups. this can be a driving factor for loan defaulting.
- Applicants applying for long term has applied for more loan