

Lending Club Case Study

Data analysis and Conclusions

Contribution:
Ashish Kumar Sharma
Swamy Satish Patil

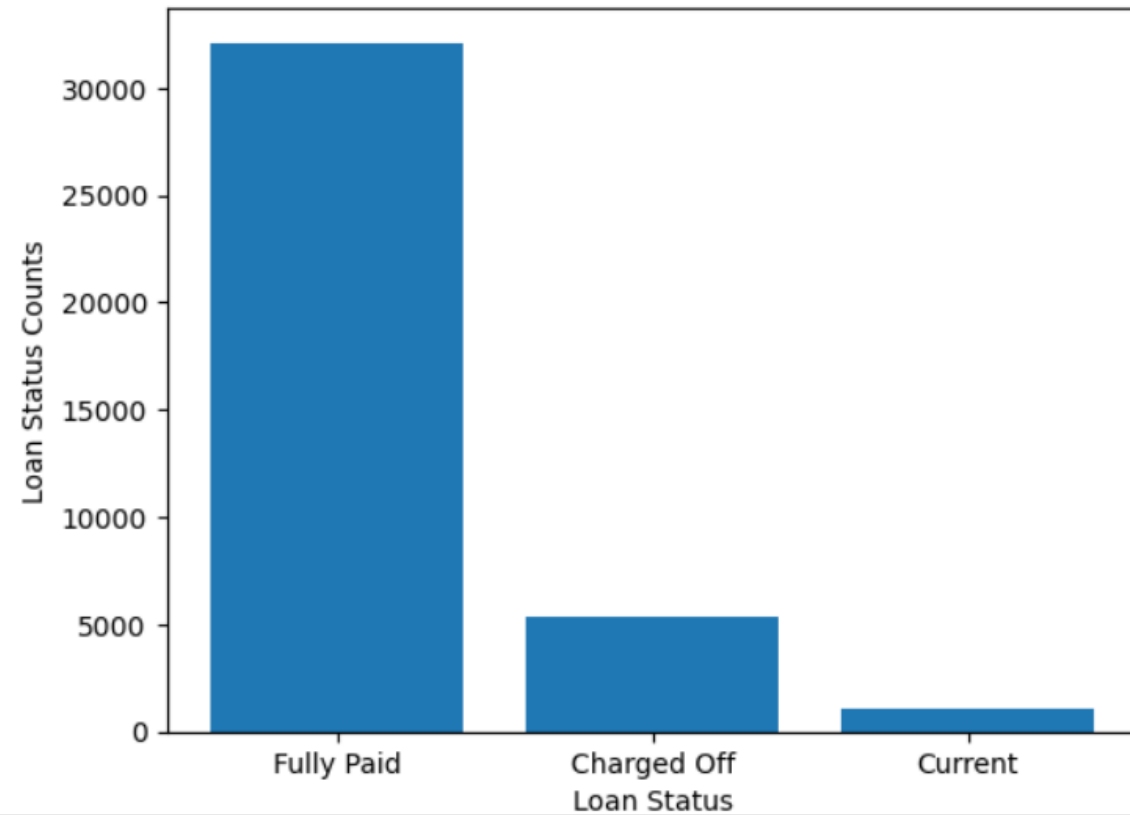
Objective:

1. Analyze the provided dataset to identify variables that are indicators of a loan default
2. Identify any patterns that may exist in the dataset between the loan_status variable and any of the other variables

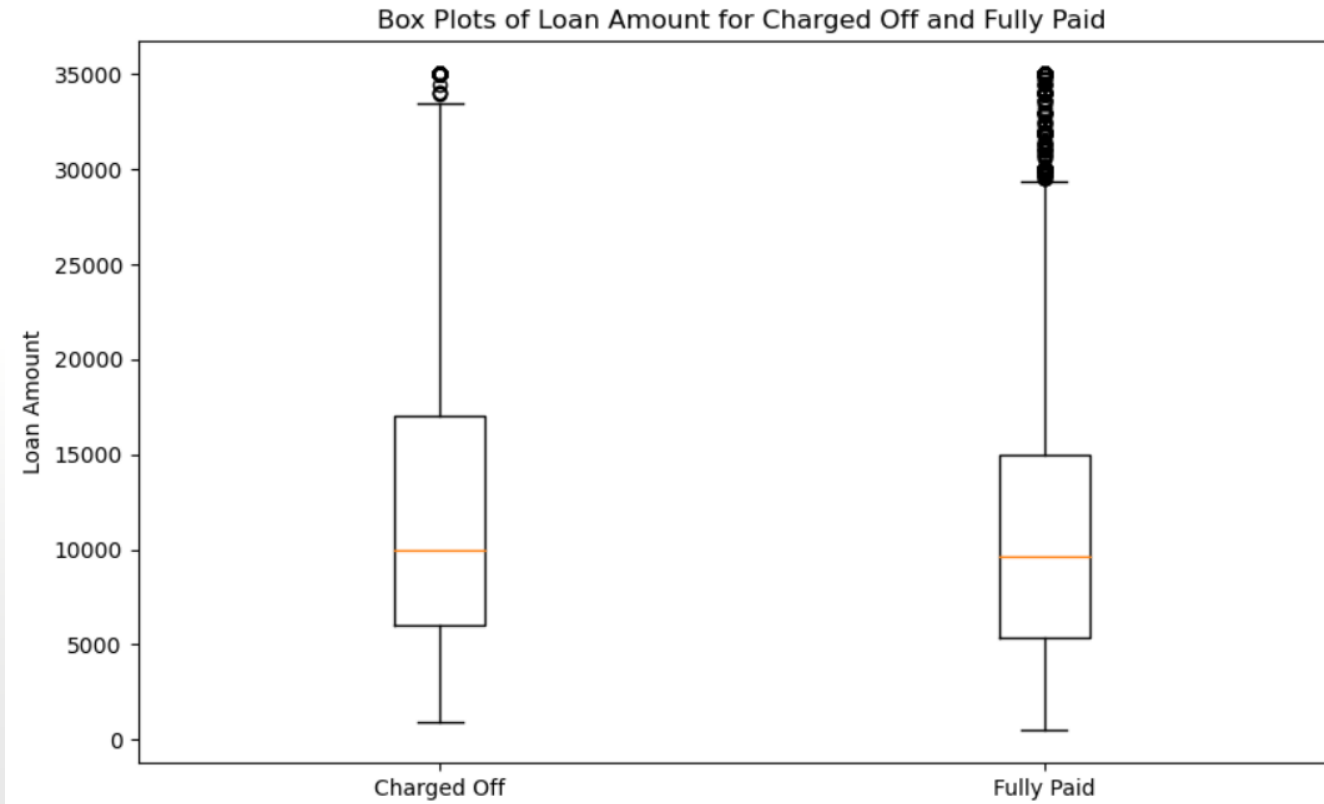
Data Understanding and Data Cleaning

1. Remove Columns which have more than 90% Null Values
2. Remove Unwanted Columns
3. Fill Null Values in Columns with appropriate values if applicable using data understanding
4. Drop some rows having Null values if can't fill with appropriate value
5. Convert Data Type to Numeric wherever applicable
6. Examine Outliers for numeric columns using Box Plot

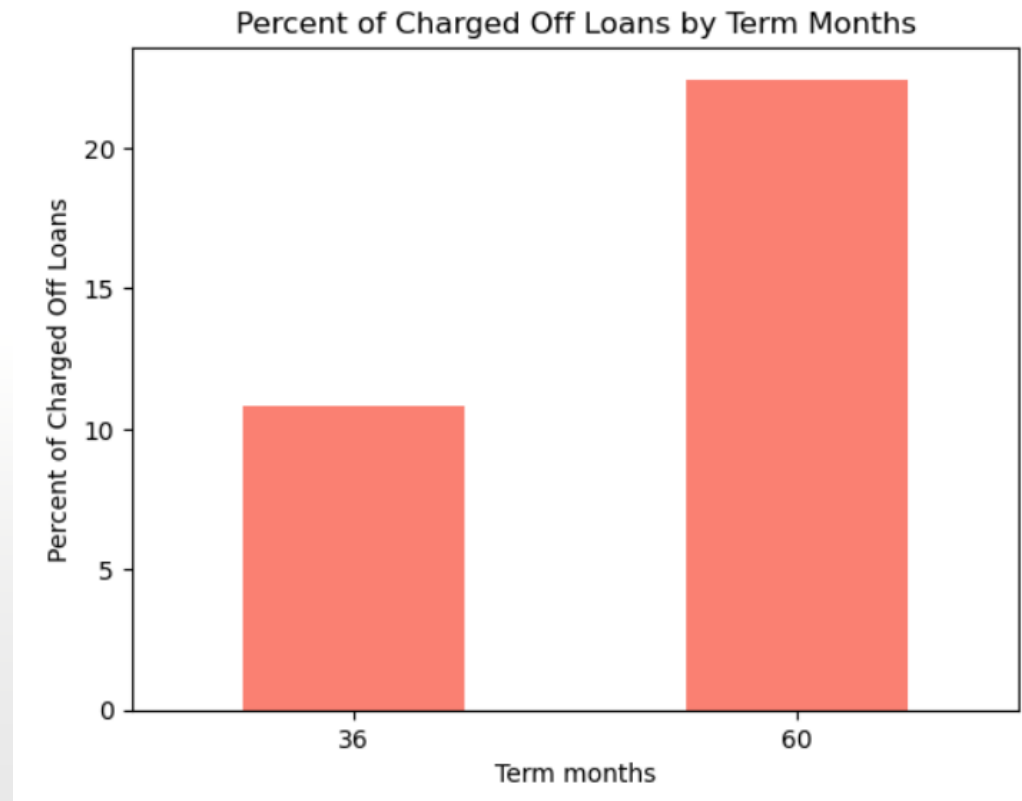
Finding I:
Proportion of
charged off loan
is 14 %



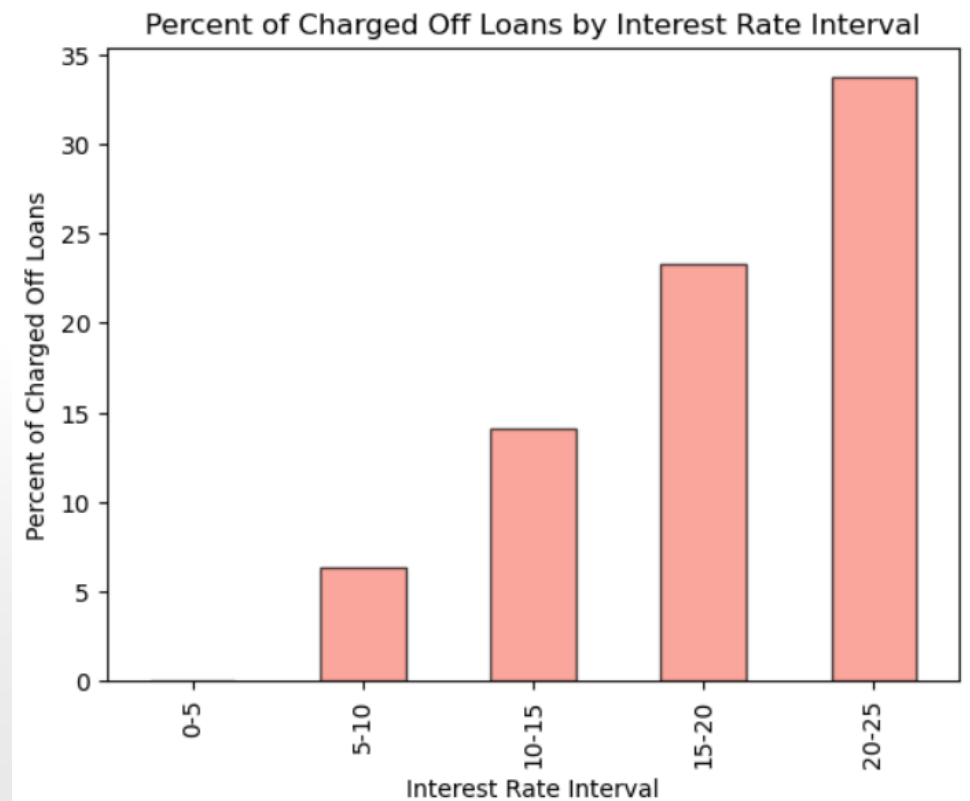
Finding 2: There is no relation with respect to loan amount for loan defaults



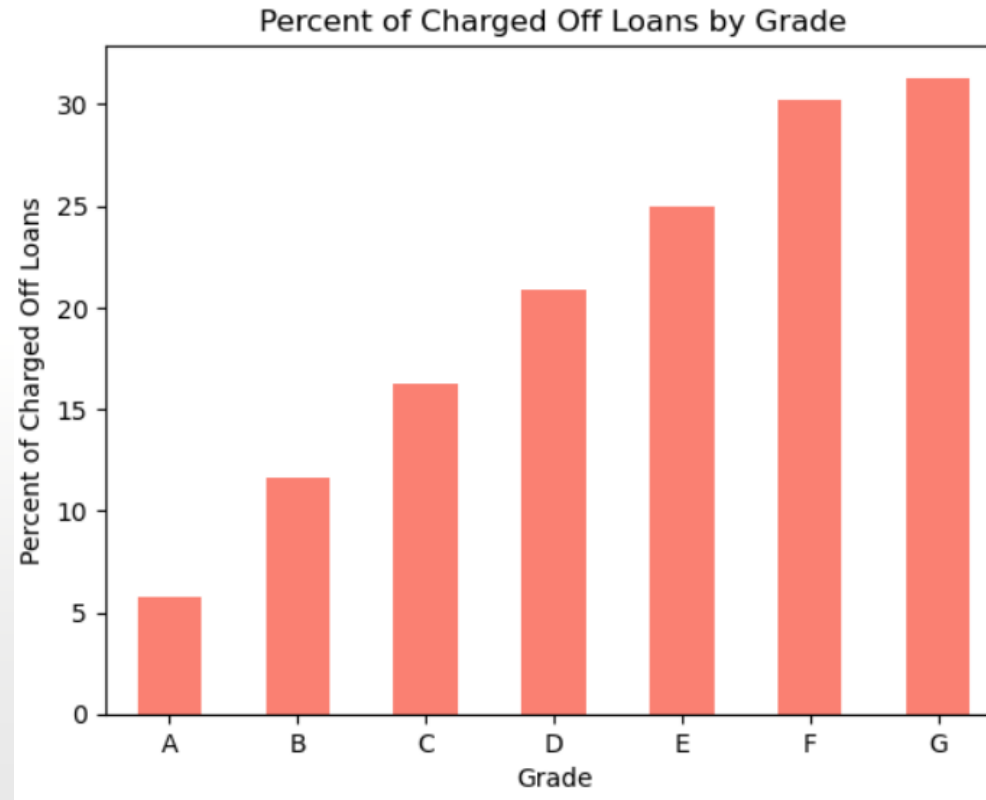
Finding 3: Likelihood of Loan Status as Charged Off is higher for Loan Term 60 months.



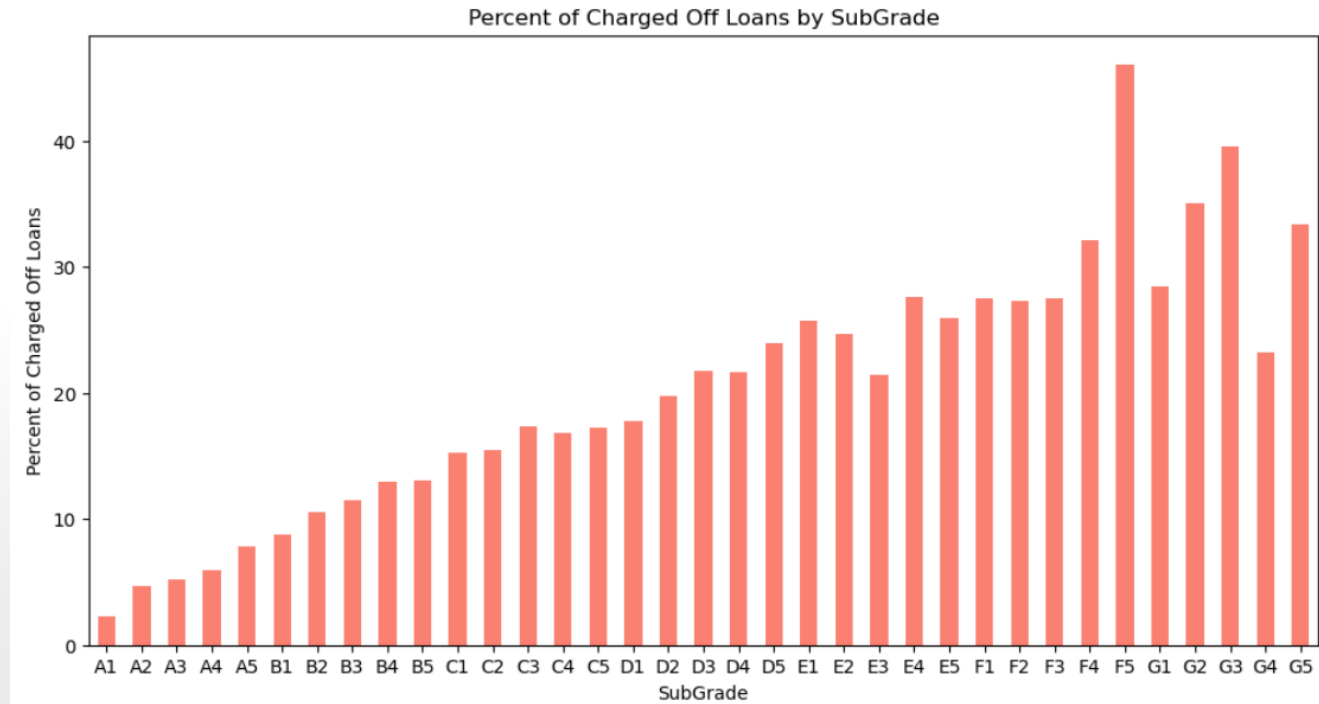
Finding 4: Interest Rates is a strong indicator of loan defaults . Higher the interest Rate , higher is the chance of loan default



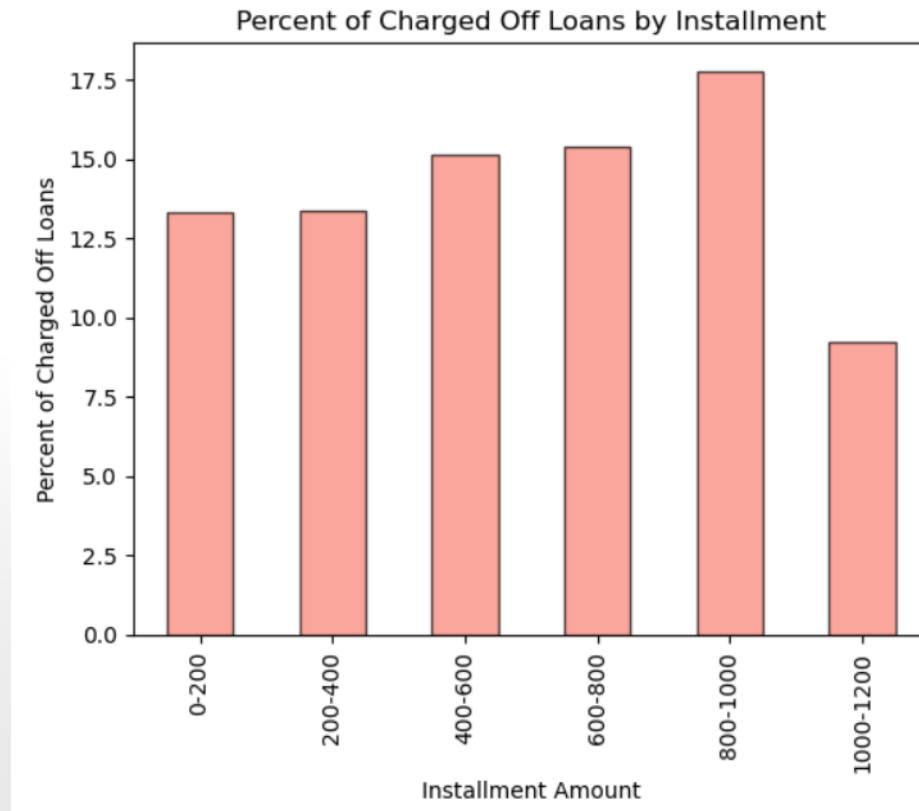
Finding 5: Grade is a strong indicator of loan default. Lower Grades have higher chances of loan default



Finding 6: SubGrade is a strong indicator of loan default. Lower SubGrades have higher chances of loan default

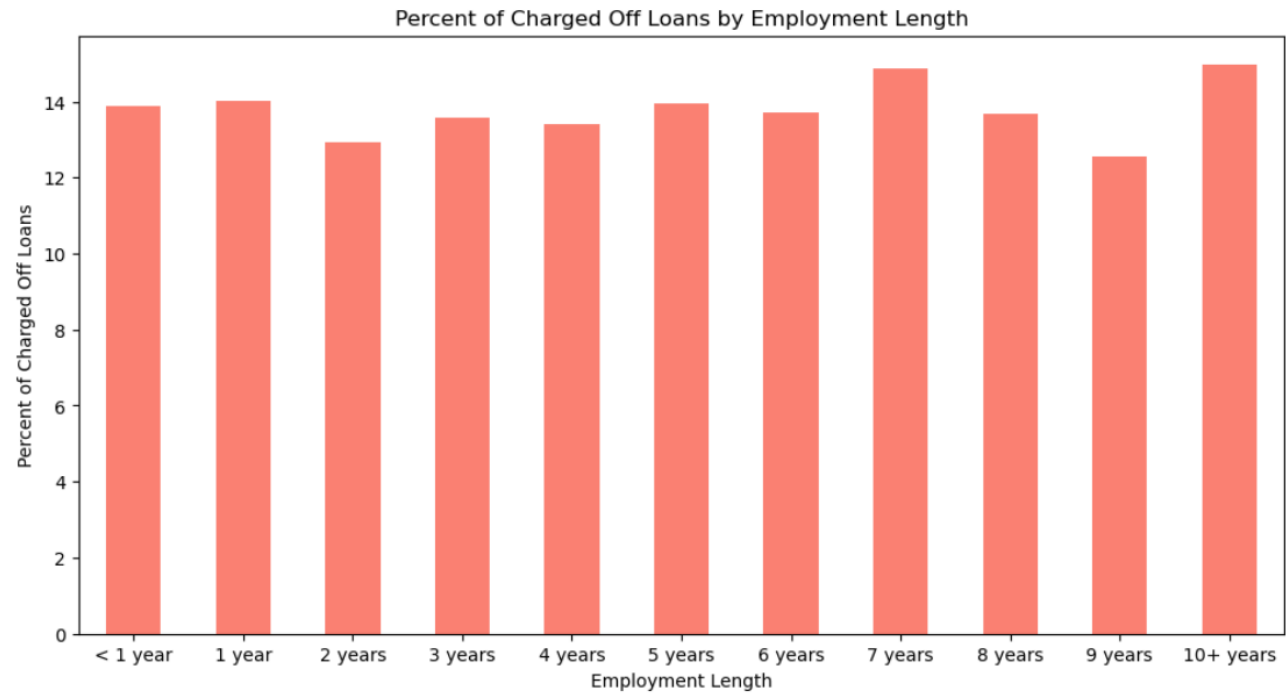


Finding 7: Likelihood of loan default increases with higher installment which can be observed when installment amount increases from 0 - 1000 but is lowered when installment amount goes above 1000

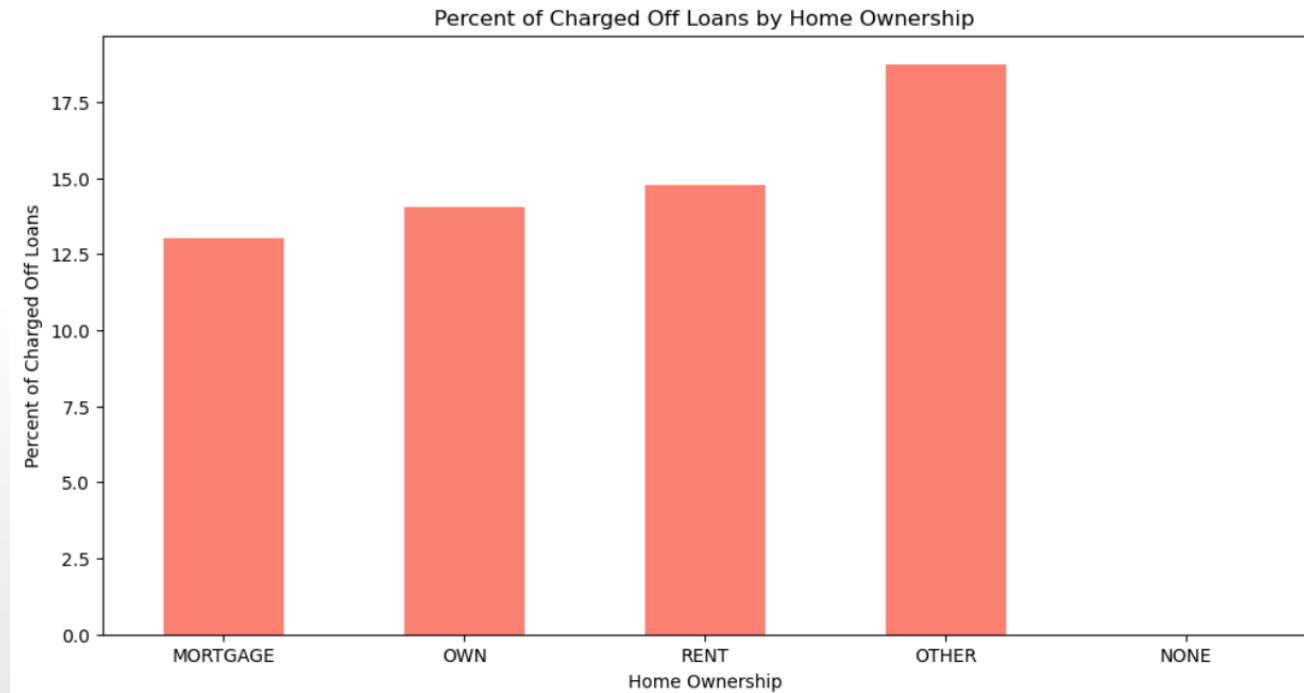


Finding 8: There is no definite trend for charged Off loans as per employment length.

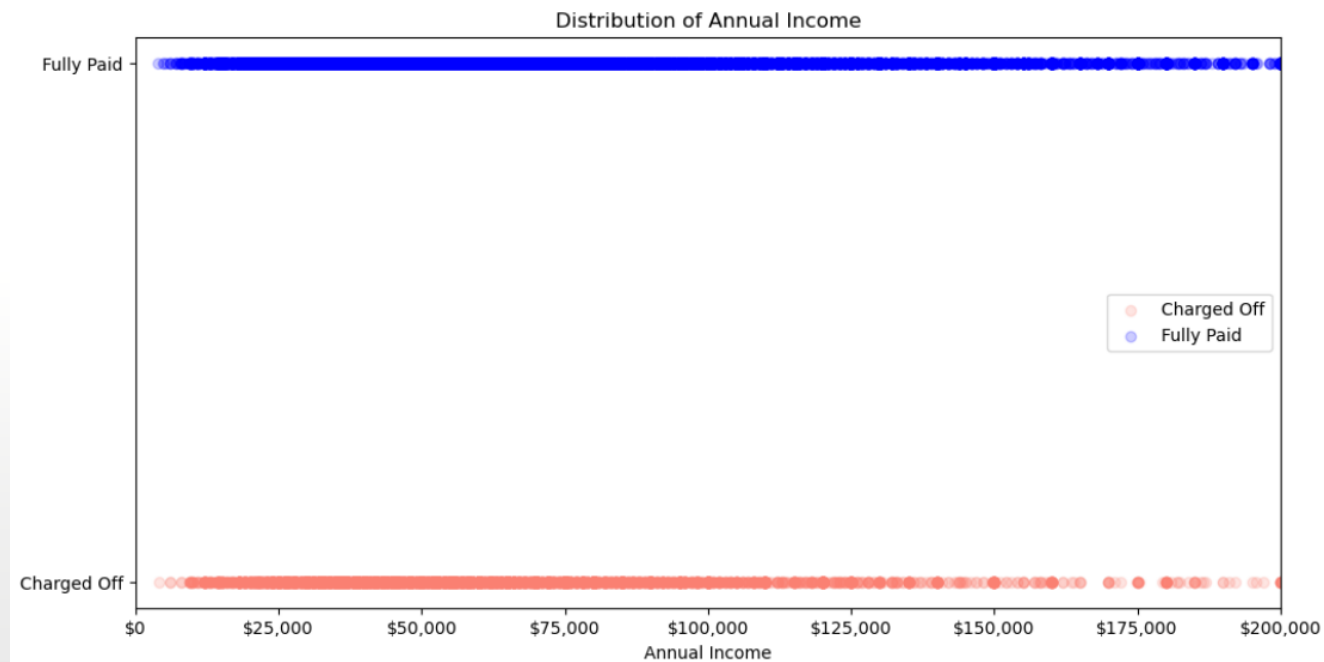
Employees having 7 and 10+ years of experience have slightly high rate of default.



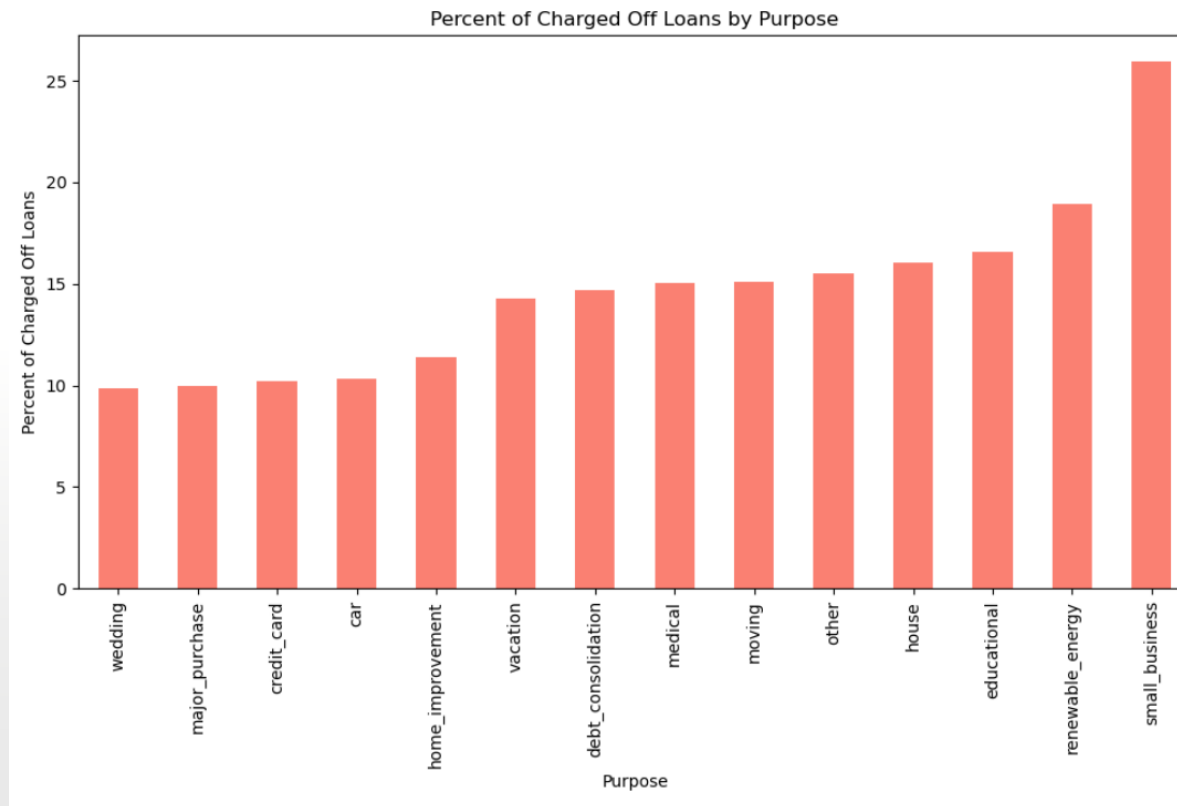
Finding 9: There is no definite trend for charged Off loans as per home ownership . Those having home ownership as other have slightly higher chances of loan default



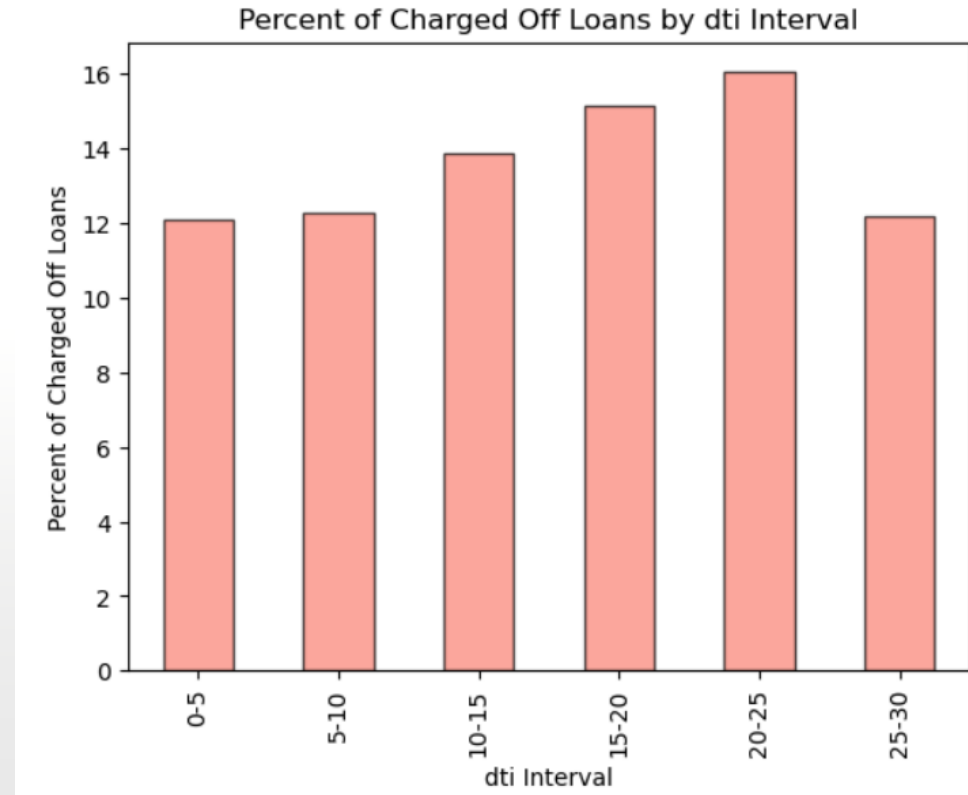
Finding 10: Charged Off Loans are more skewed towards income group below 50000.



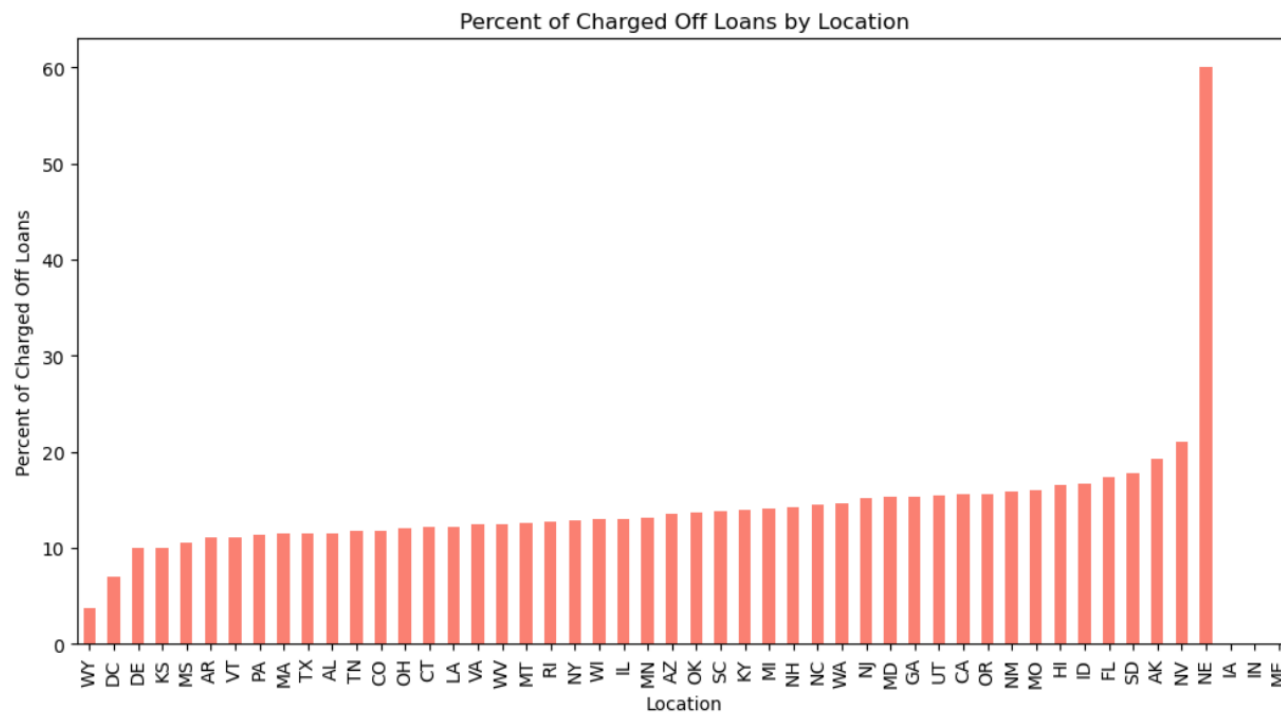
Finding 11: Purpose is a strong indicator of loan default. Some of the purposes have more bad loans



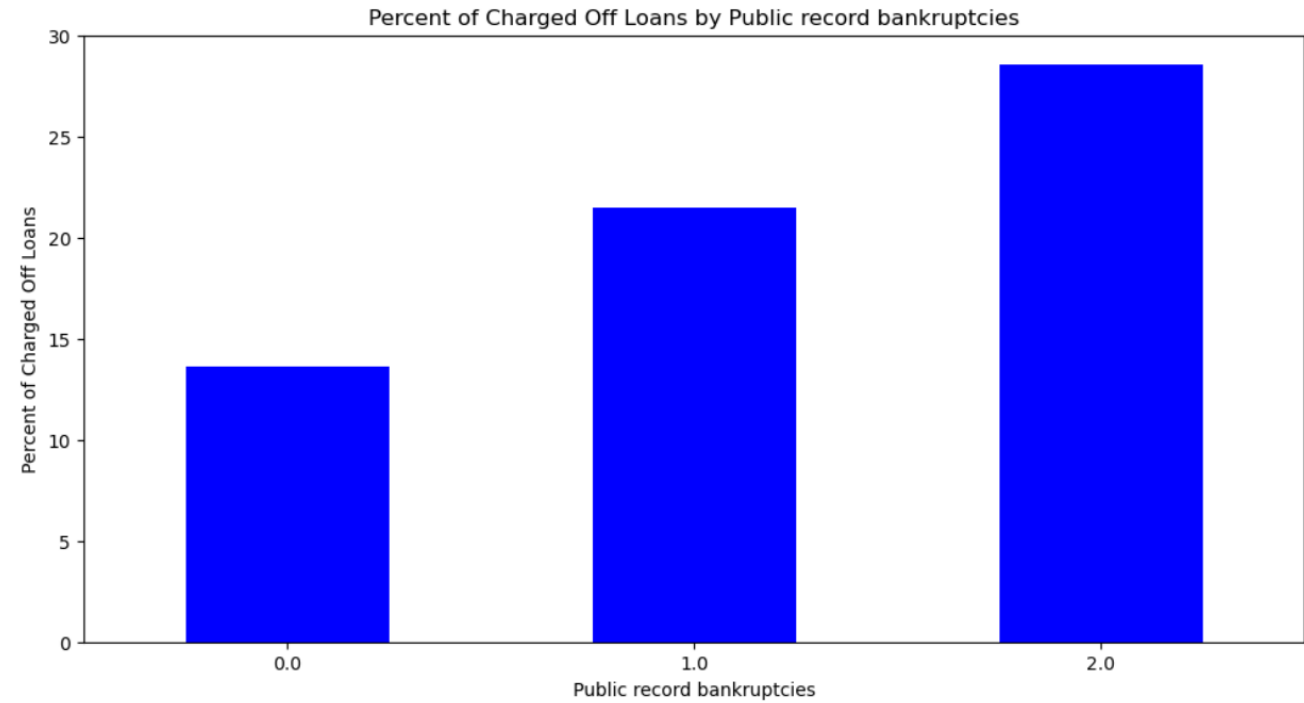
Finding 12:
Likelihood of loan
default increases
with higher dti rate
which can be
observed when dti
increases from 5 - 25
but is lowered when
dti goes above 25



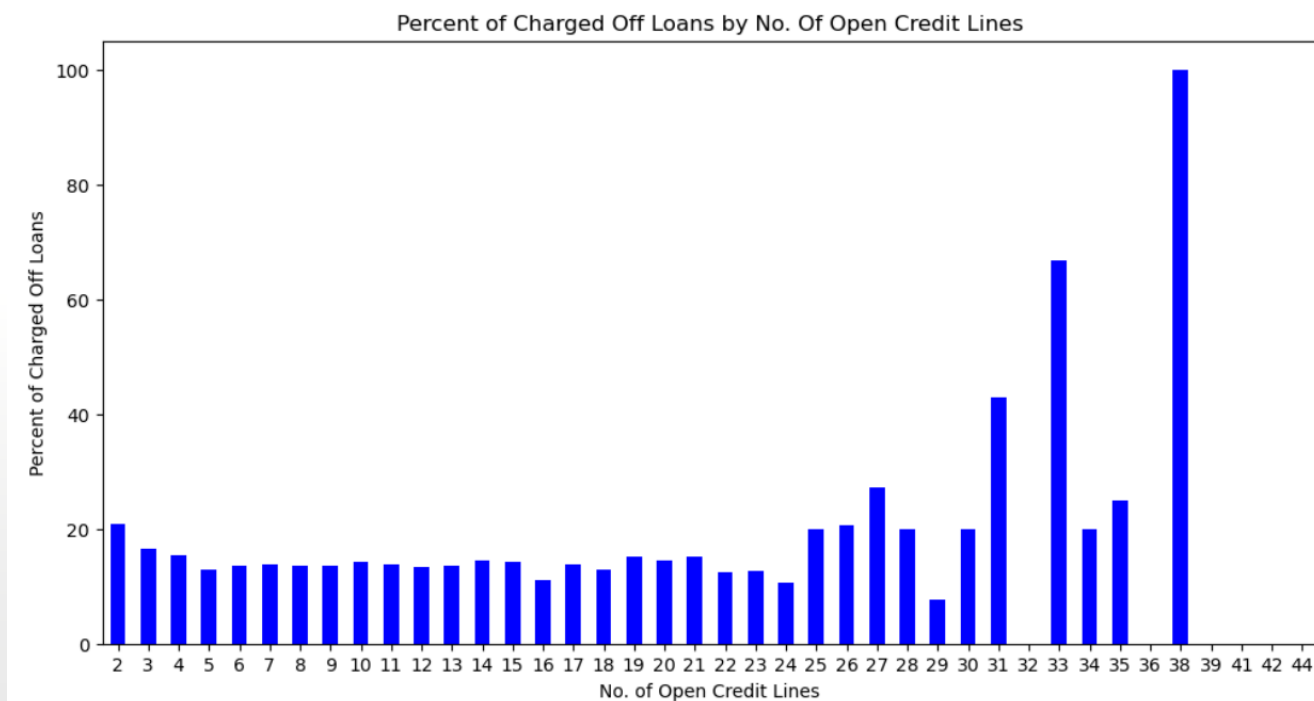
Finding 13: Some States have unexpectedly high rate of default loans such as NE(Nebraska)



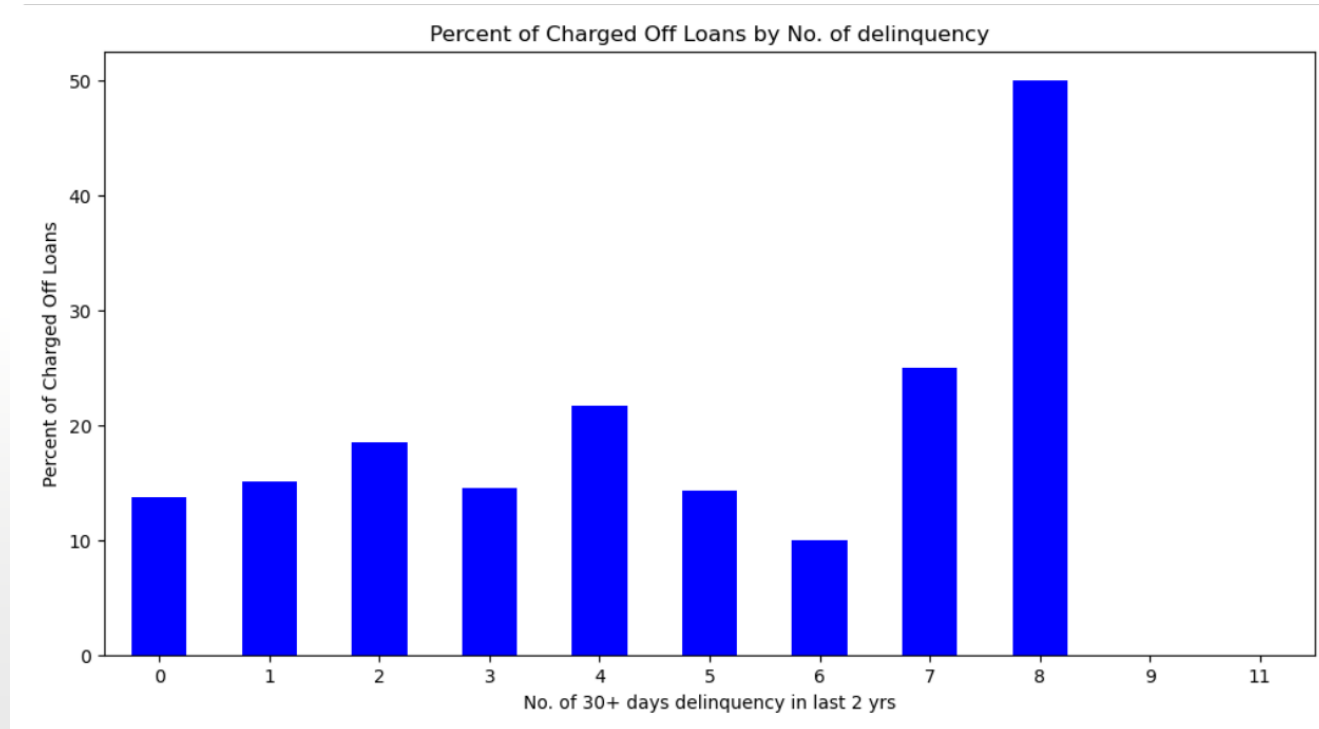
Finding 14: Public Record of Bankruptcy is a strong indicator of loan default



Finding 15: No. of open credit lines is a strong indicator of loan default



Finding 16:
delinq_2yrs is a
strong indicator of
loan default



Conclusion – Strong Indicators for Loan Default

1. Loan Term- Loan Status as Charged Off is higher for Loan Term 60 months .
2. Interest Rates is a strong indicator of loan defaults . Higher the interest Rate , higher is the chance of loan default.
3. Grade is a strong indicator of loan default. Lower Grades have higher chances of loan default.
4. SubGrade is a strong indicator of loan default. Lower SubGrades have higher chances of loan default.
5. Likelihood of loan default increases with higher installment which can be observed when installment amount increases from 0 - 1000 but is lowered when installment amount goes above 1000
6. Charged Off Loans are more skewed towards income group below 50000.
7. Purpose is a strong indicator of loan default. Some of the purposes have more percent of bad loans.
8. Likelihood of loan default increases with higher dti rate which can be observed when dti increases from 5 - 25 but is lowered when dti goes above 25
9. Some States have unexpectedly high rate of default loans such as NE(Nebraska)
10. Public Record of Bankruptcy is a strong indicator of loan default. Higher the number , higher is the chance of loan default
11. No. of open credit lines is a strong indicator of loan default. Higher the number , higher is the chance of loan default
12. delinq_2yrs is a strong indicator of loan default . Higher the number, higher is the chance of loan default



Conclusion – Other Findings

1. Proportion of Charged Off loans is 14\$.
2. Loan Amount is not a strong indicator of loan default.
3. There is no definite trend for charged Off loans as per employment length.
Employees having 7 and 10+ years of experience have slightly high rate of default.
4. There is no definite trend for charged Off loans as per home ownership .
Those having home ownership as other have slightly higher chances of loan default.

