LENDING CLUB CASE STUDY

SUBMISSION

Name: Kiran Joy

Introduction

Business Understanding

- You work for a consumer finance company which specialises in lending various types of loans to urban customers. 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:
  - 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

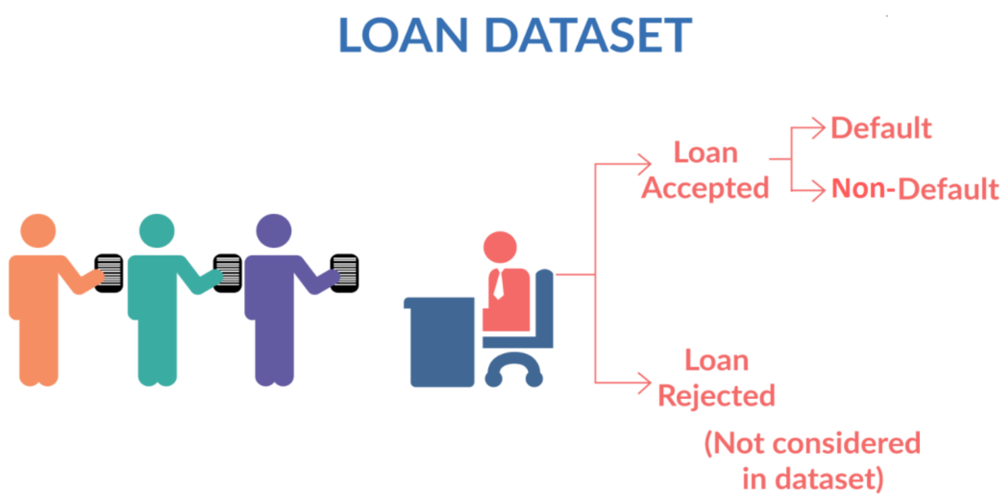
Business Objectives

- 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 utilize this knowledge for its portfolio and risk assessment.

Data Understanding

- Loan Data Set: This dataset contains the complete loan data for all loans issued through the time period 2007 t0 2011.
- Data Dictionary: This describes the meaning of these variables

Description of Data



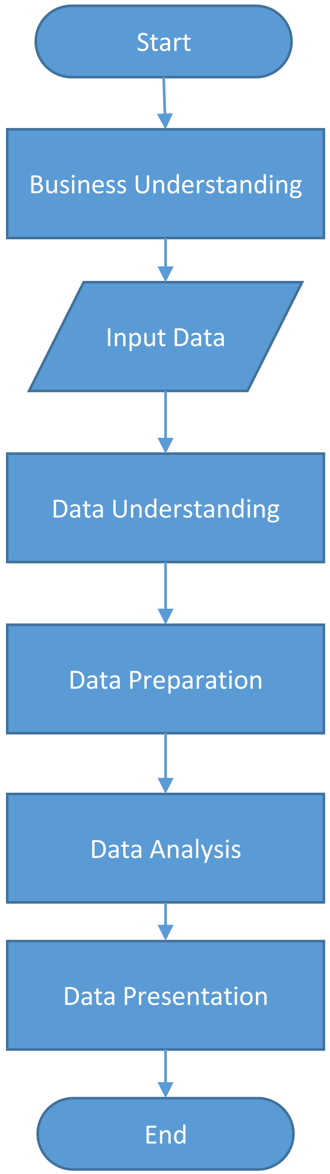
When a person applies for a loan, there are **two types of decisions** that could be taken by the company:

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.). 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)

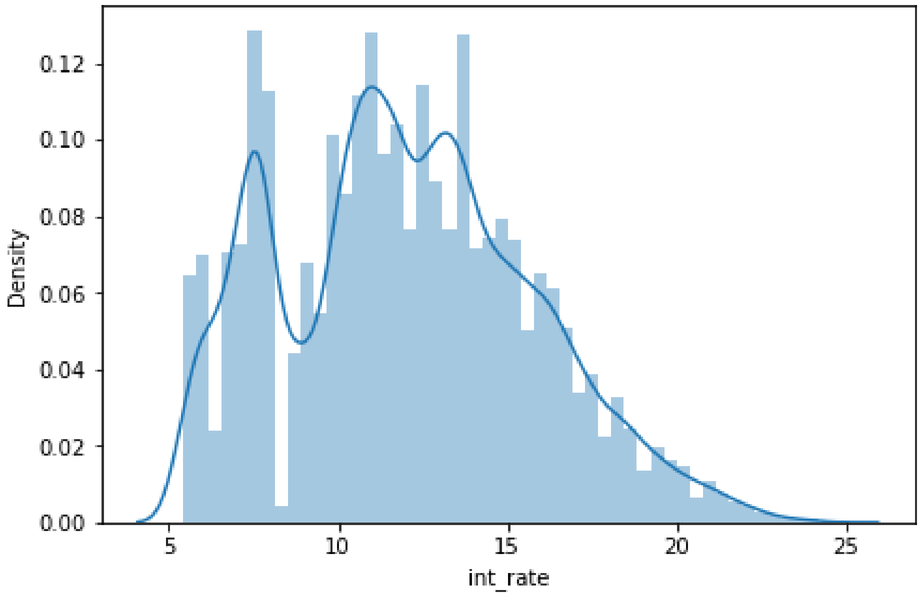
Problem solving methodology: CRISP-DM

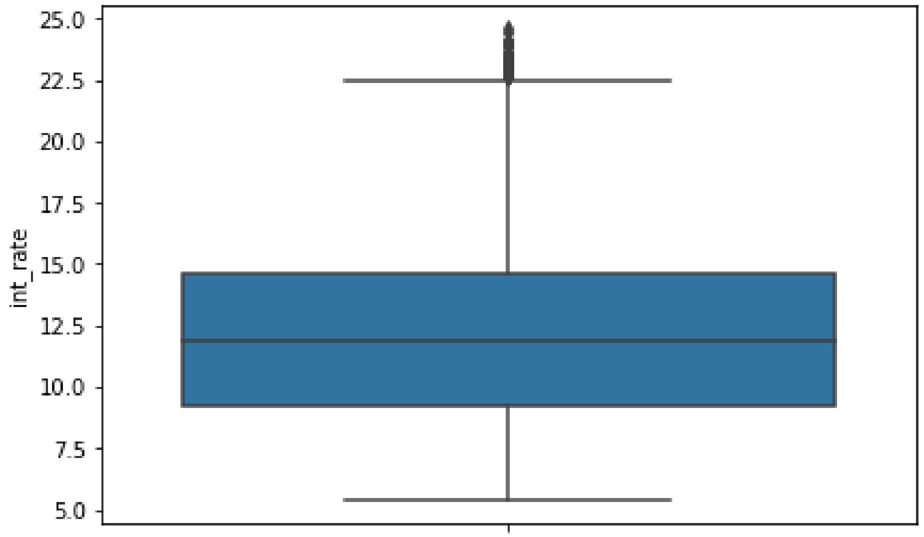


Data Cleaning

- Fix Duplicates
- Fix Missing Values
- Delete unwanted data
- Standardise values
- Filter data

Univariate Analysis: Interest Rate

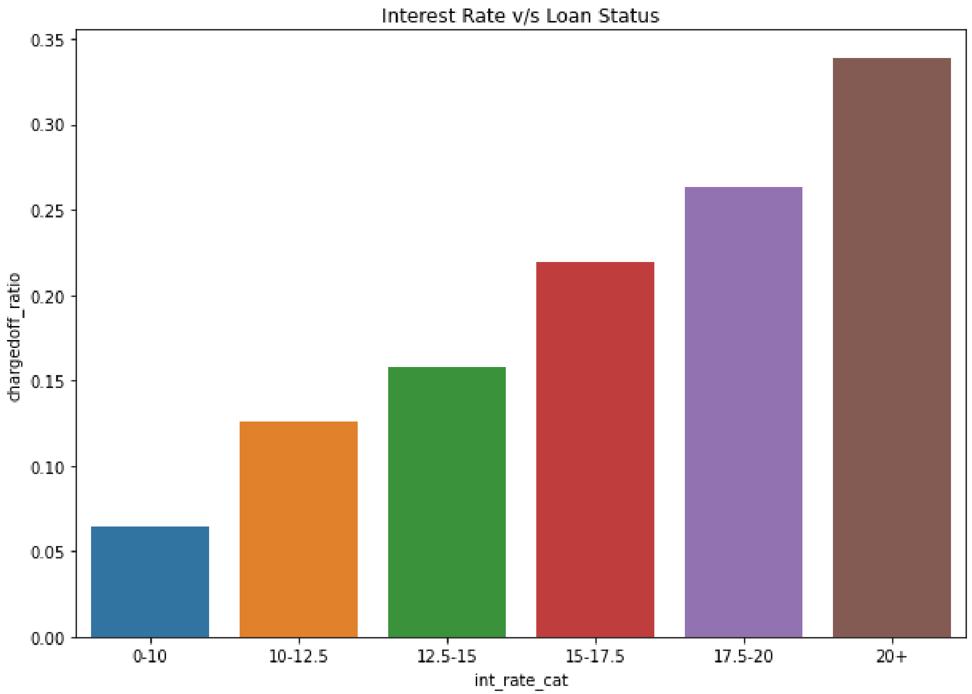




Observation:

- The above distribution plot shows that most of the interest rate are in the range of 9%-15%

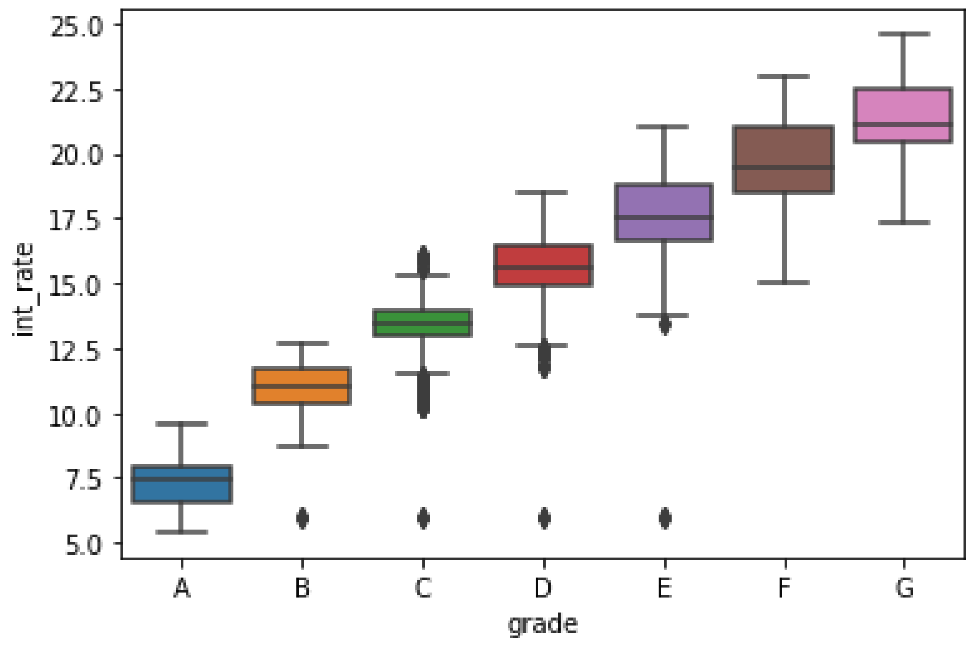
Bivariate Analysis: Interest Rate v/s Loan Status



Observation

- This clearly shows that the charged off rate increases with increase in interest rate. This means that people having loans with high interest rate are having more tendency of defaulting the loan.
- 34% of the loans are charged off incase of 20+ interest rate.
- Only 6% of the loas are charged off in case of less than 10% interest rate.

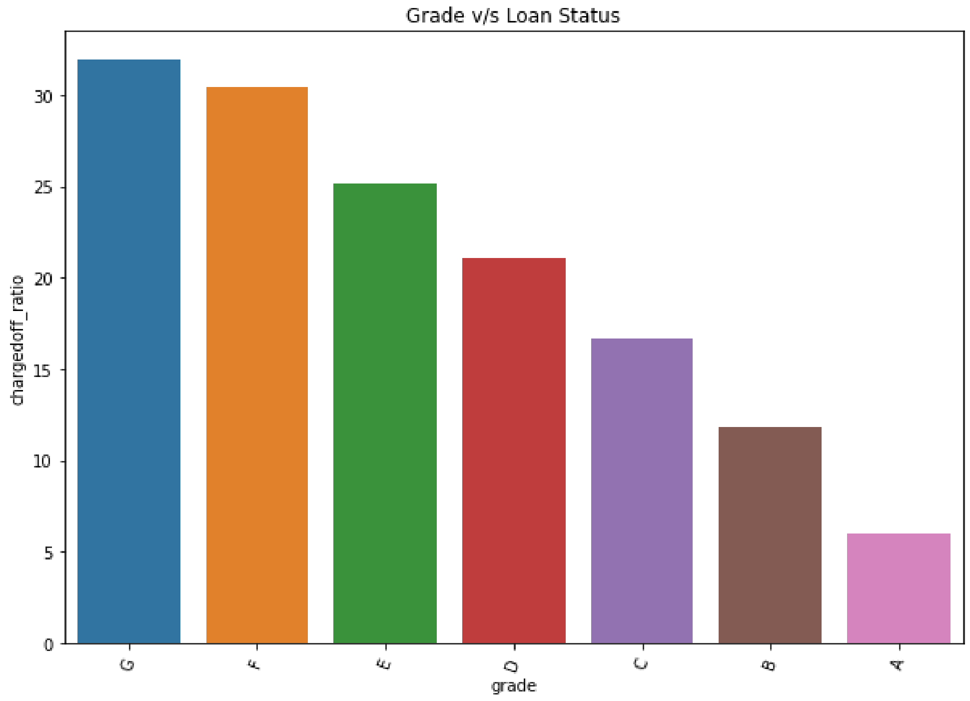
Bivariate Analysis: Interest Rate v/s Grade



Observation

- Interest rate is high for high risk borrowers. As per the information from lending club website, Lending Club's interest rates take into account credit risk and market conditions.
- The final interest rate for each loan grade is the result of the following equation:
  Lending Club Base Rate + Adjustment for Risk & Volatility
- Ref: https://www.lendingclub.com/foliofn/rateDetail.action

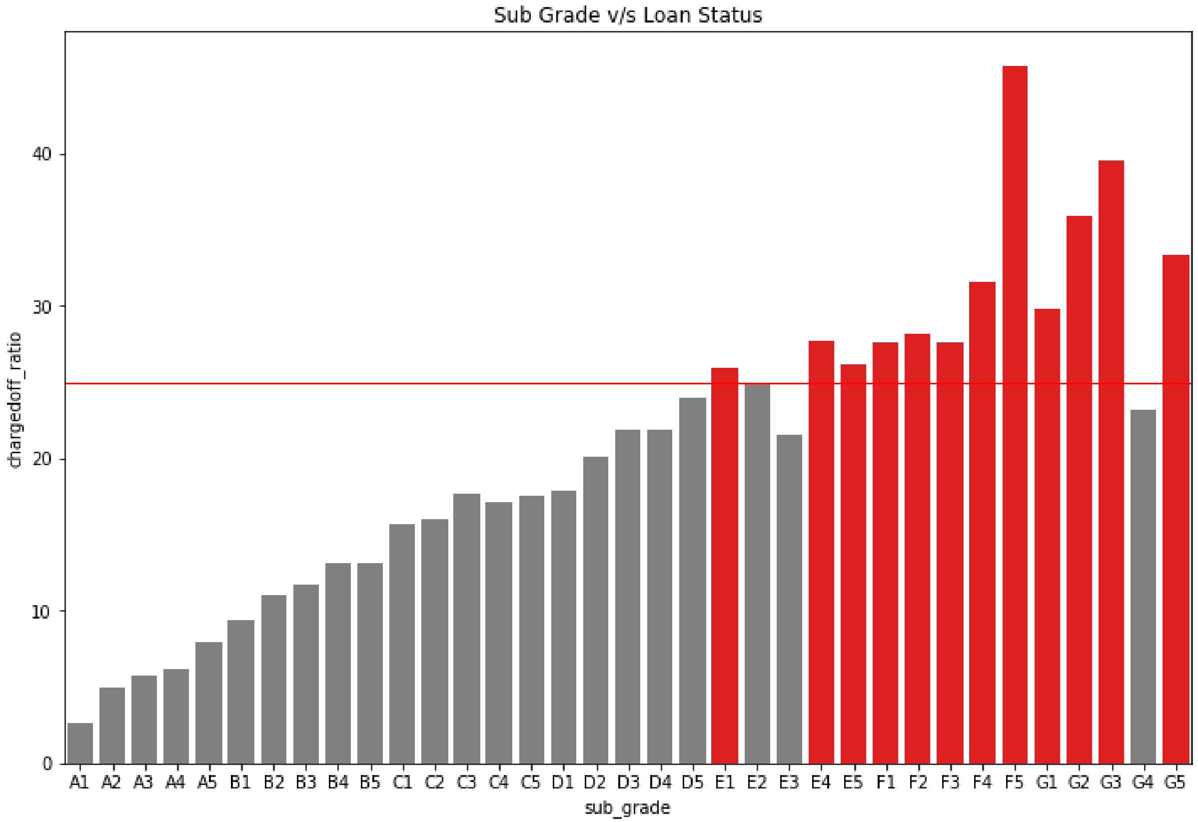
Bivariate Analysis: Grade v/s Loan Status

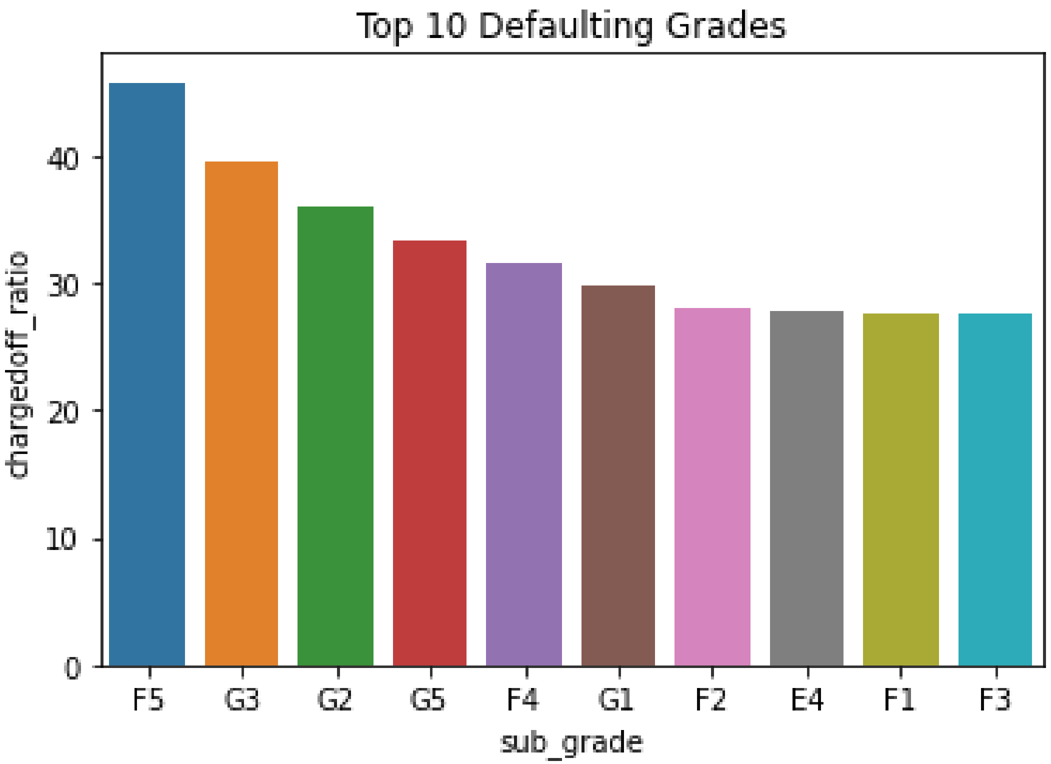


Observation

- We can see that high risk borrwers has a high tendency of defaulting the loan.
- It is better to use a higher interest rate for high-risk category to reduce the loan amount and to reduce the loss from this category.

Bivariate Analysis: Sub Grade v/s Loan Status

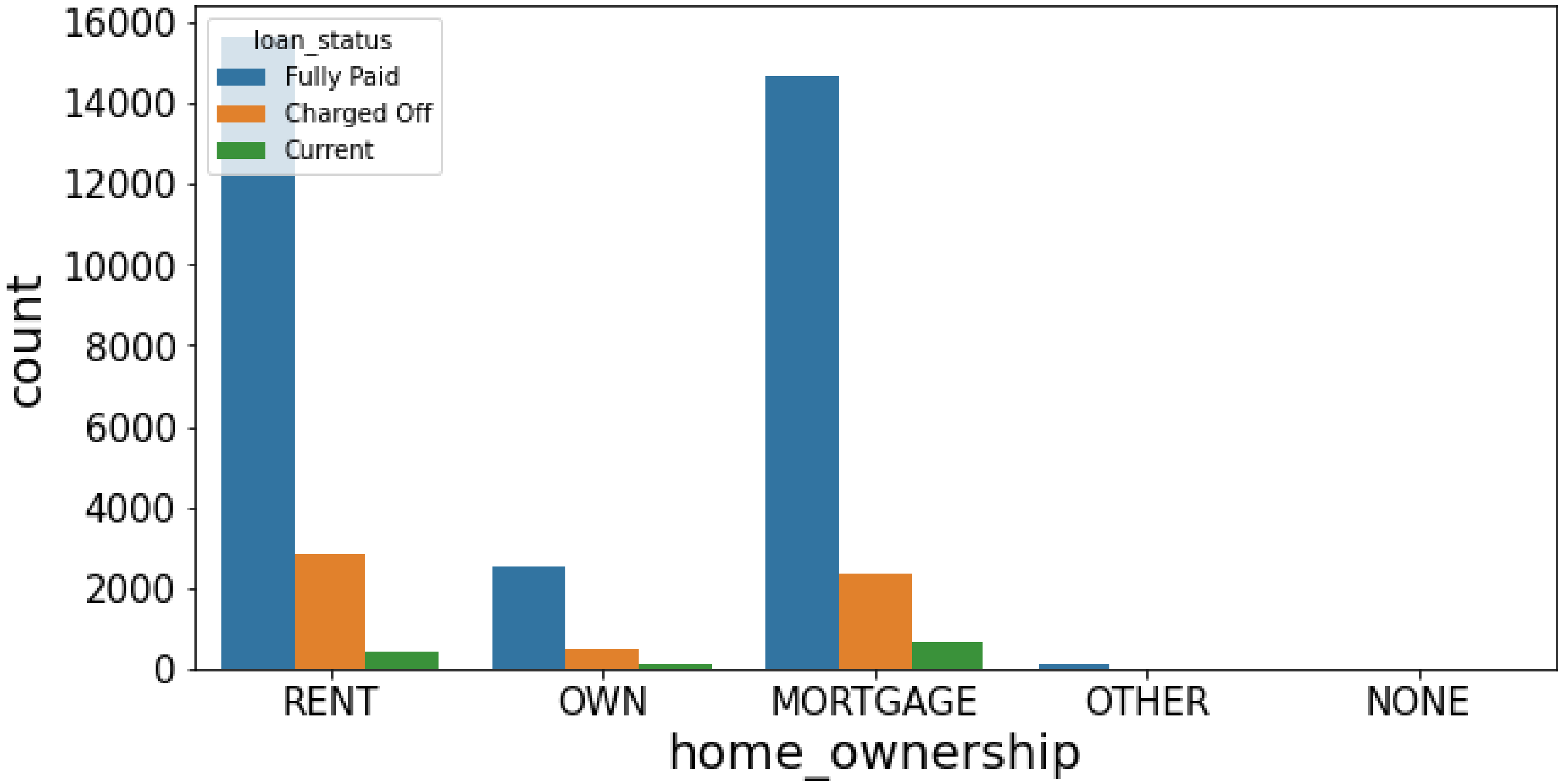




Observation

- F5 is the grade that has most charged off ratio. amlost 48% are defaulters. We should consider moving this to the highest grade/sub_grade and reduce the number of loans given to this subgrade.
- Borrowers in G4 seems to be doing good considering other subgrades of F and G. May be we can be more lenient towards this sub_grade.
- Company should be very strict when it comes to subgrades that has more than 25% default rate
- top 10 defaulting grade would be a good information for the company to make decisions while approving loans.
- Top 10 defaulters in the order or defaulting rate are, F5, G3, G2, G5, F4, , F2, E4, F1, F3

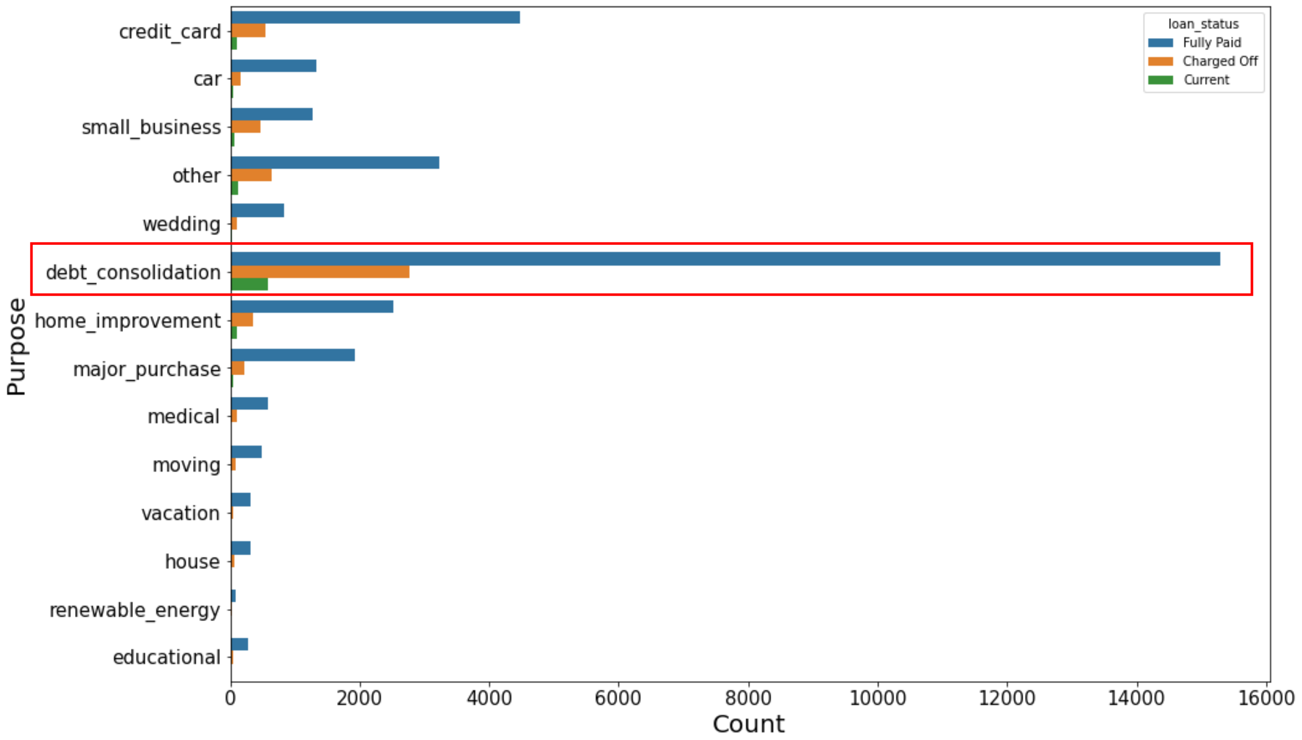
Bivariate Analysis: Home Ownership v/s Loan Status

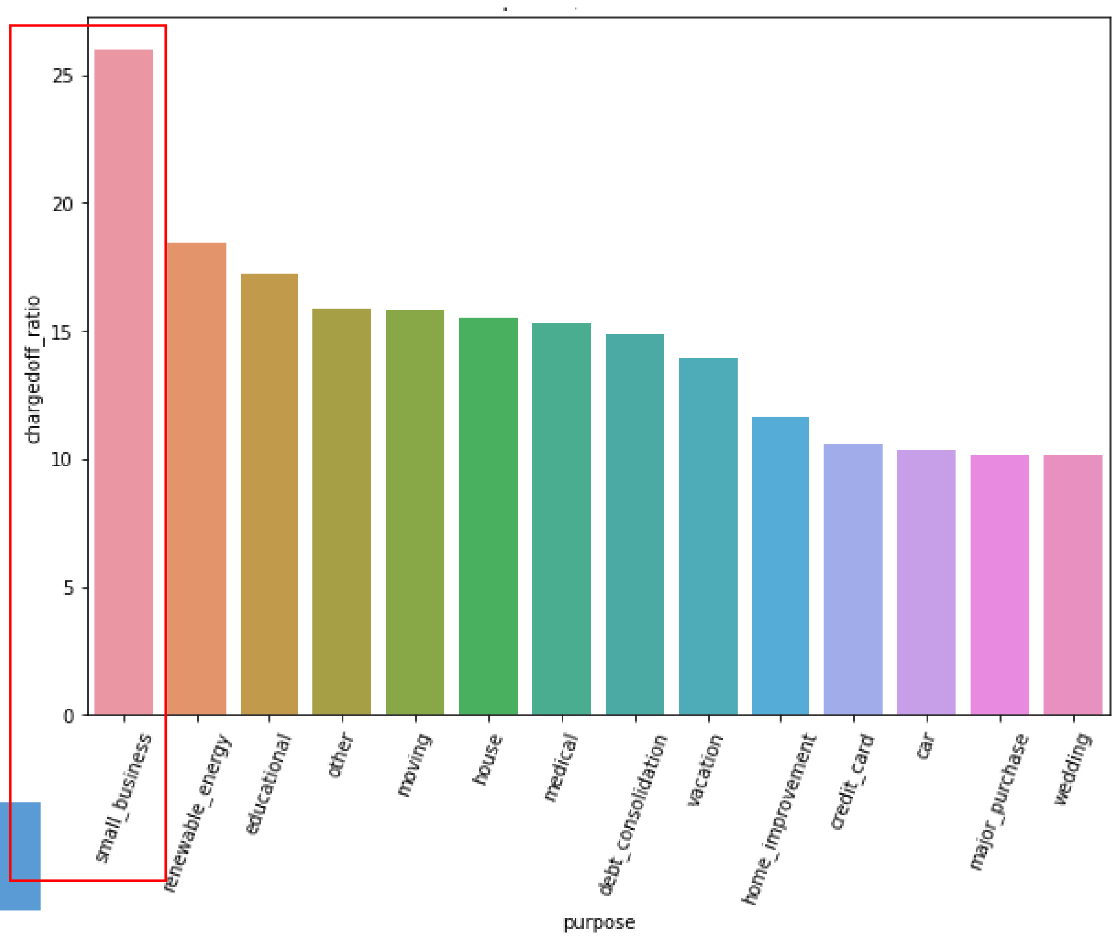


Observation

- Most of the borrowers are having rented or mortage homes.
- People who has own house are less in number comapred to rented mortage.

Bivariate Analysis: Purpose v/s Loan Status

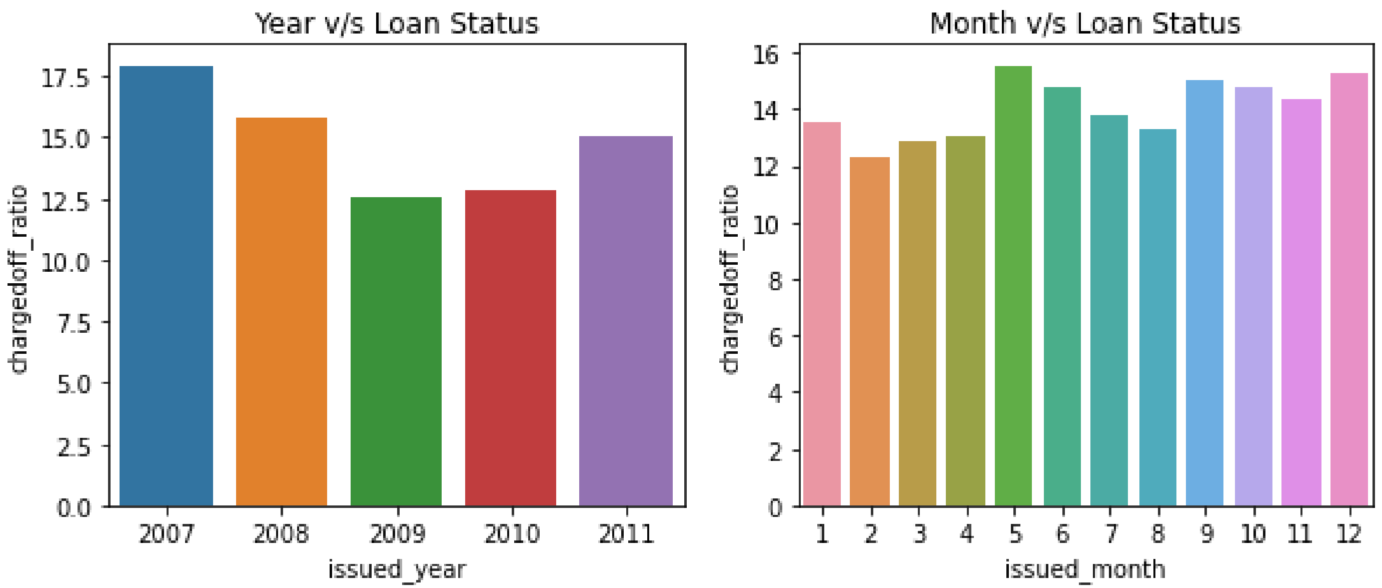




Observation

- most of the loans are taken for debt consolidation.
- most number of defaulters are also from this group.
- loans taken for the purpose of small business has a high ratio of defaulting. It is better to check the borrower's business skills while giving loan, whether he has any prior business experience or not. Was he successful in his previous business?

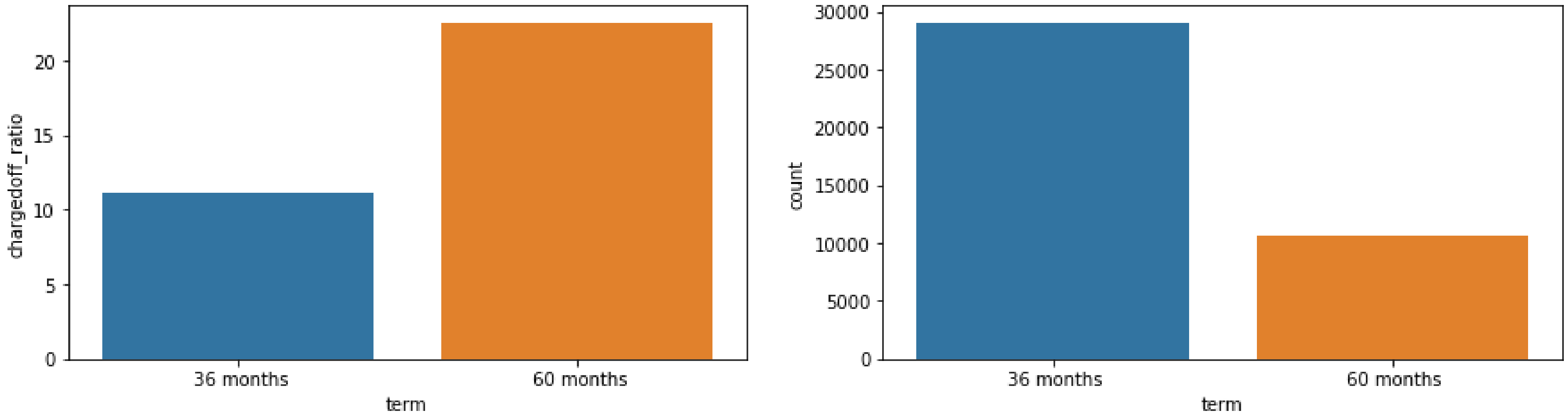
Bivariate Analysis: Year v/s Loan Status



Observation

- chargedoff_ratio was continuosuly reducing since 2007 until 2009 and then it has increased in 2011.
- People who take loan in the month of Feb has low chance of defaulting whereas the chance of defaulter are high in the month of May, Sept, Oct and Dec.

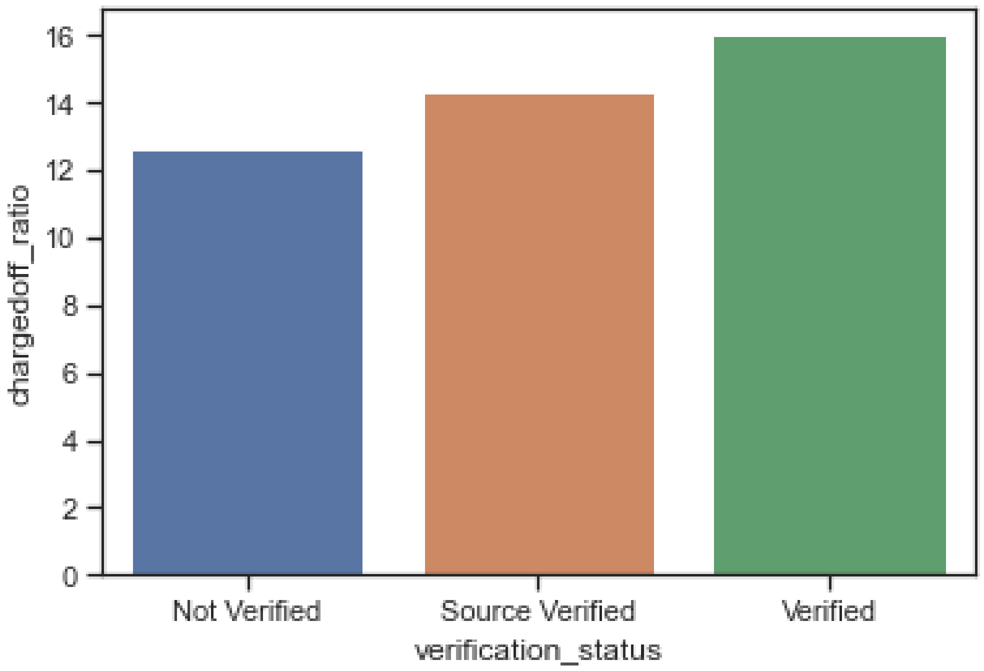
Bivariate Analysis: Term v/s Loan Status



Observation

- It is very evident that the number of defaulters are very high in 60 months term compared to 36months.
- And more than 25% of the total loans are 60 months term. Company has to be more careful when giving 60 months loan term.

Bivariate Analysis: Verification Status v/s Loan Status



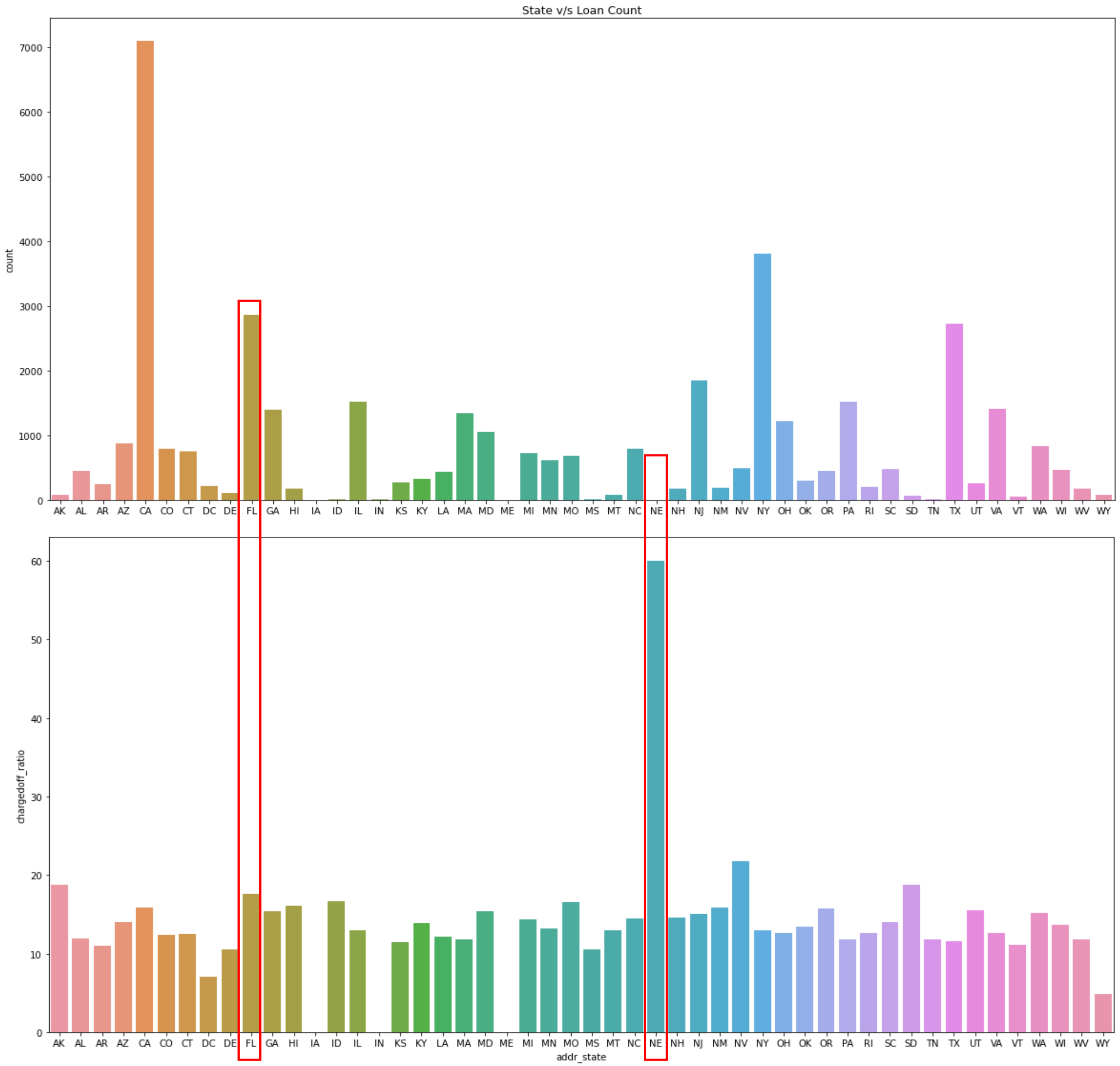
Observation

- Income verification has little to do with charged off ratio as verified income has higher charged off ratio compared to Not Verified

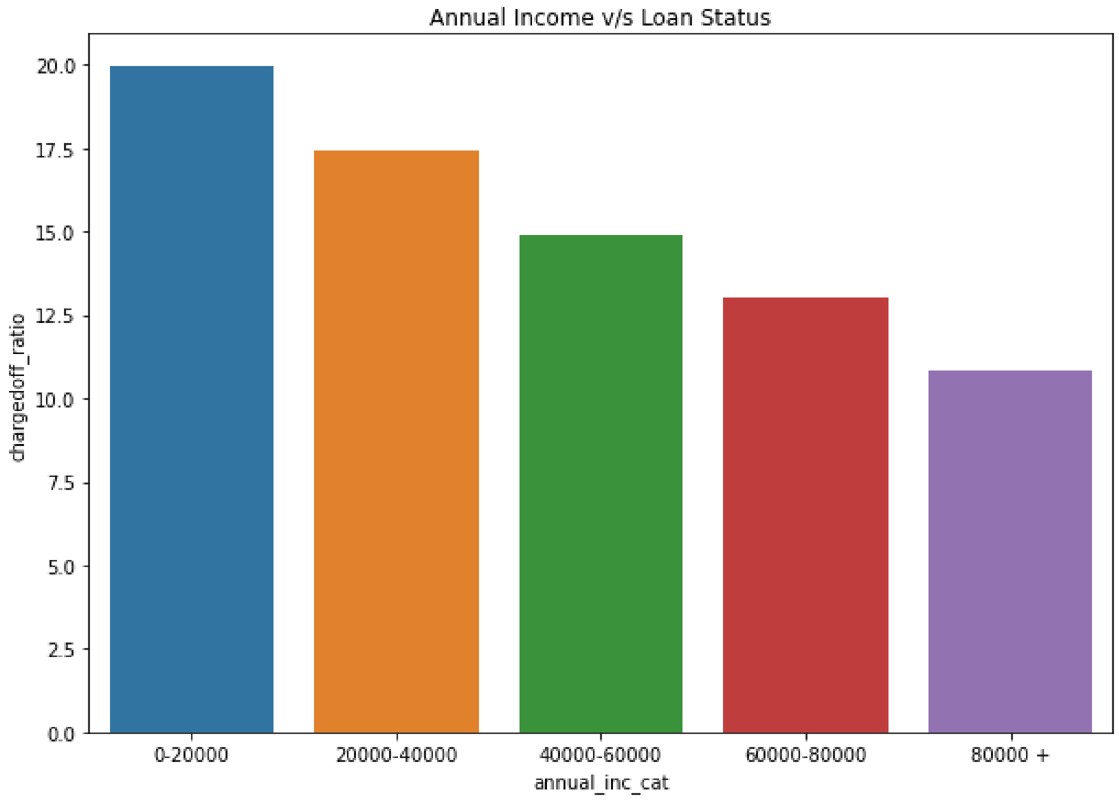
Bivariate Analysis: State v/s Loan Status

Observation

- CA, NY, TX, FL and NJ are the top 5 states in terms of number of loans and out of that FL is having the highest charged off rate. Company has to take extra caution when they give loans in these states as the charged off rate will directly affect the P&L since the numbers are more here.
- NE has the highest rate of charged off loans. However, the number of loans given in this state is only 5. So, this is not going to affect the overall P&L as of now. But need to implement strict guidelines in NE if company has plans to expand here in future.



Bivariate Analysis: Annual Income v/s Loan Status



Observation

- Charged off Ratio is high when the annual income of the borrower is low.
- The company must check the income of borrower more carefully to understand whether the borrower has a contant income to pay the loan without defaulting.

Correlation Matrix

	funded_amnt_inv	int_rate	emp_length	annual_inc	dti	open_acc	total_acc	issued_year	issued_month
funded_amnt_inv	1.00	0.31	0.15	0.25	0.07	0.16	0.24	0.26	0.08
int_rate	0.31	1.00	-0.01	0.05	0.11	0.01	-0.04	0.06	0.04
emp_length	0.15	-0.01	1.00	0.09	0.05	0.09	0.19	0.13	0.03
annual_inc	0.25	0.05	0.09	1.00	-0.12	0.16	0.24	0.01	0.01
dti	0.07	0.11	0.05	-0.12	1.00	0.29	0.23	0.10	0.02
open_acc	0.16	0.01	0.09	0.16	0.29	1.00	0.69	0.01	0.00
total_acc	0.24	-0.04	0.19	0.24	0.23	0.69	1.00	0.05	0.00
issued_year	0.26	0.06	0.13	0.01	0.10	0.01	0.05	1.00	-0.02
issued_month	0.08	0.04	0.03	0.01	0.02	0.00	0.00	-0.02	1.00

Observation

- We have completed the analysis of all variables that are having high correlation with loan status and funded amount.

Conclusions - 1

Based on the data provided, and the exploratory data analysis done on the data, below given are the observations and driving factors behind loan default.

1. People having loans with high interest rate are having more tendency of defaulting the loan
2. High risk borrowers (Low Grade) has a high tendency of defaulting the loan. This means that the Grading System is working as expected.
3. F5 is the grade that has most charged off ratio. amlost 48% are defaulters. We should consider moving this to the highest grade/sub_grade and reduce the number of loans given to this subgrade.
4. Borrowers in G4 seems to be doing good considering other subgrades under F and G. Maybe, we can be more lenient towards this sub_grade.
5. Top 10 defaulters in the order or defaulting rate are, F5, G3, G2, G5, F4, , F2, E4, F1, F3
6. It is better to keep the interest rate high for high-risk borrowers to reduce the loss, because most of the defaulters are from Low Grade with High interest rate.
7. Next Slide…

Conclusions - 2

Based on the data provided, and the exploratory data analysis done on the data, below given are the observations and driving factors behind loan default.

7. Most number of loans and number of defaulters are high when the purpose is debt_consolidation. However, loans taken for the purpose of small business has a high ratio of defaulting. It is better to check the borrower's business skills while giving loan, whether he has any prior business experience or not.
8. chargedoff_ratio was continuously reducing since 2007 until 2009 and then it has increased in 2011.
9. People who take loan in the month of Feb has low chance of defaulting whereas the chance of defaulter are high in the month of May, Sept, Oct and Dec.
10. It is very evident that the ratio of defaulters are very high in 60 months term compared to 36months. Company has to be more careful when giving 60 months loan term.
11. CA, NY, TX, FL and NJ are the top 5 states in terms of number of loans and out of that FL is having the highest charged off rate. Company has to take extra caution when they give loans in these states as the charged off rate will directly affect the P&L since the numbers are more here.
12. NE has the highest rate of charged off loans. However, the number of loans given in this state is only 5. So this is not going to affect the overall P&L. But need to implement strict guidelines in NE if company has plans to expand here in future.
13. Charged off Ratio is high when the annual income of the borrower is less. The company must ensure that the borrower is getting enough income to pay the loan.