LENDING CLUB CASE STUDY

LENDING CLUB IS A CONSUMER FINANCE COMPANY WHICH SPECIALIZES IN LENDING VARIOUS TYPES OF LOANS TO URBAN CUSTOMERS. WHEN THE COMPANY RECEIVES A LOAN APPLICATION, THE COMPANY HAS TO MAKE A DECISION FOR LOAN APPROVAL BASED ON THE APPLICANT'S PROFILE. TWO **TYPES OF RISKS** ARE ASSOCIATED WITH THE BANK'S 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

DECISION BASED ON LOAN TYPES

When a person applies for a loan, there are **two types of decisions** that could be taken by the company:

1. **Loan accepted:** If the company approves the loan, there are 3 possible scenarios described below:
    1. **Fully paid**: Applicant has fully paid the loan (the principal and the interest rate)
    2. **Current**: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
    3. **Charged-off**: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has **defaulted** on the loan
2. **Loan rejected**: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

BUSINESS OBJECTIVES

- company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

STEPS TAKEN FOR ANALYSIS

- Understanding the data
- Data Cleaning
- EDA
- Data Visualization
- Conclusion

UNDERSTANDING THE DATA

- The dataset provided contains historical data from borrowers including information about borrower's loan characteristics, employment details, loan outcomes(fully paid, charged off, current) and many other factors.
- **Below are key points have been considered**

Key Variables:

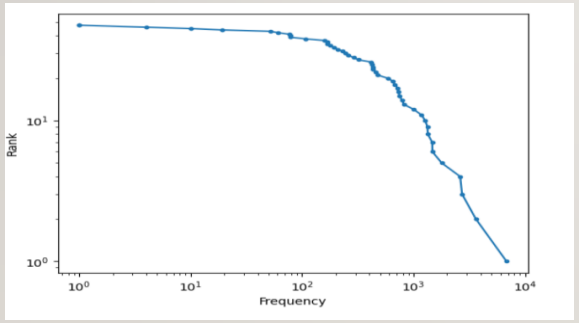
- **Loan Amount:** The amount of money borrowed.
- **Annual Income:** The borrower's annual income.
- **Verification Status:** The borrower's annual income is verified or not.
- **Purpose:** Why the loan has been taken>
- **Interest Rate:** The annual percentage rate (APR) charged on the loan.
- **Loan Term:** The duration of the loan (usually 3 or 5 years).
- **Grade:** Lending Club's internal credit grade assigned to the borrower (A-G).
- **Employment Length:** The duration of the borrower's current employment.
- **Home Ownership:** Whether the borrower owns, rents, or is mortgaged.
- **Debt-to-Income Ratio (DTI):** The ratio of total monthly debt payments to monthly gross income.
- **Credit History:** Information on the borrower's credit score, credit inquiries, and past credit behavior.
- **Loan Status:** The final status of the loan (e.g., fully paid, charged off, current).

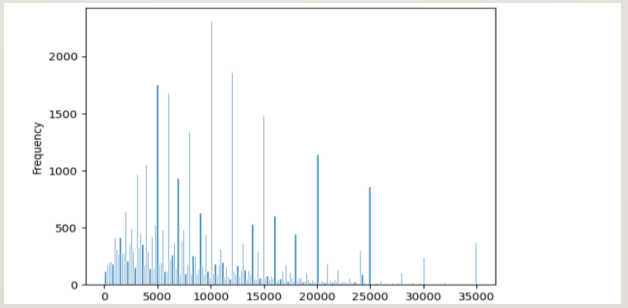
DATA CLEANING

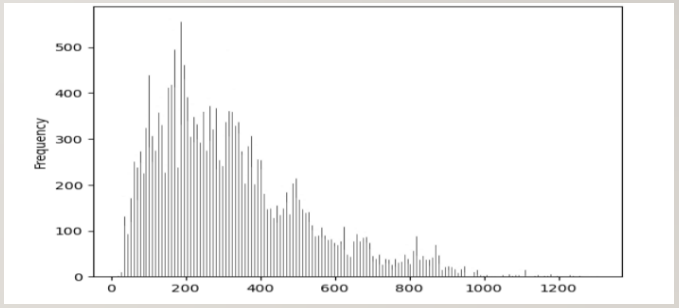
- Steps taken to clean the data:
- Dropped columns contains all or maximum NULL values
- Dropped rows contains NULL values and where the values are not huge
- Converting data to proper Datatype for analysis.

EDA-UNIVARIATE ANALYSIS

- The Rank-Frequency plot shows that it follows the power law distribution
- total_rec_prncp shows a very clear pattern of 5000, 10000,15000,20000,25000,30000,35000
- In Installment also we see similar pattern of higher frequency near round off values like 100, 200, 300, 400, 500, 600,800

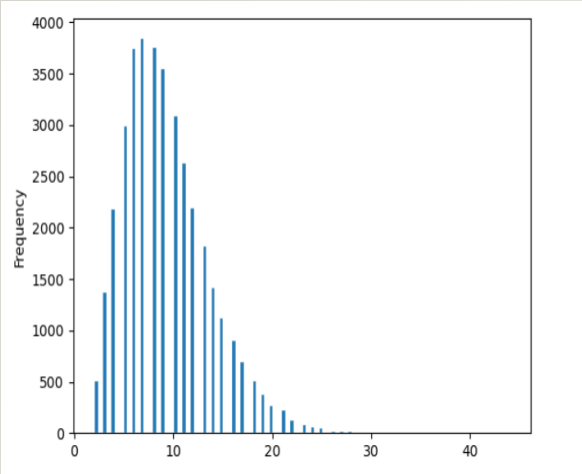


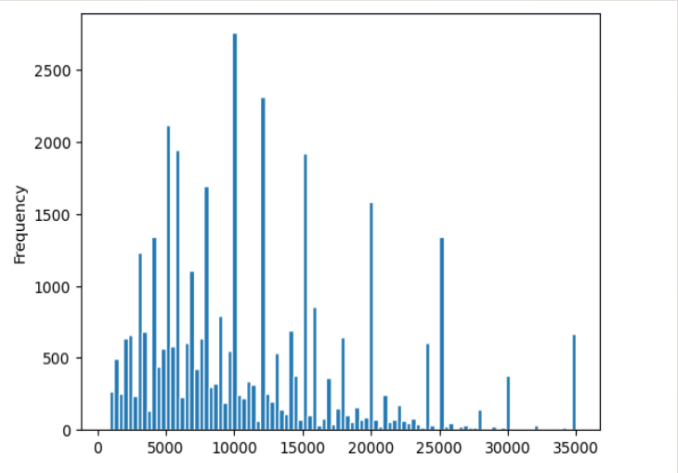


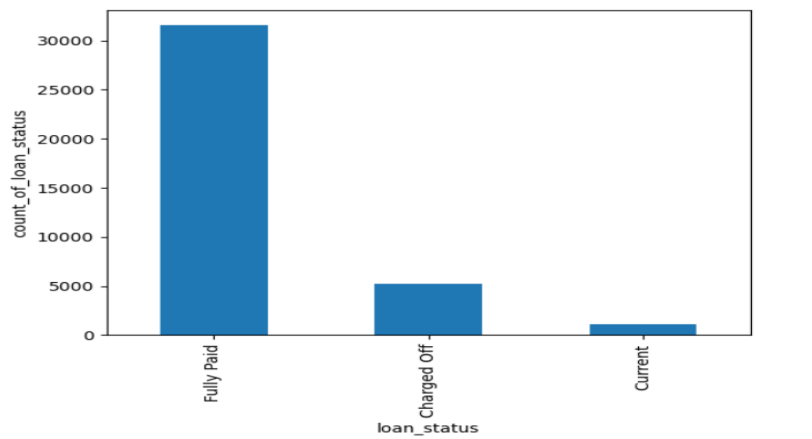


EDA-UNIVARIATE ANALYSIS

- Open Accounts shows a very nice curve
- Loan Amount, shows a spike at 5000,10000,15000,20000,25000,30000,35000, which means that either people take loans of round figures or banks give round figure loans
- For loan_status=Charged_off have high possibility to be defaulters.Sepecially where loan amount is in higher range and applicant status is Charged_off the risk is very high.



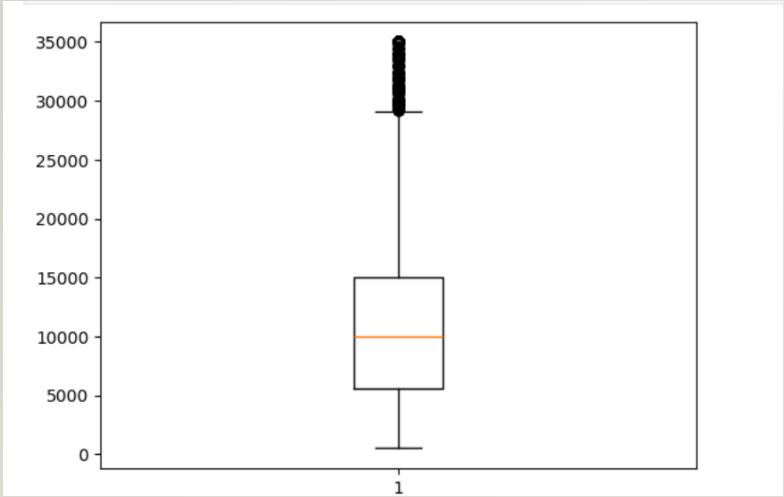




EDA AND DATAVISUALIZATION- LOAN AMOUNT AND FUNDED AMOUNT

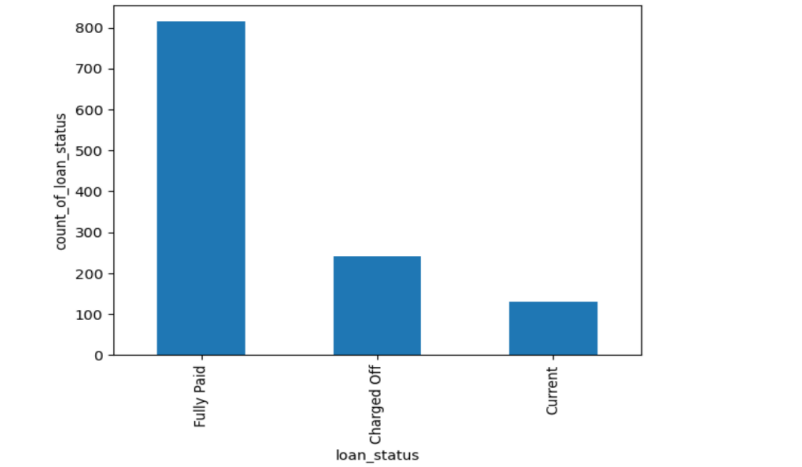
- We considered the key variables for EDA and Data Visualization.
- Correlation between loan amount and funded amount is 0.98 which is very high.

The outliners for loan amount **ranges from 30000 and 35000.**



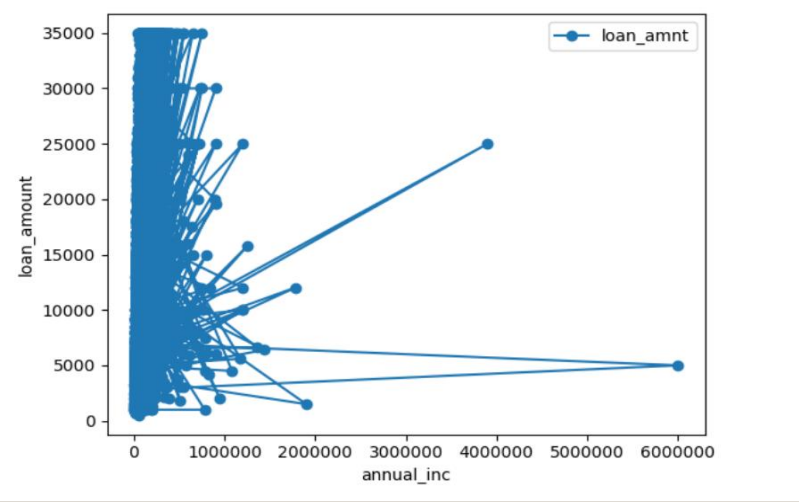
EDA AND DATAVISUALIZATION- LOAN AMOUNT AND FUNDED AMOUNT

- Even though we have some outliner range for loan_amount but it does not have a huge impact on defaulters. As we see from below presentation most of the cases where loan_amnt is between 30000 and 40000 loan_status is fully_paid. Same goes for funded amount.



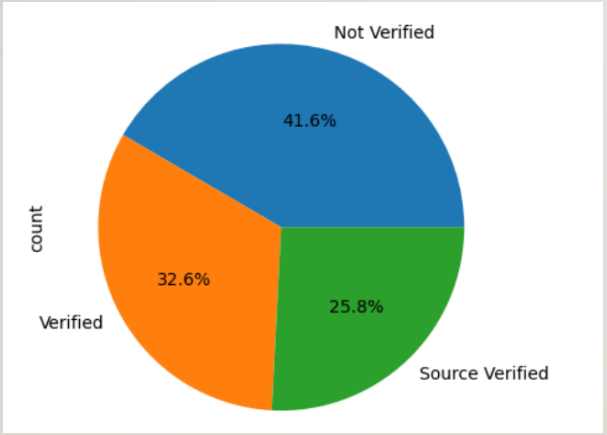
EDA AND DATAVISUALIZATION- ANNUAL INCOME

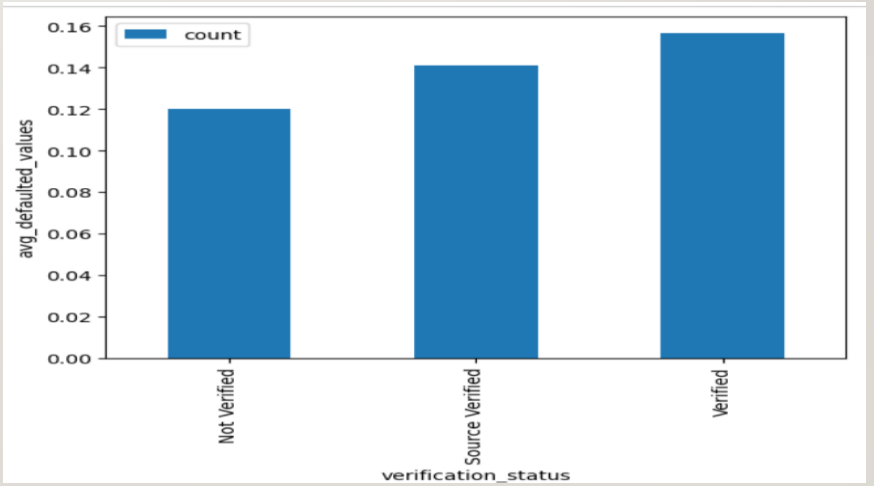
- some cases where annual income is low the loan_amnt is in higher range. In this case if previous loan_status is not Fully paid then approving other loan may be risky. And we found some borrowers like that. Though the percentage is very less.(around 0.6 percent)



EDA AND DATAVISUALIZATION- VERIFICATION STATUS

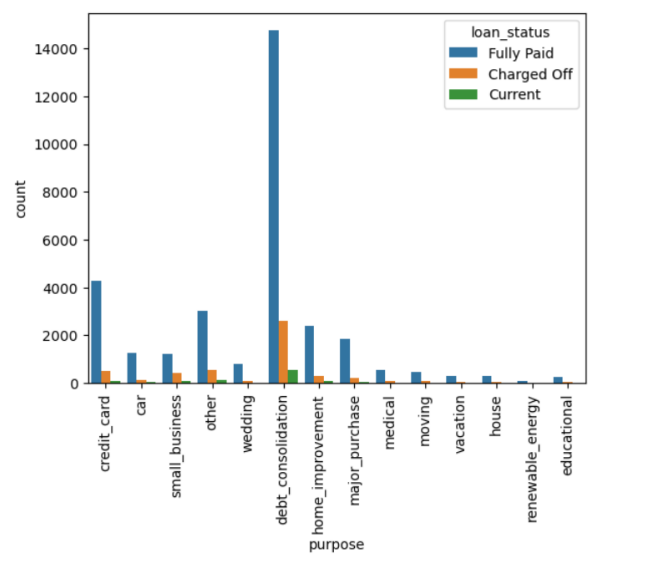
- The percentage of non verified borrowers were the highest(around 41.6%) which is unusual scenario. But surprisingly the percentage of defaulters are most in case of verified borrowers-16%,followed by Source Verified-14%. But 12% non verified applicants seems to be defaulted.

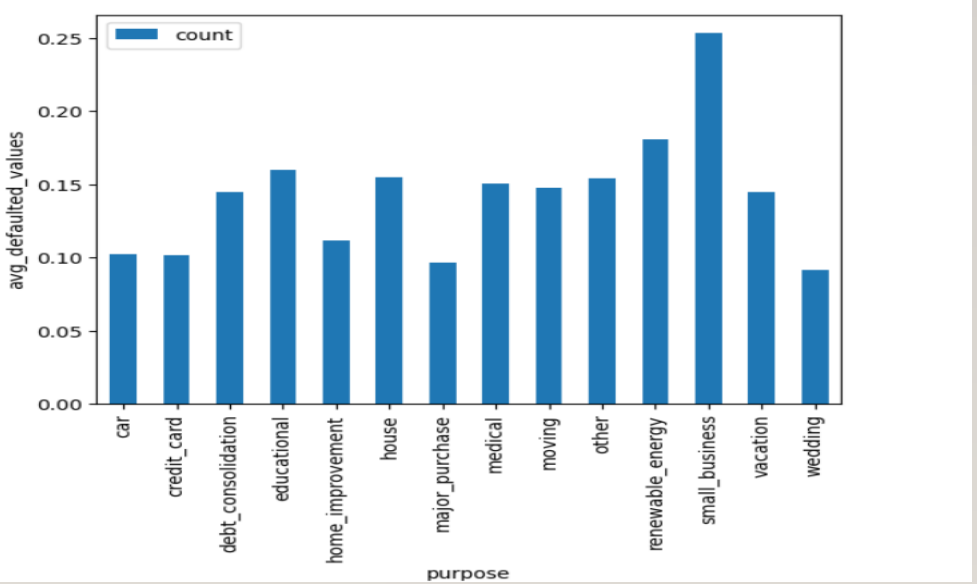




EDA AND DATAVISUALIZATION- PURPOSE

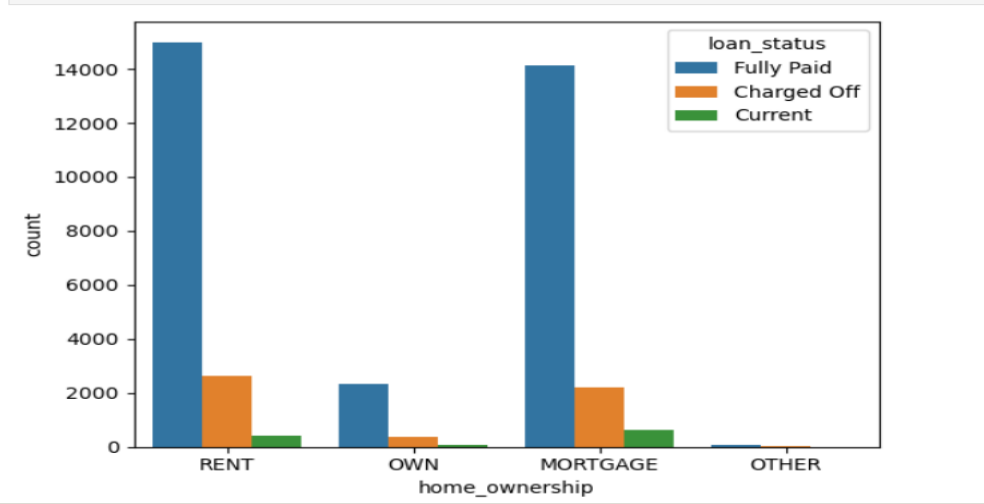
- Most people borrowed loan for debt_consolidation followed by credit card.
- People taking loan for small_business has the highest defaulted rate followed by people taking loan for renewable energy.

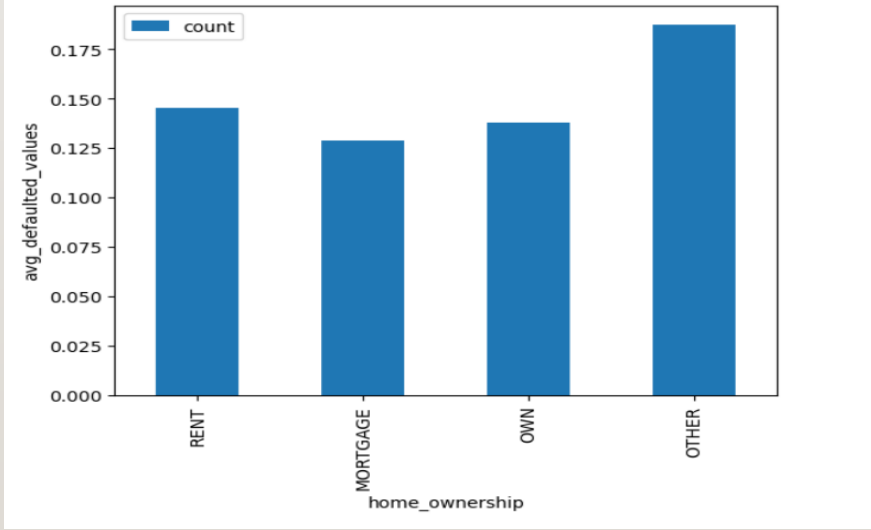




EDA AND DATAVISUALIZATION- HOME OWNERSHIP

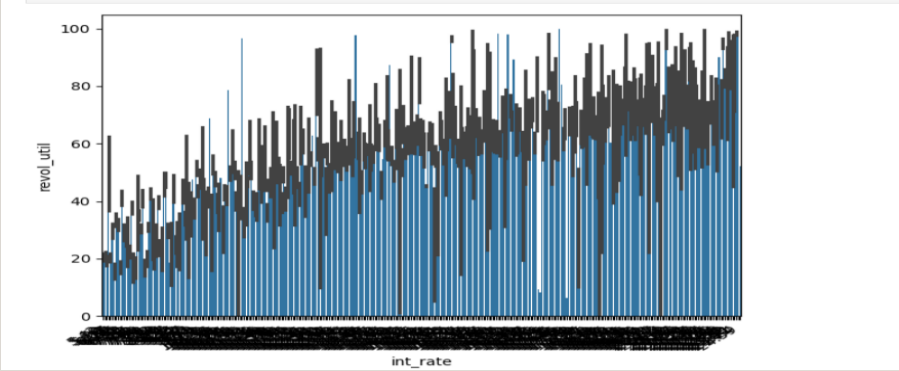
- Whether the borrower owns, rents, or is mortgaged in every cases we are seeing defaulters.
- Surprisingly percentage of defaulters are lowest in case of home_ownership is Mortage.

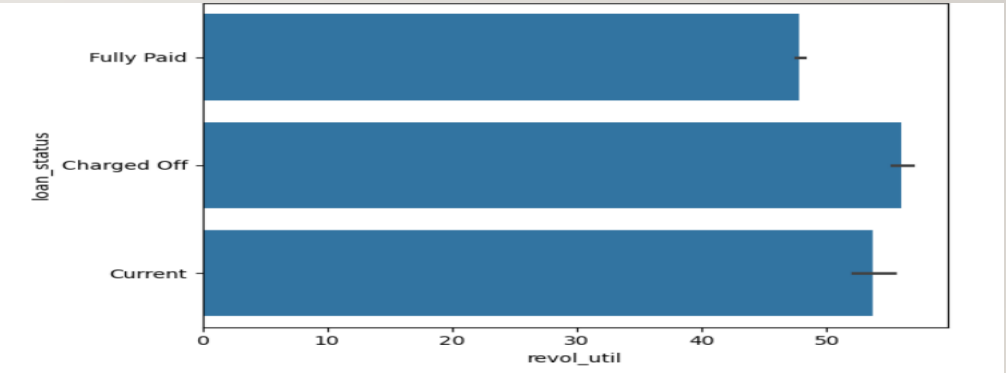




EDA AND DATAVISUALIZATION- REVOL_UTIL AND INTEREST RATE

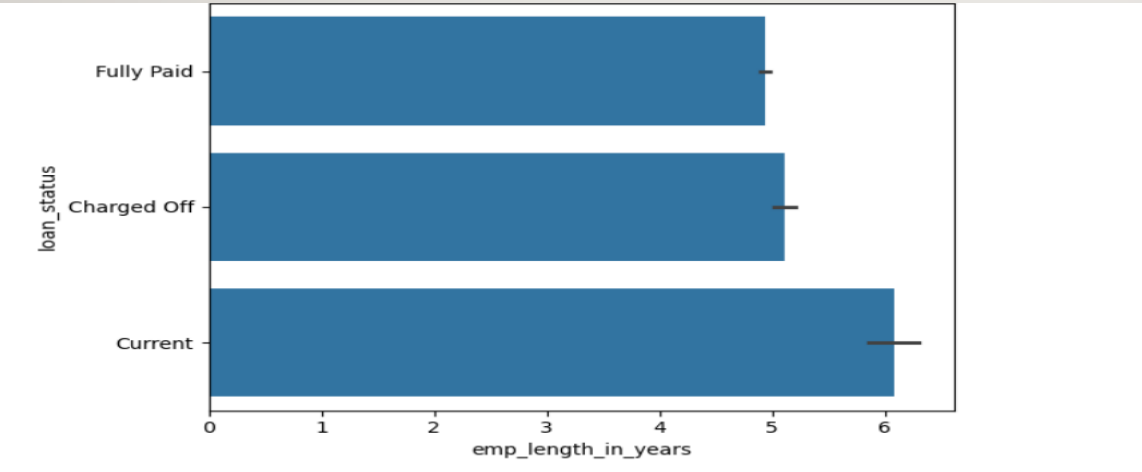
- There is a clear pattern between revol_util and int_rate, with increase in amount of credit the borrower is using the interest rate increases, which means risk is higher for people who are closer to their credit limit.
- Borrower who have used their credit limit more are more likely to have Charged off status that means they are more likely to default.





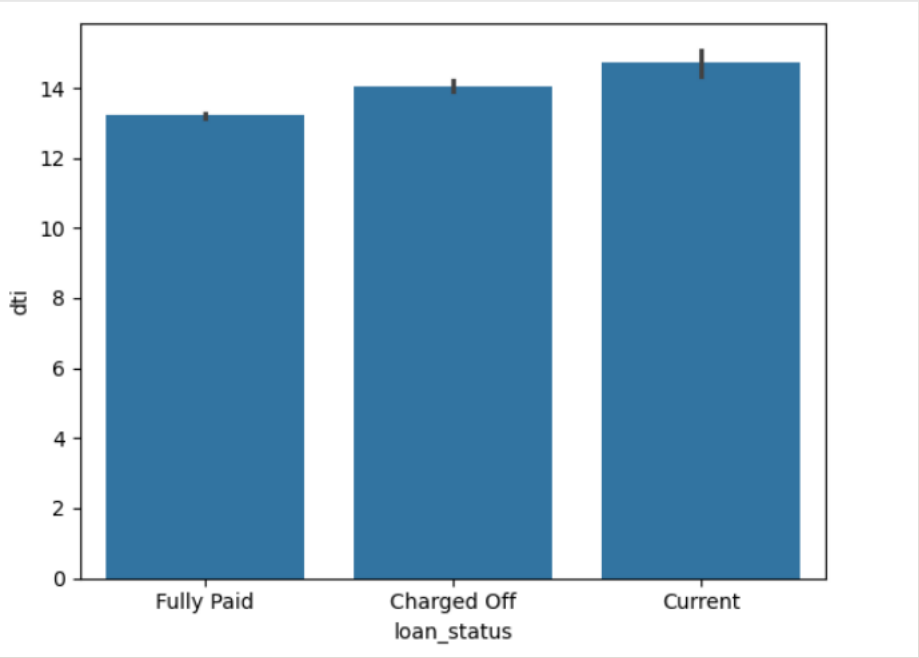
EDA AND DATAVISUALIZATION- EMPLOYMENT_LENGTH

- Employment length doesn't really have a huge impact on repayment of loan.



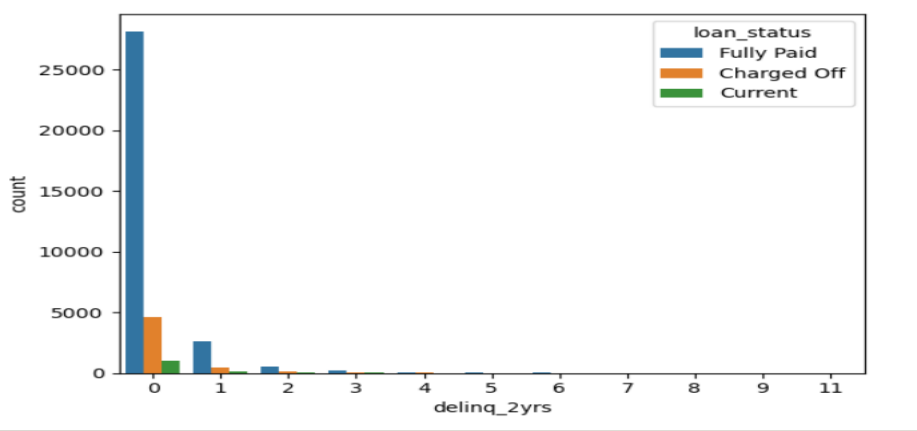
EDA AND DATAVISUALIZATION- DTI

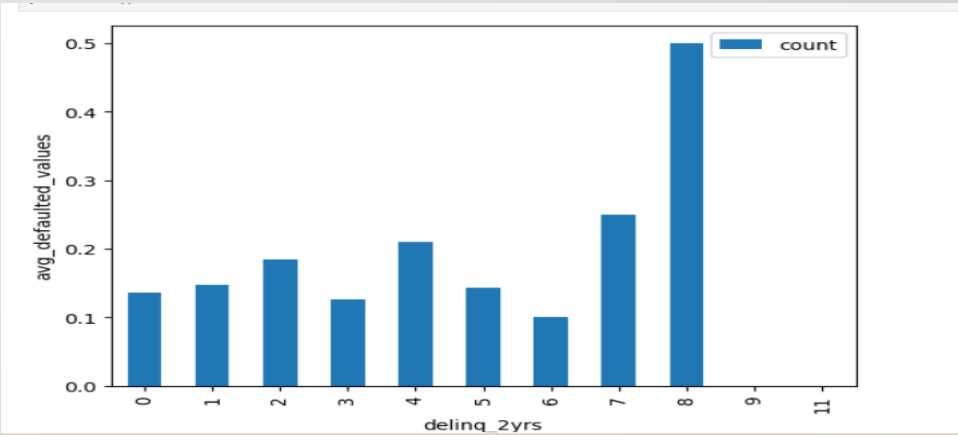
- DTI doesn't really have a huge impact on repayment of loan.



EDA AND DATAVISUALIZATION- DELINQ_2_YR

- delinq_2yr value is consistently resucing in the dataset which is good.
- When delinq_2yrs is 8 the defaulters value is 50%. From the representation we cannot predict a specific pattern to find the defaulters considering delinq_2yrs as the pattern in not consistent.





EDA AND DATAVISUALIZATION- GRADE

- It is a consistent pattern. The chance of finding defaulters is highest for grade G and lowest in grade A. Company should consider D,E,F and G grade as risky and implement proper risk assessment for those grades.