Lending Club Case Study

Exploratory Data Analysis on Loan Defaulters

Problem Statement

- Use Exploratory Data Analysis to understand how consumer attributes and loan attributes influence the tendency of default.

Data Understanding

- The data given to us contains information about past loan applicants and whether they ‘defaulted’ or not.
- The dataset has 111 attributes out of which we picked the following attributes for our analysis
  - 'loan_amnt', 'funded_amnt', 'term', 'int_rate', 'installment', 'grade’, 'sub_grade', 'emp_title', 'emp_length', 'home_ownership', 'annual_inc', 'verification_status', 'loan_status', 'dti', 'pub_rec', 'revol_bal’, 'revol_util', 'total_acc'

Data Cleaning

- **Approach with missing values:** Column emp_title, emp_length and revol_util contains missing values (6.2%, 2.7& and 0.126% respectively). The percentages of missing values are not significant and also, we do not have any reliable source to fill them. Hence, we are leaving the missing values as it is.
- Corrected data type of interest rate and revol_util
- Rows containing missing values of revol_util were dropped to carry out operations necessary to correct the data type

Findings from Basic summary metrics

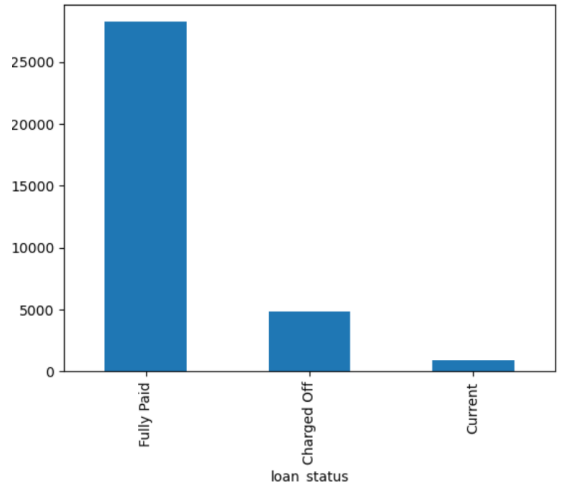
Most of the applicants are unverified

Most of the applicants fully pay their loans

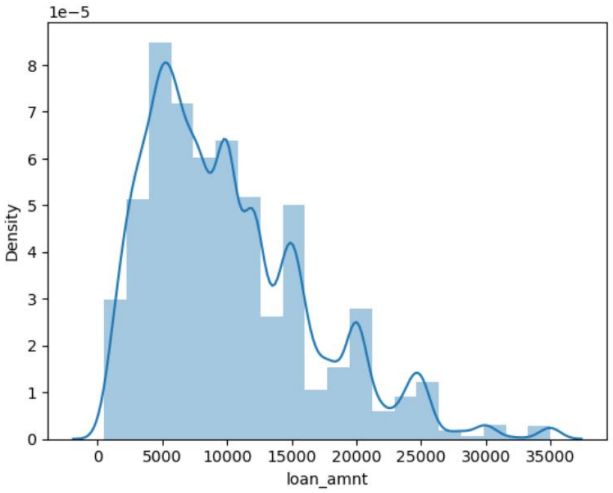
Most of the applicants are from US army

Most of the applicants live in rented accomodation

Findings from Univariate Analysis

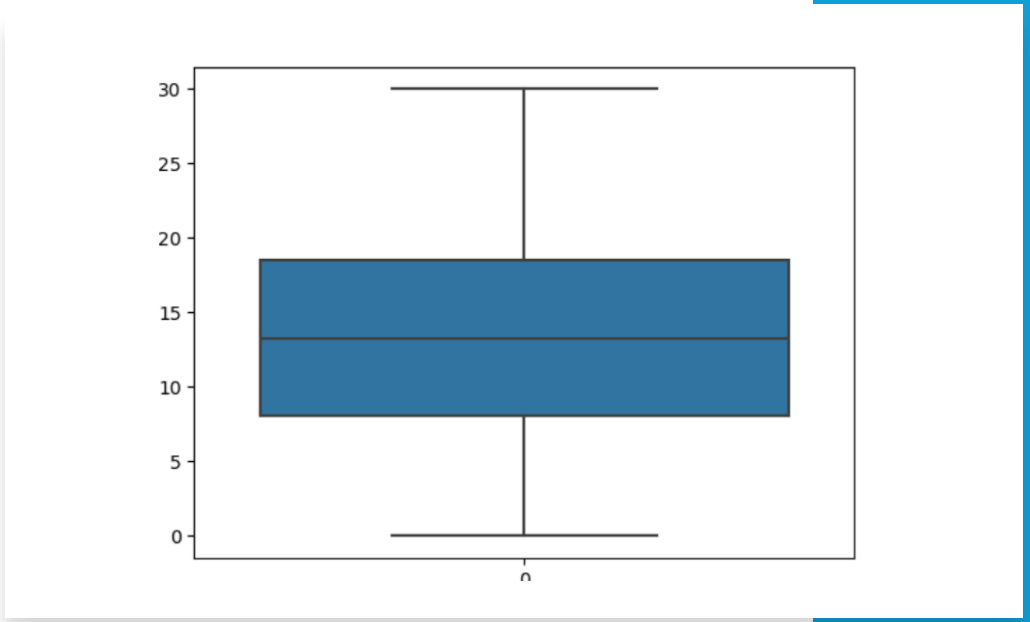


- About 14% of the applicants defaulted in their loans.
- Majority of the loans are skewed between 5000 to 10000 dollars.

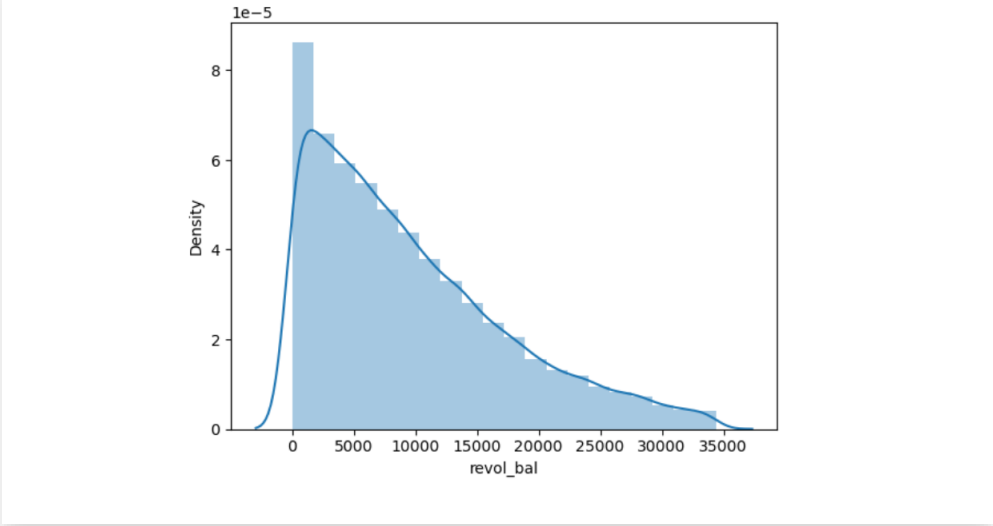


Findings from Univariate Analysis

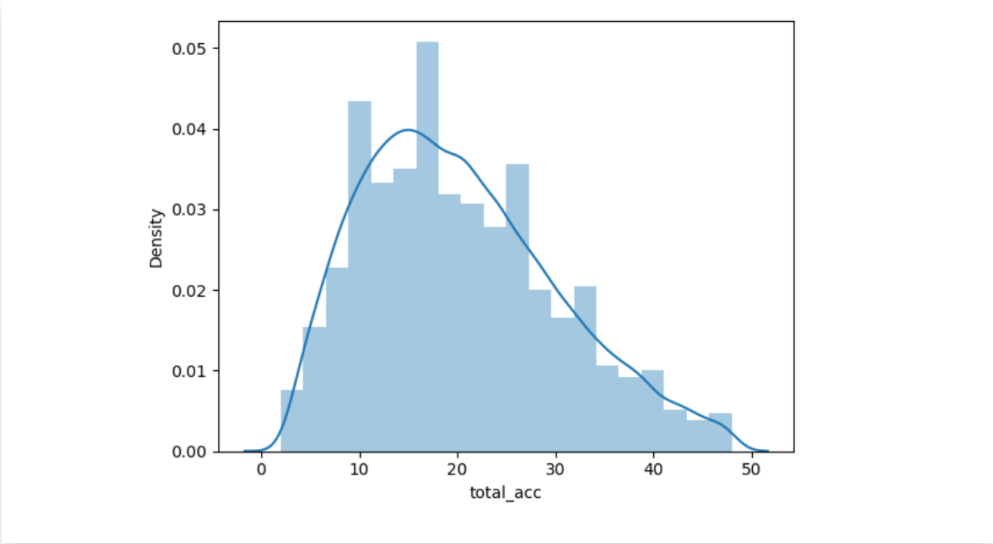
- A high median value of debt-to-income ratio (14) and a possible indication of wrongful declaration of income
- High number of applicants with unverified income



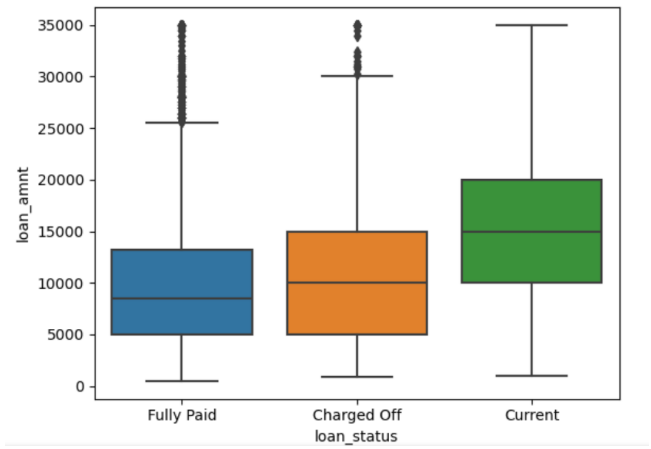
Findings from Univariate Analysis

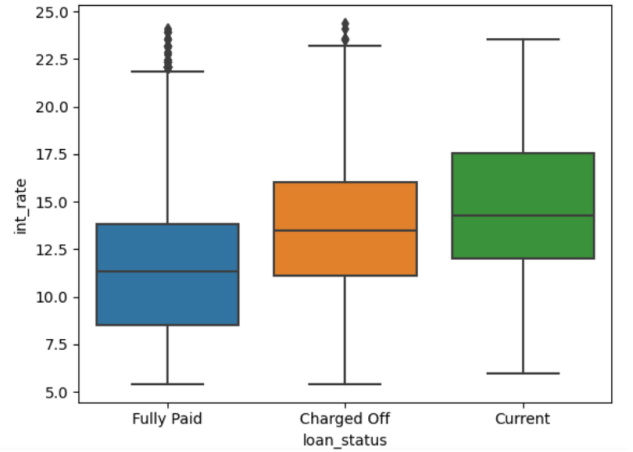


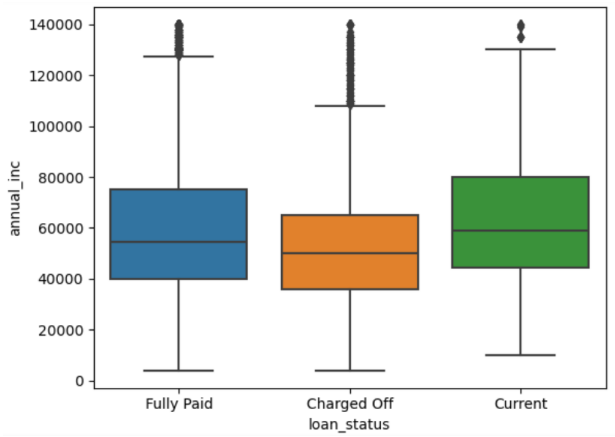
- Wide variation in revol_bal values with the distribution being skewed towards lower values showing that most borrowers are not relying too much on credit.
- The distribution of total accounts is a bit left screwed indicating that borrowers are not overly relying on credit.



Findings from Bivariate Analysis

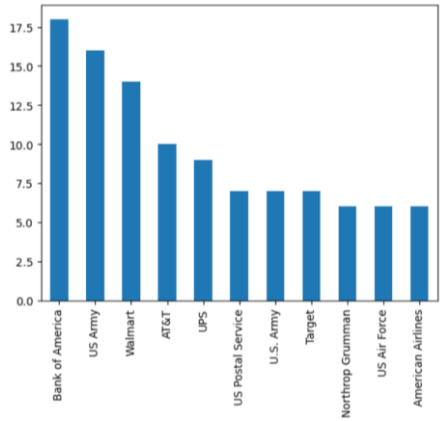


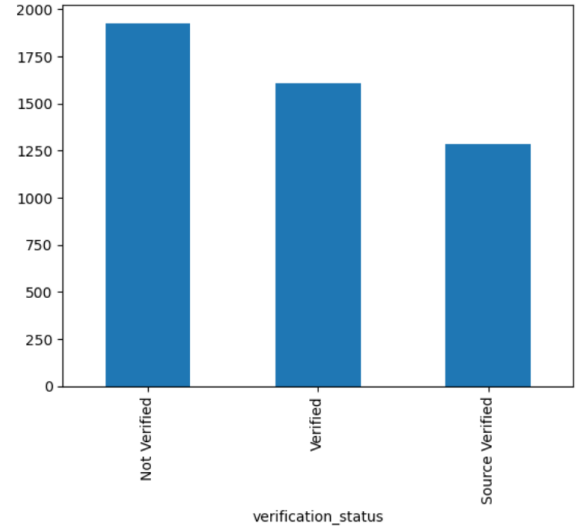




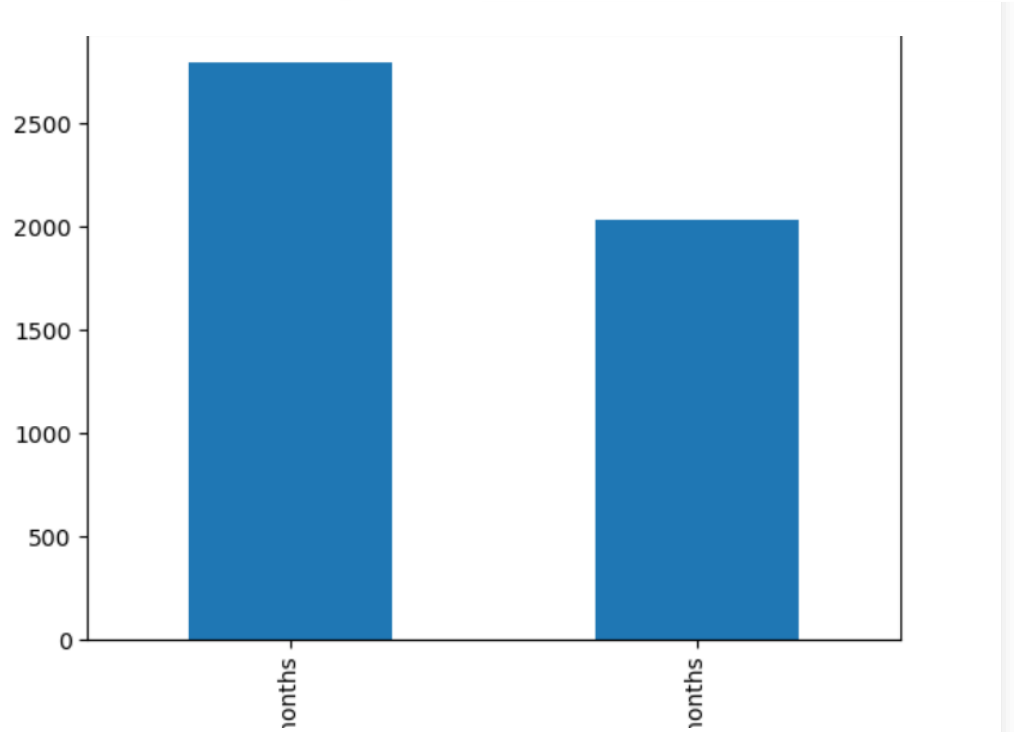
- Loan amounst of defaulted applicants is slightly higher than fully paid applicants.
- Interest rate of defaulted applicants is significantly greater than fully paid ones.
- Defaulters had a significantly less annual income than the others

Findings from Bivariate Analysis



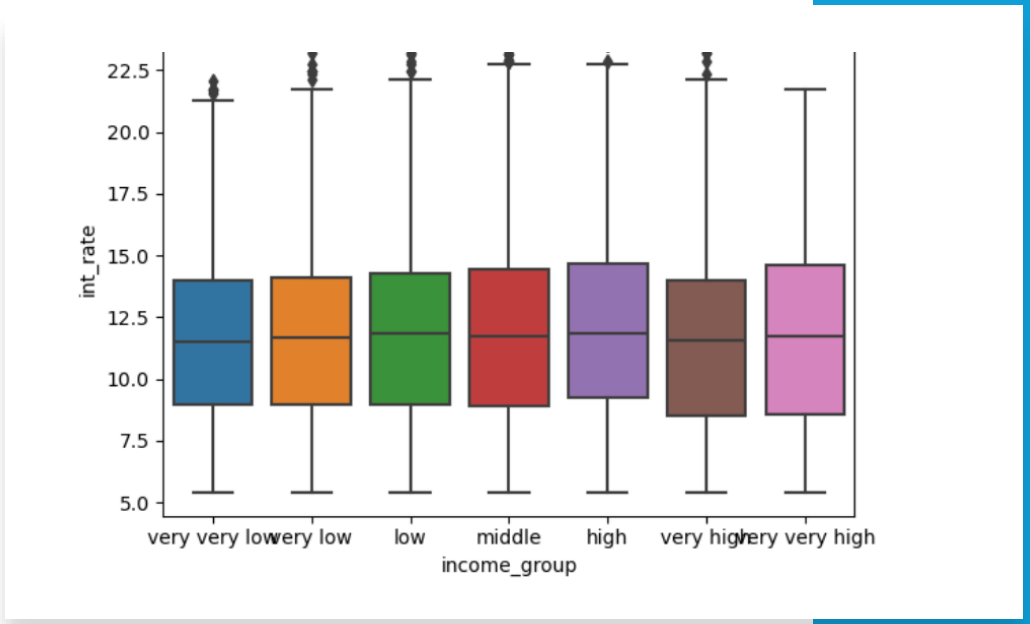


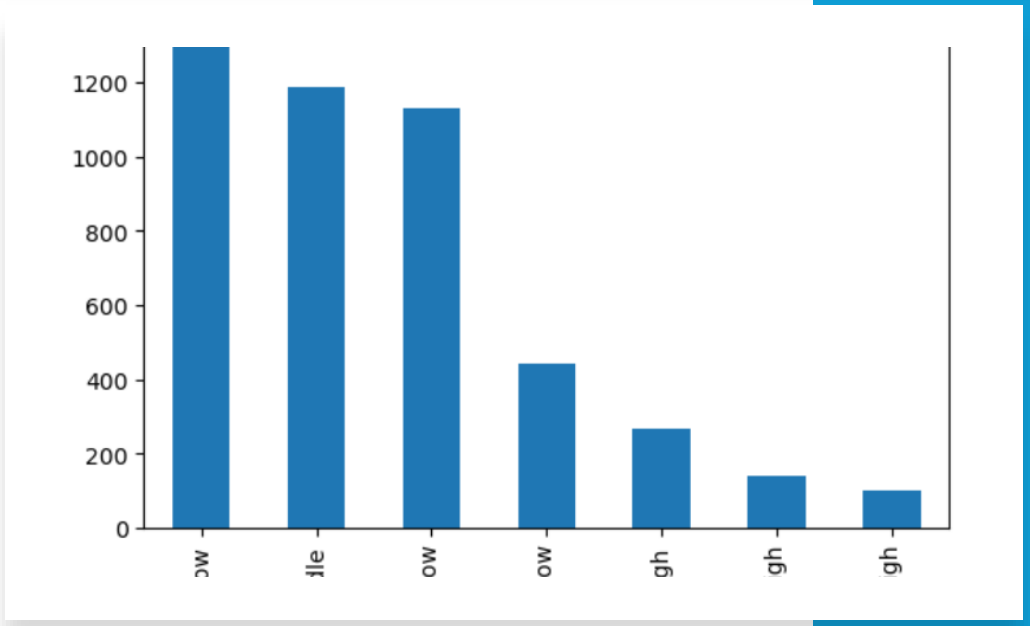
- Applicants with 36 months term default more than 60 months
- Unverified applicants have a higher default frequency
- Amongst top 10 employment types, employees of Bank of America default most followed by applicants of US Army



Findings from Bivariate analysis

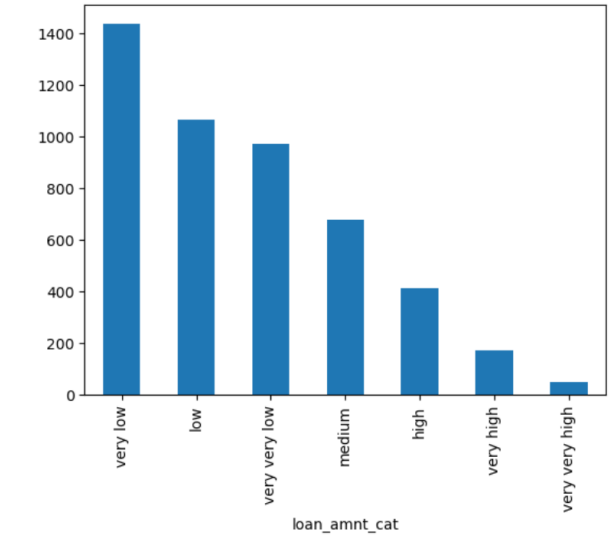
- Most of the applicants from low (40000 to 60000 dollars) and middle (60000 to 80000 dollars) and consequently they default the most. These applicants are levied similar interest rates as others.

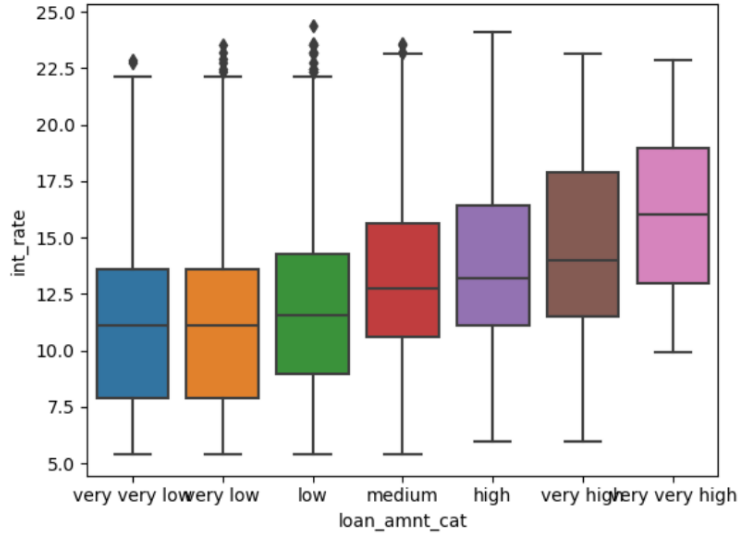




Findings from Bivariate analysis

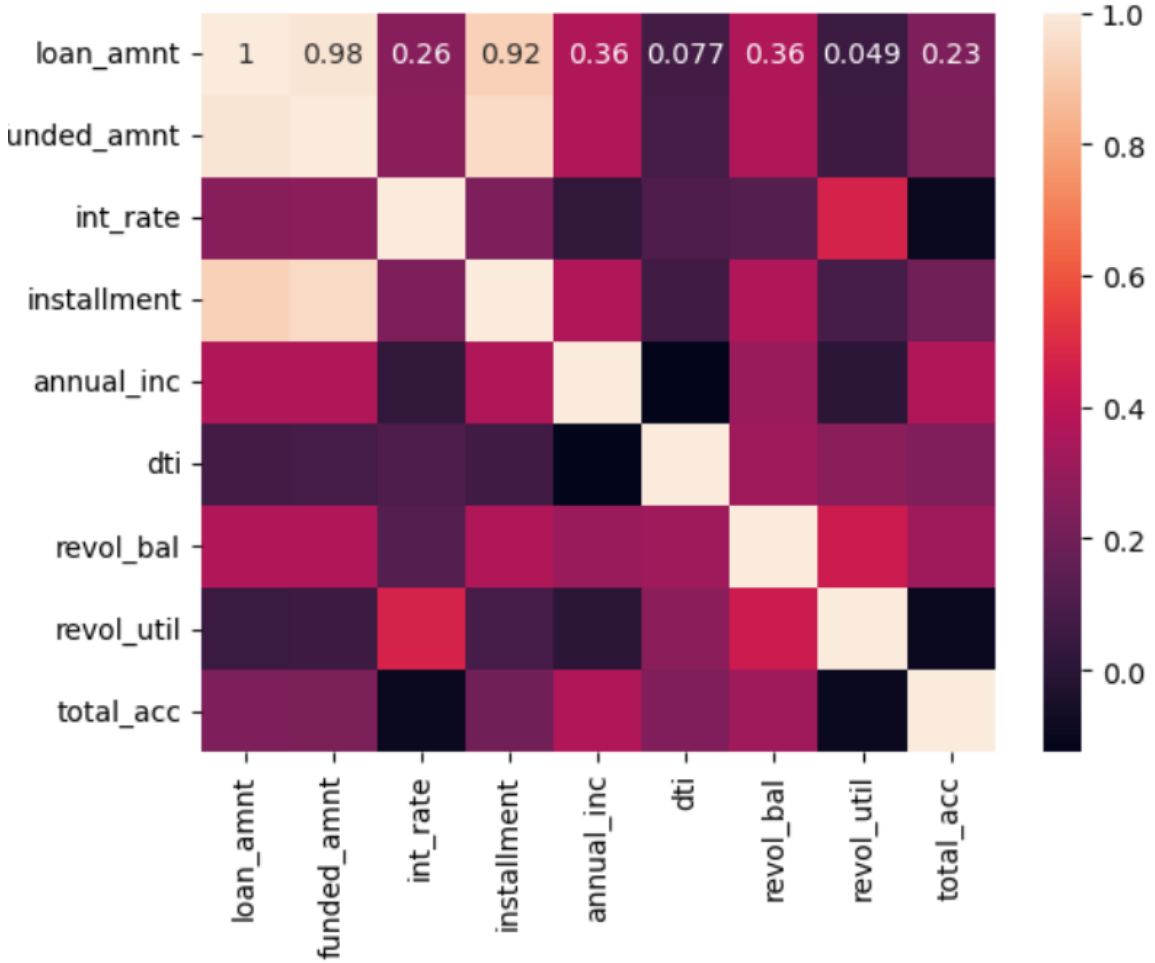
- Most of the loans are very low(5000 to10000 dollars) or low value (10000 to 15000 dollars) loans. Interest rate charged is comensurate with the loan amount.





Findings from multivariate analysis

- Loan amount is having a mildly positive correlation with annual income and revolving balancea
- Interest rates have a mildly positive correlation with loan amount, debt to income ratio and revolving balance utilization
- Total account is having a mildly positive correlation with annual income and debt to income ratio



Recommendations

1. Refine loan approval criteria based on identified risk factors.
2. Consider offering lower amounts or higher interest rates to risky applicants.
3. Limit loans to borrowers whose income is unverified. Charge a greater interest rate for unverified borrowers
4. Offer flexible payment plans for short term loans
5. Include debt-to-income ratio as a factor for approving the loan and deciding the interest rates