

# Lending Club Case Study

Parth Gandhi



# Content

- Problem Statement
- Risk factor
- Decision Factor
- Analysis Pedagogy
- Univariate Analysis
- Bivariate Analysis
- Conclusion

# Problem Statement

- Lending club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.
- The 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 utilise this knowledge for its portfolio and risk assessment.

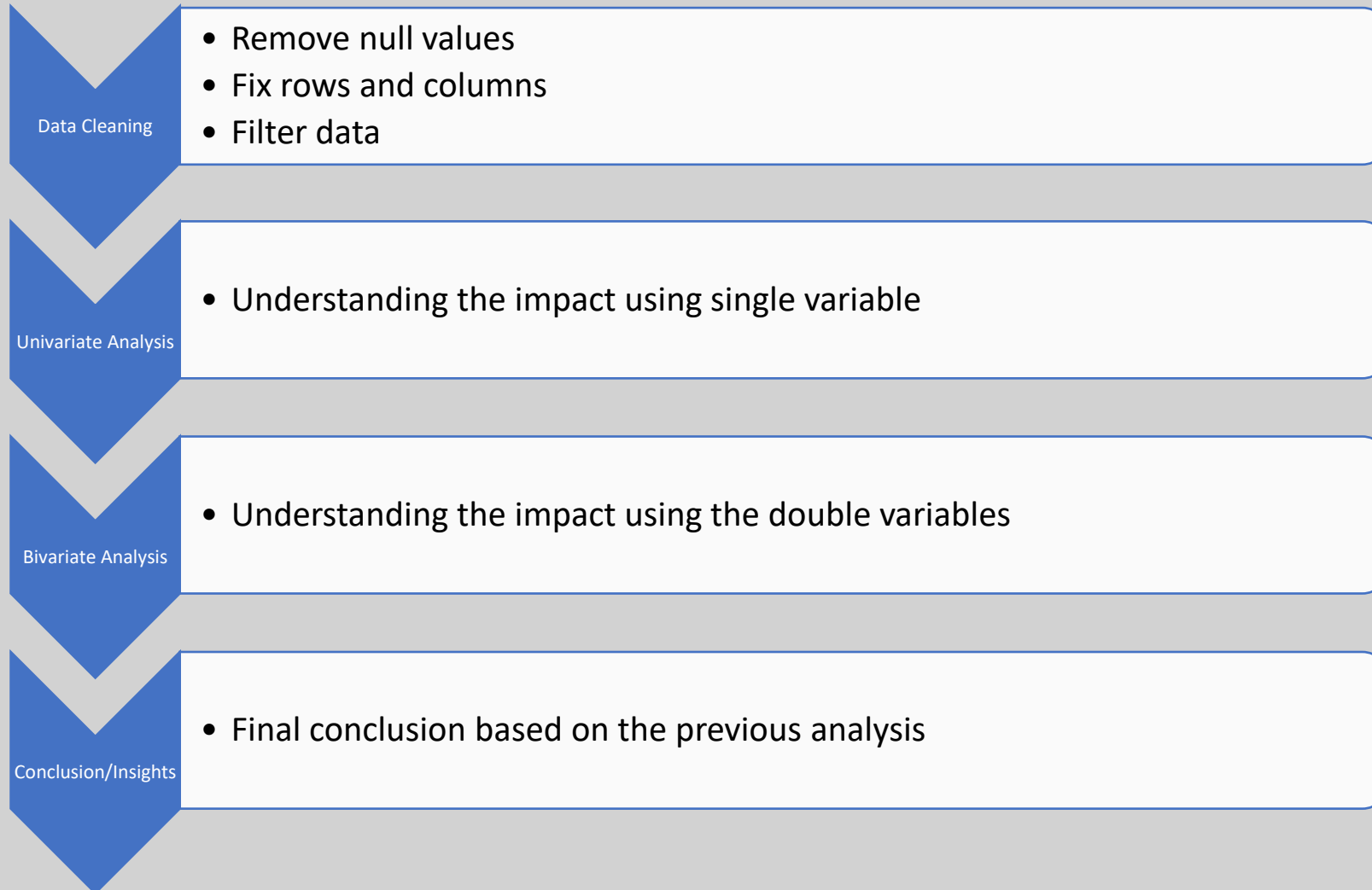
# Risk Factor

- 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:
  1. If the applicant is likely to **repay the loan**, then not approving the loan results in a **loss of business** to the company
  2. 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 Factor

- 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.).

# Analysis Pedagogy

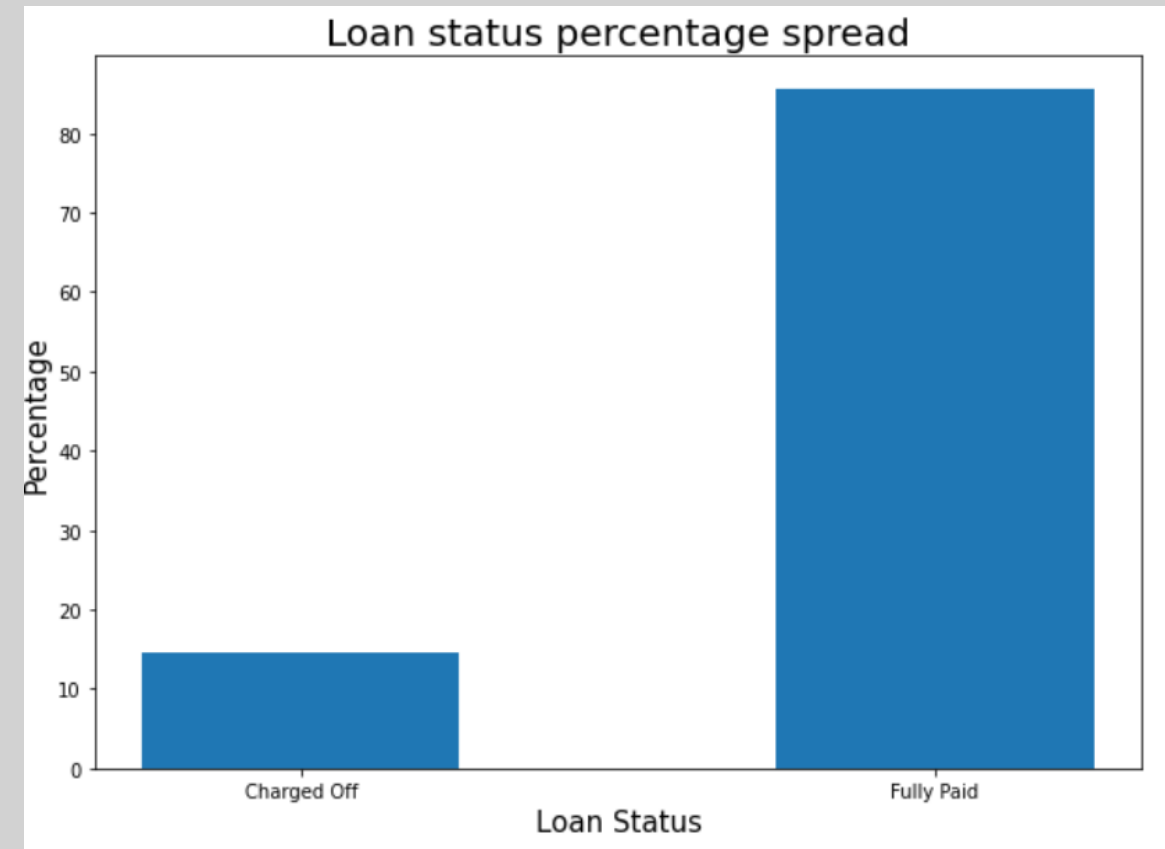


# Defaulter vs Non-defaulter Percentage



Out of the total Loan application

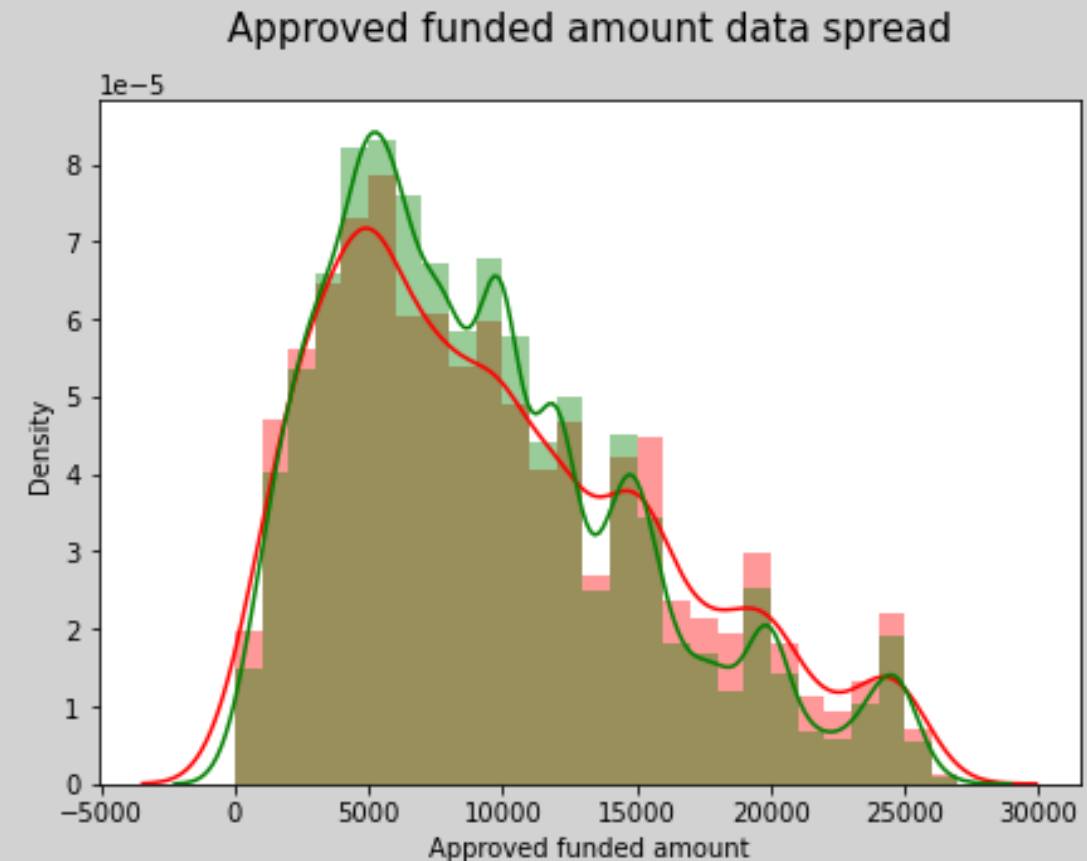
- Charged Off(Defaulter) Loan percentage ~ 15%
- Fully Paid Loan percentage ~ 85%



# Analysis based on Approved Loan Amount



- People with approved loan amount ranging from 4k to 12k has high chance of repaying the loan.
- People with approved Loan amount beyond 12k, has high chance of defaulting

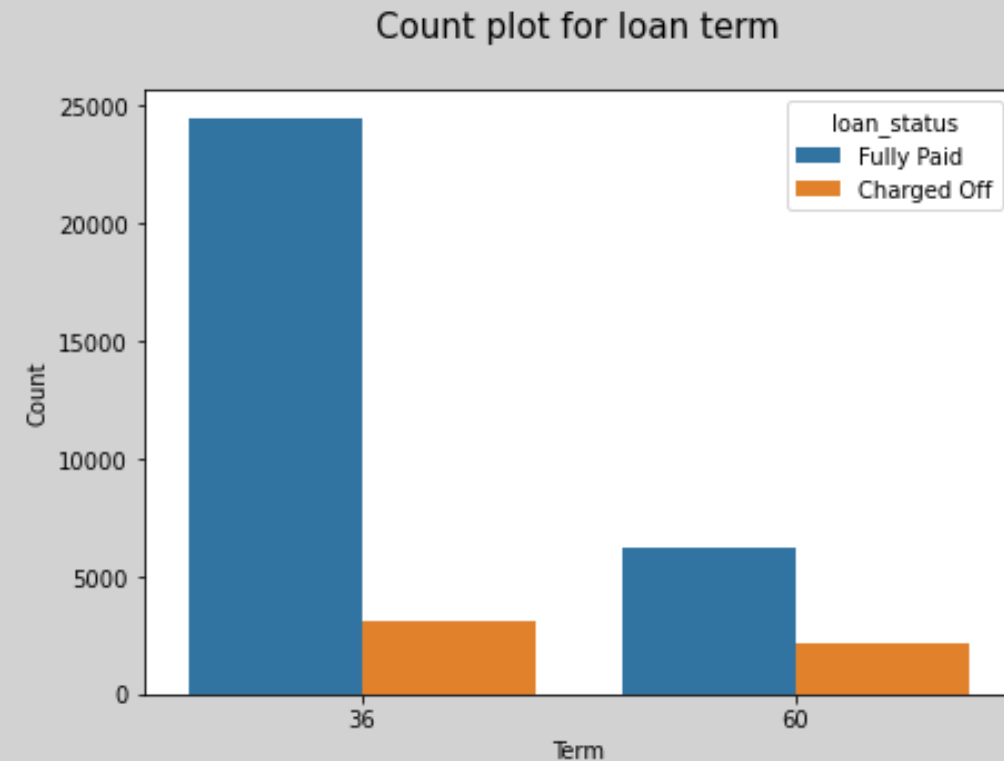




# Loan Term relationship with Defaulter



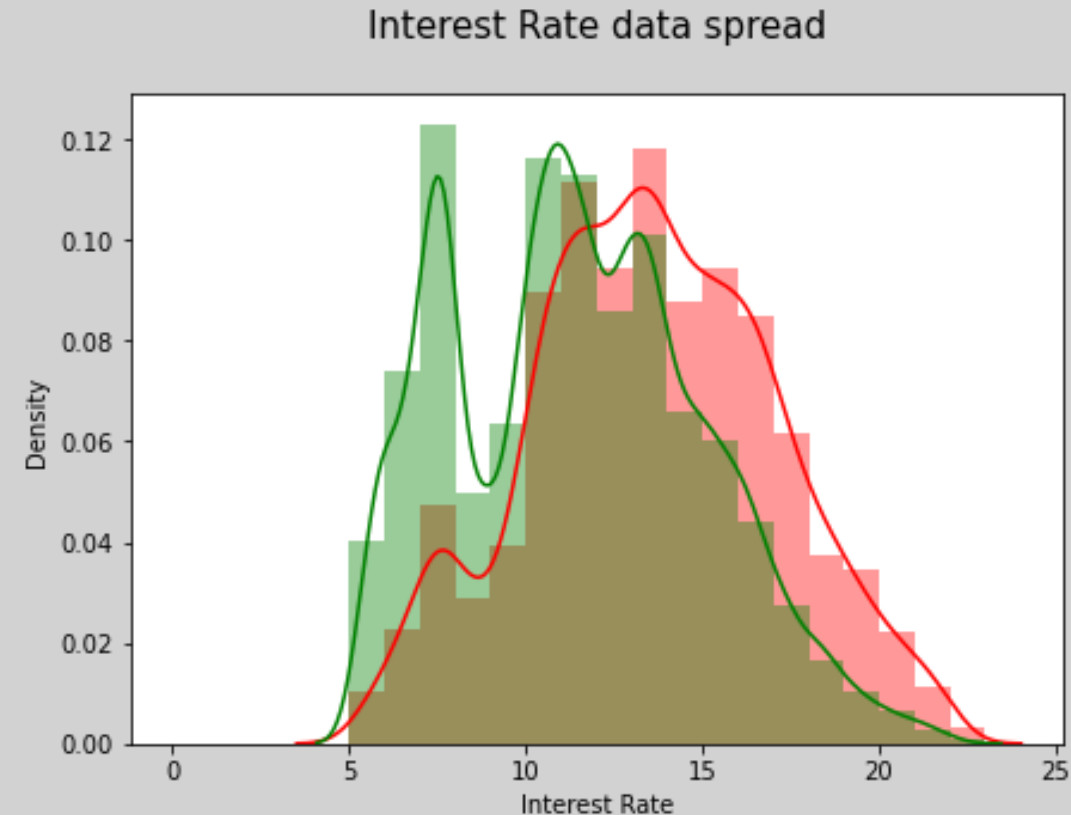
- More people opt for 36 months term than 60 months term
- People opting loan for 36 months has high chance of re-paying up the loan as against the one opting loan for 60 months



# Distribution of Interest rate



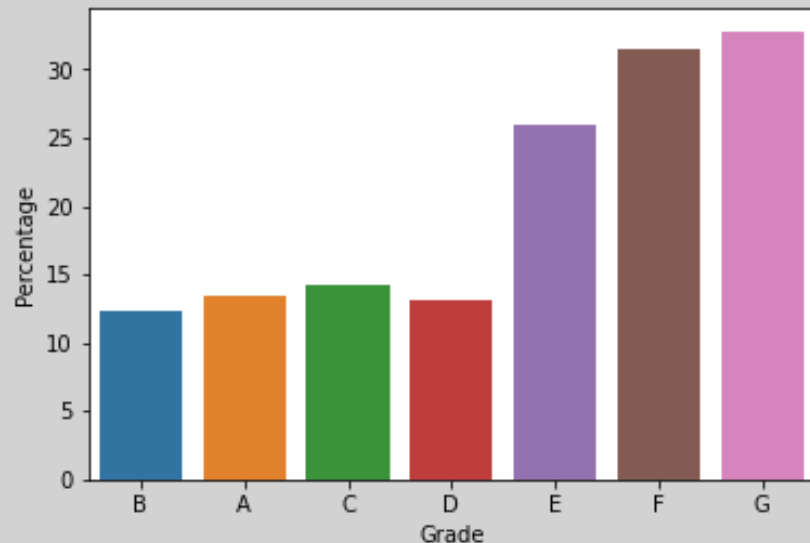
- People with interest rate less than 12 has high chance of re-paying the loan
- However, people going for high interest rate (More than 12) has high chance of defaulting



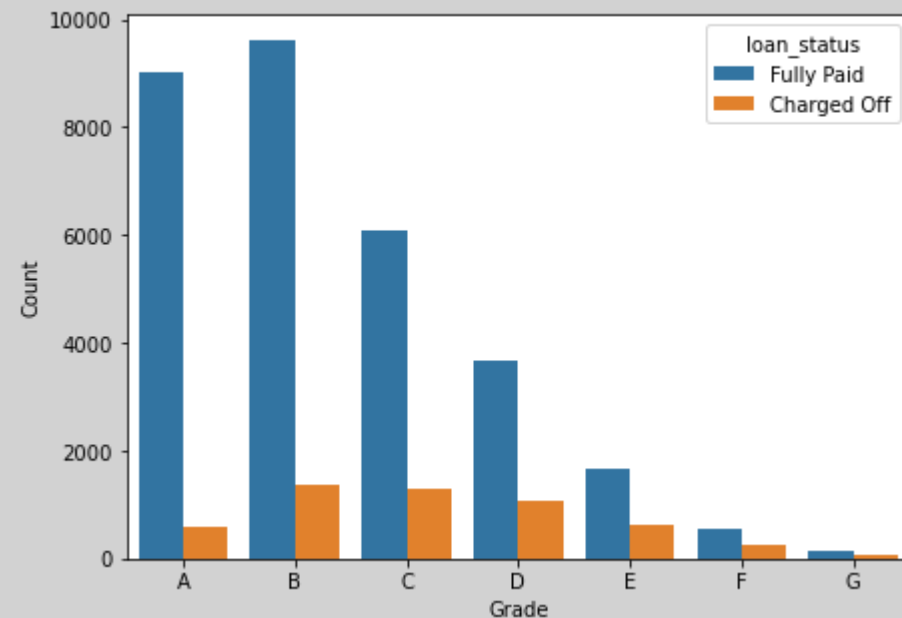
# Based on LC Grade

- Grade A, B, C has high number of applications
- At a high level If Grade B, C, D has high number of defaulters.
- However, the percentage of getting defaulting is high when the application moving from D to G

Percentage of Defaulted applications Vs Grades



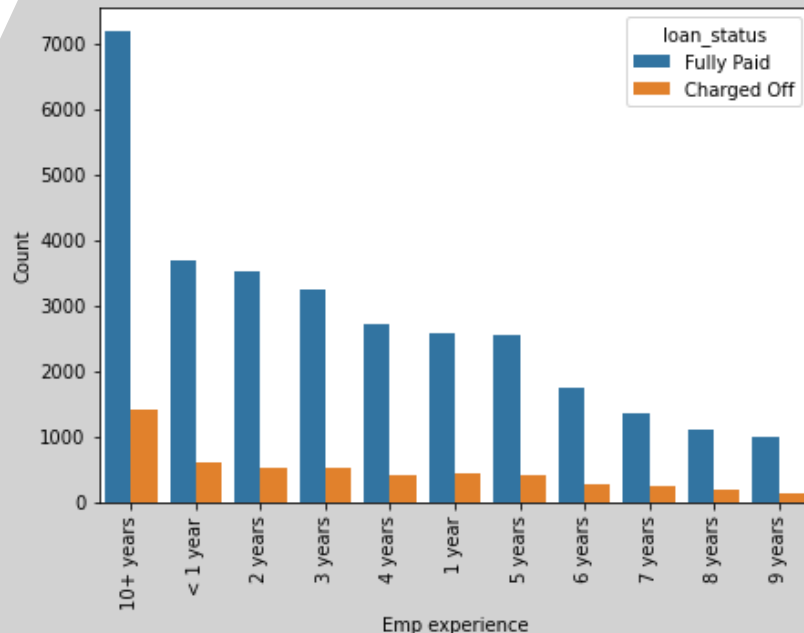
Count plot for Grade data spread



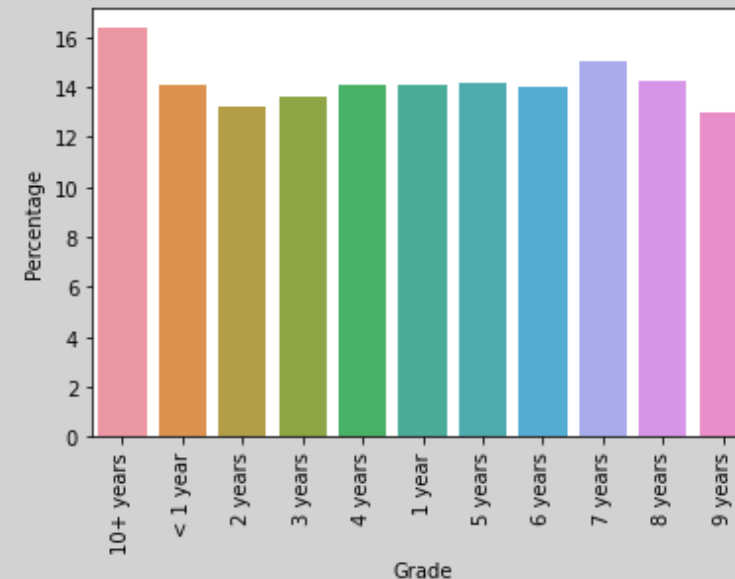
# Employment Experience Analysis

- We are getting more applications with 10+ years of experience and they also have high percentage of loan default
- With experience of 1, 4, 5, 6 years have almost similar percentage of defaulting

Count plot for employee experience



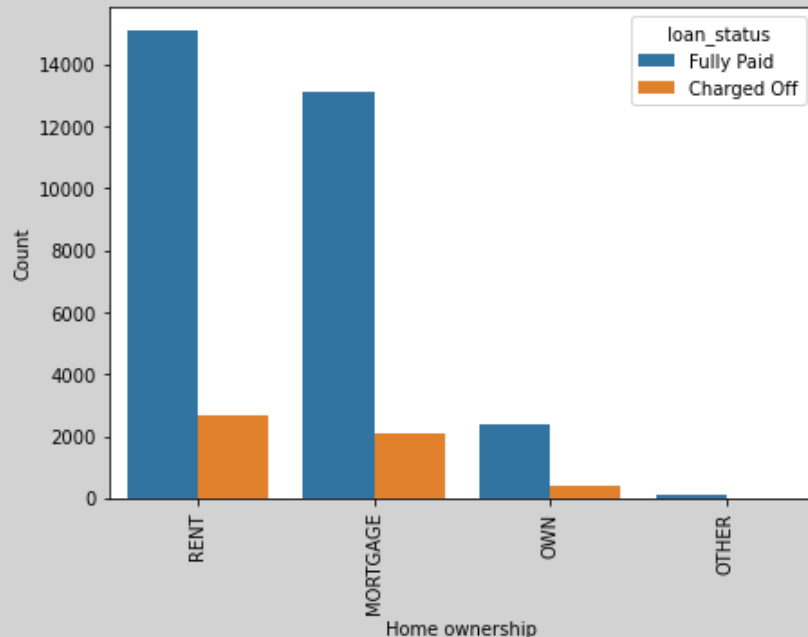
Percentage of default Vs Emp experience data spread



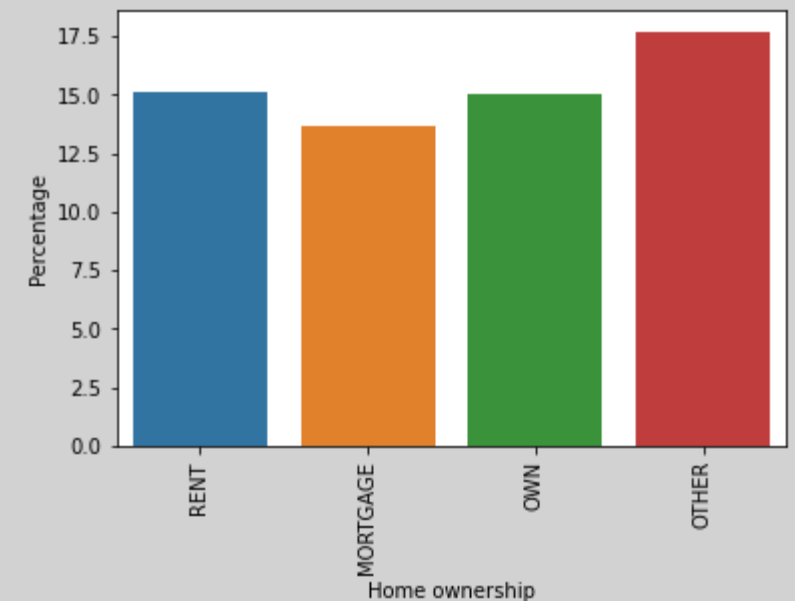
# Home Ownership Analysis

- People staying in rented house or mortgage has higher percentage of loan application
- However, people with OTHER category has the highest defaulter percentage, followed by RENT, OWN and MORTGAGE

Count plot for Home ownership data spread



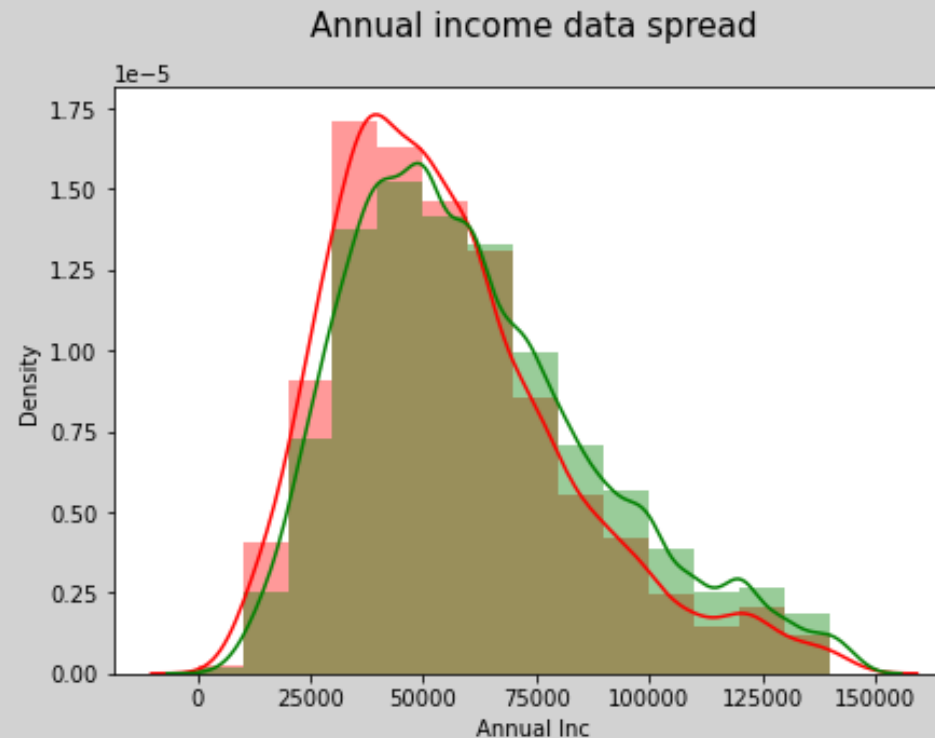
Percentage of default Vs Home ownership data spread



# Annual Income Analysis



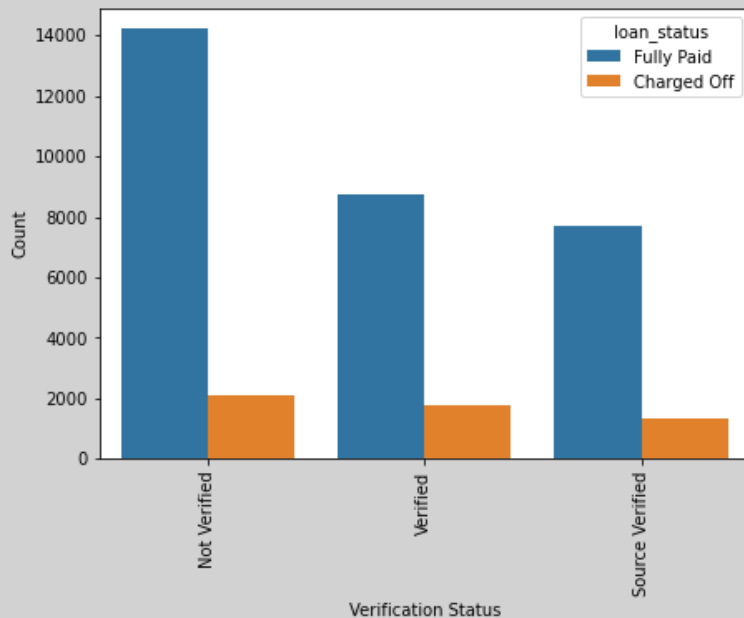
- People with annual income between 30k to 50k has high chance of getting default than the people with income more than 50k



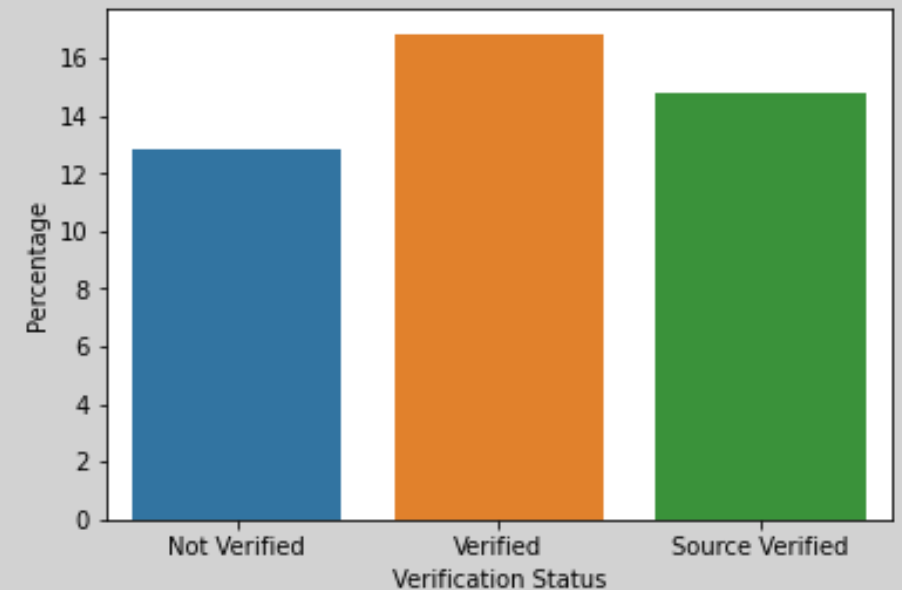
# Verification Status Analysis

- People with Not verified income status has higher number of loan applications
- On the other hand, people with Verified income status has more chance of being defaulter

Count plot for Verification status data spread



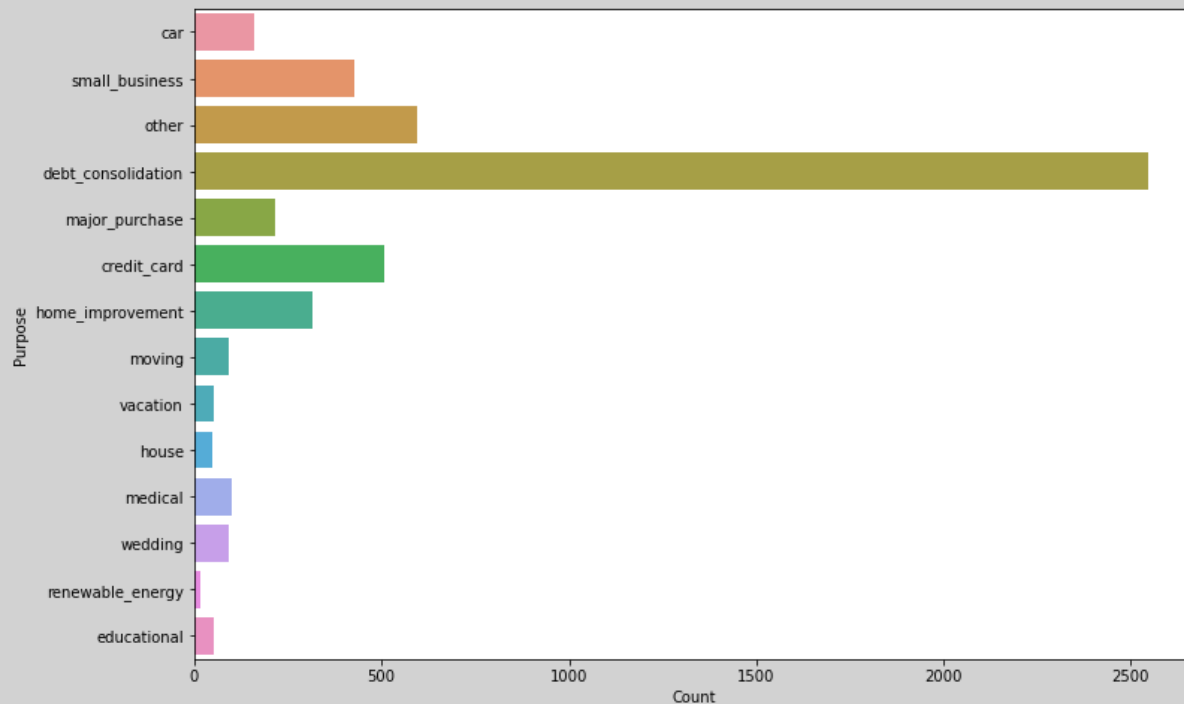
Percentage of Defaulted applications Vs Verification Status



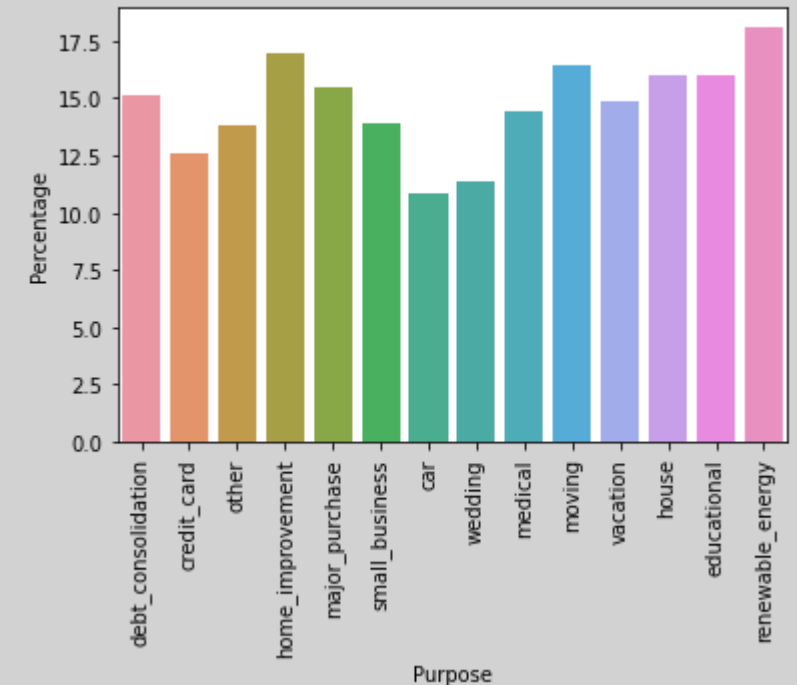
# Loan Purpose Analysis

- Highest number of borrower took loan for the purpose of "**debt\_consolidation**" and they have the highest number of defaulters
- However, the defaulter chances are more for the loan taken for "**renewable\_energy**" and "**home\_improvement**"

Count plot for loan purpose



Percentage of Defaulted applications Vs Loan purpose

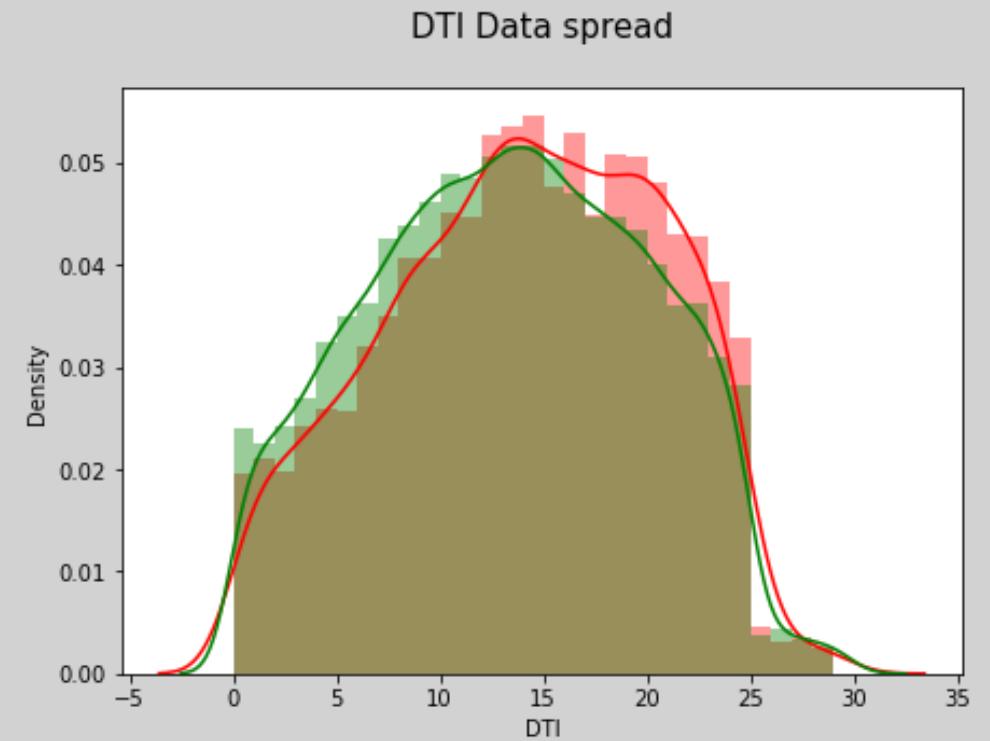




# Loan DTI analysis

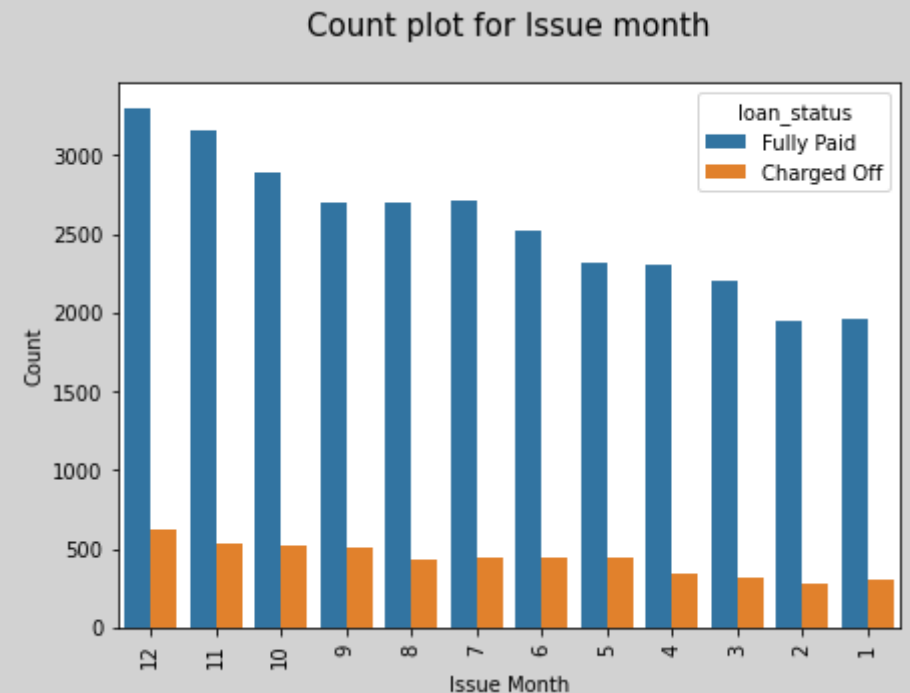


- Borrower with dti less than 12 has high chance of re-paying the loan
- Borrower with more dti ( $>12$ ) has high change of defaulting



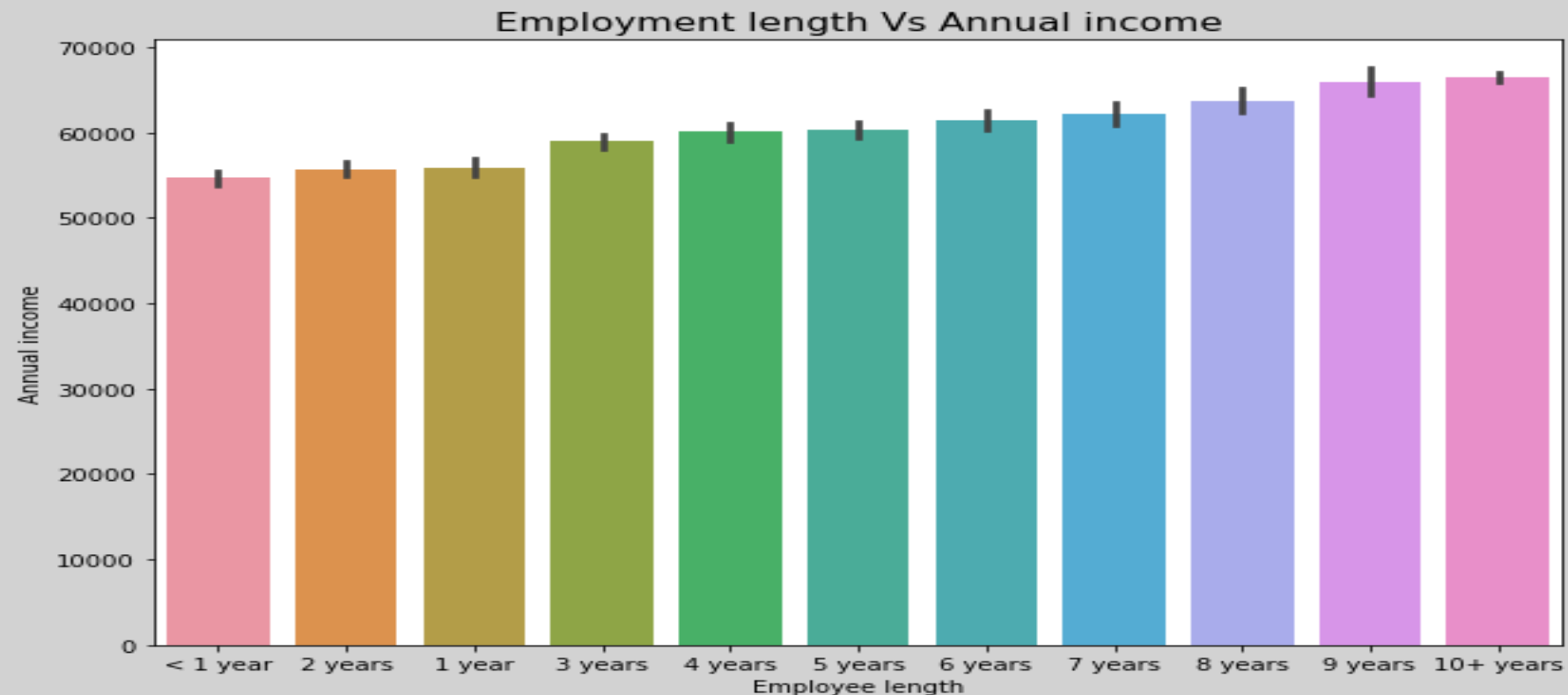
# Loan Issue Month Analysis

- There is a slight upward trend in the number of applications from Jan to Dec.
- December has highest number of applications and high number of defaulters.



# Emp length Vs Annual income (Bi-variate)

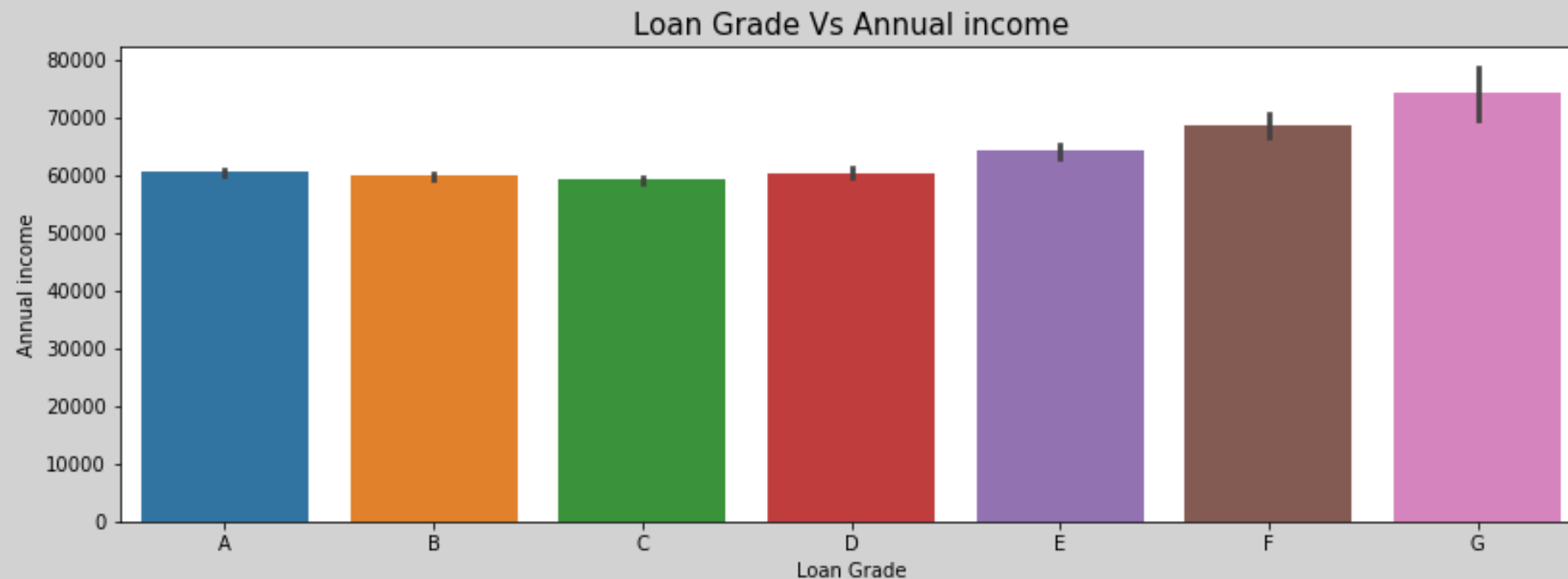
- As the experience is increasing the average annual income is also increasing, but at a slow rate



# Grade Vs Annual income (Bi-variate)

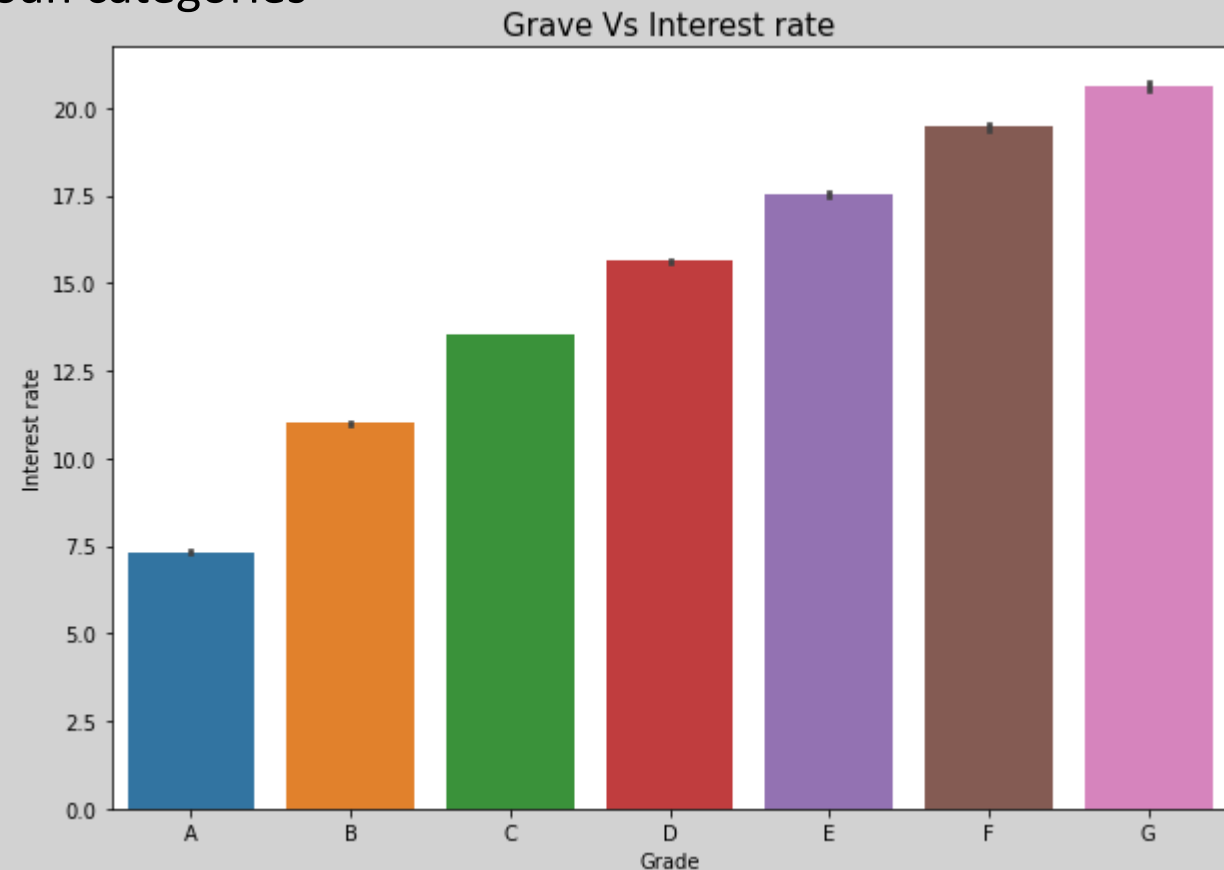


- As the grade moves from A to G, annual income(average) is increasing.



# Grade Vs interest rate (Bi-variate)

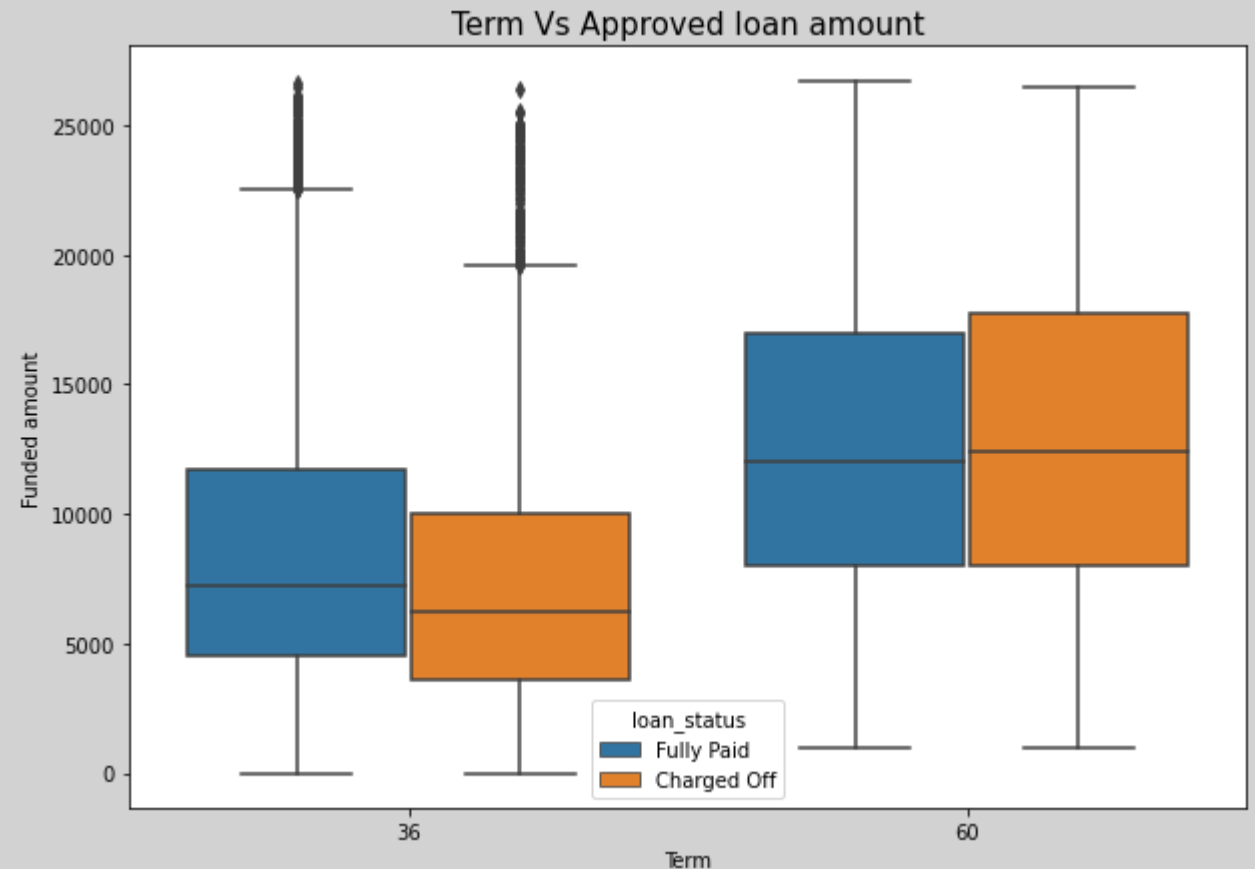
- LC Grade F & G are attracting high interest rates
- Grade A has low interest rates among all the loan categories



# Term Vs Funded amount approved



- Borrower with higher loan approved amount will go for higher loan term.
- More Borrower with funded approved amount are opting for 36 months as term and they are re-paying the loan as well

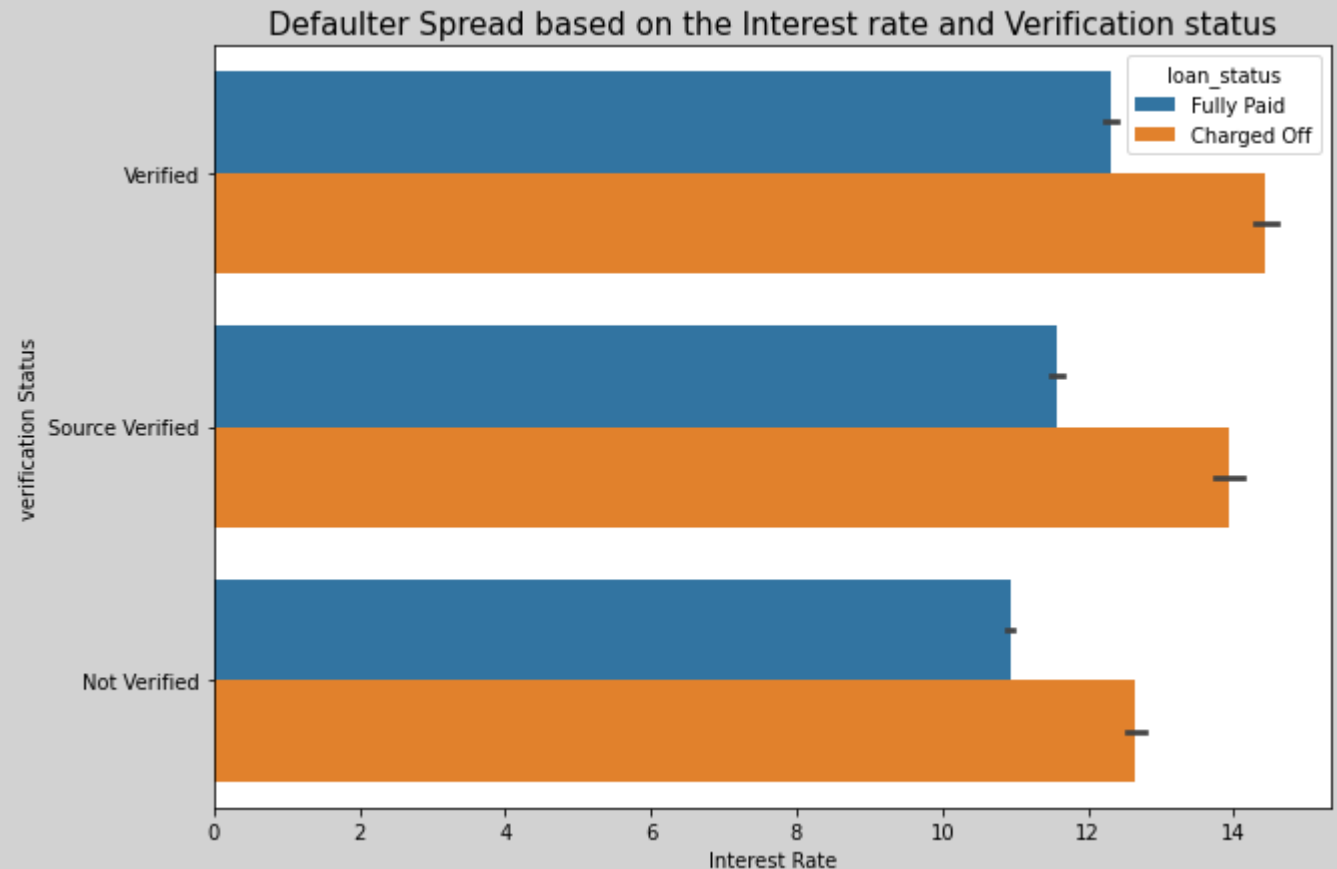


# Interest rate Vs Verification status (Bi-variate)



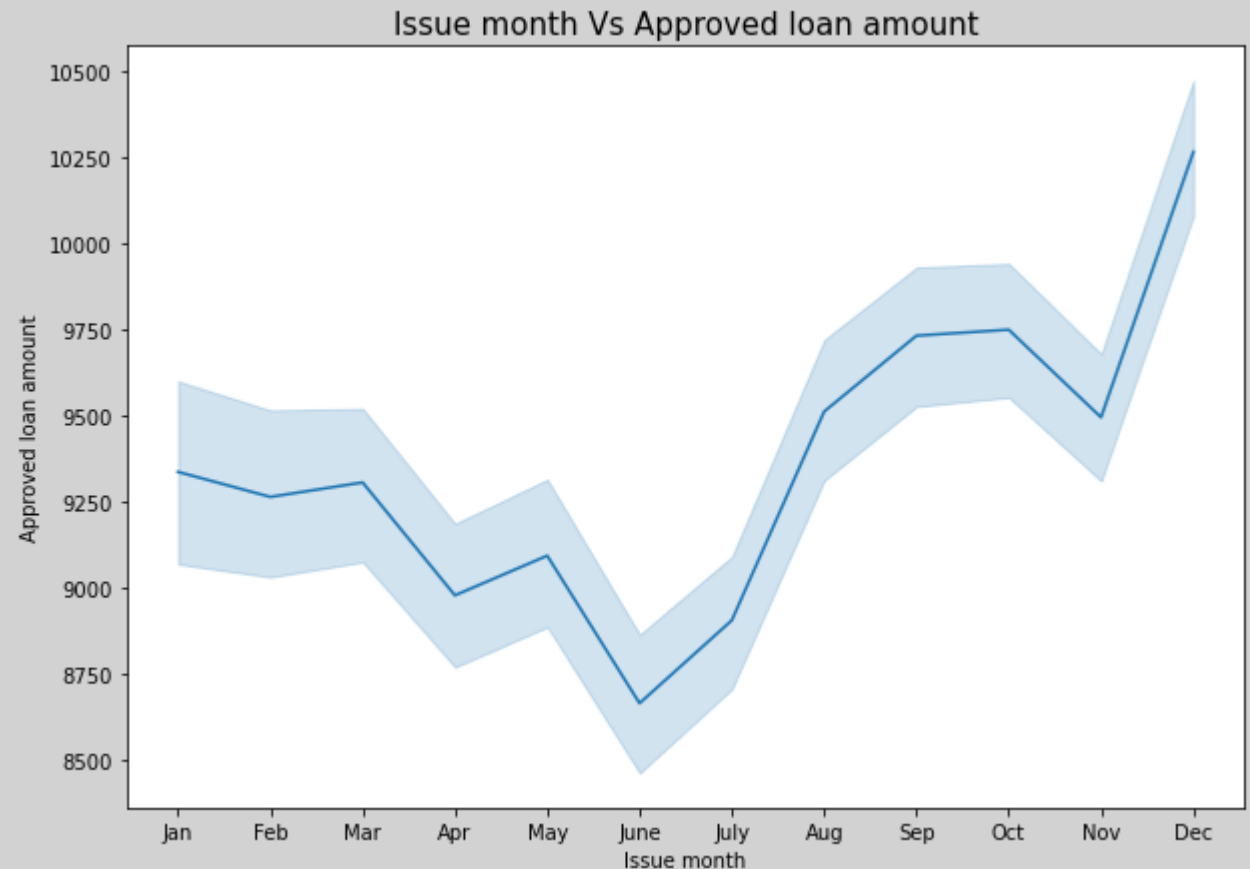
- If borrower with income verified is going for high interest rate (Average), chances of getting defaulted is high

- In all categories, if the borrower is going with low interest rates, chances of re-paying the loan is high



# Issue month Vs funded amount (Bi-variate)

- The average approved loan amount is very less in middle of the year (June month).
- From June, again we can see an upward trend with another dip in Nov.
- December has the highest approved loan amount (Average).

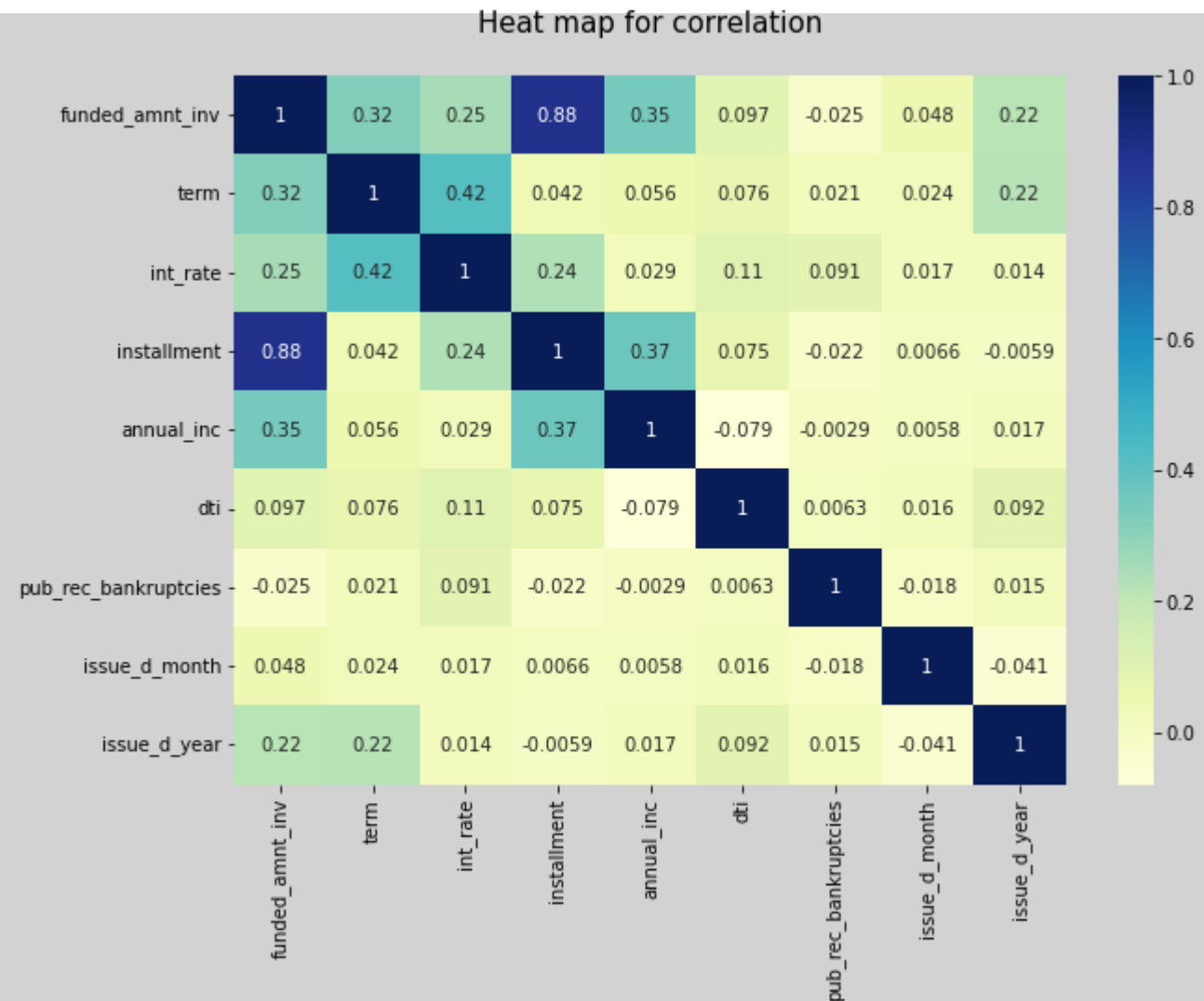




# Correlation Matrix



- Approved funded amount has a positive correlation with installments (0.88). As the funded amount increases, installment increases.
- Interest rate to term has positive correlation(0.42). As the loan term increases, interest rate increases



# Conclusions

Driving factor (Variables) in determining whether the borrower can be defaulter are as below:

- Term(**term**) 60 months (Categorical Variable)
- Loan amount(**funded\_amnt\_inv**) > 12k (Continuous Variable)
- Interest rate(**int\_rate**) > 12% (Continuous Variable)
- LC Grade(**grade**) D,E,G (Categorical Variable)
- Employment Length(**emp\_length**) >= 10 years (Continuous Variable)
- Home Ownership (**home\_ownership**) is OTHER (Categorical Variable)
- Annual Income (**annual\_inc**) between 30K to 50K (Continuous Variable)
- Verification status(**verification\_status**) is Verified (Categorical Variable)
- Purpose of Loan(**purpose**) is for "renewable energy" of "home improvement" (Categorical Variable)
- DTI(**dti**) > 12 (Continuous Variable)
- Loan disbursed/approved in December (Derived Variable)
- Borrower with verified income and high interest rate

# Thank You

Parth Gandhi

