



# EDA

For

---

# Leading Club Case Study

Date:- 06/07/2022

**Participated By:-**

1. Prakash Shrivastav
2. Pooja Maini

# Content

- Problem Statement and Approach
- Assumptions
- Univariate Analysis: Term
- Univariate Analysis: Employment Length
- Univariate Analysis: Grade
- Univariate Analysis: Home Ownership
- Univariate Analysis: Purpose
- Univariate Analysis: Verification status
- Bi-Variate Analysis: Loan Status vs Funded Amount
- Bi-Variate Analysis: Grade vs Funded Amount
  
- Bi-Variate Analysis: Sub-Grade vs Loan Status
- Bi-Variate Analysis: Verification Status vs Loan Status
- Bi-Variate Analysis: Correlation
- Recommendations



## Problem Statement

- Identify the driving factors behind loan default.

## Approach



## Assumptions

**Grade :** It refers to loan grade which is used to assign a quality score to a loan based on the credit history of the borrower, quality of the collateral, and the likelihood of the repayment.

Likelihood of loan repayment

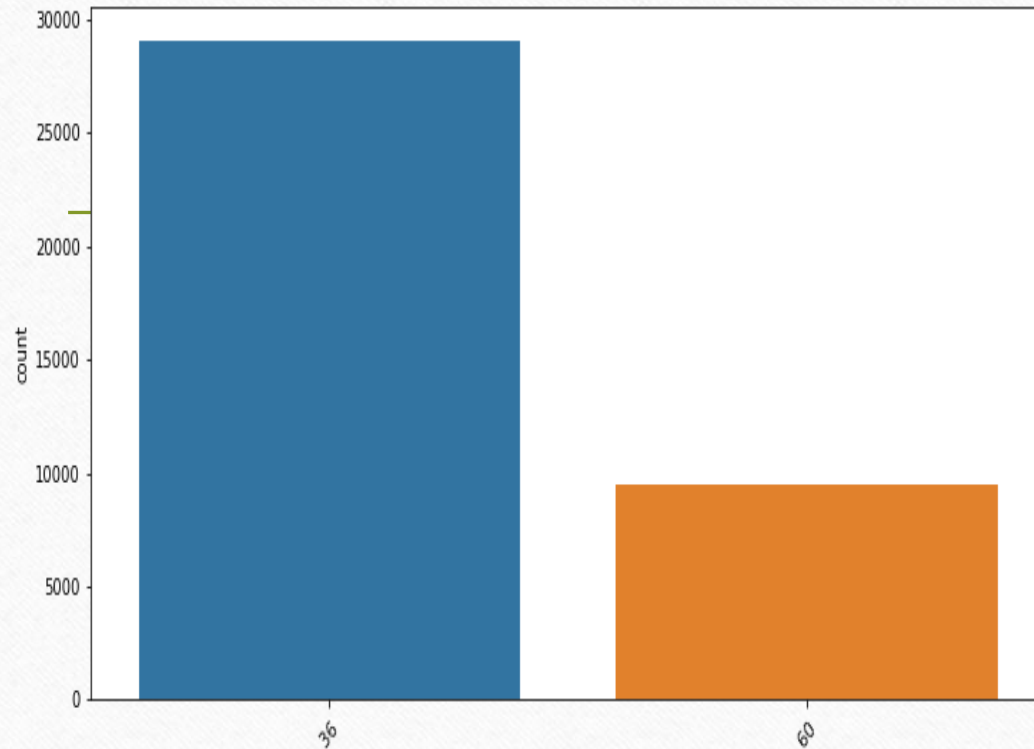
A > B > C > D > E > F > G ( A has greater likelihood of loan repayment than B)

**Debt consolidation:** It refers to the act of taking out a new loan to pay off other liabilities and consumer debts.

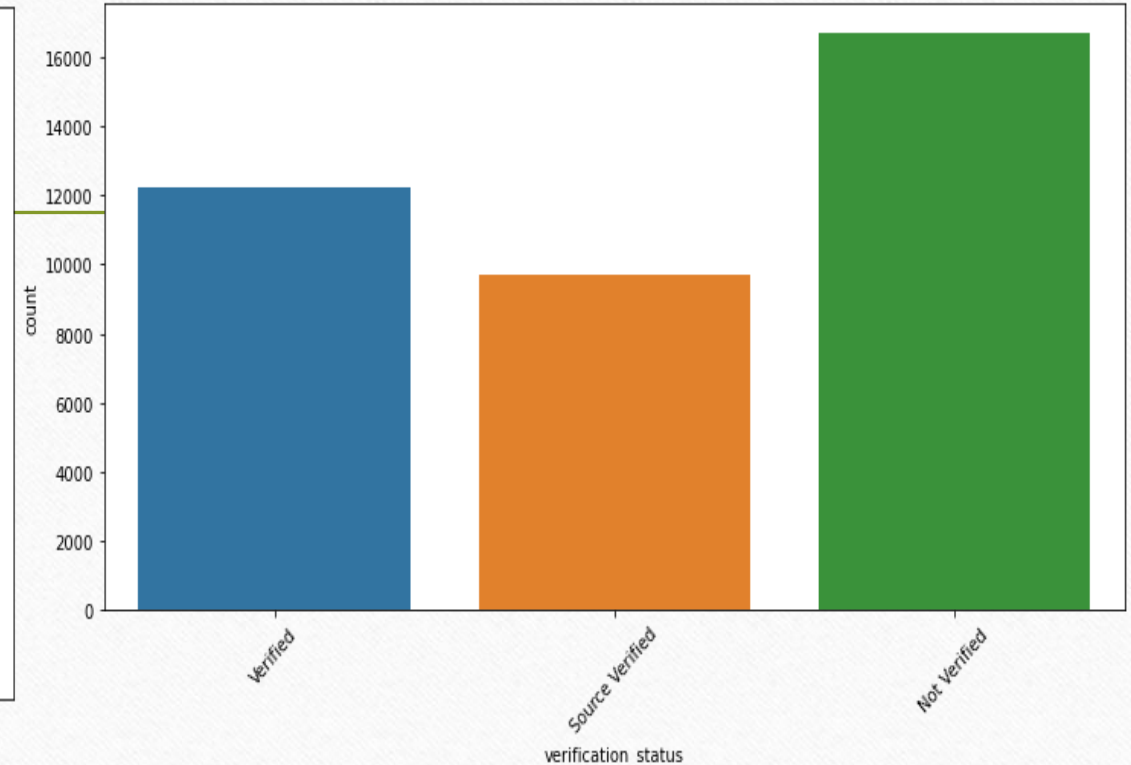
**Verification status**

Indicates if income was verified by LC, not verified, or if the income source was verified

## Univariate analysis : Term



## Univariate analysis : Verification status

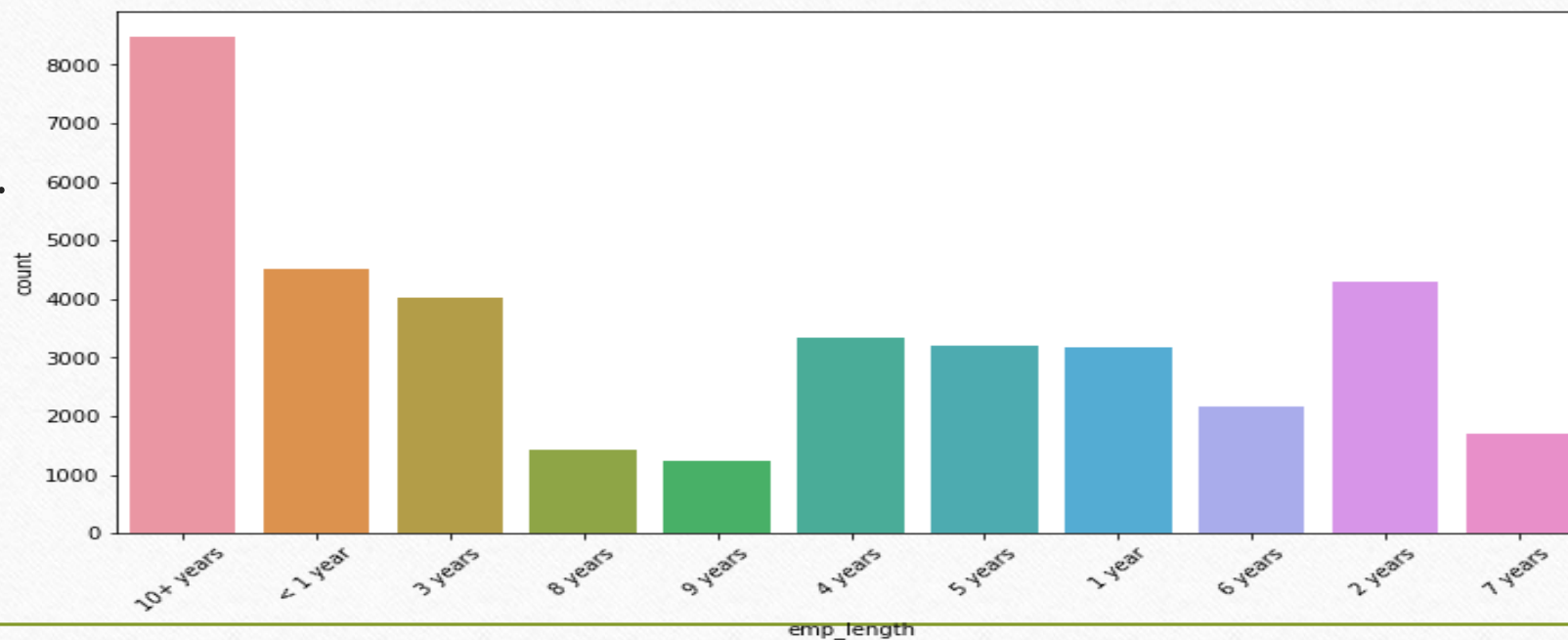


- Term : It is observed that customer is preferring to take loan for term for 36 months ( 3 Years ) compared to 60 Months ( 5 Years )
- Verification status : It is observed that significant number of applications are not verified and still loan has been approved for them.



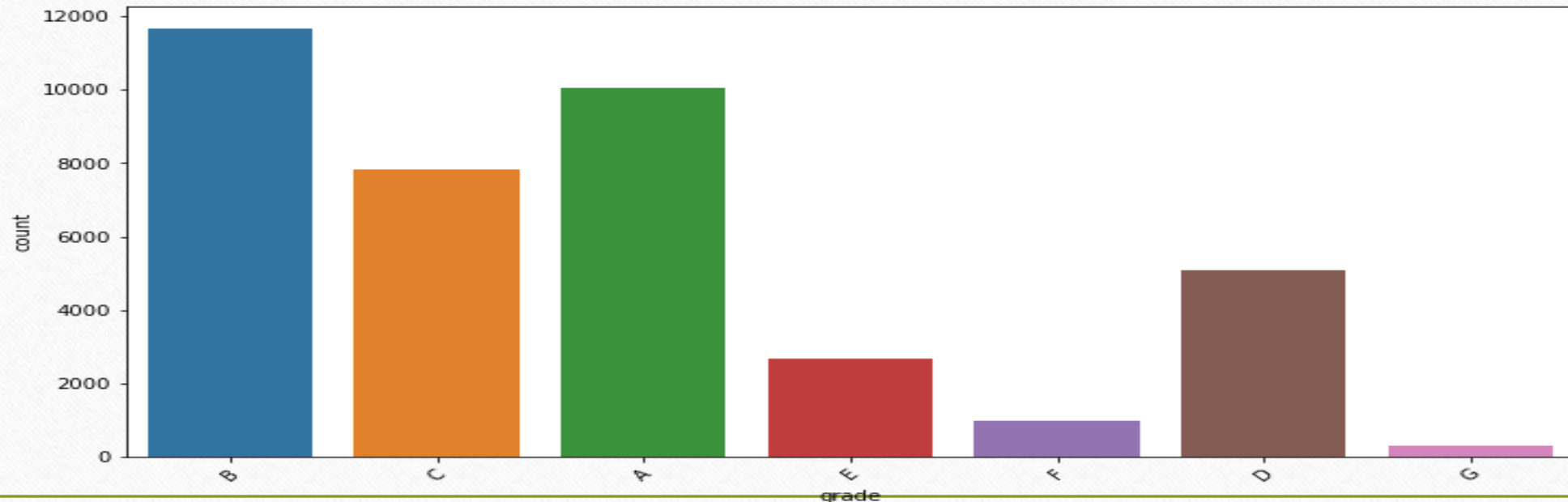
## Univariate Analysis: Employment Length

- It is observed that loan demand is higher among those customers who are having employment length greater than 10 years.



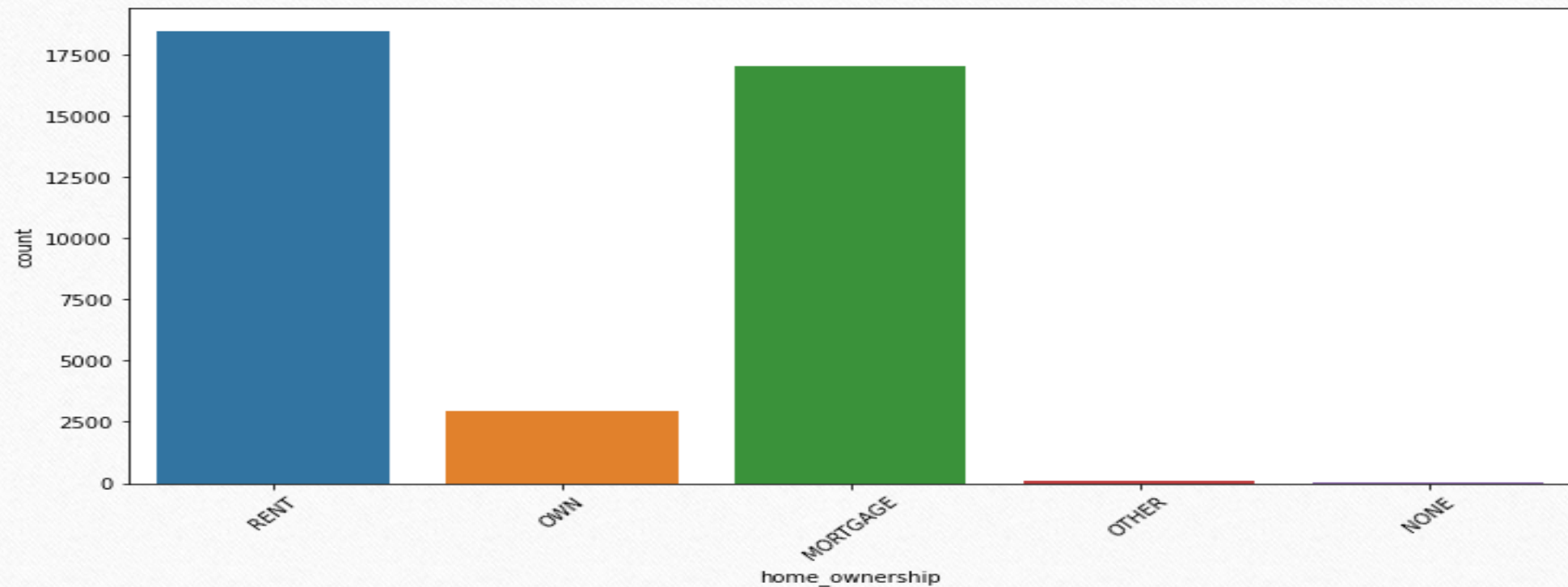
## Univariate Analysis: Grade

- It is observed that grade B type applicants are highest compared to other grade applicants. Grade “G” type applicants is the lowest.
- Grade “A” and Grade “B” customer count is comparable, difference between them is low.
- However, we still have some applicants which fall under lower grade ( E, F and G ) , hence are at higher risk of repayment.



## Univariate Analysis: Home Ownership

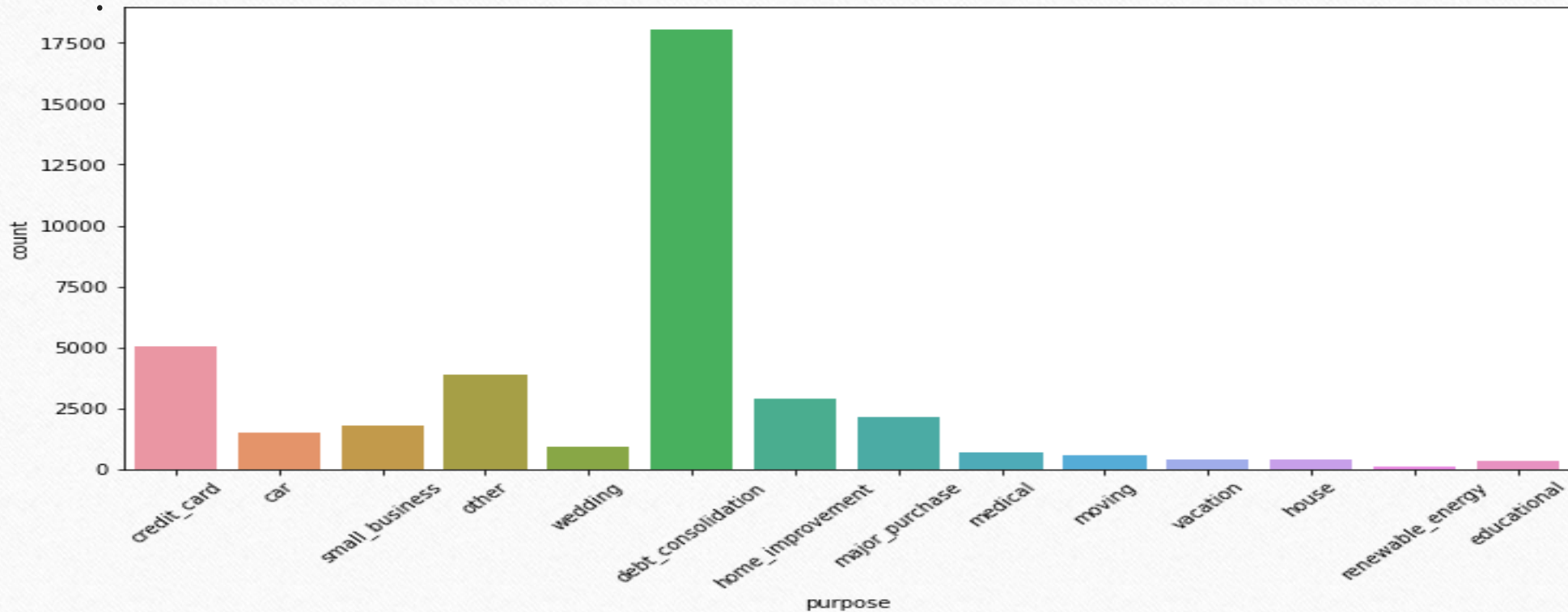
- Customer count for home ownership type RENT & MORTGAGE is higher compared to other types.





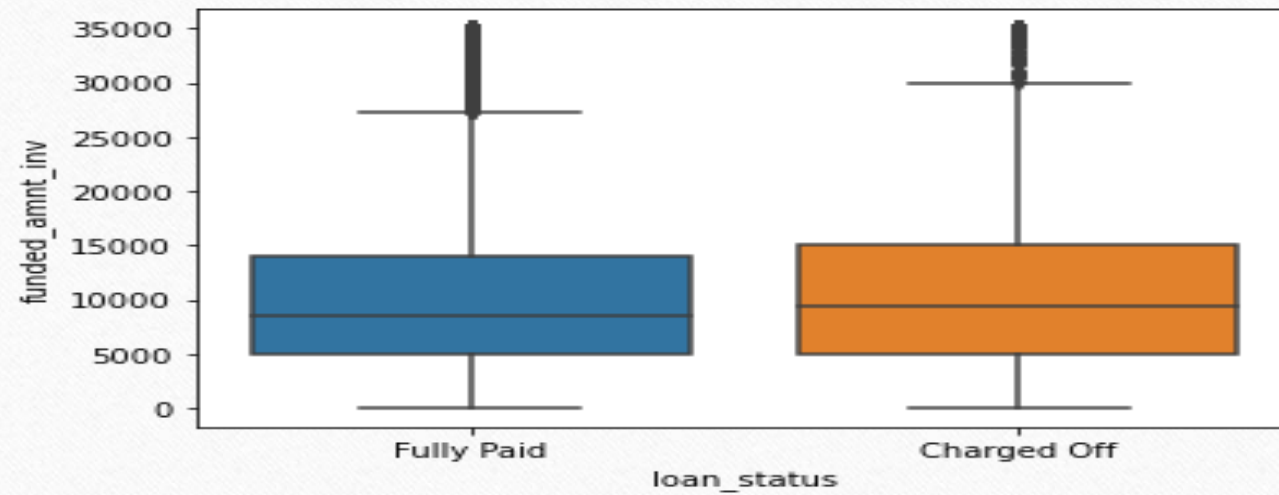
## Univariate Analysis: Purpose

- It is observed that significant rise in bar for the customer taking loan for the purpose of debt\_consolidation



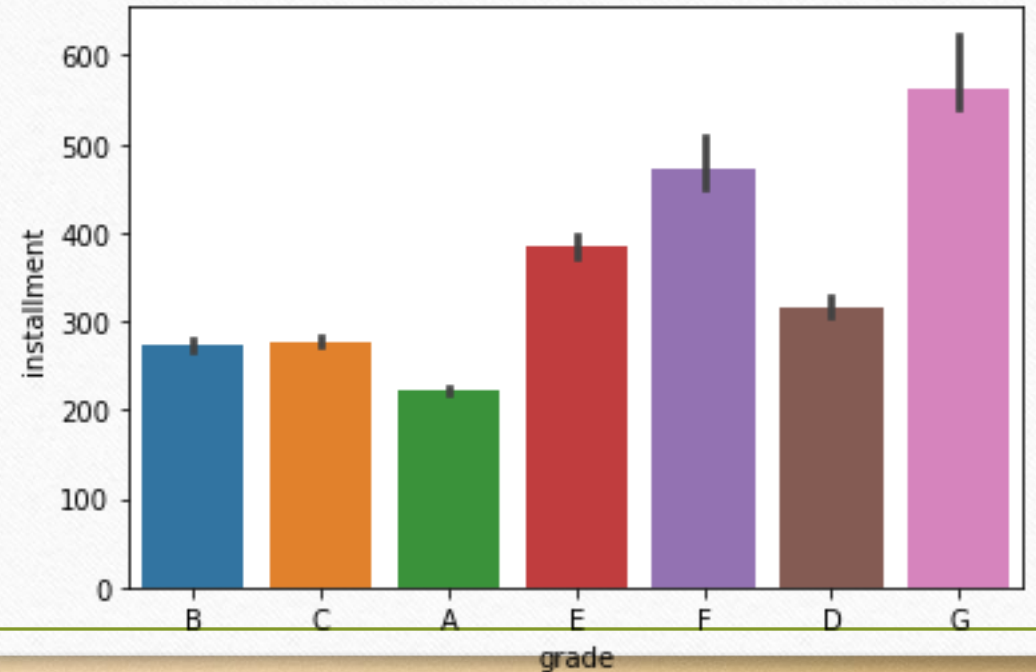
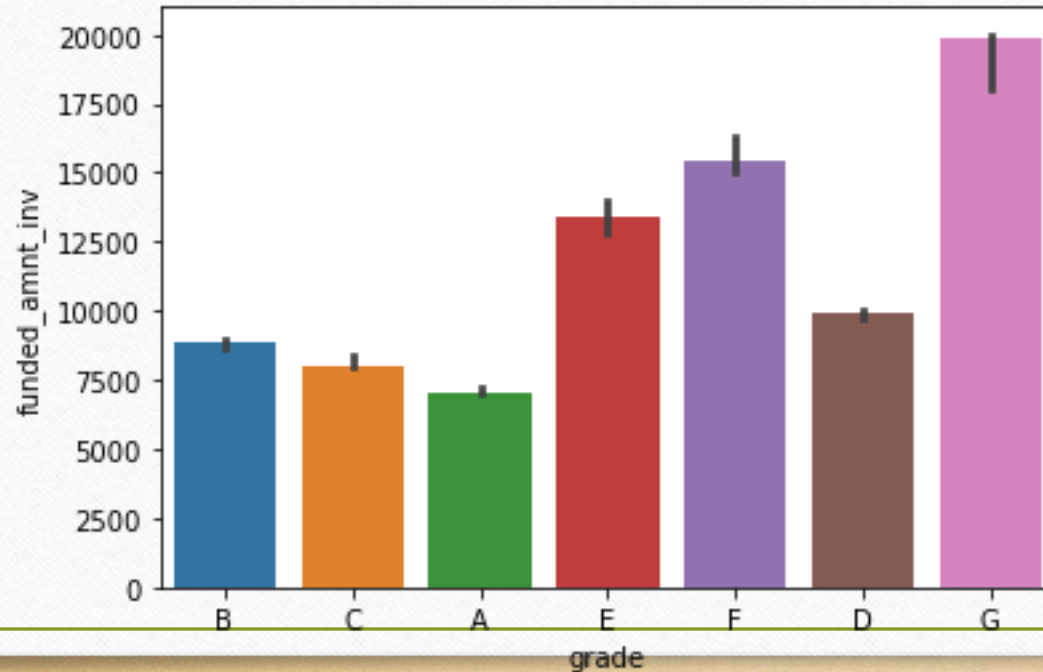
## Bi-Variate Analysis: Loan Status vs Funded Amount

- Interquartile range for charged off is higher than fully paid.
- It suggests that more if funded amount is high then there is more chance of Charged-off



## Bi-Variate Analysis: Grade vs Funded amount

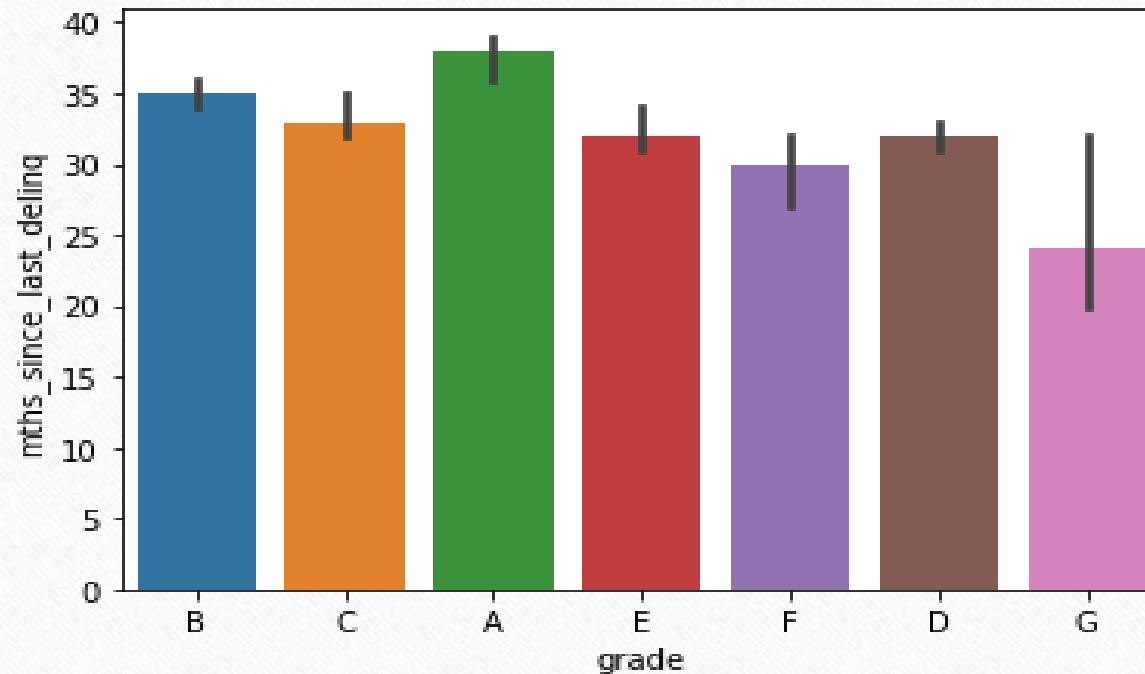
- It is observed that grades ( E, F and G ) who are at higher risks of repayment have higher funded amounts/installments.





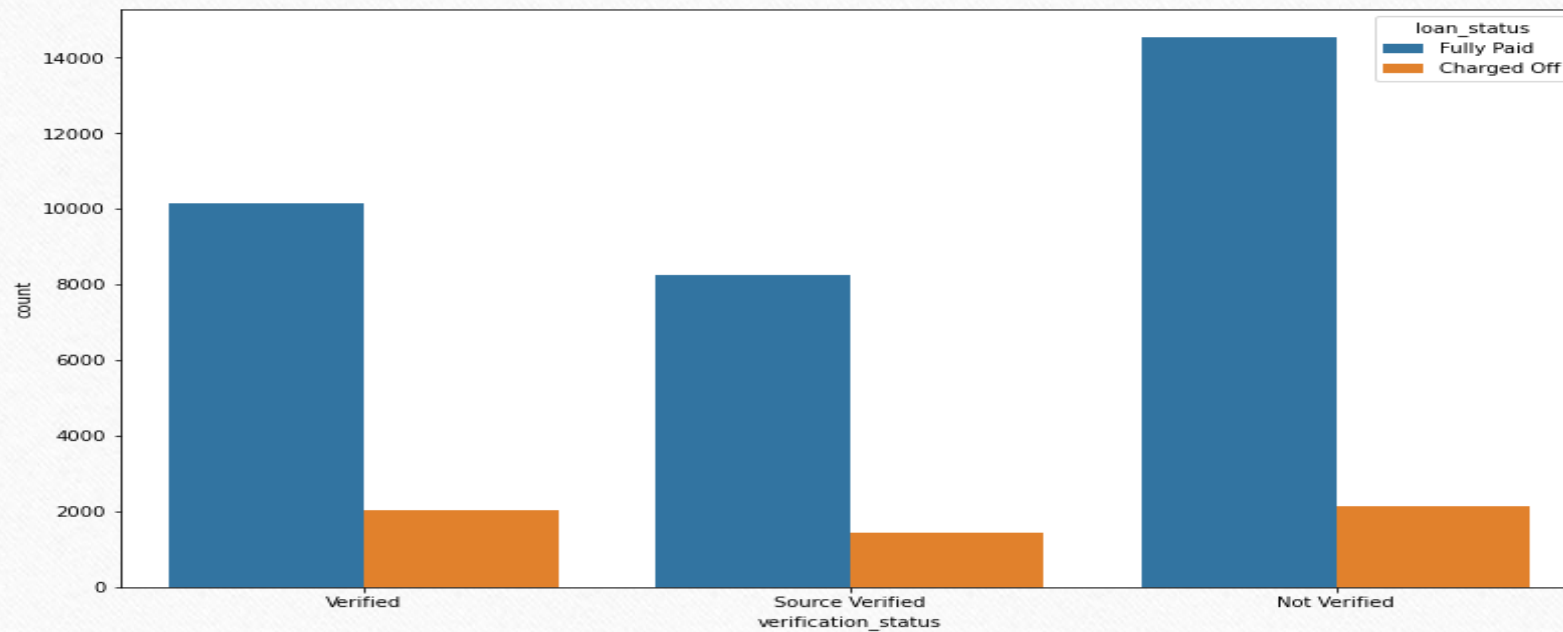
## Bi-Variate Analysis: Grade vs Months since last delinq

- The number of months since the borrower's last delinquency is lower for grades ( E, F and G ) hence are more delinquent.



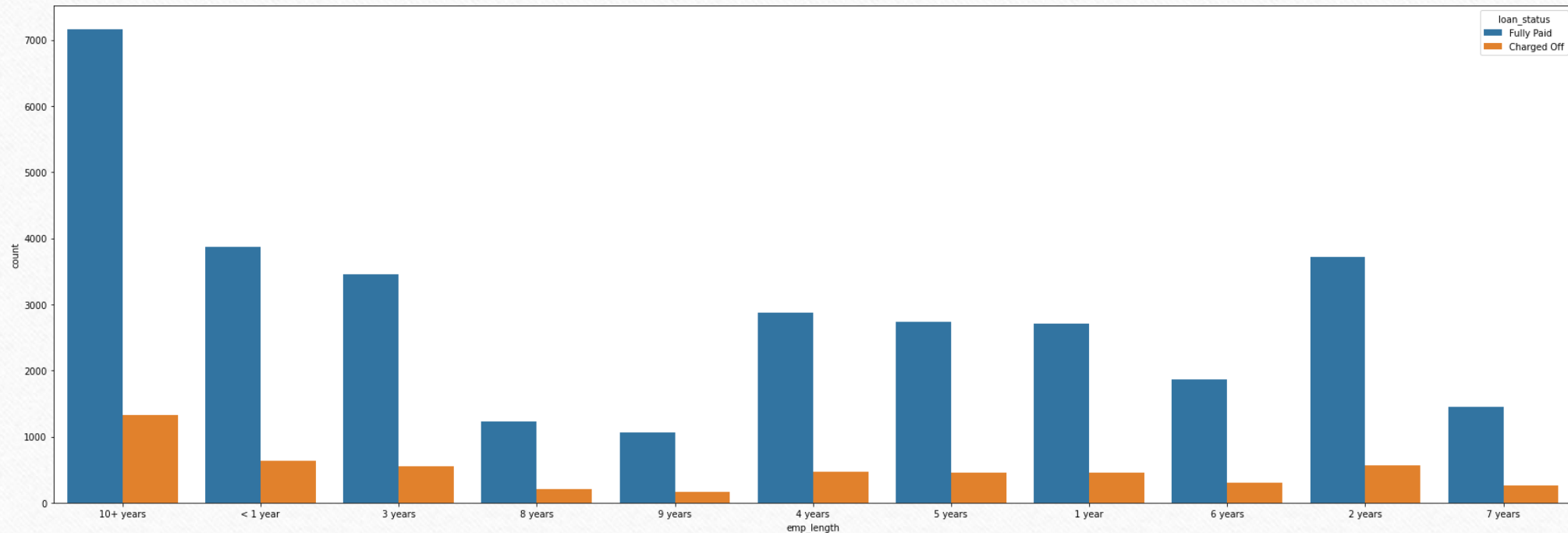
## Bi-Variate Analysis: Loan Status vs Verification Status

- It is observed that charged off count is comparable for verified and source verified.
- It is clear indication for the improvement of the verification process of the organization as charged off count is the highest for Not Verified



## Bi-Variate Analysis: Loan Status vs Employment Length

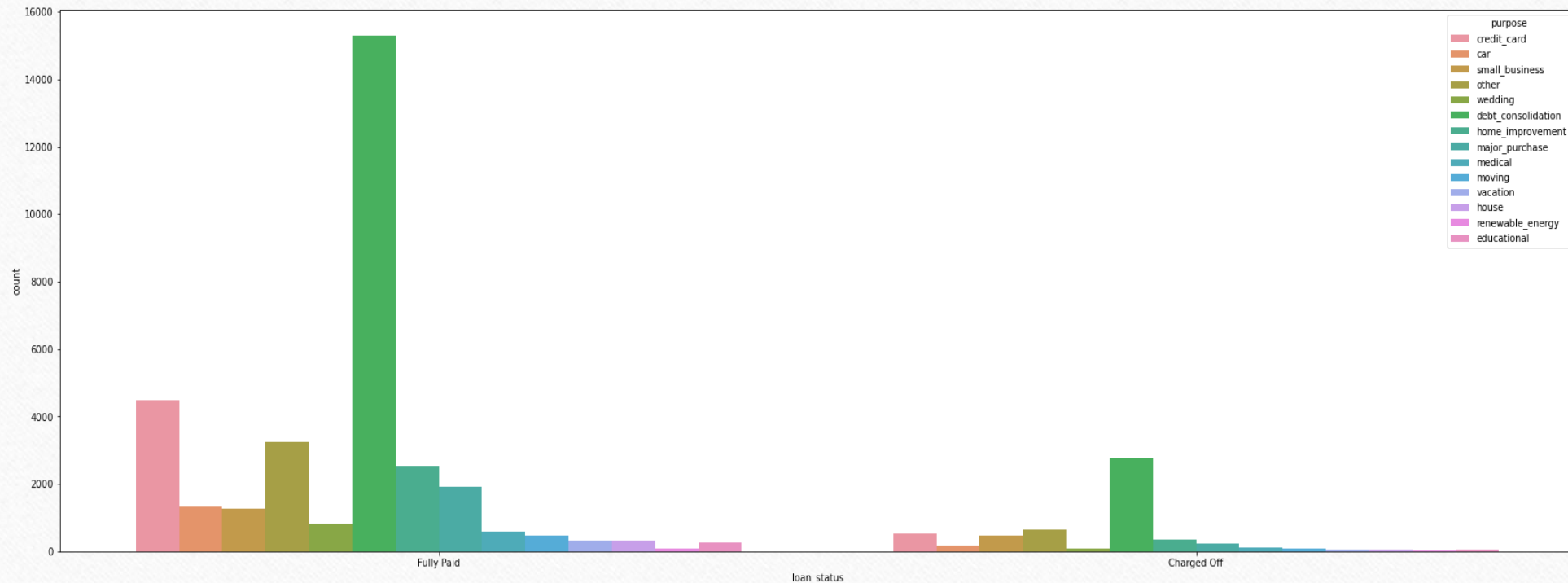
- It is observed that charged off count is higher for those who have employment length greater than 10+.
- Though our dataset ( univariate analysis for employment length ) indicates that applicants with 10+ years experience are more in number compared to any employment length.
- This indicates a strong need to ensure relevant checks are in place to check eligibility for 10+ year experienced applicants.





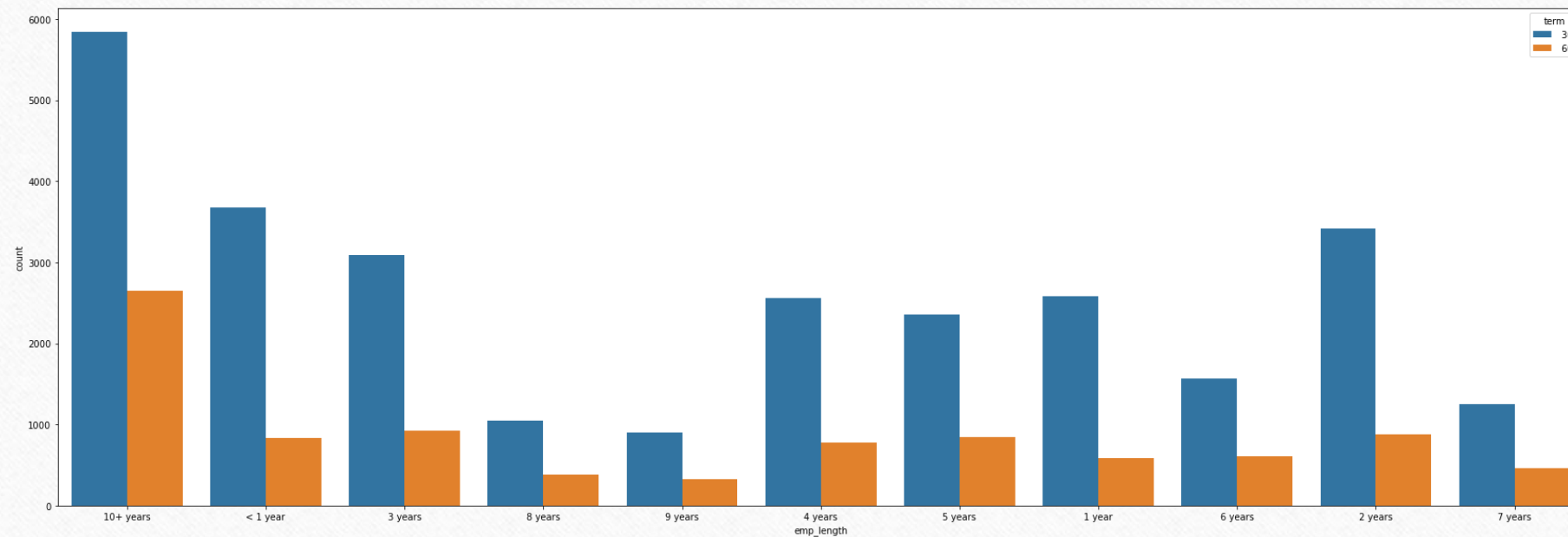
## Bi-Variate Analysis: Loan Status vs Purpose

- Maximum applicants have charged off / defaulted as the loan applied is for debt consolidation.
- Debt consolidation is a form of debt refinancing that entails taking out one loan to pay off many others.



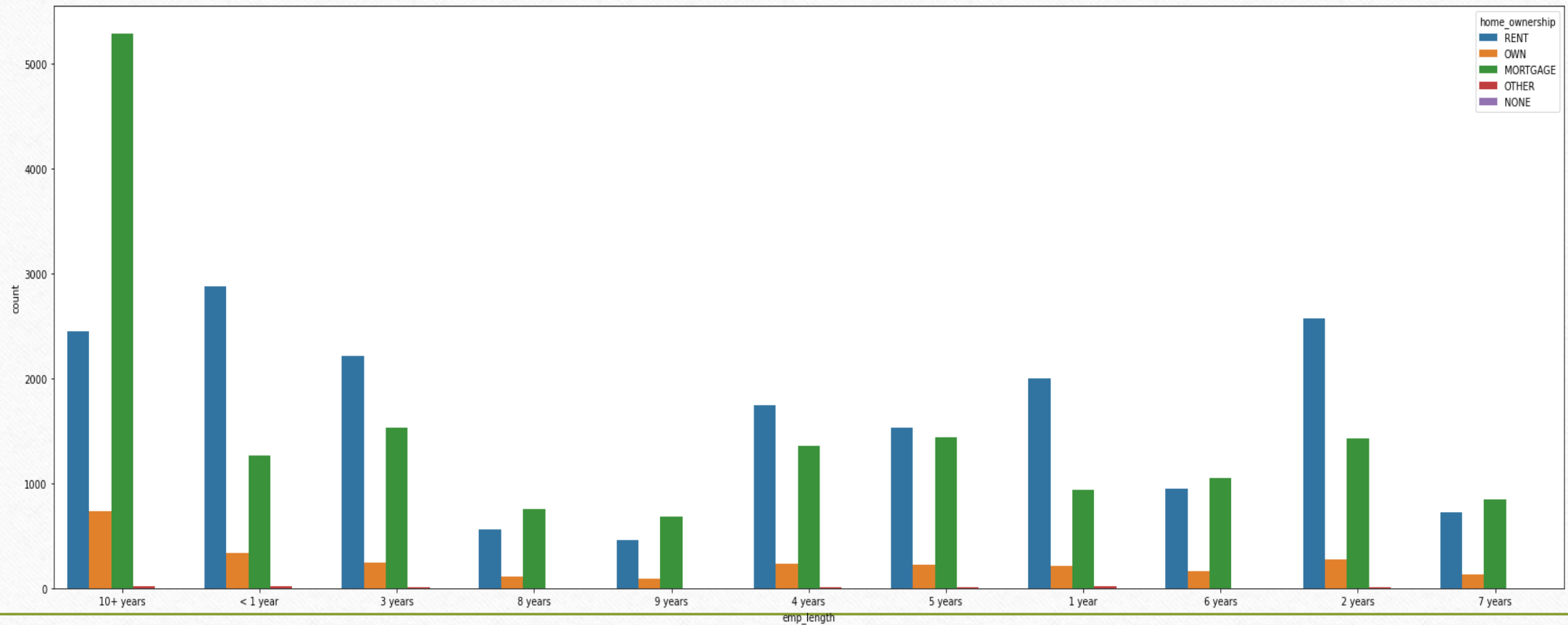
## Bi-Variate Analysis: Employment length vs Term

Funded amount is more for higher experienced applicants and the term to repay is 36 months which is lesser compared to other terms. Hence chances for them to default is higher.



## Bi-Variate Analysis: Employment length vs home ownership

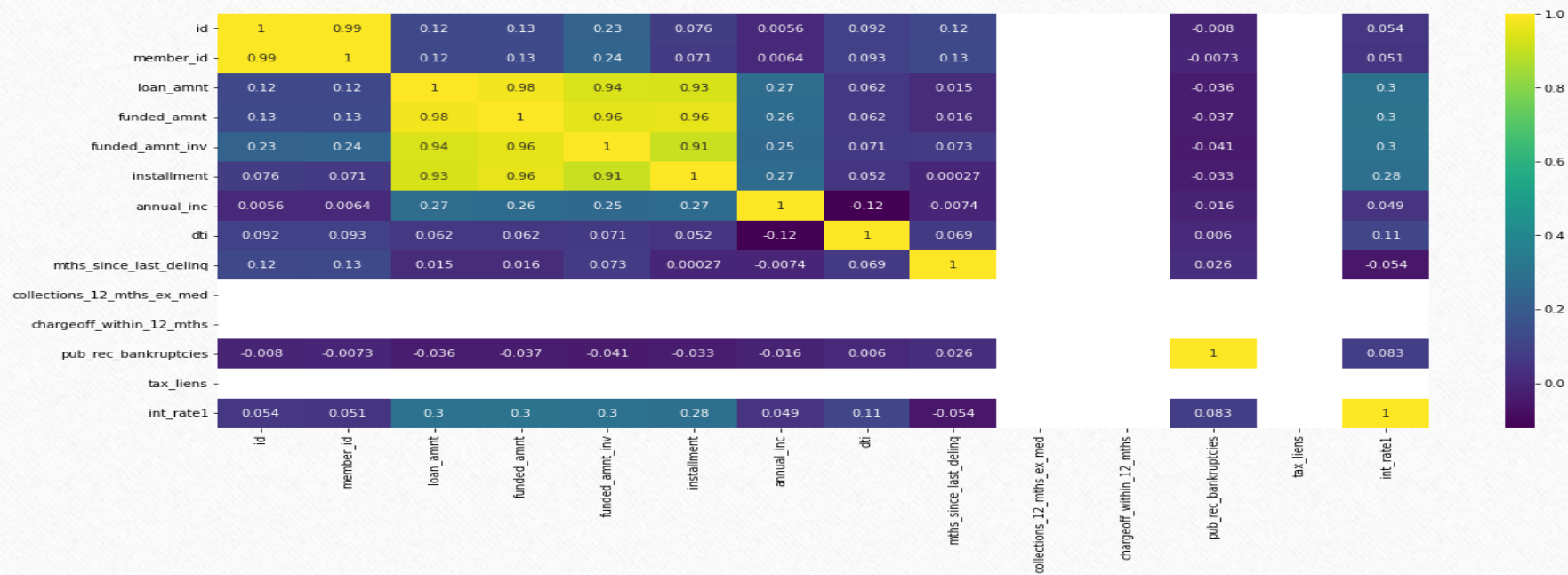
Mostly applicants in the experience of 10+ years have a mortgage, hence pushing them towards defaulter category as they already have existing debts.





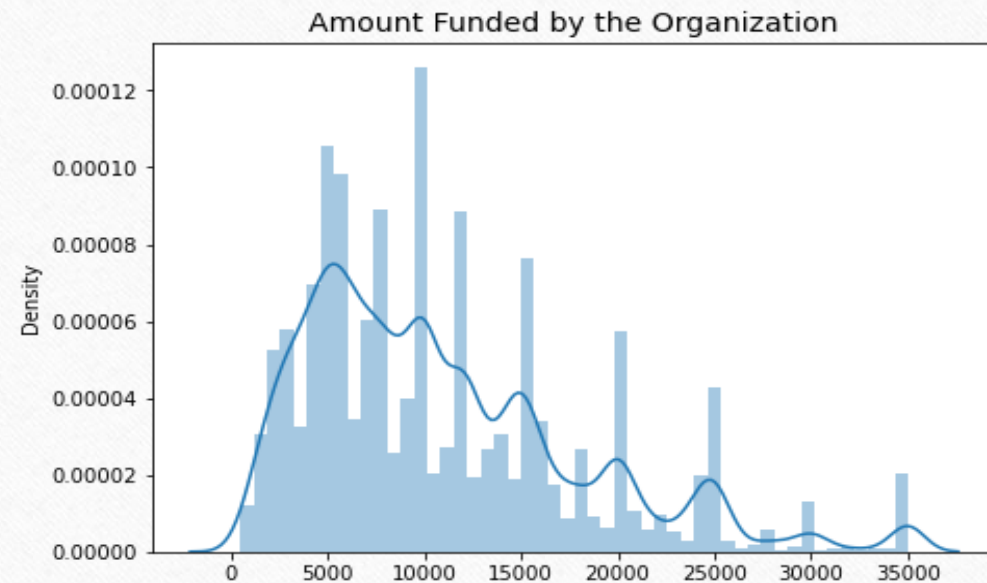
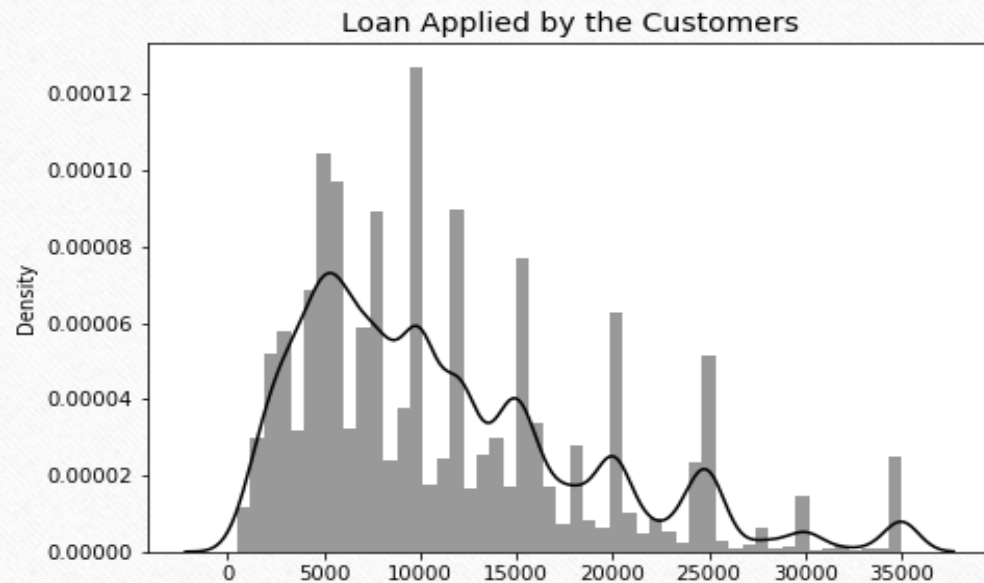
## Bi-Variate Analysis: Correlation

- Correlation factor for loan amount requested and funded amount is high which reflects that most of the requested loans are getting accepted.
- Correlation factor for loan amount and annual reported income of the customer is low . It represents customer is most likely to default and recommend for improvement of verification process of the organization.



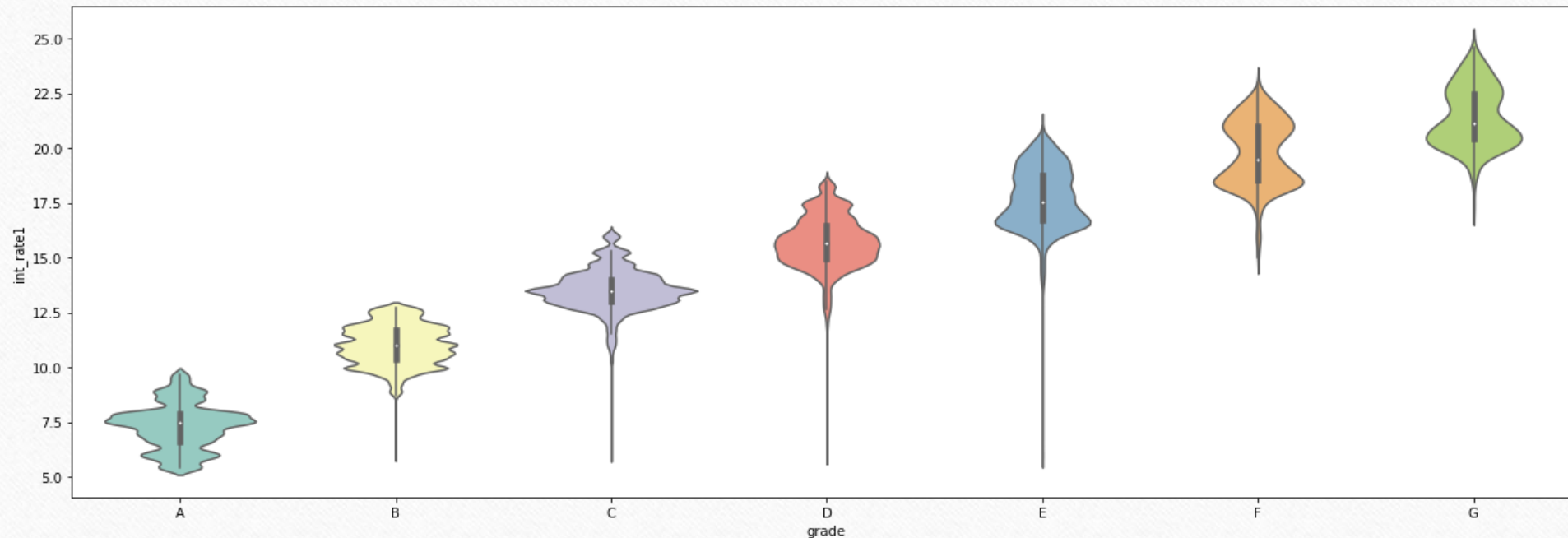
## Bi-Variate Analysis: Loan Status vs Grade

- It is observed graph follow the same pattern for the Loan applied by customer to the Amount funded by the organization.
- It shows that almost all the loans are accepted by the organization and it recommend to improve verification process for approving loan.



# Bi-Variate Analysis: Interest Rate vs Grade

- It is observed that loans for the customers Grade type “G” who is more prone to default in future are getting approved with high rate of interest.
- Only considering high rate interest for approving loan for these customers would incurred huge loss for the organization.





# Recommendations

- **Verification status** : Charged off count is comparable across category Verified and Source verified (Refer slide 13). Which suggest that improvement is required for verification process as defaulters are reported despite verification.
- **Months since last delinq** : The number of months since the borrower's last delinquency (Refer slide 12) is nearest for grades ( E, F and G ). Hence are more delinquent. Hence application belonging to such categories should be assessed carefully before approving as they are at higher risk of becoming defaulters. Greater Funded amounts should not be approved for them.
- **Purpose** : Purpose of the loan should be assessed carefully as mostly applicants fall into the category of debt consolidation ( Refer slide 15 )
- **Employee length** : Though our dataset ( univariate analysis for employment length ) indicates that applicants with 10+ years experience are more in number compared to any employment length. Due to large data set around 10+ years , we observe that count of defaulters is high in this category. **But why at all are they defaulting?** Higher experienced applicants are opting for more funded amount and the term to repay is 36 months which is lesser compared to other terms. Purpose for opting for loan is Debt consolidation Hence chances for them to default is higher ( Refer slide 16,17 ). Chances are that more experienced applicants are greater liabilities.
- **Term** : As mentioned above, large funded amounts and shorter terms ( with existing debts ) might result in an applicant defaulting. ( Refer slide 16,17 )
- **Interest Rate**: As our analysis shows that for the customer type Garde “G” we are approving loan with high rate of interest but these customers are more prone to default. So approving loan for higher rate of interest will not help.

Thank you

---