GRAMENER CASE STUDY

SUBMISSION

Group Name:

1. Naga Sai Aditya B
2. Najaf
3. Pradeep
4. Shalvi

Gramener Case Study– A case study to identify risky applicants to a Loan company

Objective of the analysis : The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Methodology and Deliverables:

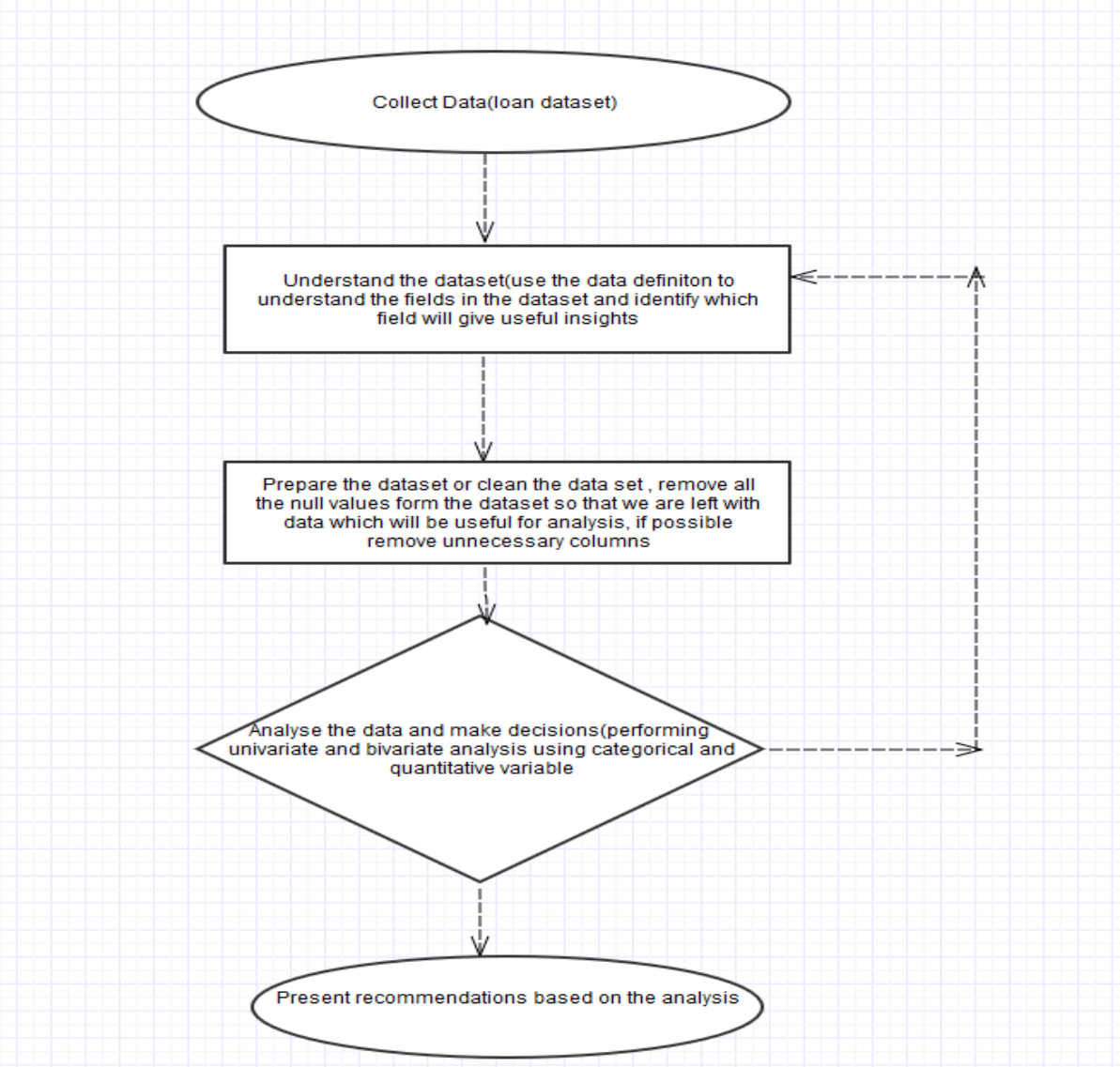
Implement exploratory data analysis on the input dataset to gain an understanding of risk analytics in a business environment.

[1] Perform Uni-variate and Multi-variate analysis of variables to identify patterns and inconsistencies in the dataset.

[2] Determine driving factors which are strong indicators of 'risky applicants'.

[3] Present observation through neat visualizations.

Problem Solving method



Assumptions and Data Handling

- Data Cleaning is performed on the loan dataset. In total we had 111 columns but few columns had 100% null values hence we identify those columns and drop them.
- Then we have columns like tax_liens which has 39 Null values and all 0.
- Then we also drop those columns where null value percentage is more than 50.
- We would also remove few columns which we would not need in our analysis.
- We would remove the loan_status = current values as that is not going to give us any insight.
- The data structure of the int_rate and revol_util are represented as character type due to the presence of the % symbol. We will remove the symbol and represent it as a numerical value representative of the interest percentage. i.e. 10.75% will be converted into 10.75.
- We would also create derived metrics like **profit and loss** and **loan_inc_ratio** which is ratio of funded amount and annual income.
- Following set of analysis done after cleaning the data:

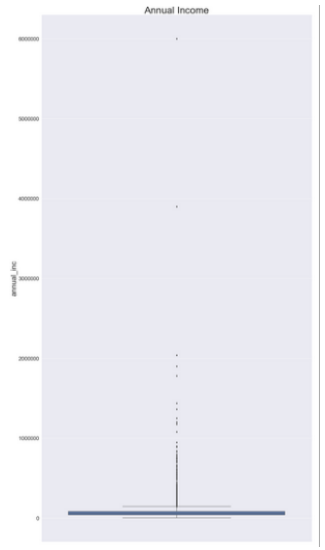
.

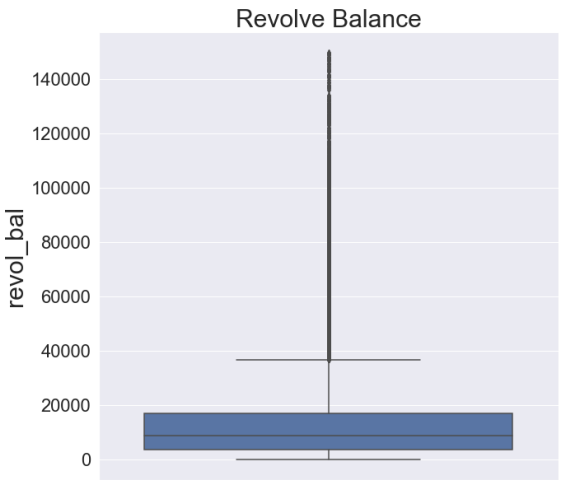
Assumptions and Data Handling

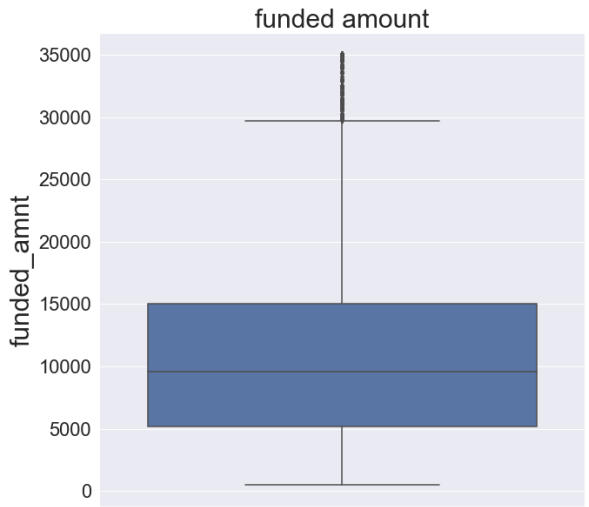
1. After removing 100% null values we were left with 57 columns.
2. Then we removed columns like tax_liens and also columns where null percentage was higher than 50% hence we got 53 columns .

3. We also removed columns that were not going to be used in our analysis like **Loan Title** which is a drilldown of the loan purpose attribute and is specific to the applicants loan needs. Since it is a text heavy column with numerous entries branching from a main group it will notbe essential for EDA. We will exclude this column. Loan desc is again a text based description of the purpose of the loan. Since text analysis isn't under the purview of this case study we will not consider this attribute as well.**URL** leads to a web address specific to a particular loan application record. It is again a non-essential attribute in analyzing driver attributes to credit loss or fraudulent loan applicants. We will disregard this column as well. The attribute **collections_12_mths_ex_med** representing number of collections in the past year excluding medical collections contains records with 0 or Na values. Therefore we will disregard this attribute.

4. Now lest consider the outliers in the analysis for the columns funded_amnt, annual_inc and revol_bal



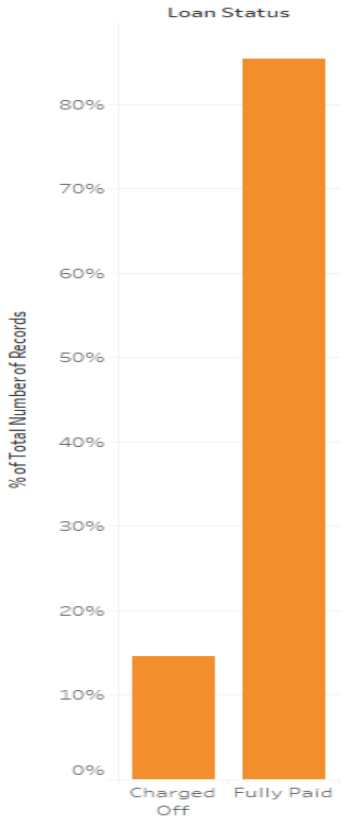




Since there are so many outliers we need to remove them as that would give us incorrect analysis or would affect the overall percentage used in the analysis.

Analysis and Conclusion

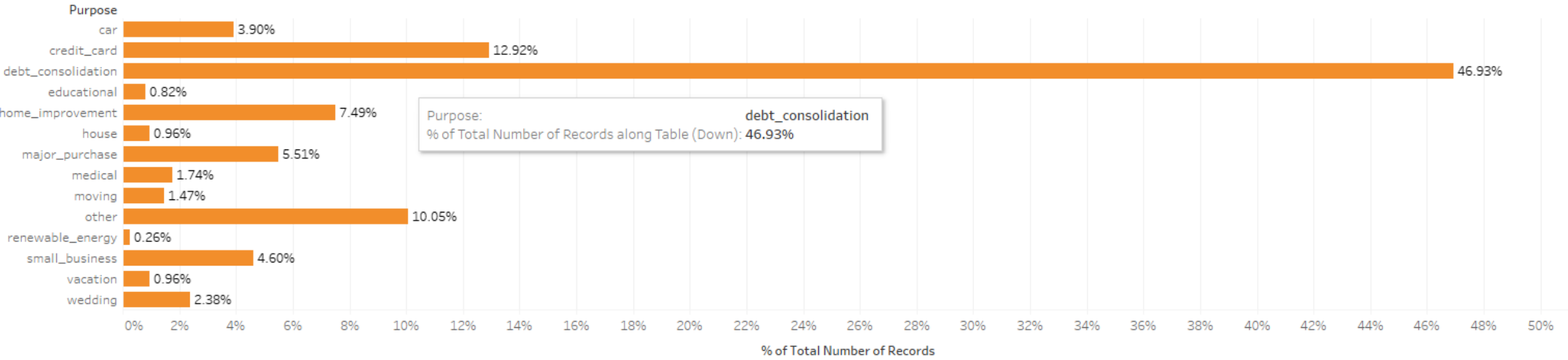
We will now see percentage of applicants that defaulted or were “Charged-off”



We can see from the graph that approx. 15% of the approved loans has resulted in credit loss, so we will have to analyze this issue in detail as this is the major concern for the lending company

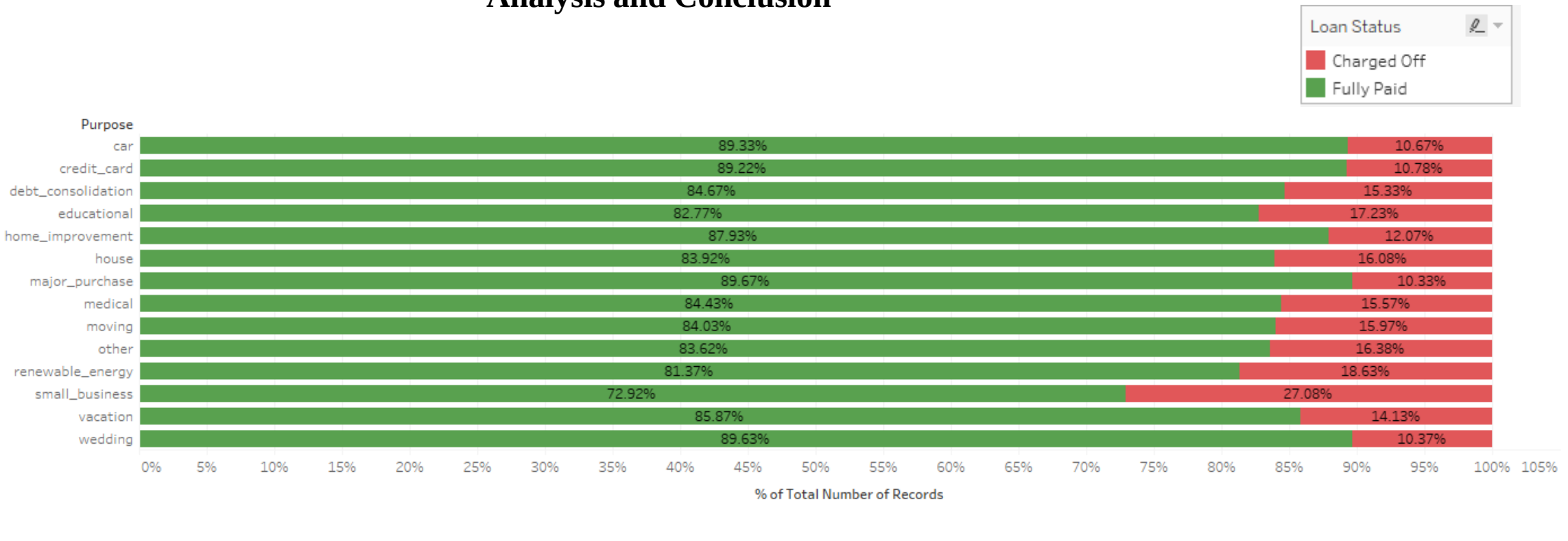
Analysis and Conclusion

Now lets look what was the purpose for which the loan was applied for



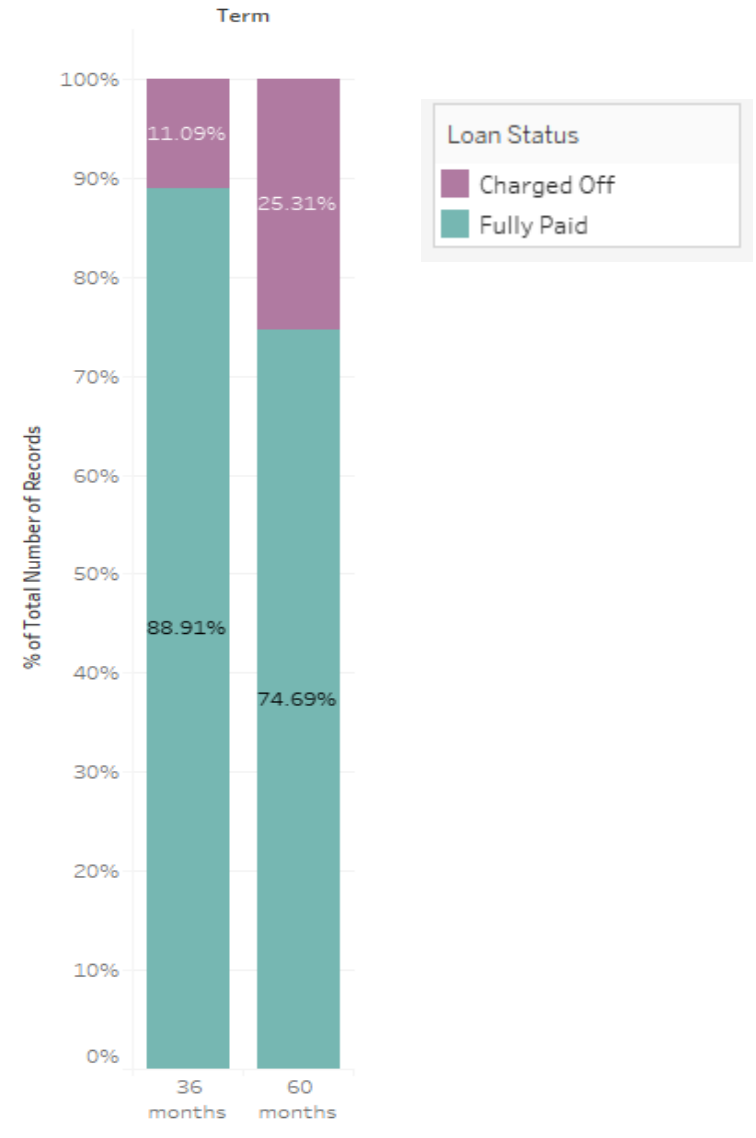
Insight: Debt_consolidation was the purpose for which maximum people applied for loan i.e. almost 46.93% of the total applicants , now lets look at the purpose for which the credit loss was high

Analysis and Conclusion



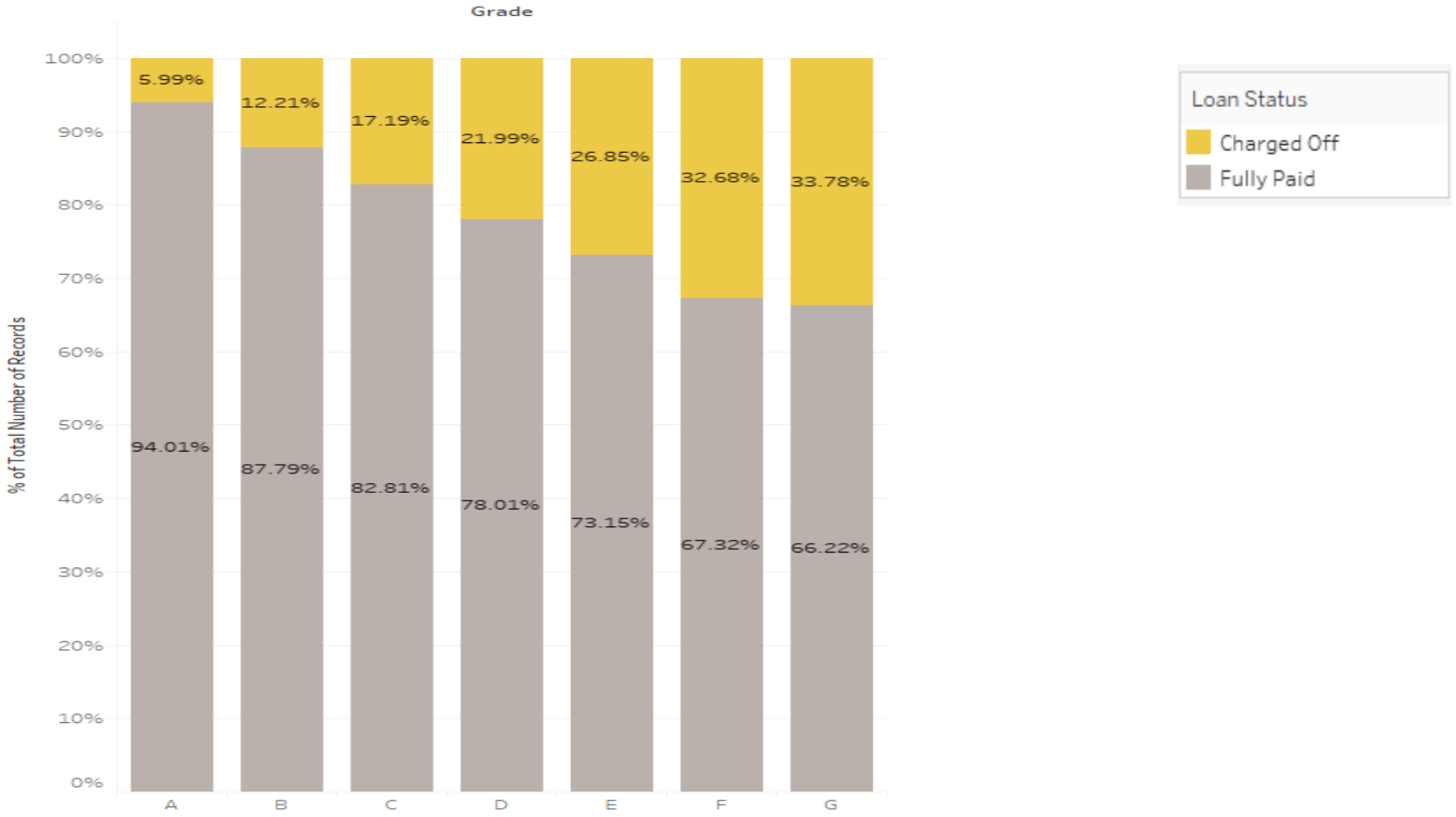
Insight: Although the highest number of loan applications are received for debt consolidation From the plot it is clear that the highest percentage of defaulters which is 27.08% is from **small business.** The company should avoid giving loans for small business or should verify the income sources of the small business rigorously.

Analysis and Conclusion



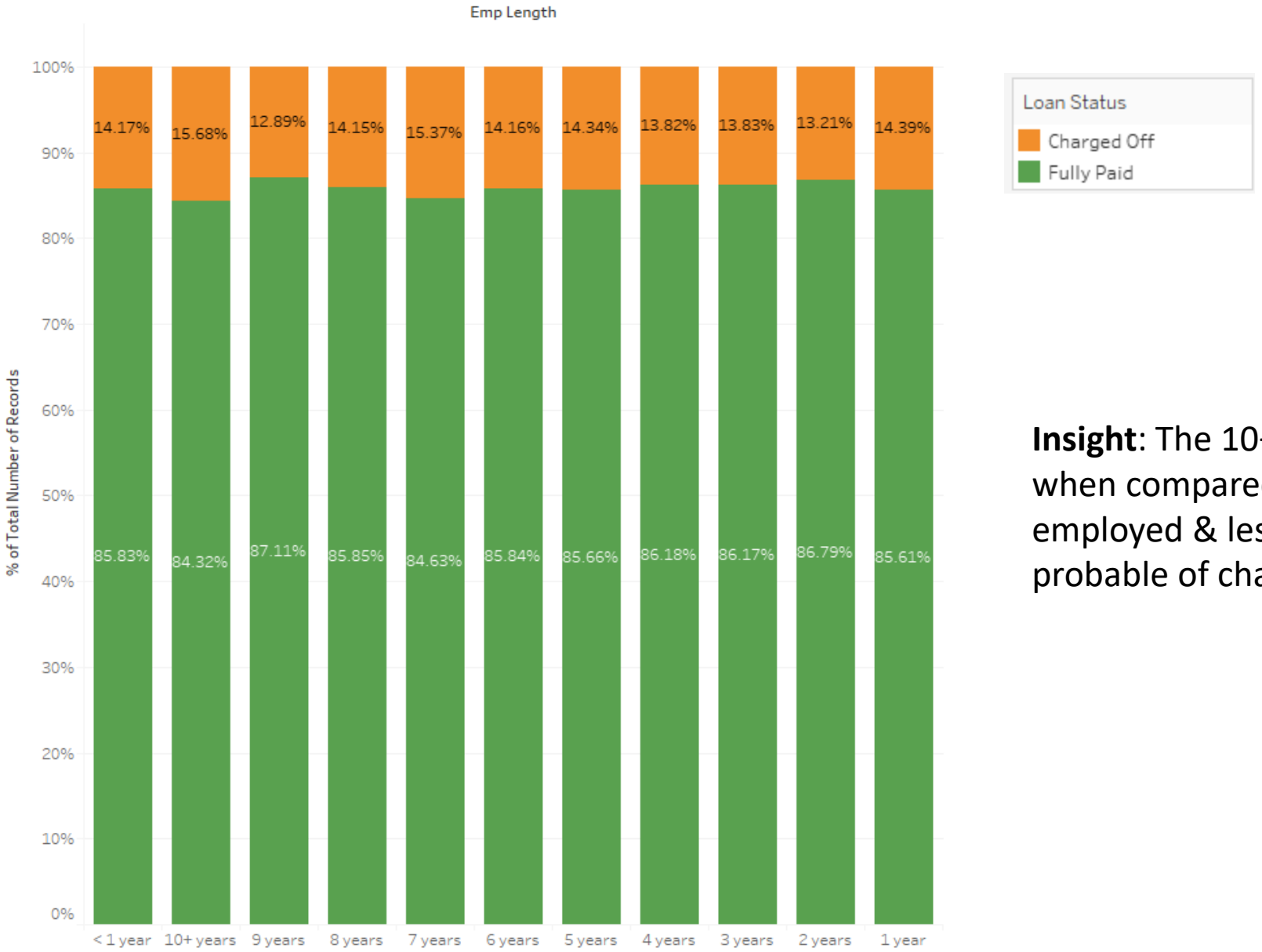
Insight: The credit loss is high for 60 months term of loan(25.31%) and is considerably higher than the short term of 36 months, which means the company should provide a shorter term loan and avoid giving long term loan.

Analysis and Conclusion



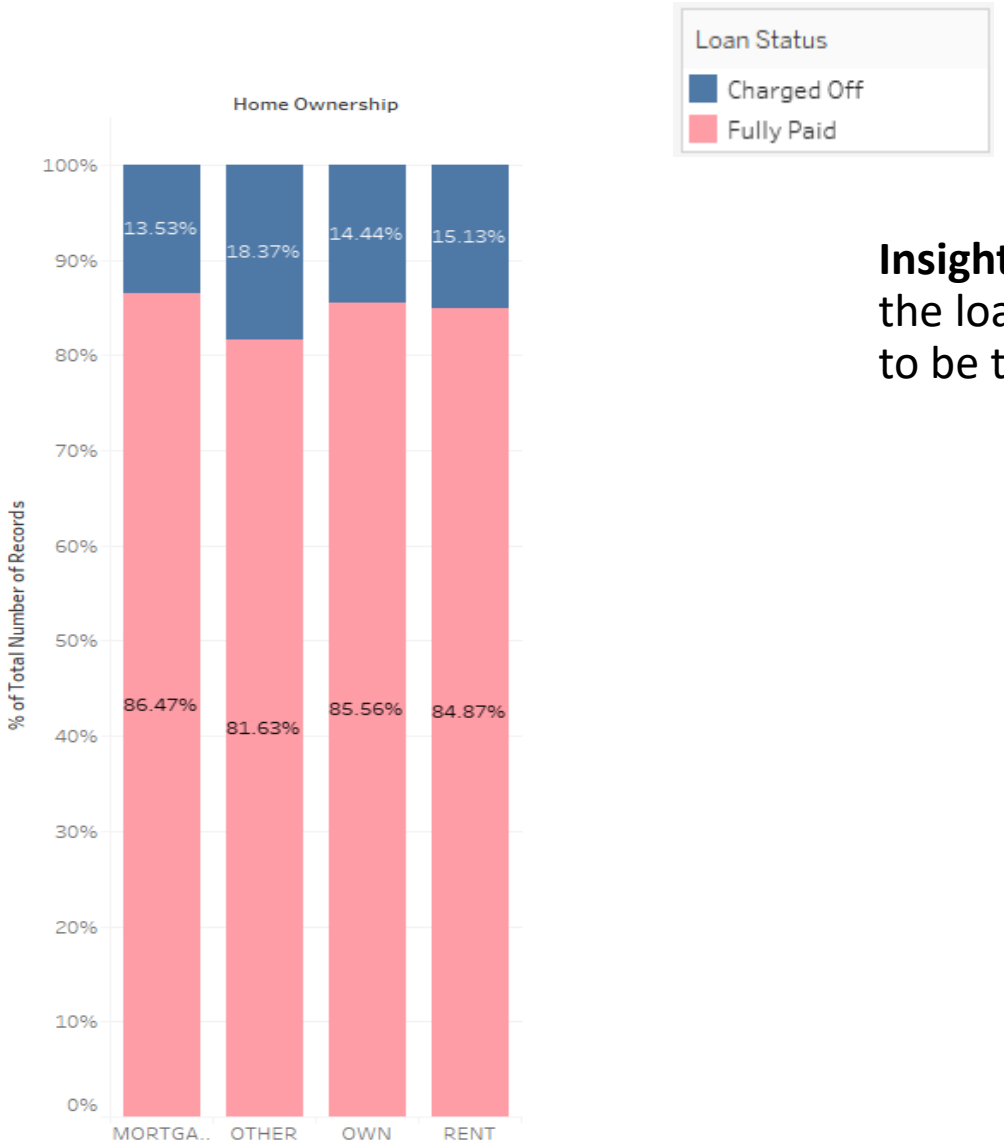
Insight: As we move from Grade A to G, probability that person will charged off is increasing.The grade G and grade F has higher chance of defaulting(33.78% & 32.68% respectively) .

Analysis and Conclusion



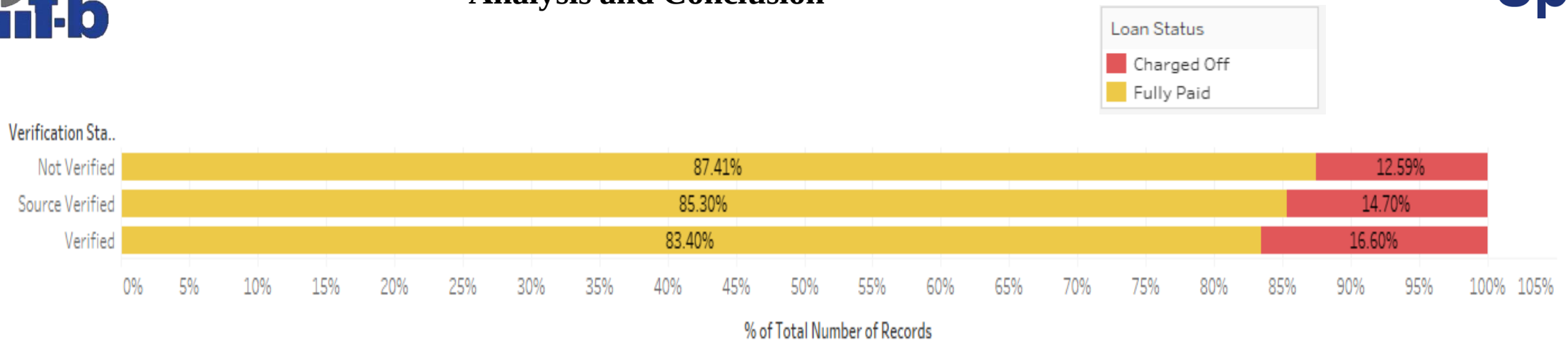
Insight: The 10+ and 7 years group is more likely to default when compared to other groups. Applicants who are self employed & less than 1 year of experience are more probable of charged off.

Analysis and Conclusion



Insight: We see that there is no impact of home ownership on the loan servicing on customers, hence this is not contributing to be the driving factors.

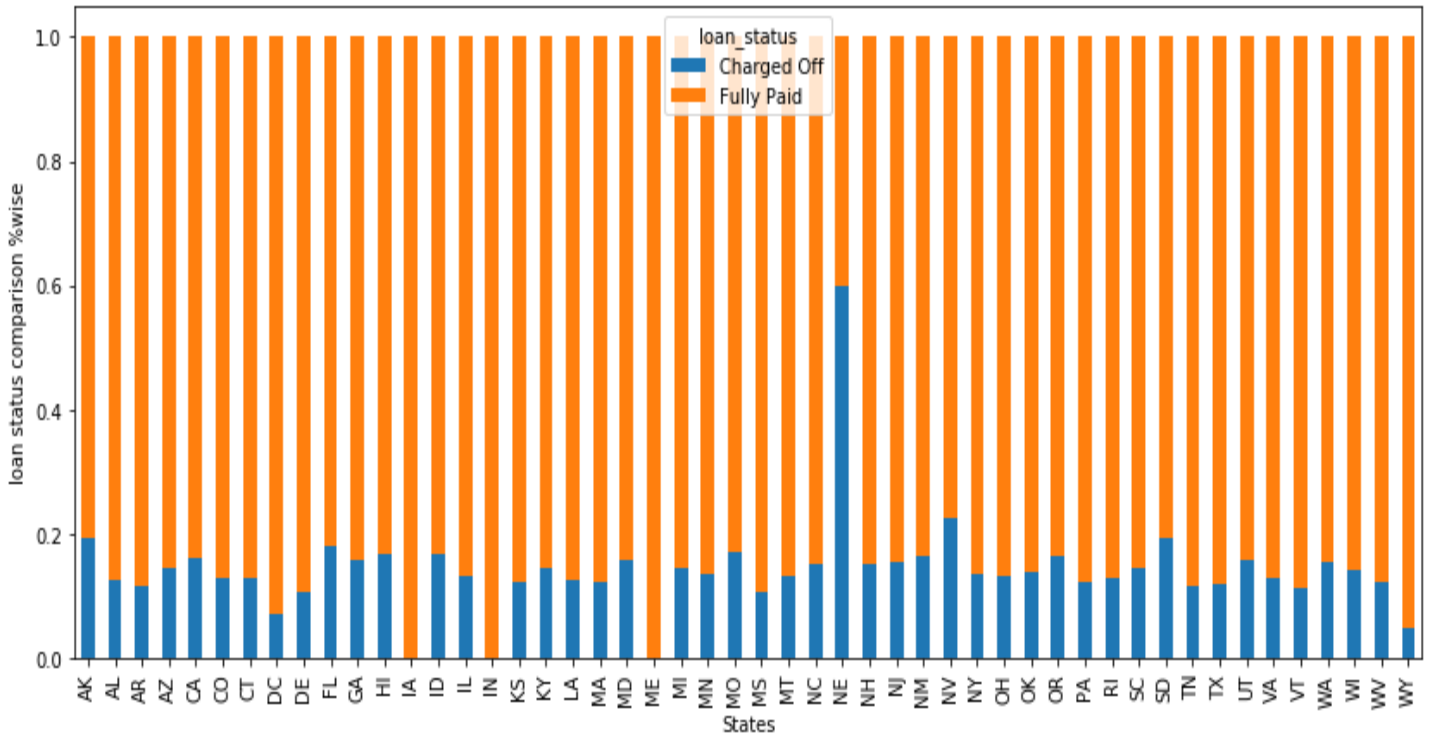
Analysis and Conclusion



Insight: The company uses a methodology to verify the income sources of the applicant but we can clearly see from the graph that verification model is not working properly as the defaulters are mostly those individual whose income source has been verified.

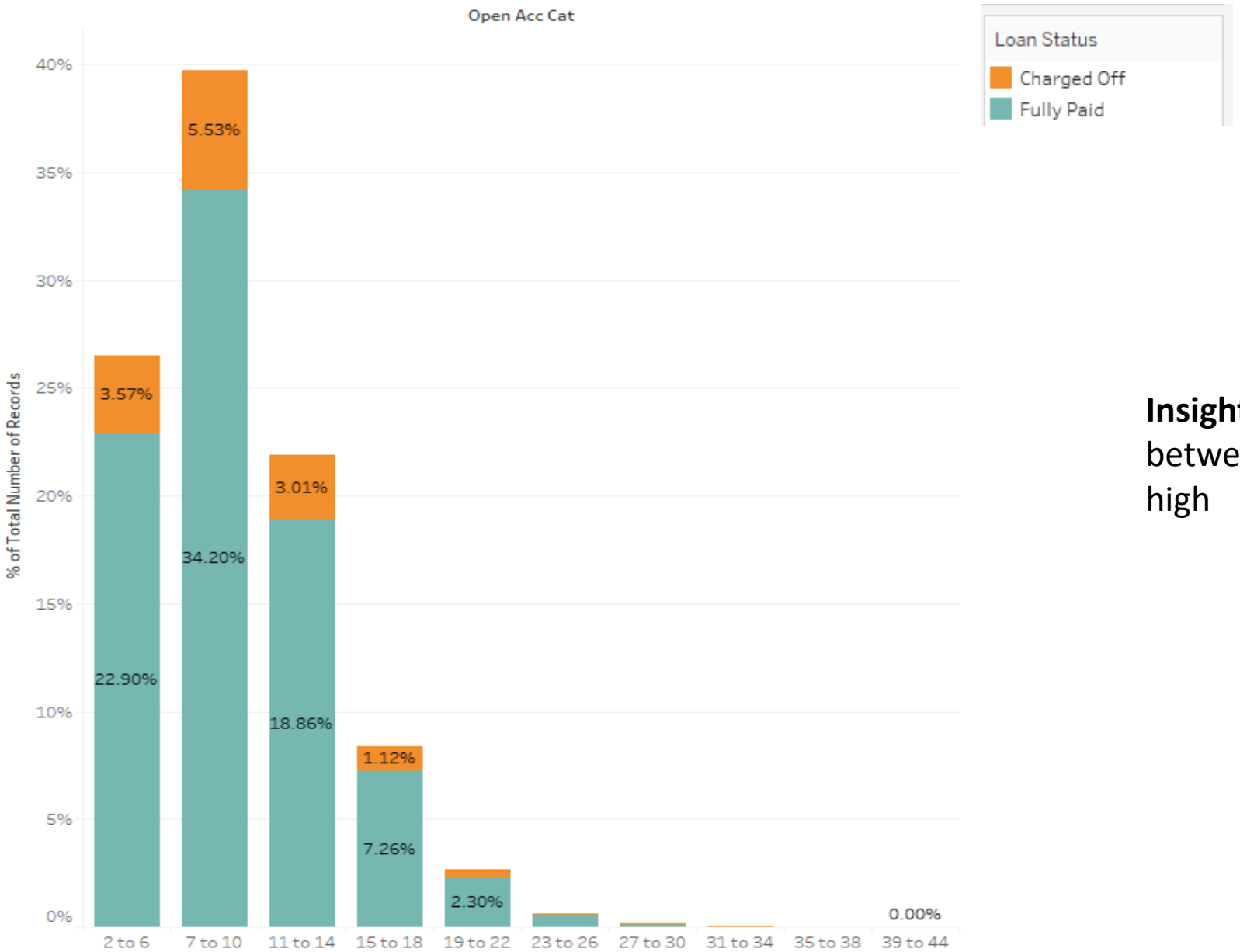
Hence the company needs to be more vigilant about the verification process.

Analysis and Conclusion



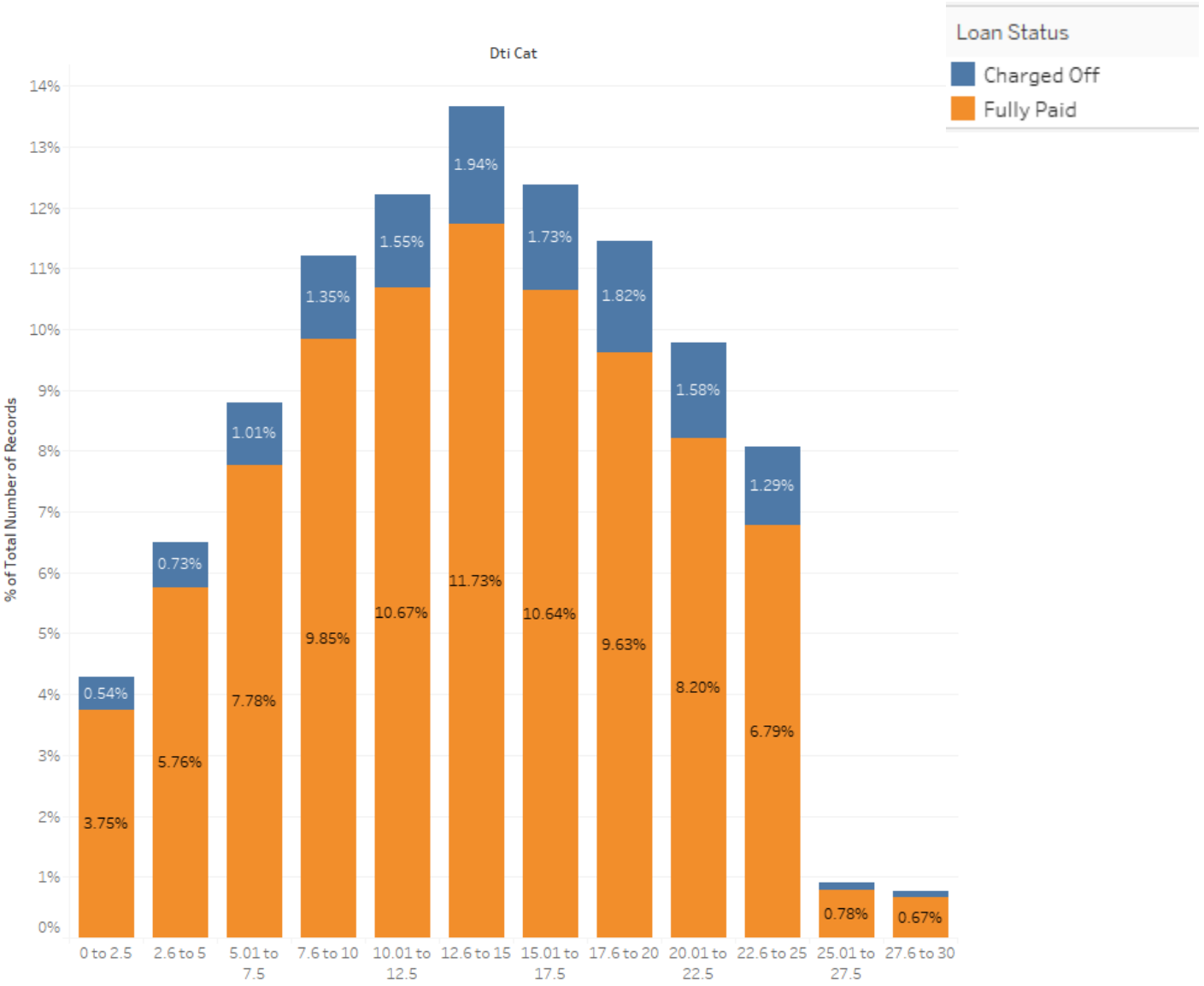
Insight: We see NE (New England),NV (New Virginia) is a stand out in terms of defaulters, company has to treat NE requests with caution
States - IA,IN,ME have been stable in paying off the loans, can be considered as a good area to provide loans.

Analysis and Conclusion



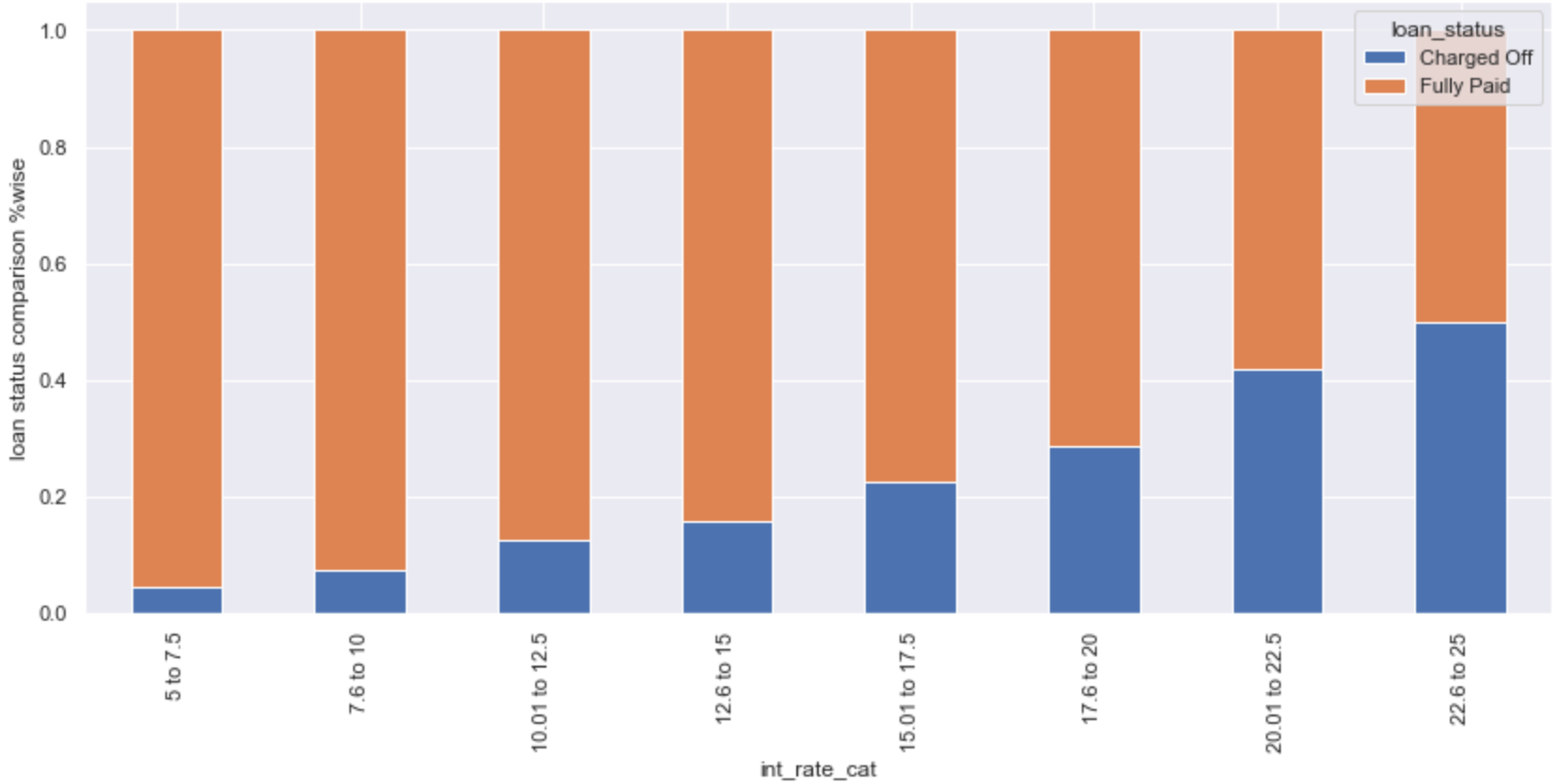
Insight: If the open credit lines of the borrower is between 7 to 10 then the chances of charge off is high

Analysis and Conclusion



Insight: The highest chargeoff was where dti is 12.6 to 15

Analysis and Conclusion



Insight: We can see from the graph that as interest rate is increasing the credit loss chance or the likeliness of being charged off is increasing.

Conclusion

Top Major variables to consider for loan prediction:

1.Purpose of Loan – The small business had higher percentage of charge off , hence the company should consider before lending loan to small businesses.

2.Employment Length - The 10+ and 7 years employment length group is more likely to default when compared to others.

3.Grade- **:** As we move from Grade A to G, probability that person will charged off is increasing.The grade G and grade F has higher chance of defaulting

4.Interest Rate- As the interest rate is increasing the credit loss chance or the likeliness of being charged off is increasing.

5.Term - The credit loss is high for 60 months term of loan and is considerably higher than the short term of 36 months, which means the company should provide a shorter term loan and avoid giving long term loan.

6.Loan Status : Which is used to compare with the variables for effective analysis and decision making.

7.State : NE (New England),NV (New Virginia) is a stand out in terms of defaulters, company has to treat NE requests with caution
States - IA,IN,ME have been stable in paying off the loans, can be considered as a good area to provide loans