GRAMENER CASE STUDY

SUBMISSION

Group Name:
1. Vikram Arvapalli
2. Purvaa Reddy
3. Sai Varshitha
4. S. Thanusha

Brief Overview of case study

Gramener Banks, an asset banking company wants to find why there is more defaults and to find whether a person defaults or not.

Two Problems:

Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:\

- Fully paid: Applicant has fully paid the loan (the principal and the interest rate)

- 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'.

- 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

 Loan rejected:The company had rejected the loan (because the candidate does not meet their requirements). 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).

LOAN DATASET

Loan Accepted → Default
Loan Accepted → Non-Default

Loan Rejected
(Not considered in dataset)
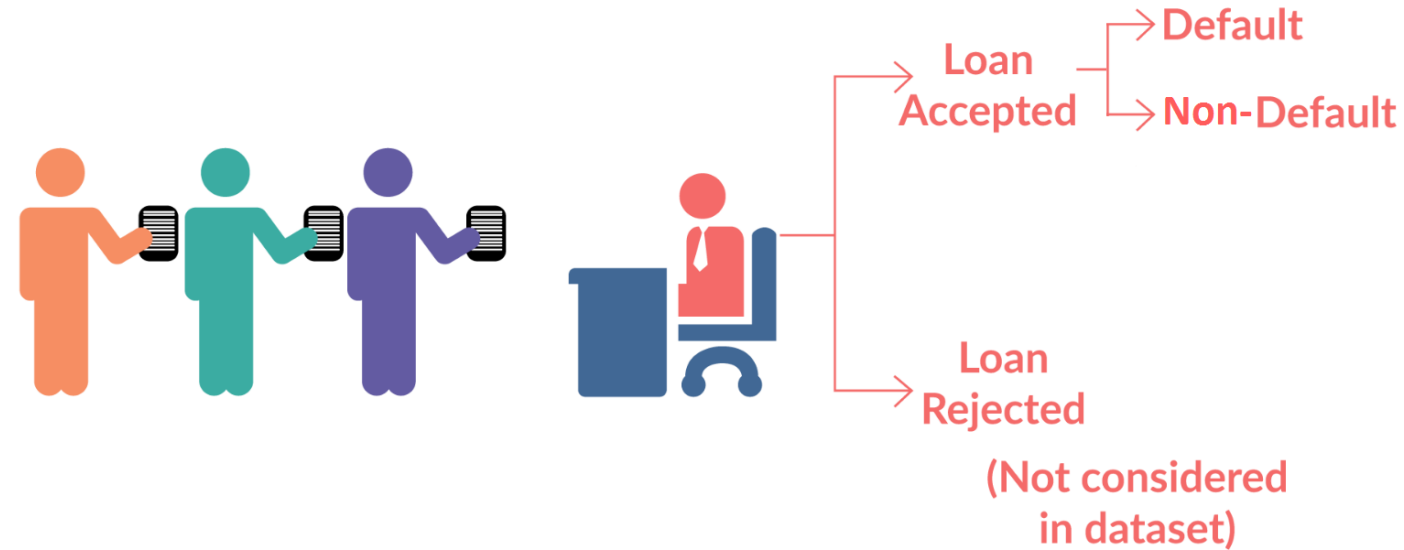
loan accepted or rejected

There is less problem if loan is rejected but it is high if banker face problem when they are many defaults they need to identify customers before approval

Objectives

- 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

- **Objective:**

- •Identification of Loan Applicant traits that tend to 'default' paying back

- •Understand the 'Driving Factors' or 'Driver Variables' behind Loan Default phenomena

- •Gramener may choose to utilize this knowledge for its portfolio and risk assessment of new loan applicants

Metadata Summary: Variables considered in analysis

Categorical variables	Numerical Variable	Categorical variables
Term	Loan Amount	pub_rec
Grade and Sub Grade	Funded Loan Amount	pub_rec_bankruptcies
Employee Title	Interest Rate	
Employee Length	Instalment	
Home Ownership	Annual Income	
Verification Status	DTI	
Purpose		
Addr State		
Delinq_2yr		
Inq_last_6_months		
mths_since_last_record		

Loan Status
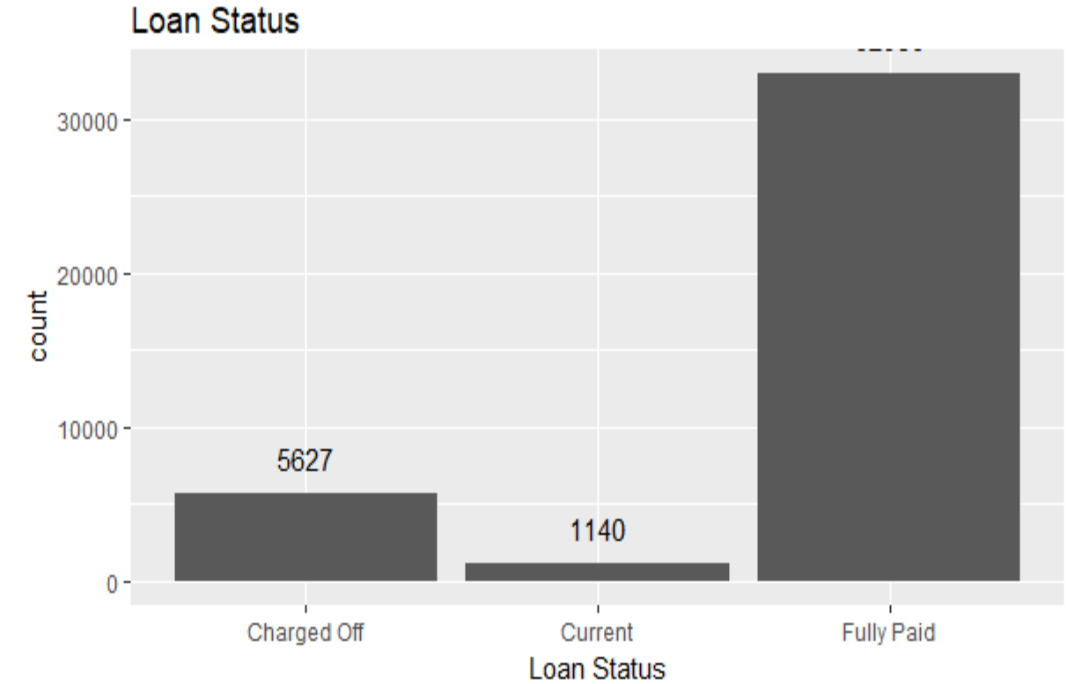
Off the total loans about 14% loans have defaulted

According to the percentage its ok because it only 14% what if that 14% defaults have taken loan with high amounts then that effects the bank

Dealing with Missing values :

- In this dataset 54 variables contain all the observations as NA so they are removed.

- 14 columns have more than 70% of 0 so they are removed.

- Some columns are replaced with 0 in the place of NA.

- 4 columns from remaining has same data in column hence they were removed.

- Some columns like title have soo many unique values so they are removed

Others:

- Rounded of some numeric variables.

- Changing datatypes to string and numeric accordingly

- Removed "months" from terms and "%" from some variables for visualization purpose.

Top 7 Deciding Parameters for Loan Defaulter

1. Loan Amount
2. Grade
3. Purpose of Loan
4. State
5. Home Ownership
6. Verification Status
7. Terms

Finding correlations
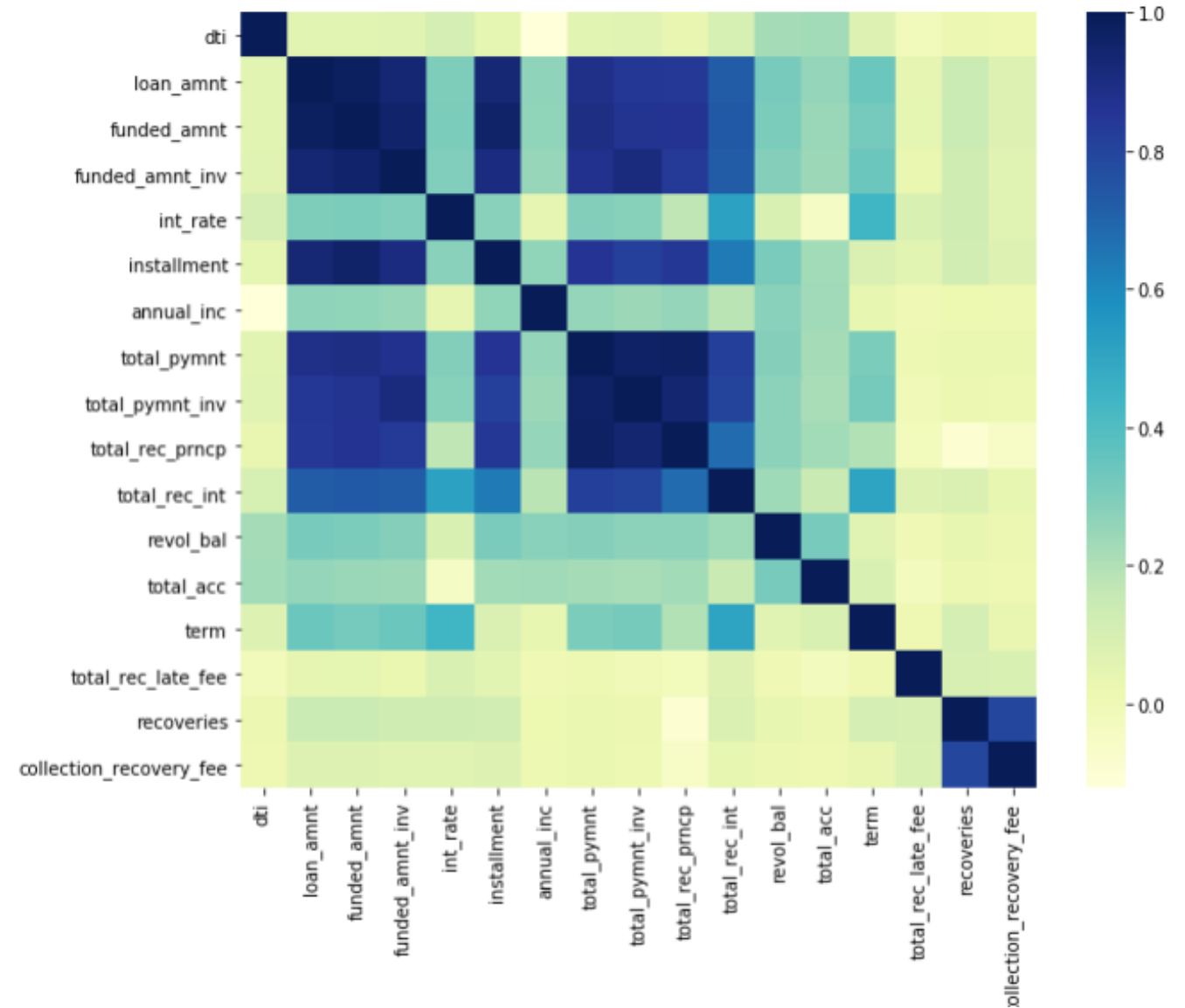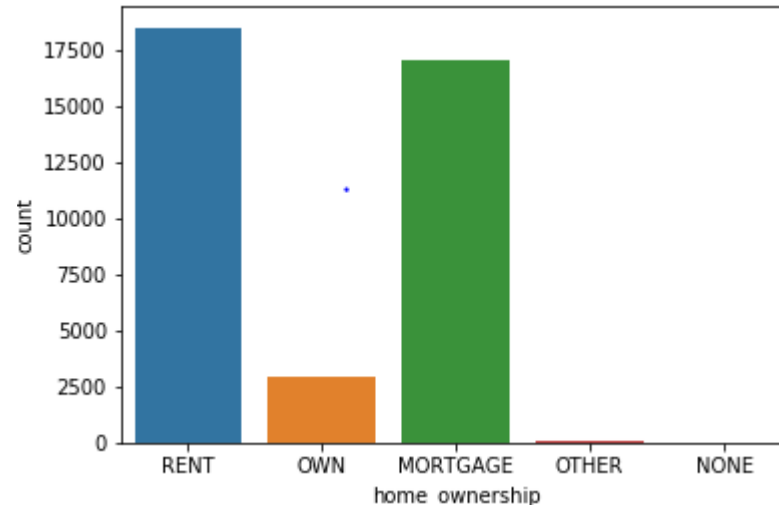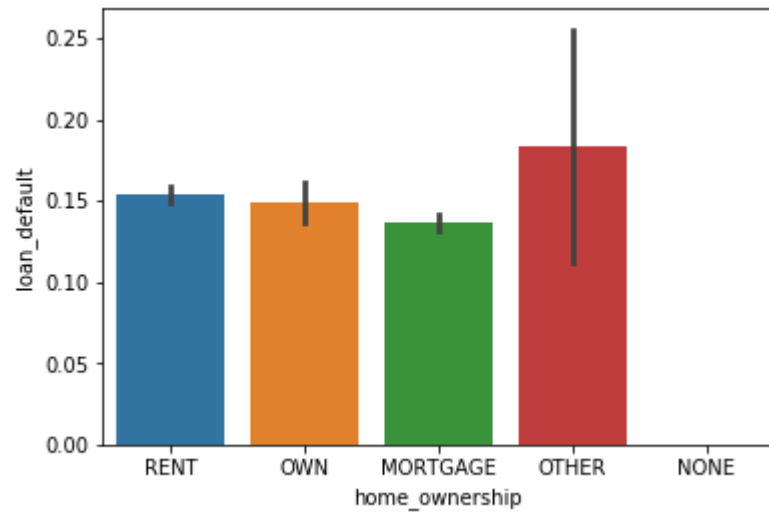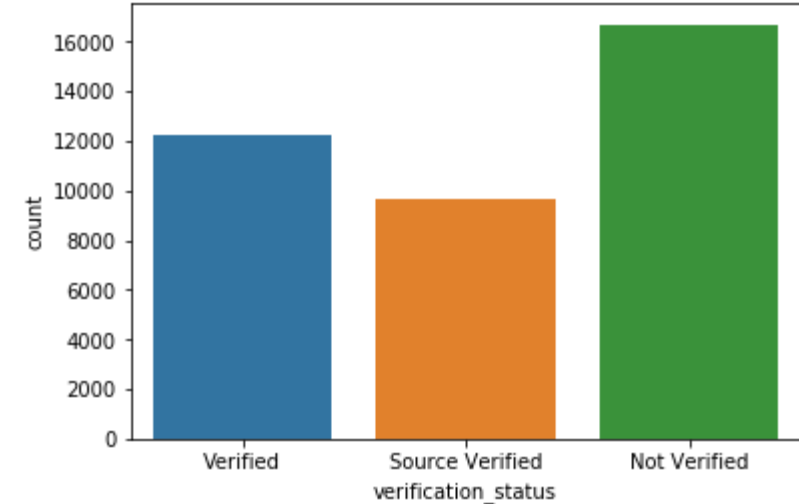
Analysis on Loan Amount

- So out of 39717 loans granted 5627 were charged off.
- For 1140 members status is current.

Loan Analysis

Loan Status

Analysis on Home Amount
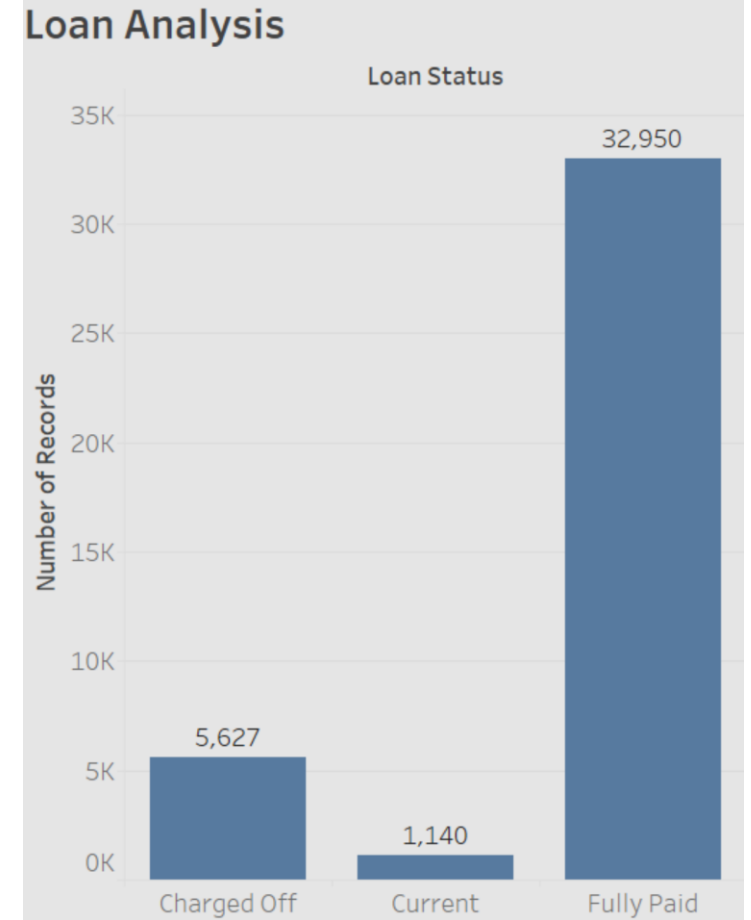
- So out of 39717 loans granted 5627 were charged off.
- For 1140 members status is current.
- But the defaults are high In other

Analysis on Verification Status
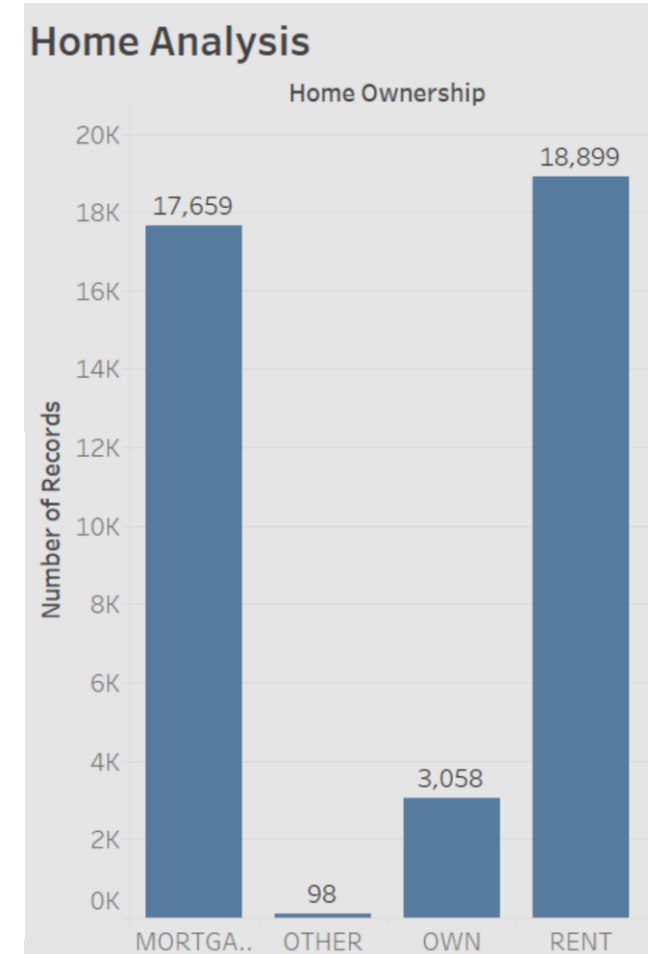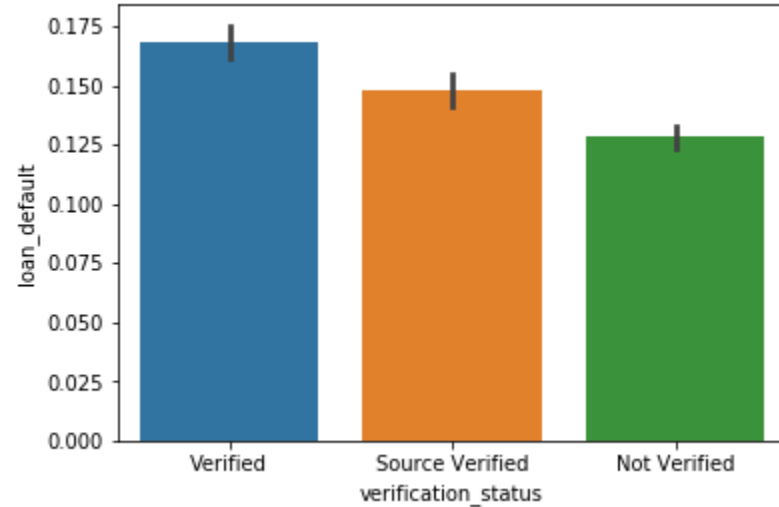
This analysis shows the bank's carelessness.

- With out verifying the income of the person they granted loans for 2142 members.
- Here as we see no of defaults are high in verified and not verified that means bankers haven't verified properly
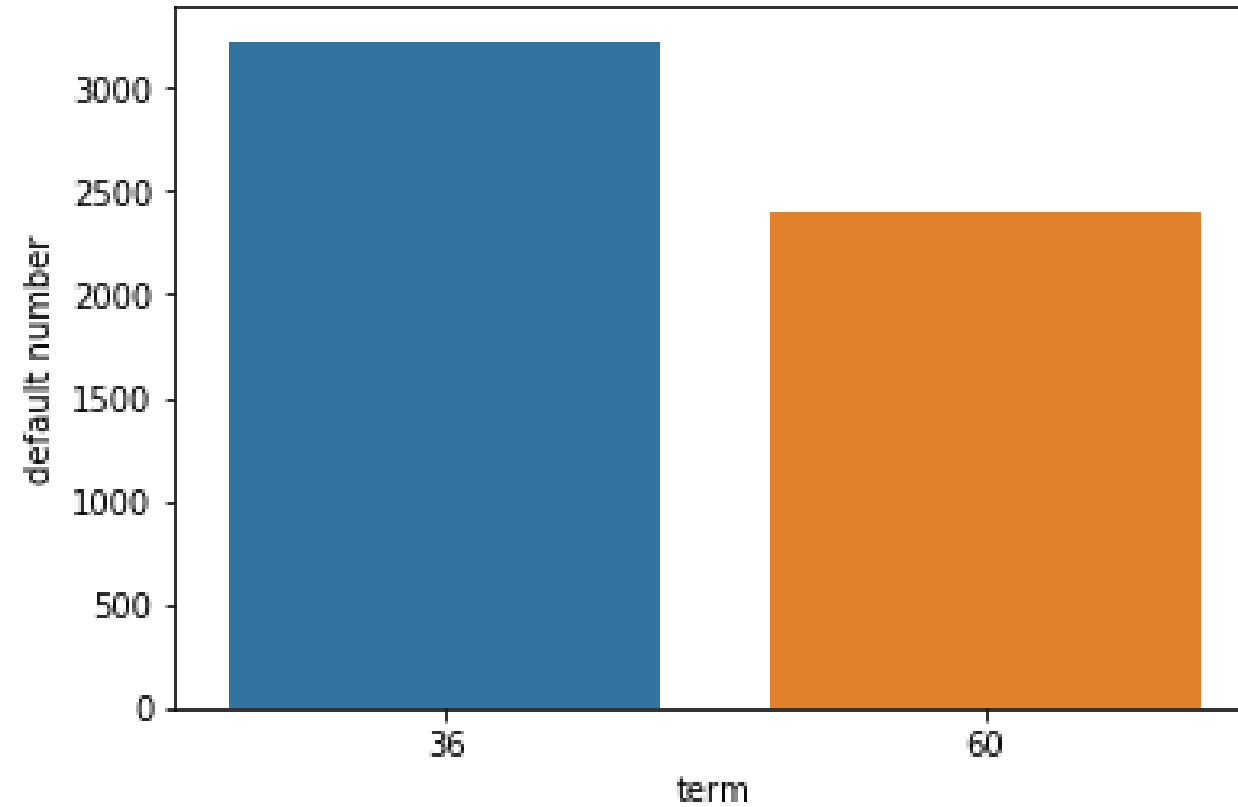- comparatively source verified are better than both
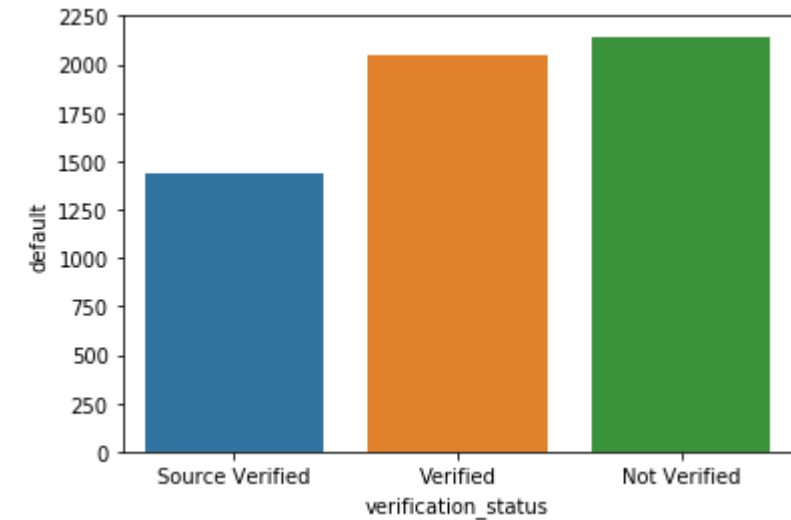
Analysis on States where most charge off taken
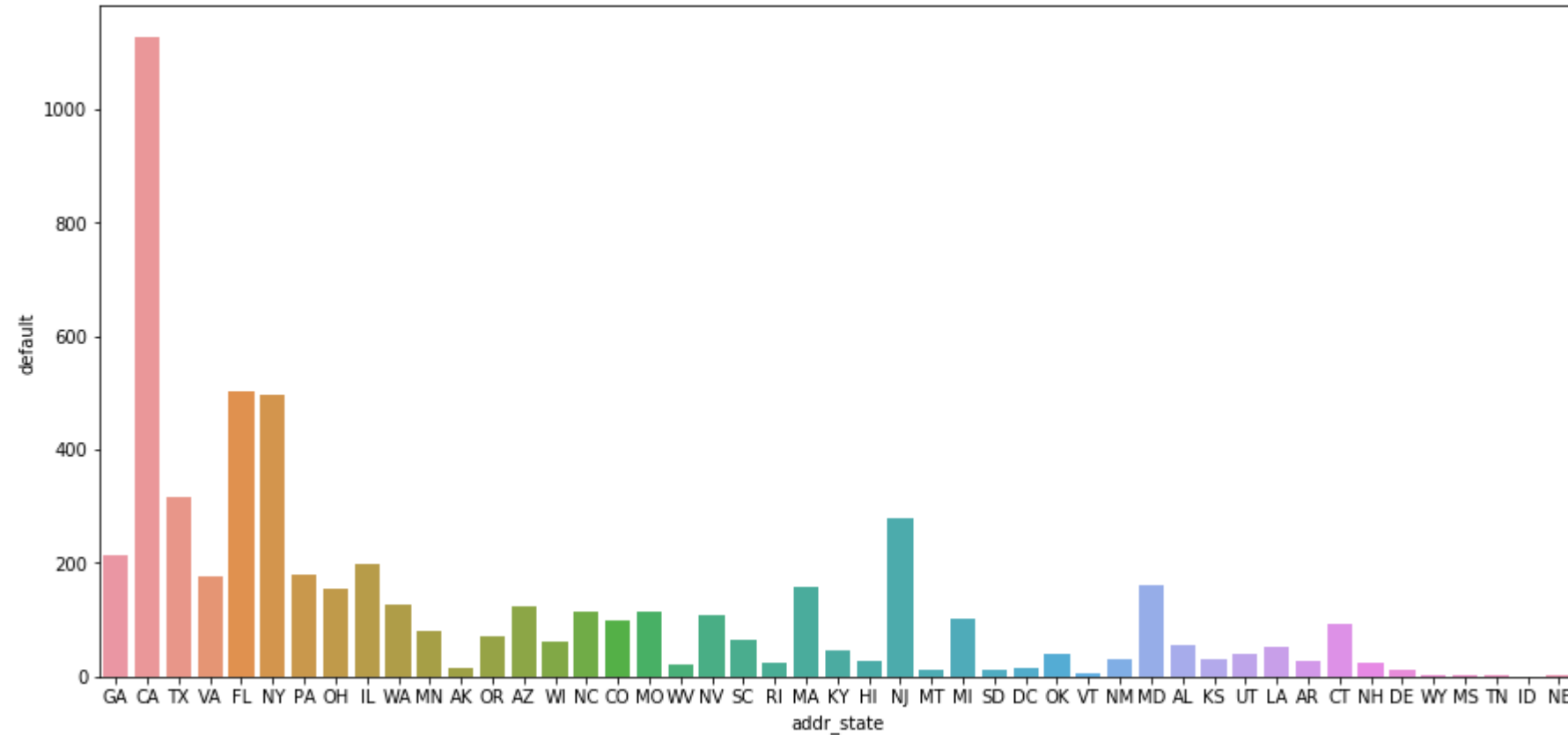
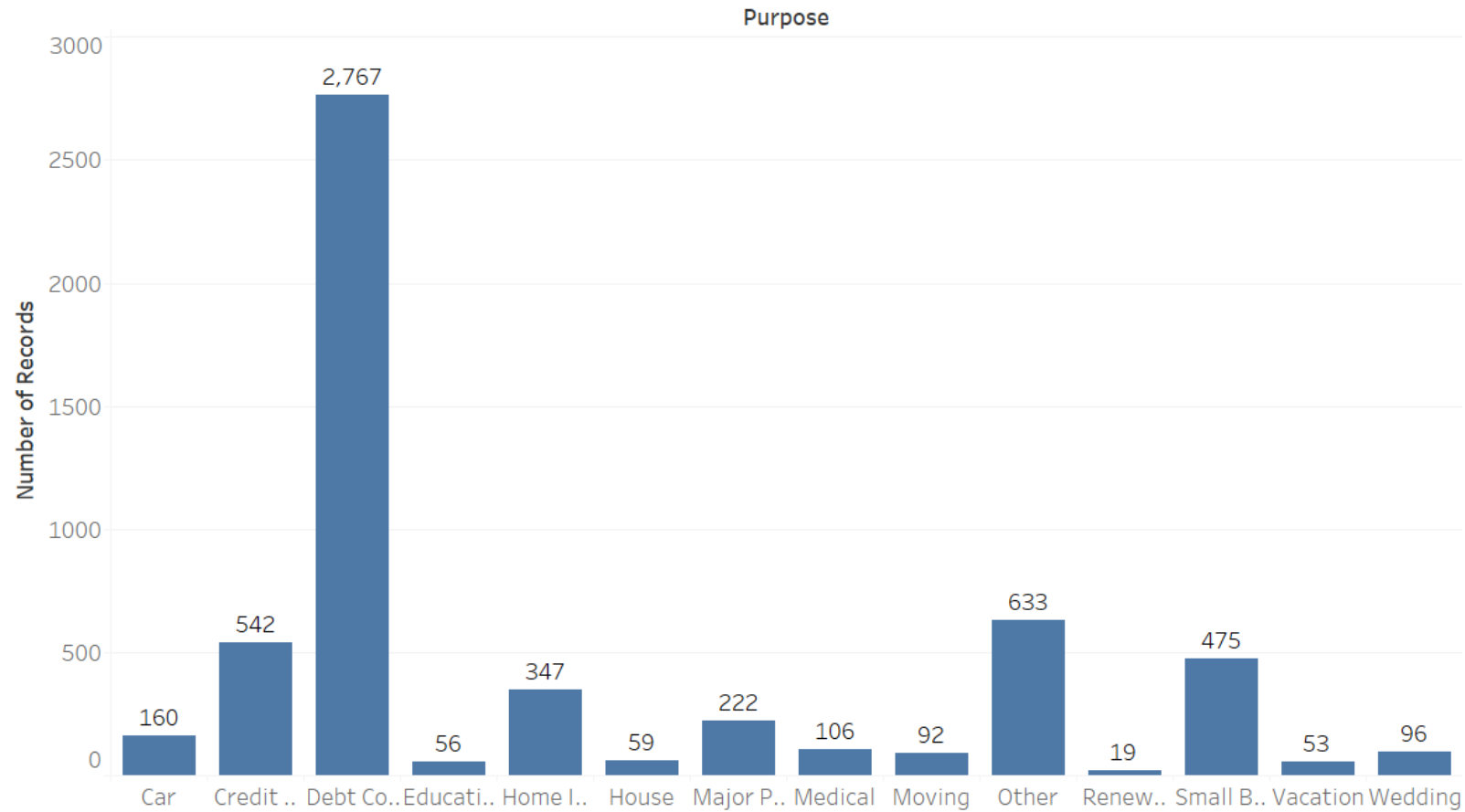California State has most number of cases which are charged off

Analysis on Grades where most charged off taken
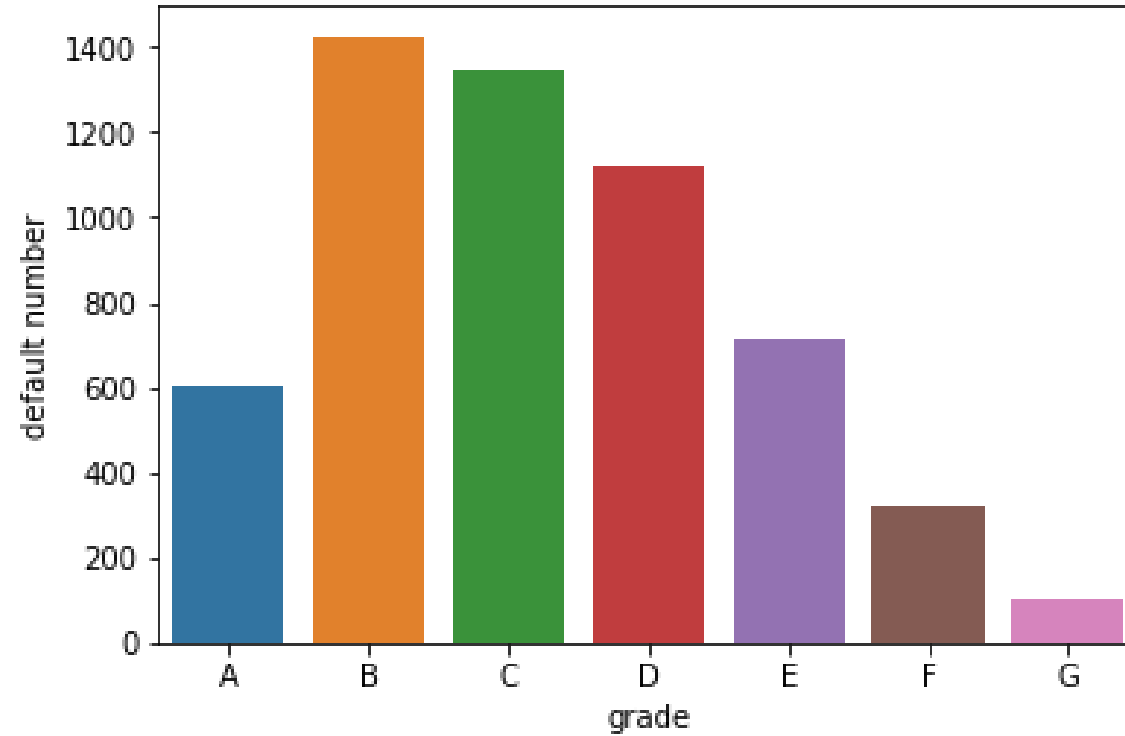
Most of the charged off occurred in B, C and D grades of loans
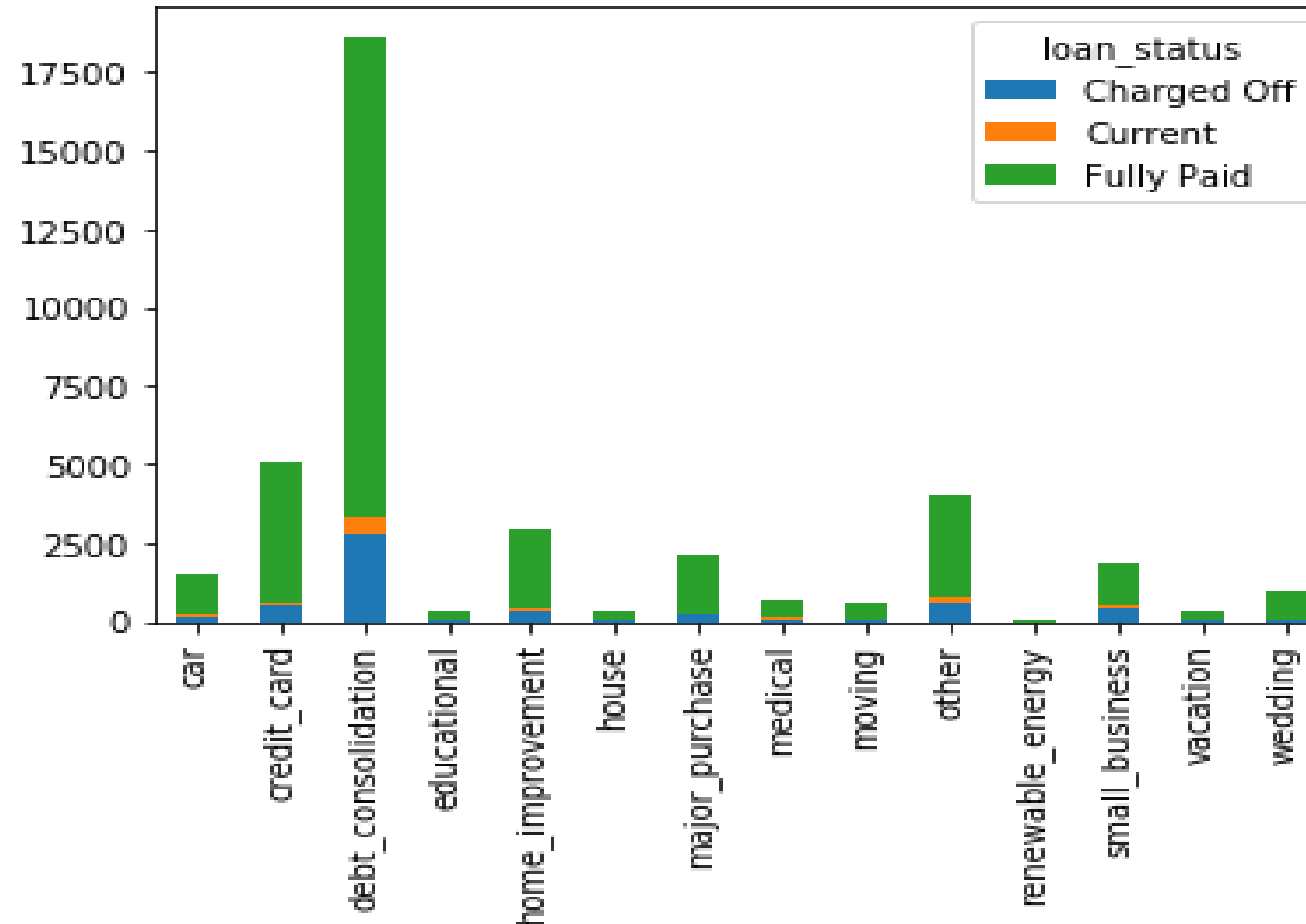
Analysis on Terms

The number of defaults are more for the people who have the term of 36 months
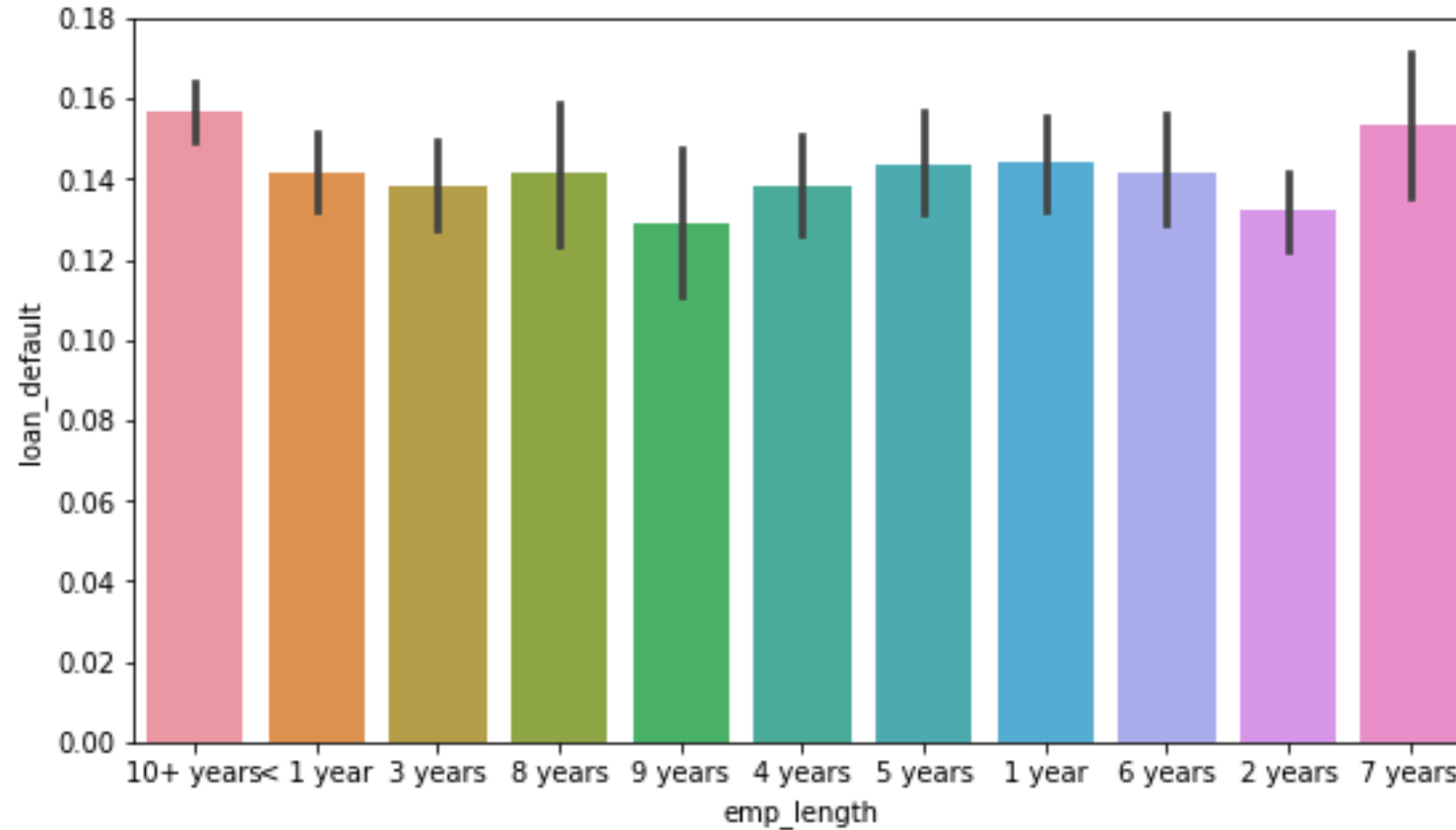
Analysis on Charged Off purpose

2767 Loan defaulted in Debt Consolidation purpose, followed by when purpose is not specified and is others
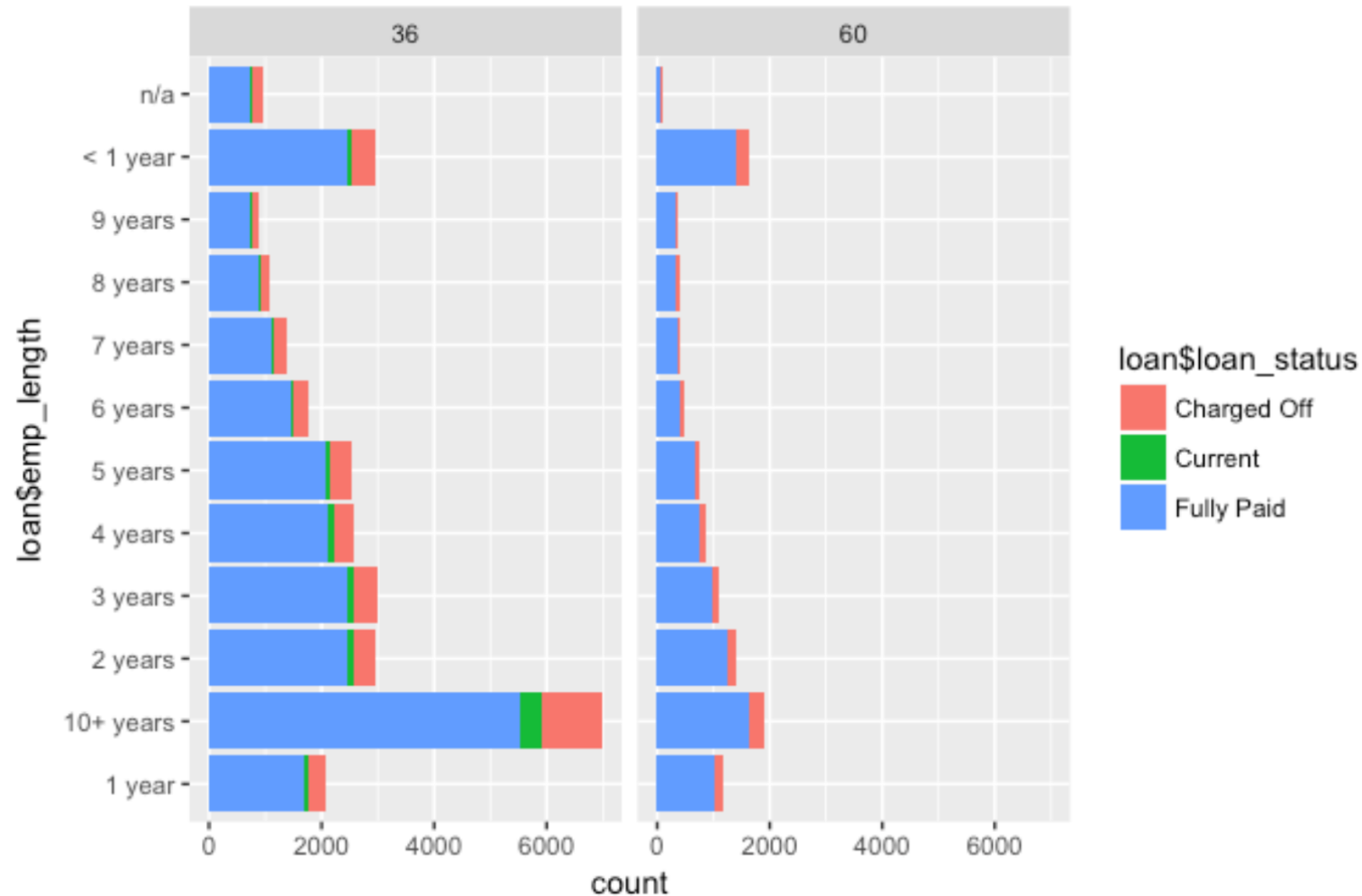
- For overall Population of loan we see that higher % of people with purpose as small business as charged off. Overall Debt Consolidation then small Business
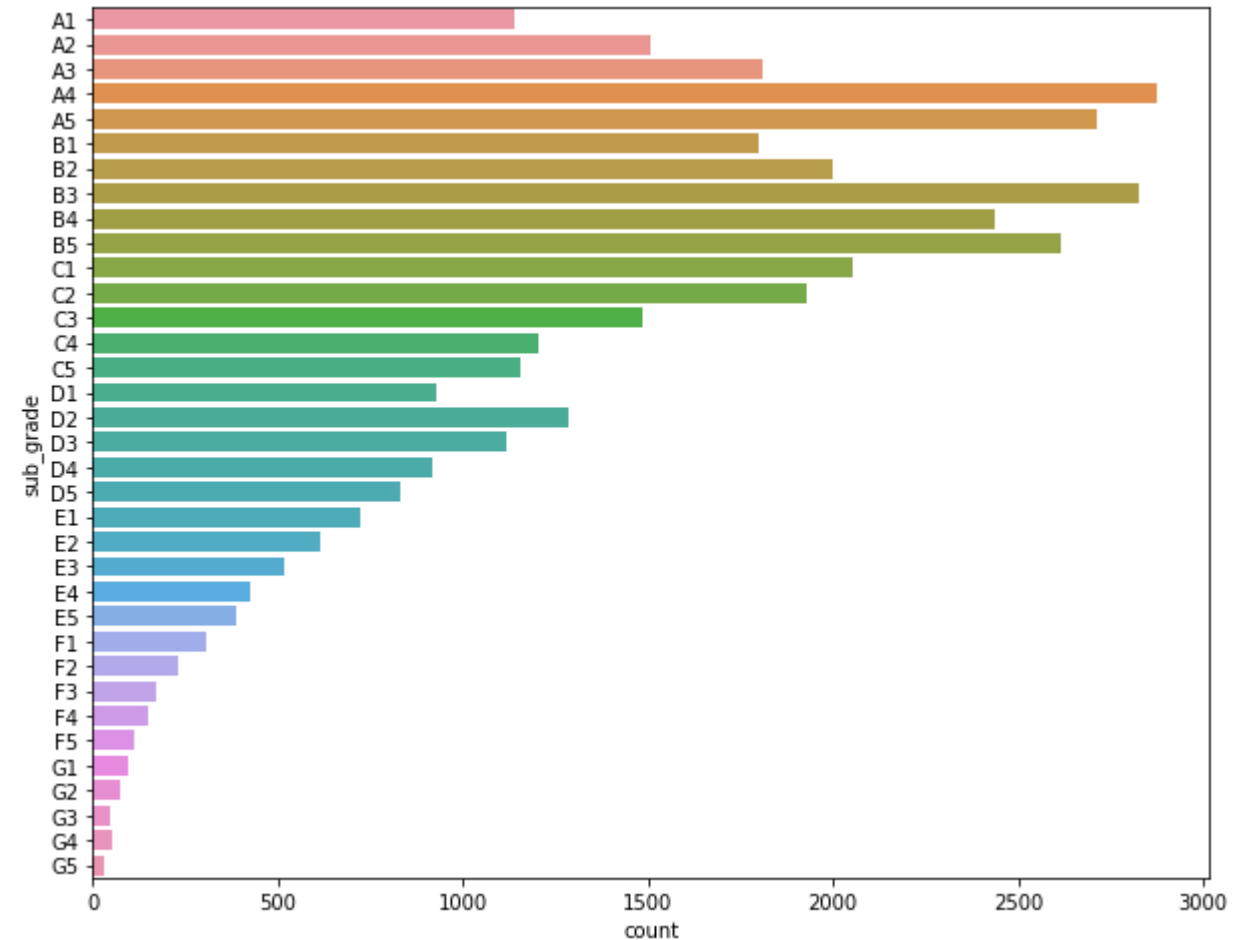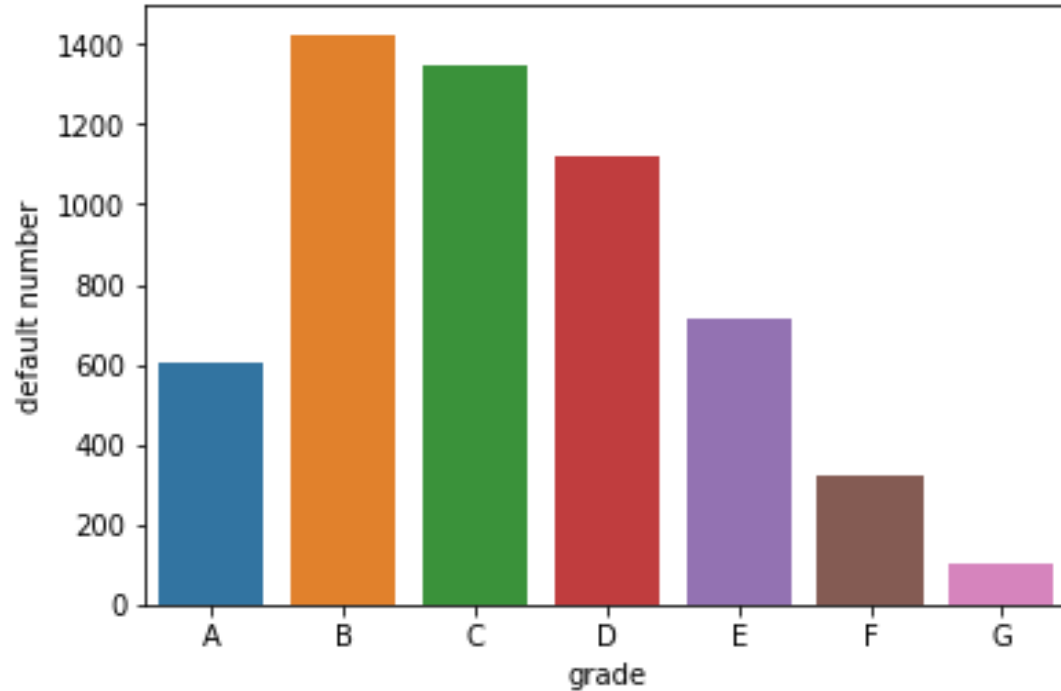
Analysis REASONS FOR emp_length

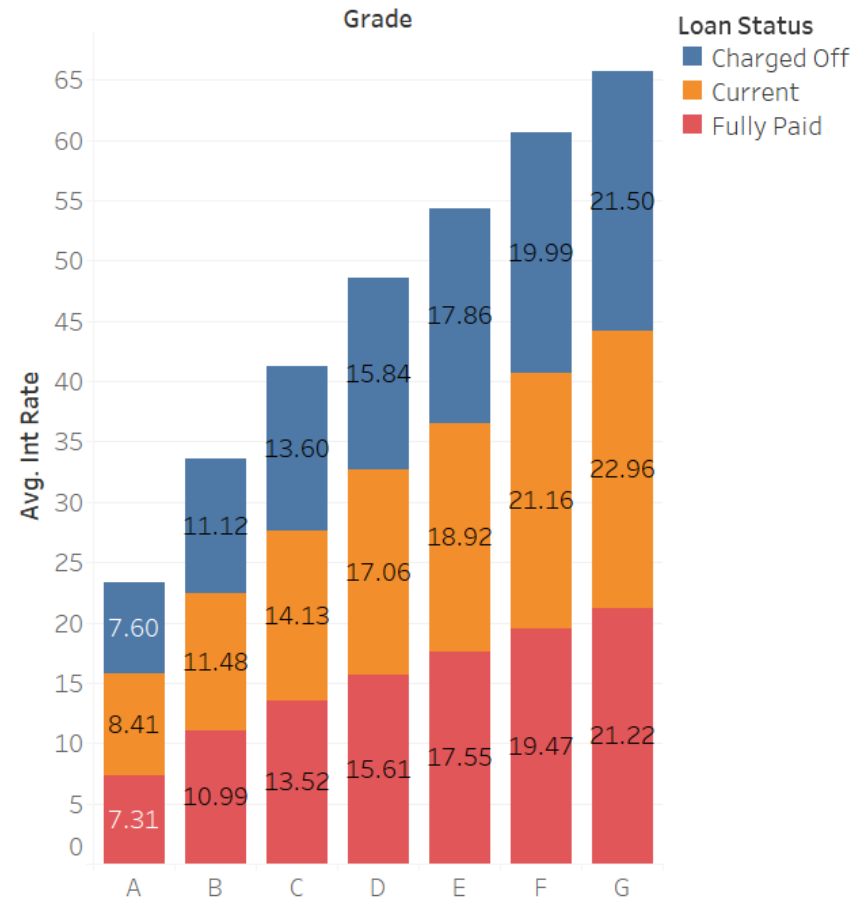LOAN STATUS WITH TERM TO EMPLOYEE LENGTH, HIGHER %
WITH 10+ YEAR EMPLOYEE LENGTH

Analysis REASONS FOR emp_length

LOAN STATUS WITH TERM TO EMPLOYEE LENGTH, HIGHER %
WITH 10+ YEAR EMPLOYEE LENGTH

Analysis Of Grades

It clearly indicates that People with Grade B have applied for maximum loans and also with highest rate who have cleared loan

Analysis Of Grades

This analysis can help in understanding the average interest rates at various grades. Also to identify the interest rate patterns

Conclusions

Let a person ask for loan at 7-8 years of experience and if he ask for 20000K
Which is ¼ of his salary in term of 36 months , let rate of interest be 15% then interest would be 3000 which is 5% of his salary per month
If this person has any check bounces or any current loan or with any mortgage then providing loan for this person would be risky

- Here Grades , Annual income , credit files are important to check for approval of loans

According to the data above bankers haven't verify properly ,verified customers tend to defaults so the bankers didn't verify .

- To approve any loans they need to check grades of customers ,any other pending loans from other banks ,any mortgages present
- bankers should check customers annual income and employee length if the time period is less than that means he may haven't saved any amount yet so providing loan my even Charged off

-> After examining annual income and grades we have found that all the grades B and C have almost similar and higher distribution.
So providing For A and G grades are better rather than B and C