THREE AND A HALF MEN

EDA CASE STUDY

LOAN DEFAULT

A look at what factors affect customer loan default

ANUPAM MAJHI
HARI NYSHADAM
LIJO GEORGE THOMAS
SNIGDH SHWETANK

INTRODUCTION

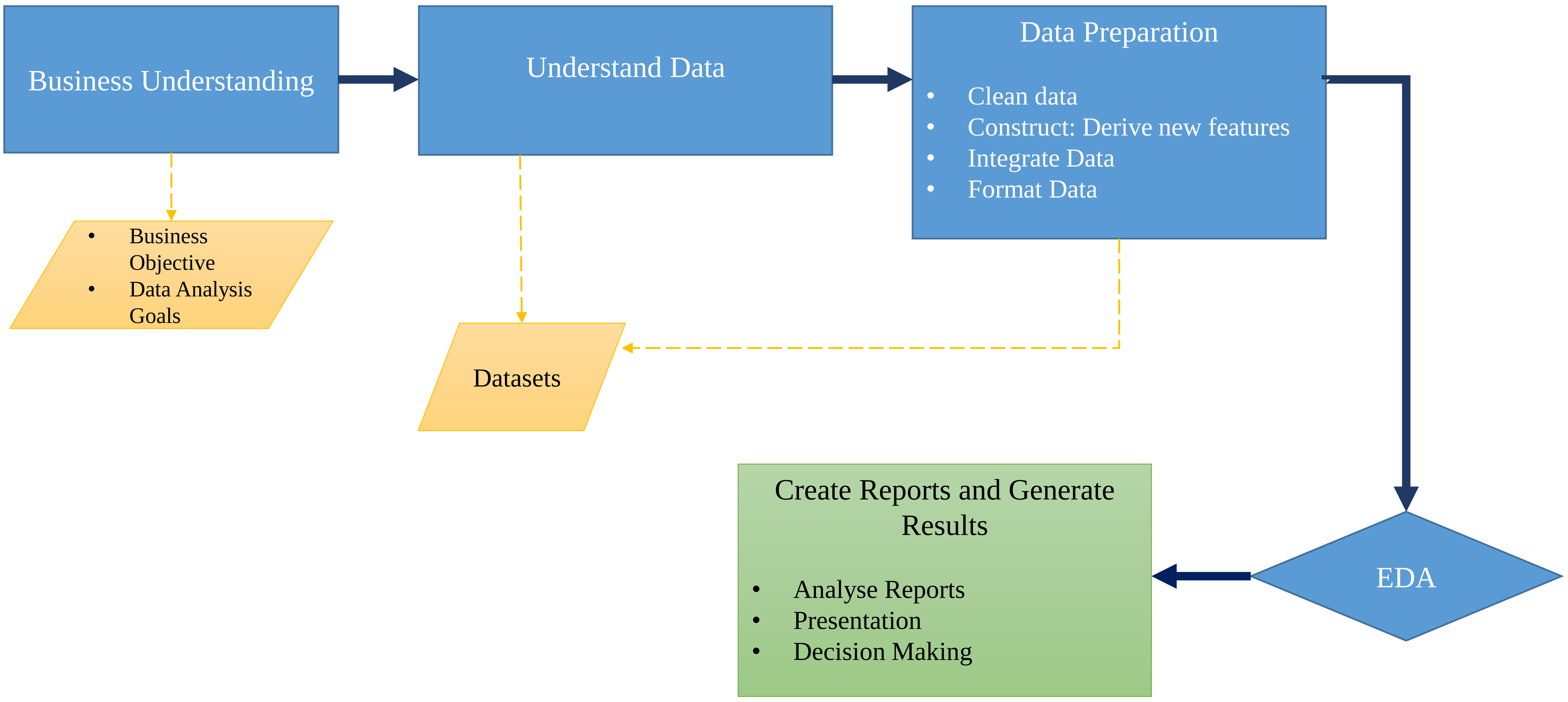
This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures

OBJECTIVES

Identify risky loan applicants, i.e, borrowers who default on their loan (labelled as "charged -off").

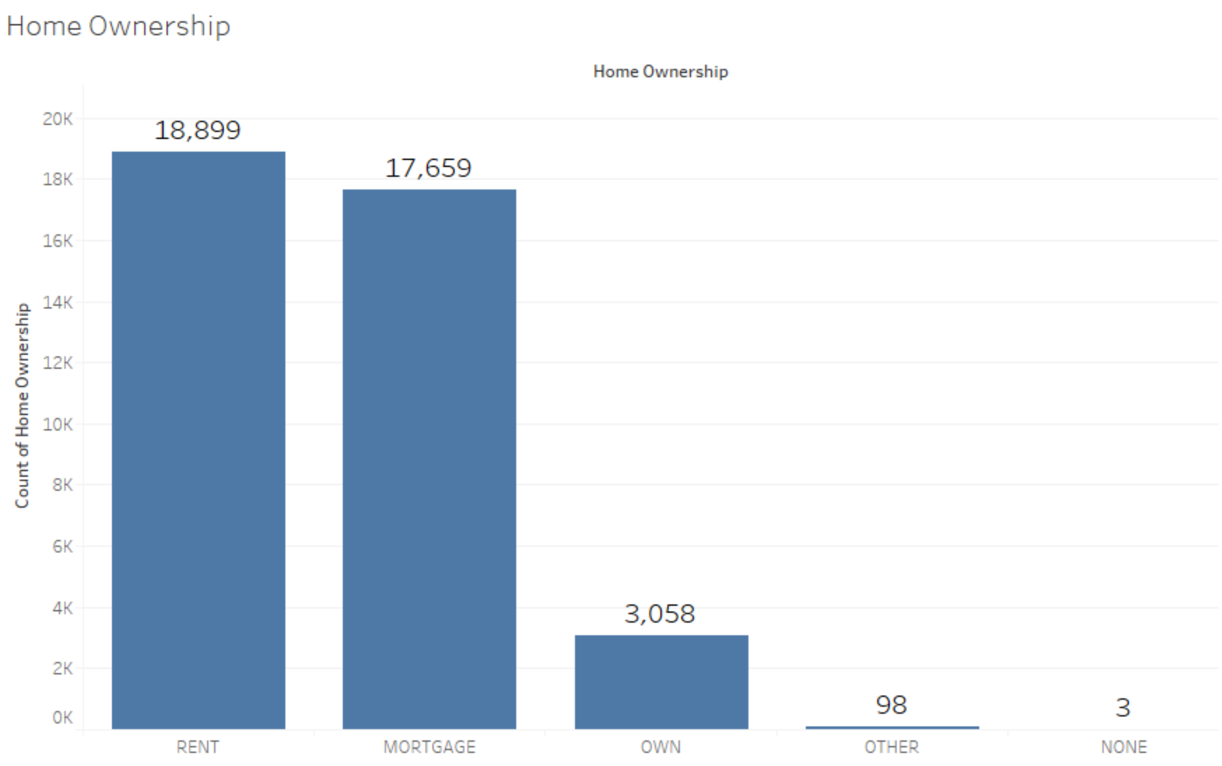
- Identify factors contributing to customer default
- Explain the results of univariate, bivariate analysis etc. in business terms
- Include visualisations and summarise the most important results in the presentation

PROBLEM SOLVING METHODOLOGY



UNIVARIATE ANALYSIS

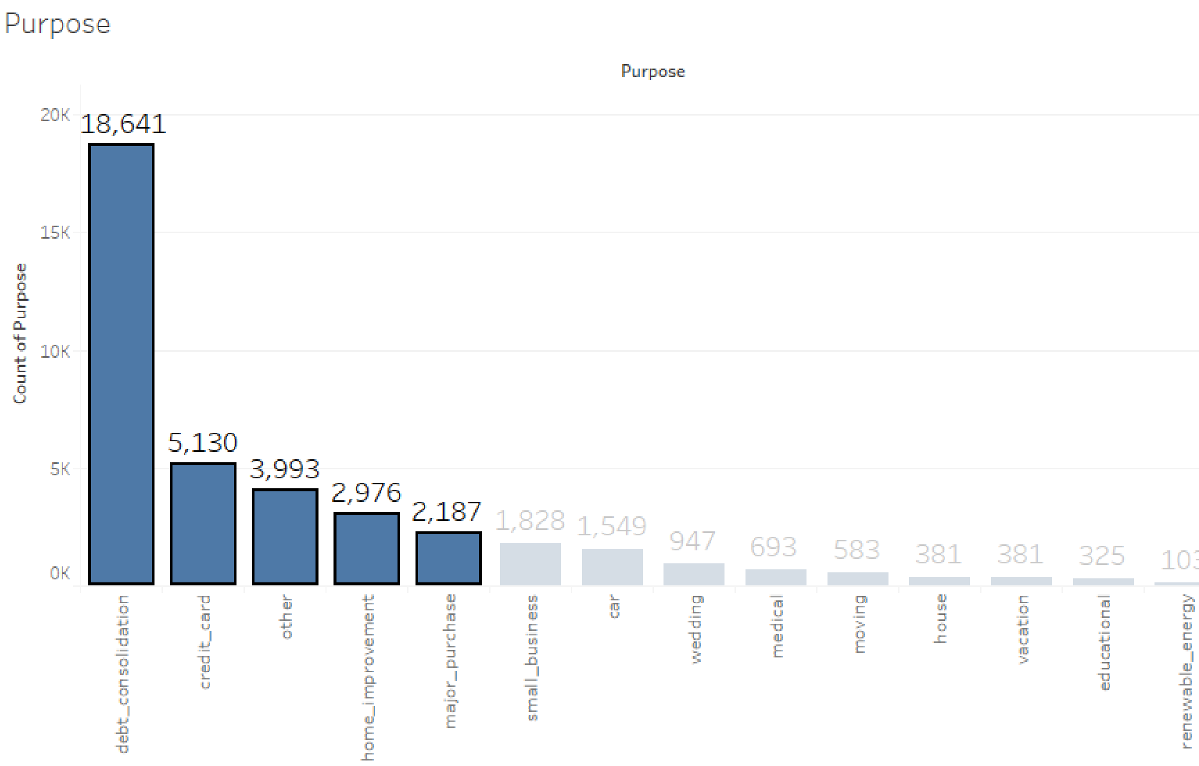
1. HOME OWNERSHIP



- The top two categories for Home Ownership among applicants is 'Rent' and 'Mortgage' .

- These two categories alone account for 92% of all loan applicants.

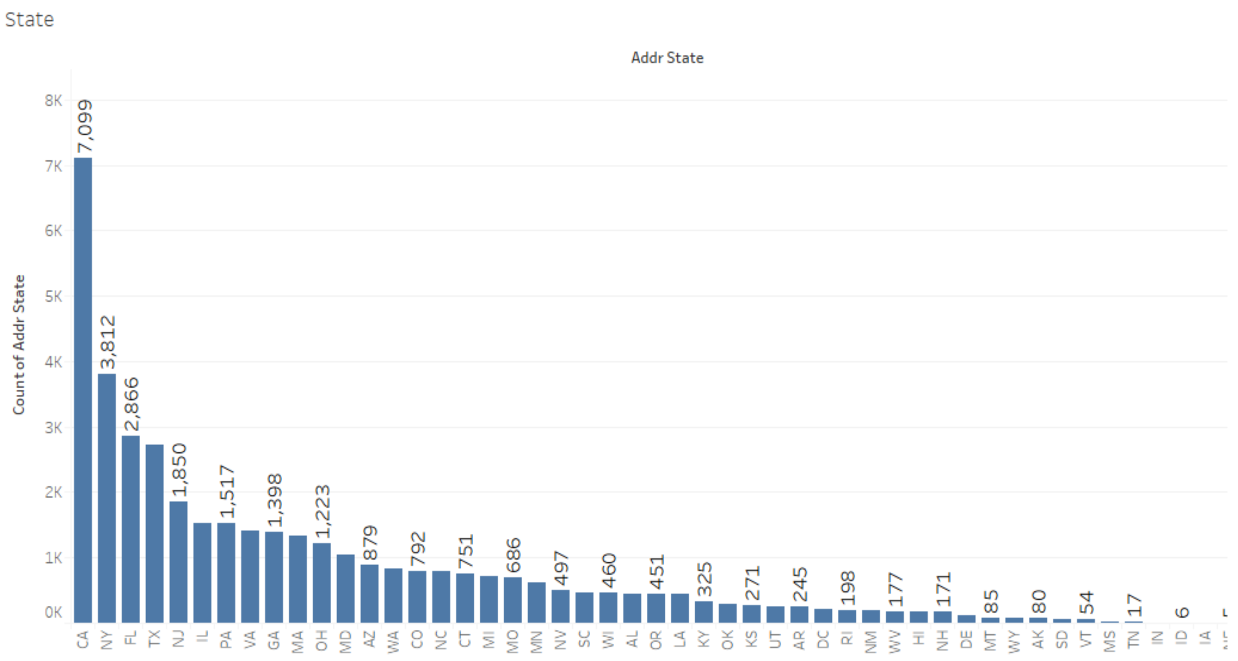
2. PURPOSE



- The top 5 categories for which loans are taken for are, 'Debt Consolidation', 'Credit Card', 'Other' ,'Home Improvement' and 'Major Purchases'.

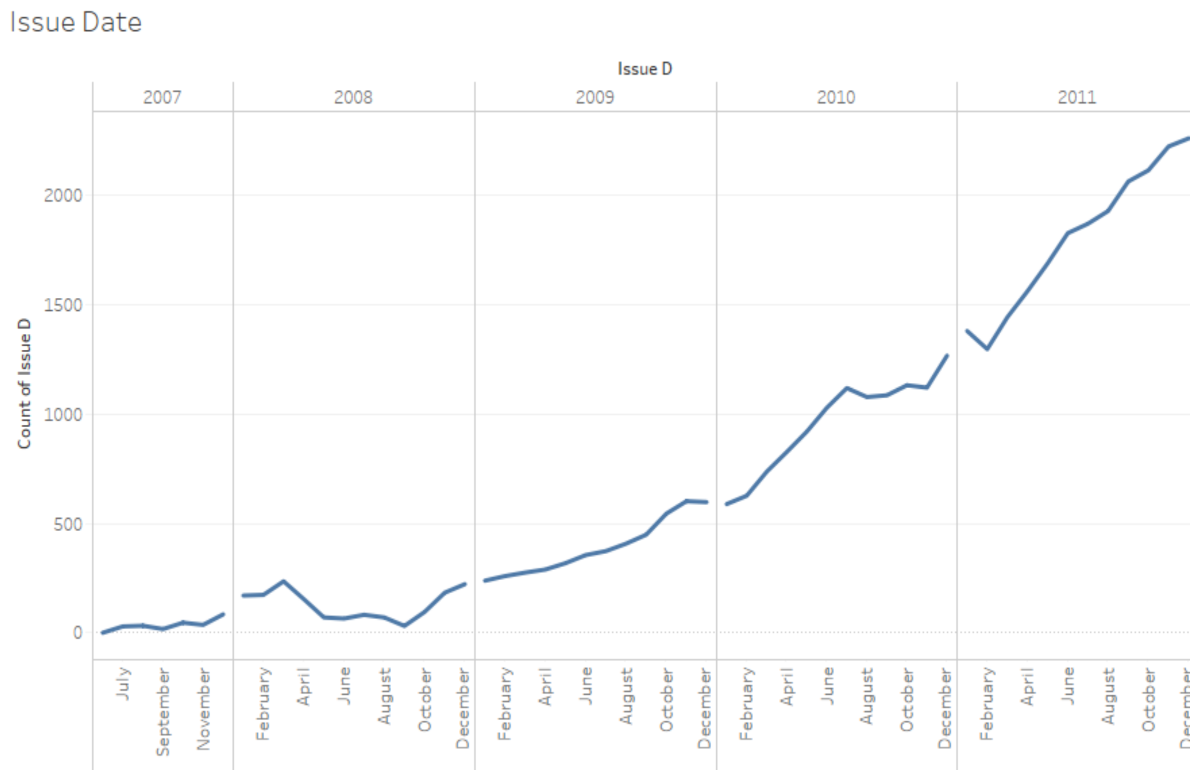
- These categories hold true for roughly 80% of all applicants

3. STATE



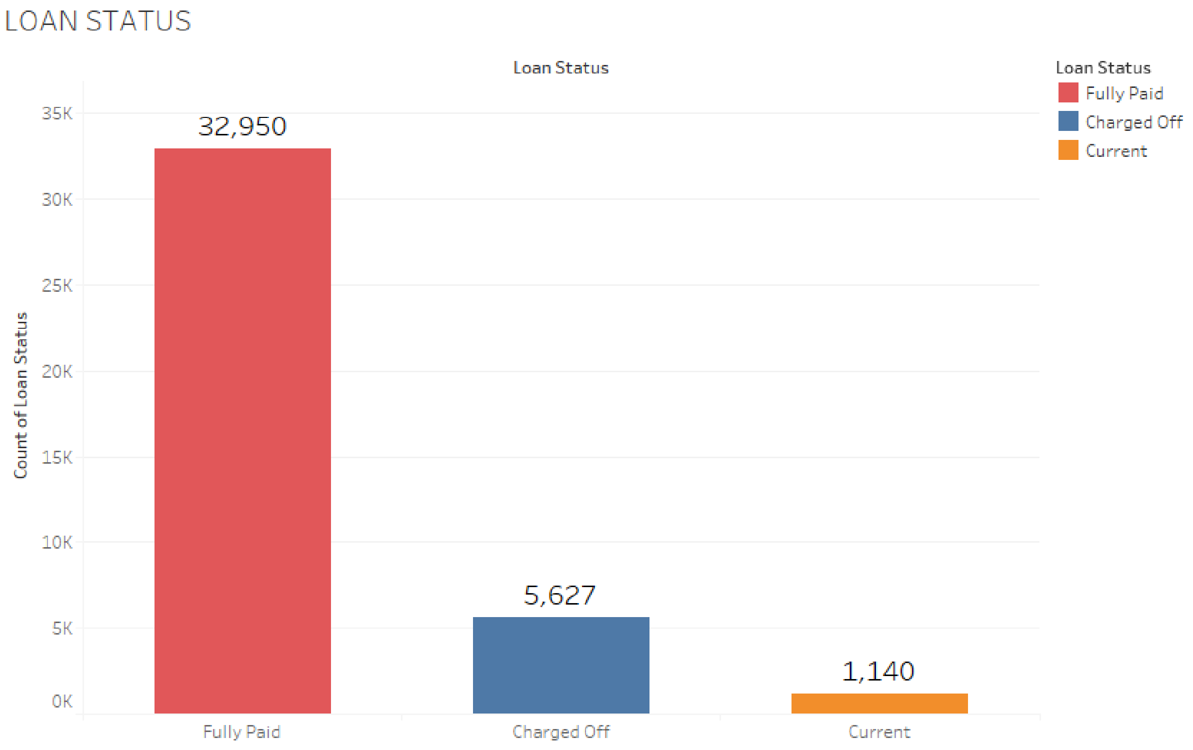
- The top 5 states which the applicants are from are , ‘CA’, ‘NY’, “FL’ , ‘TX’ and ‘NJ’.
- Roughly 44% of all applicants are from these states.

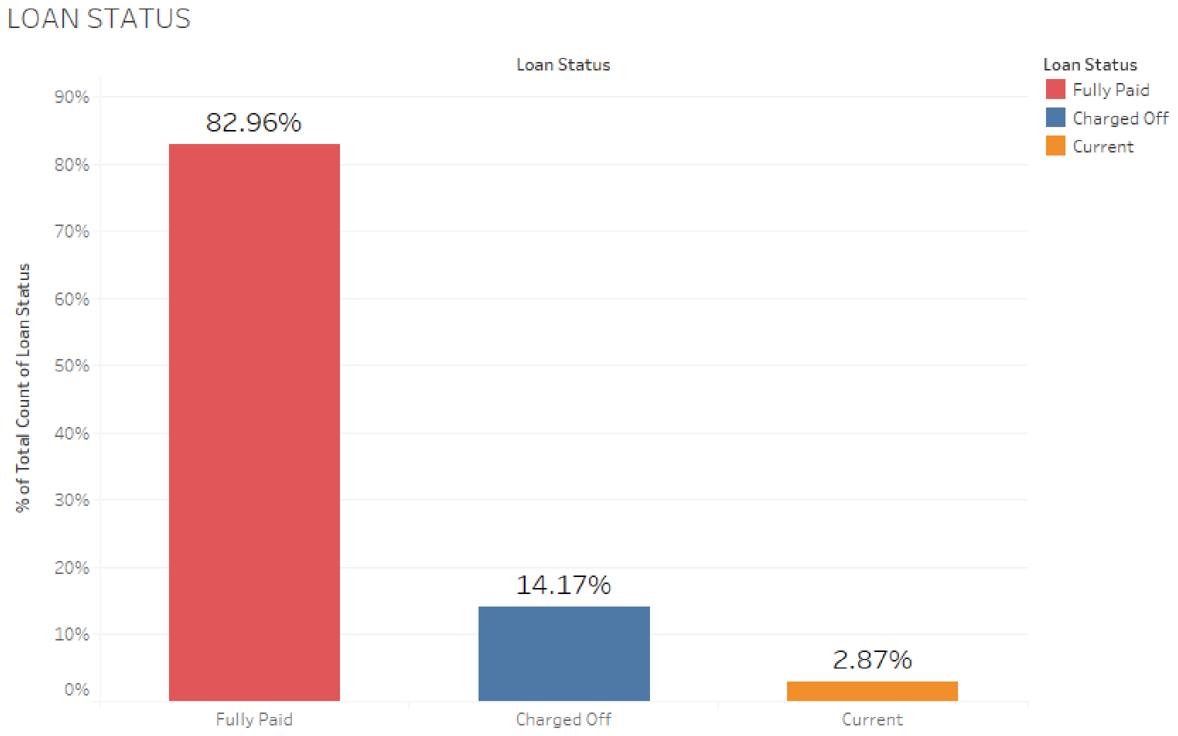
4. ISSUE DATE



- The number of loans taken steadily increases every year.
- The latest year provided , 2011, has the most loan applicants.

5. LOAN STATUS





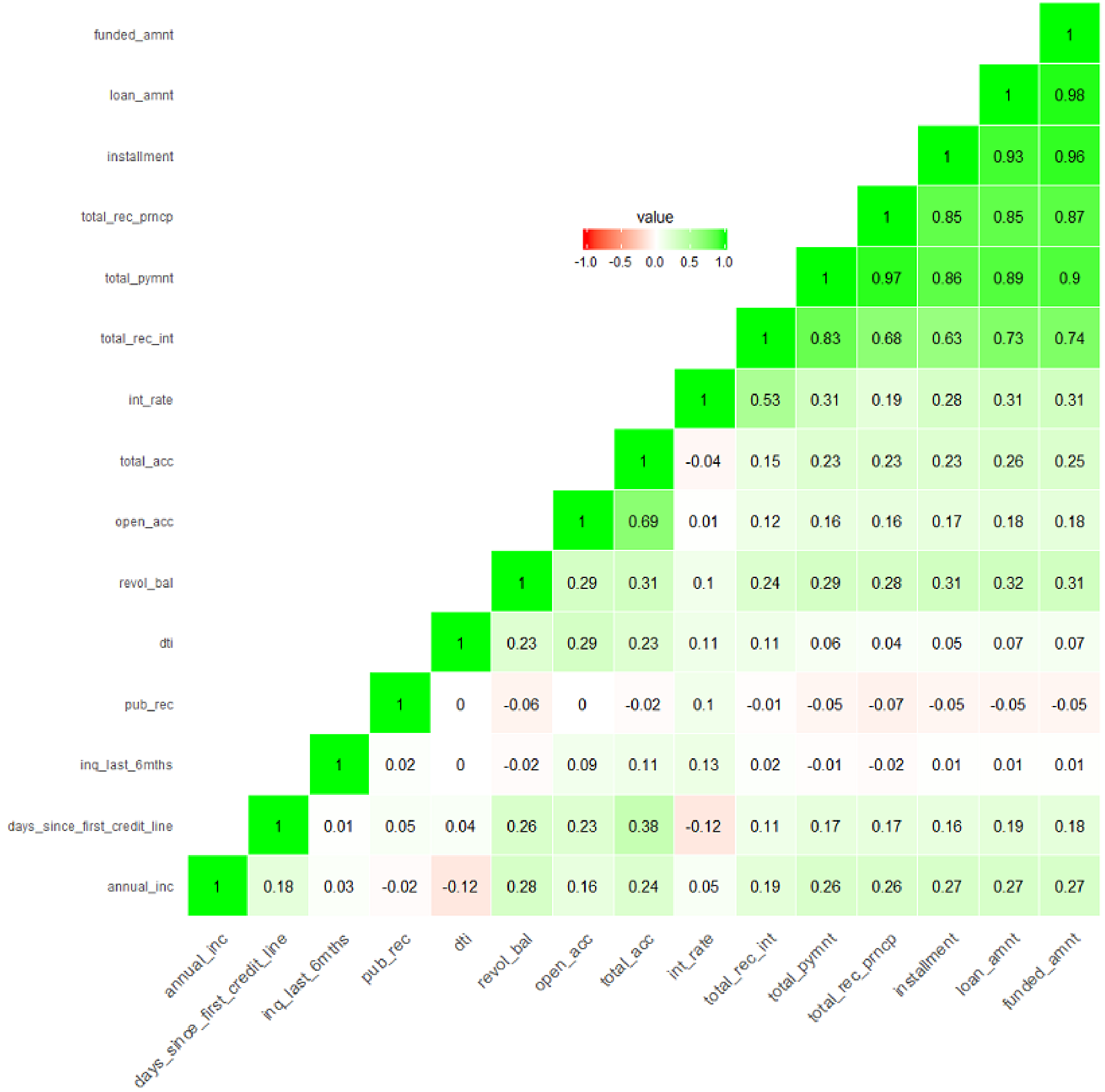
- The maximum number of loans provided , have been '*Fully Paid*'. These account for **32,950** loans, which is roughly **83%** of all loans.

- **5,627** loans have been ' *Charged Off*', which means that these applicants have defaulted. This is roughly **14%** of all loans.

- '*Current Loans*' are about **1,140** and roughly form only **3%** of all loans provided.

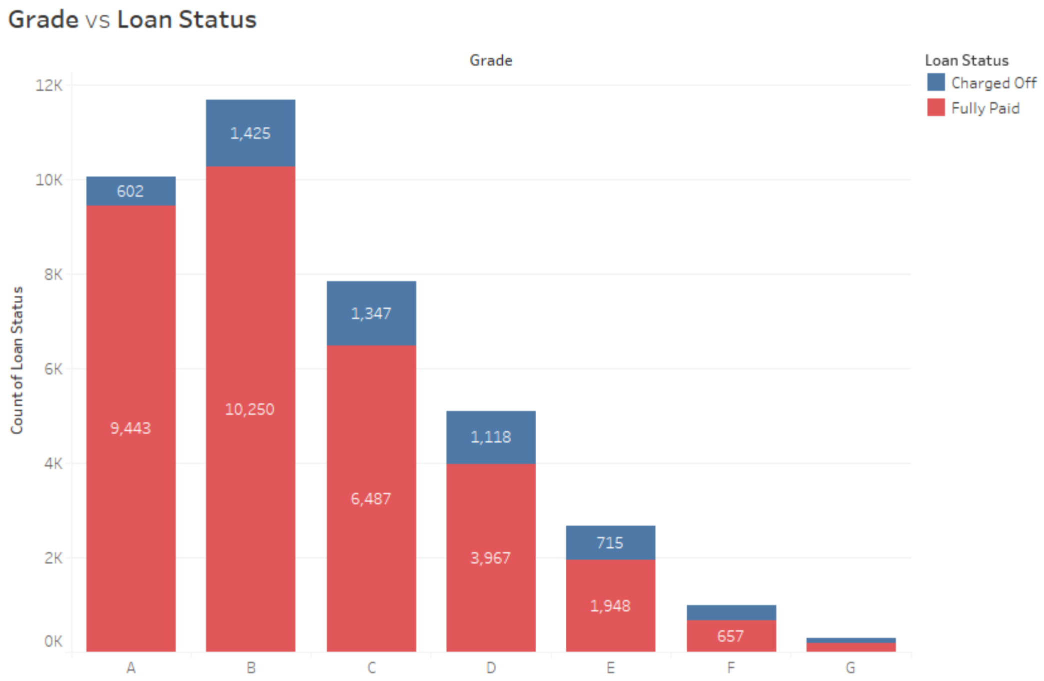
BIVARIATE ANALYSIS

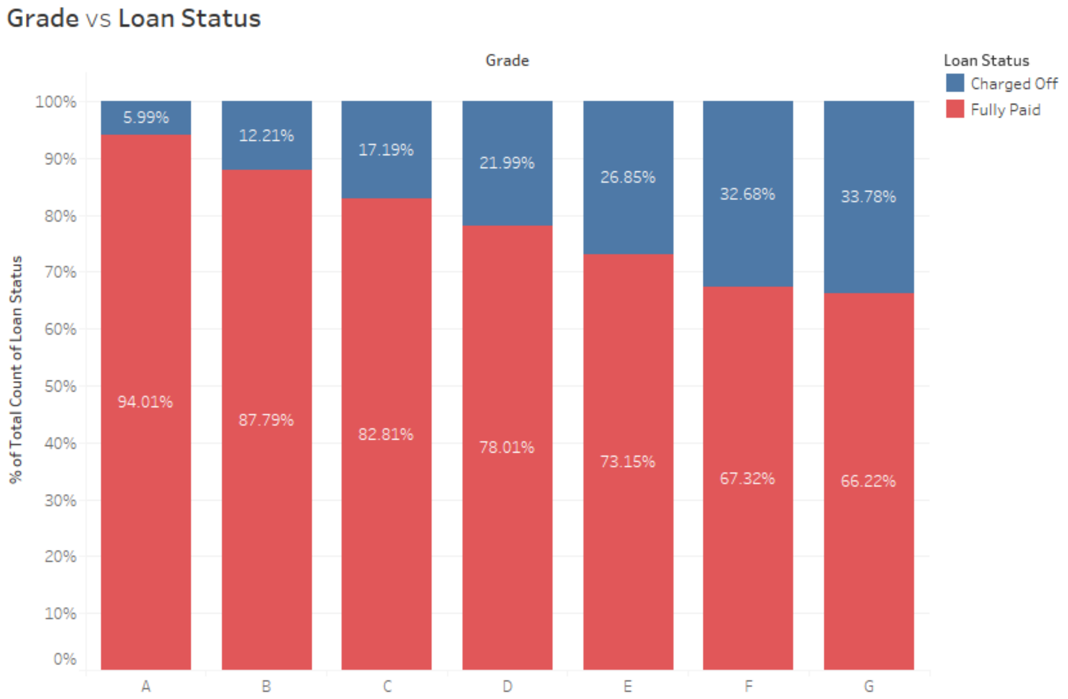
CORRELATION MATRIX



- There is a very strong positive correlation between the loan amount and funded amount because higher the loan amount higher will be the funded amount.
- A correlation is also a strong positive correlation between installment and funded amount which as the installments increase the total amount committed to the loan increases.
- So since total amount and funded amount, funded amounts and installments are highly correlated, it goes without saying that the installments and total amount are also strongly correlated.
- Total recurring principle is strongly positively correlated with installment, loan amount and funded amount which again makes logical sense.
- Total payment is strongly positively correlated with total recurring principle, installment, loan amount and funded amount.
- The total recurring interest is strongly positively correlated with total payment, funded amount, loan amount, total recurring principle and installment in the descending order of intensity.
- The interest rate is moderately positively correlated with the recurring with the total recurring interest, meaning as the interest goes up the total recurring interest also goes up but not by the same extent.
- The count open account is moderately positively correlated total accounts as some of the accounts might be closed but more the number of open account higher will the count of total accounts.
- The are also many variables that are very mildly correlated meaning as one variable increases the other variable decreases as it is evident in the corresponding plot.
- *<u>This correlational plot provides evidence for pairs of variables that are meant to related from a financial sense and also gives the strength of nature of their relationship that is whether positive or negative</u>*.

GRADE



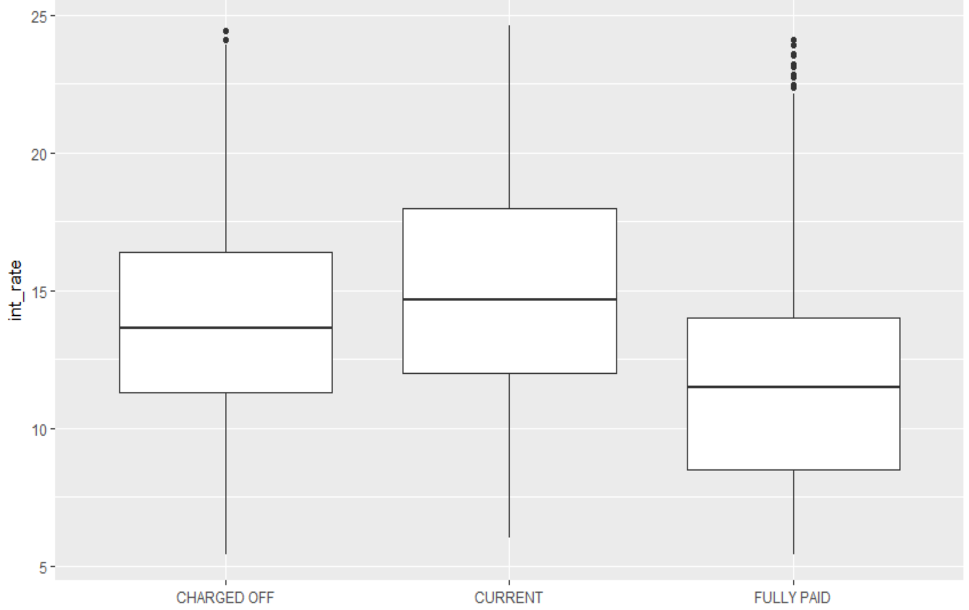


• The customer default rate increases as the interest rate becomes higher, i.e., the grade goes from A-G. Grades are assigned based on the interest rates. Higher Grades indicate a larger interest rate.

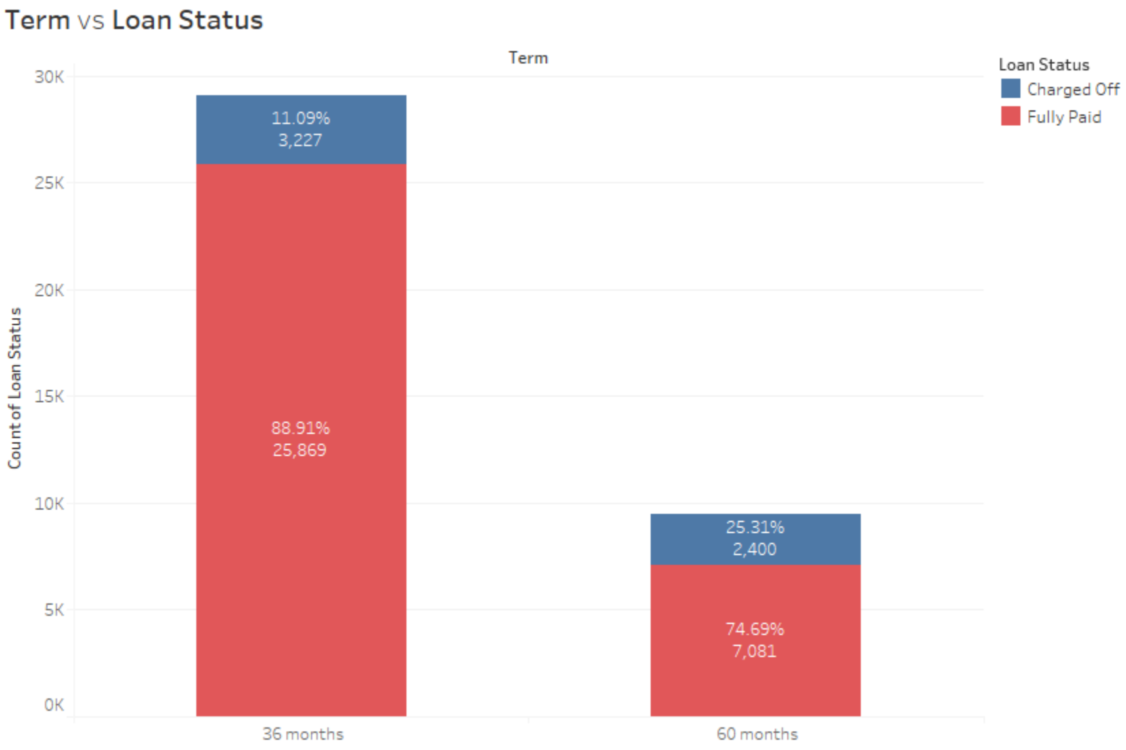
• While total customer default rate remains relatively low at **6%** of total population, at grade A, it increases steadily to a staggering **34%** of total borrowers by Grade G.

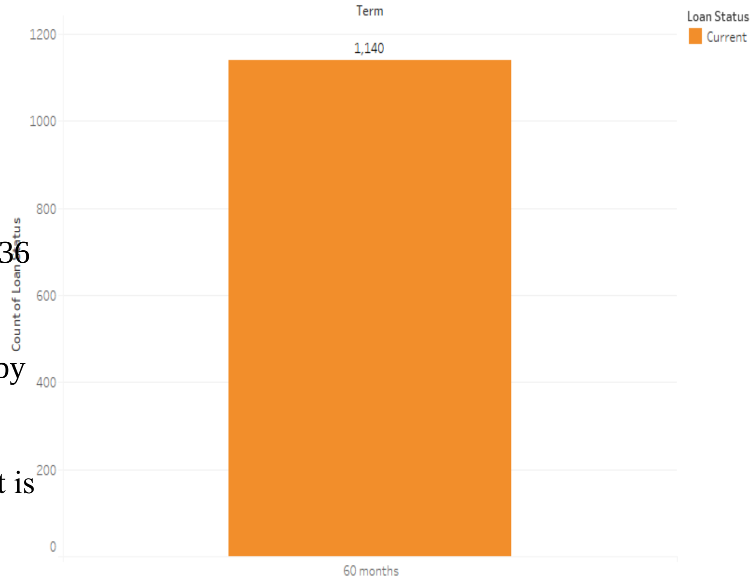
• High Interest rates are used to reduce loss to a bank, but , here in this case, it contributes directly to higher default.

•The Box Plot plotted , clearly verifies this pattern.



LOAN TERM





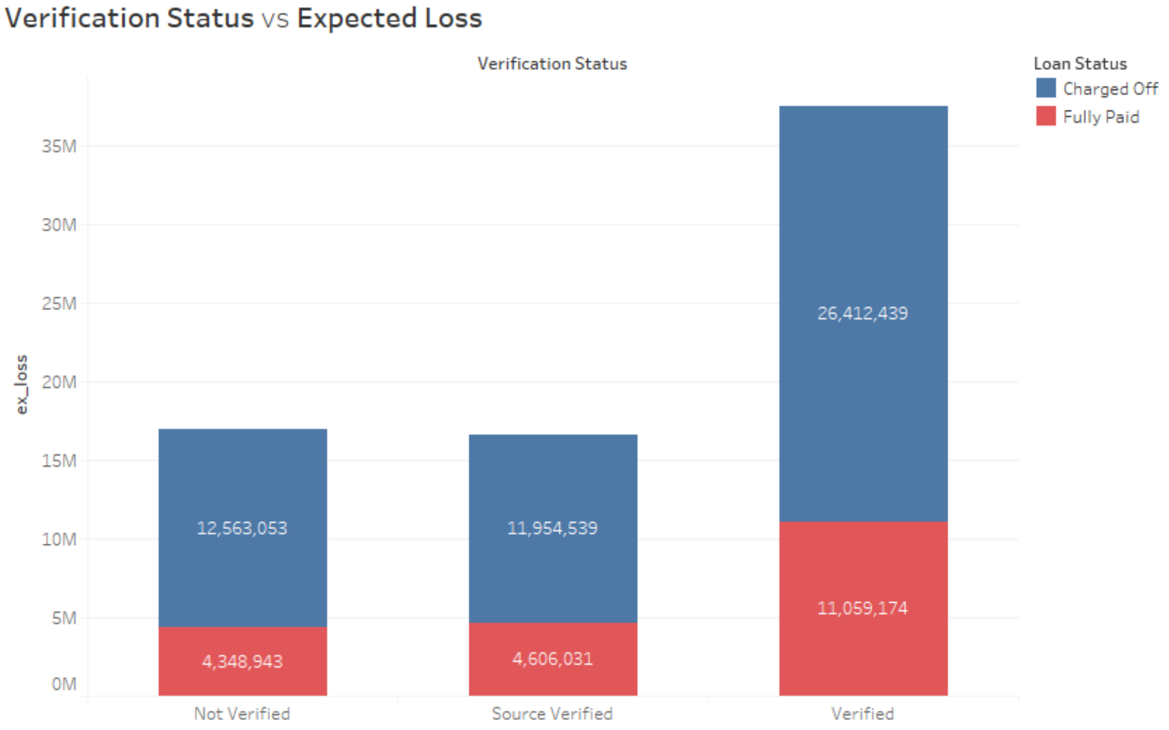
• The rate of default is higher for the higher term loans, i.e., the 60 month loans , as opposed to the 36 month loans.

• 60 month loans were introduced only in **May 2010** by the company, so the number of 60 month loans is much lesser than the 36 month loans, but, they display a higher relative rate of default.

• Expected loss from 60 month loss stands at **32** million, while expected loss for 36 month loans stand at **19** million. This thereby validates the above analysis

• Interestingly, all the 'Current' loans are 60 month loans. Since expected loss and default rate for 36 month loans is lower, it is encouraged to promote 36 month loan further.

VERIFICATION STATUS

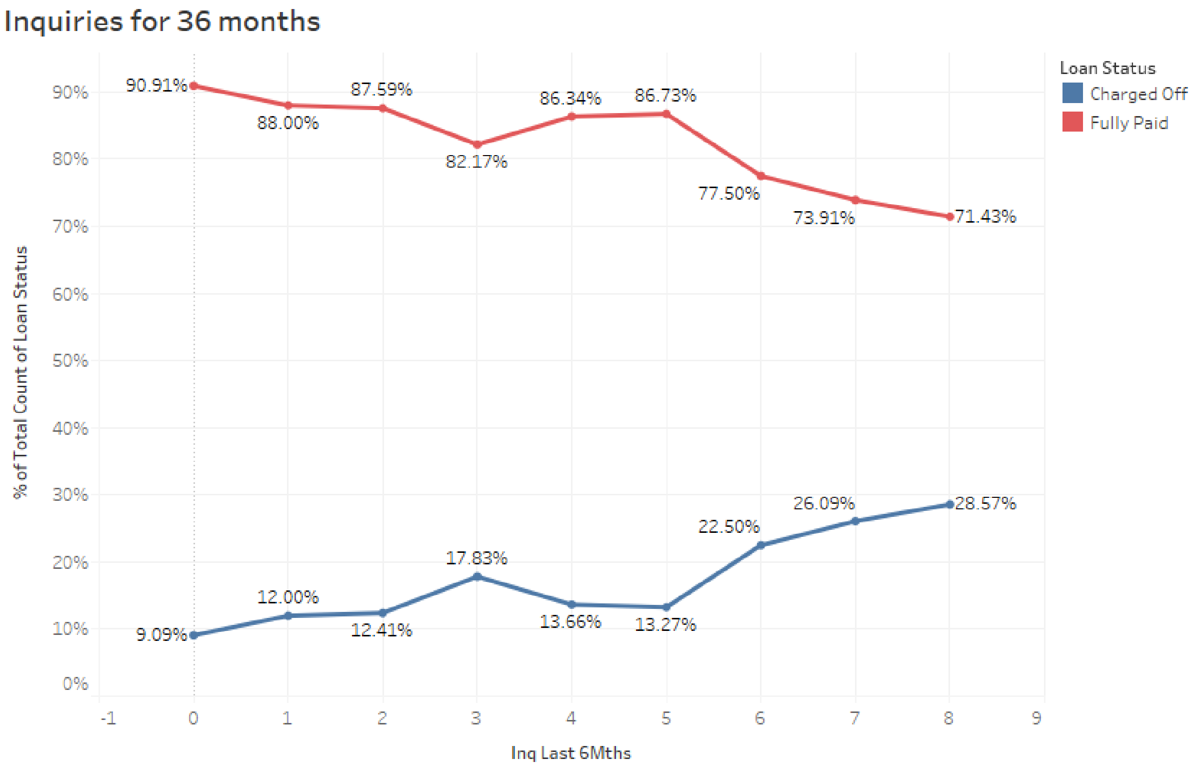


• Surprisingly, the rate of customer loan default is higher when the customer's income has been verified by LC.

• While this might not seem like a significantly higher rate, the expected loss for the three verification statuses show the problem. The expected loss for a 'Verified' loan is roughly **26** million, which is almost equal to the combined sum of the other two.

• This indicates that there is a need to improve the verification process. The best way is to verify the Source before providing the loan.

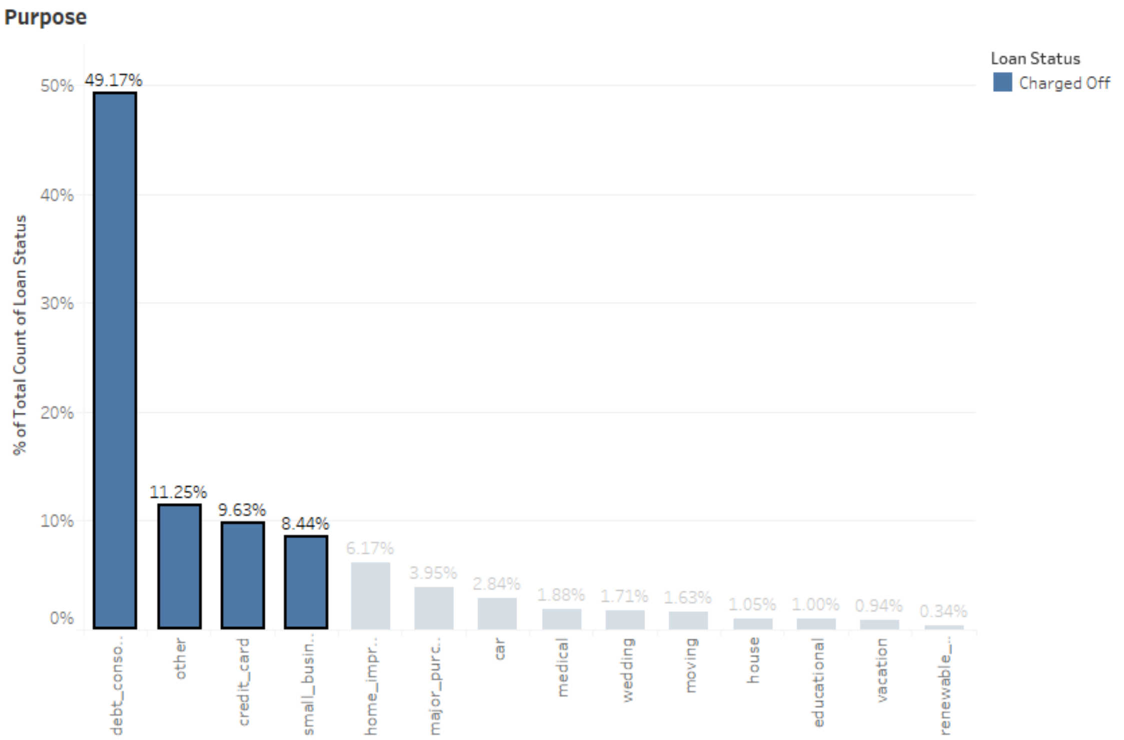
INQUIRIES in the last 6 Months

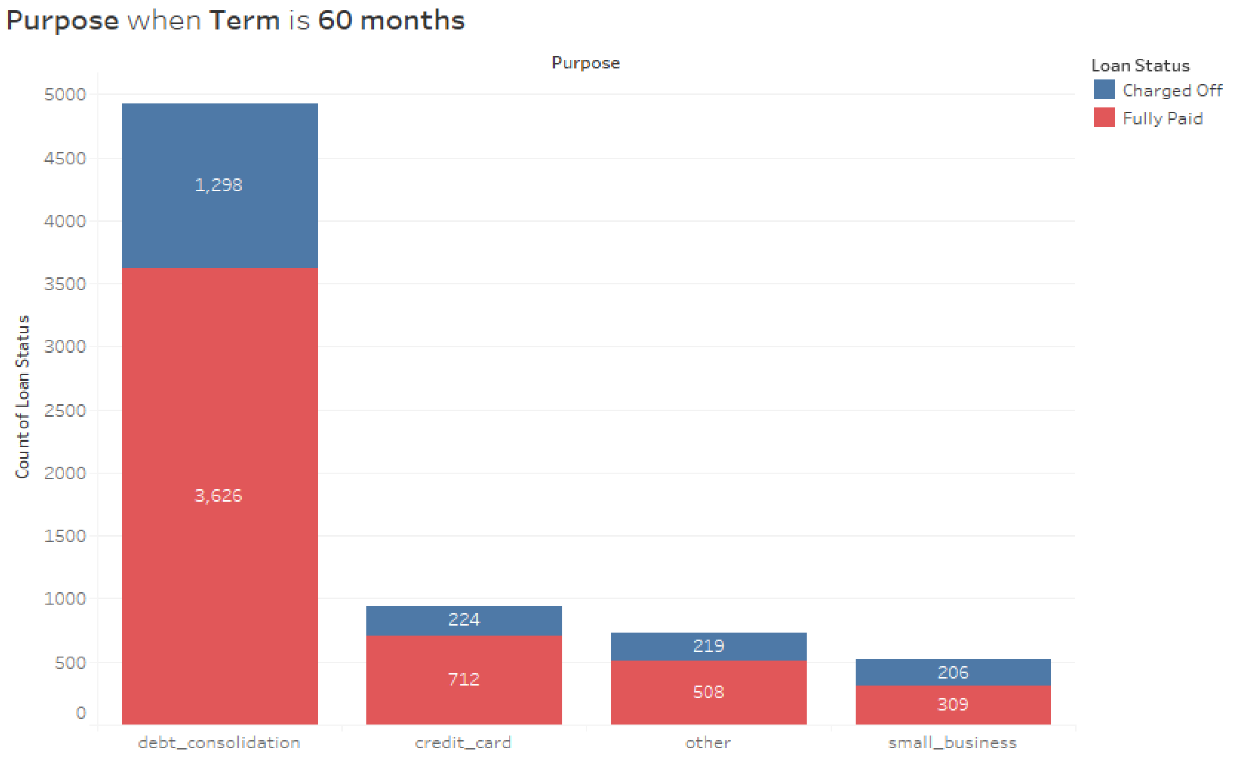


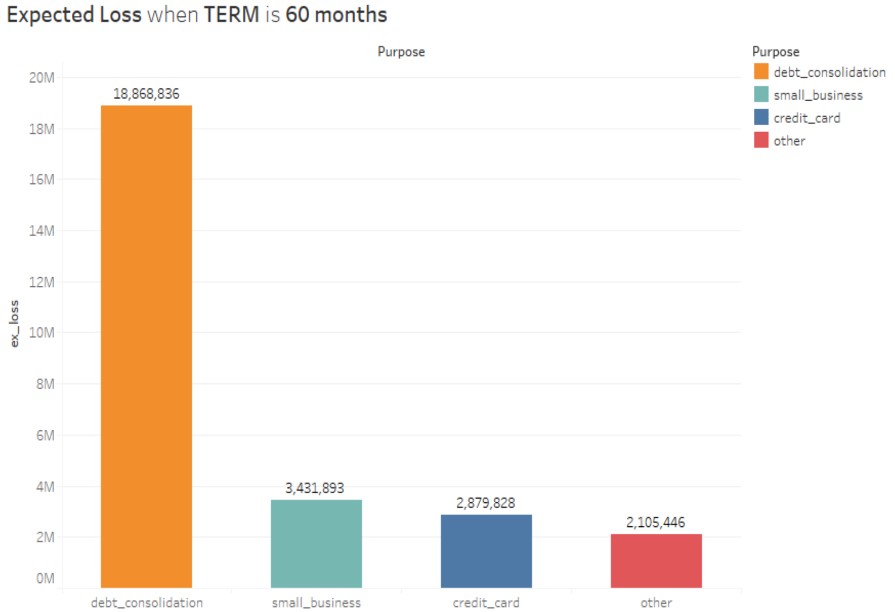
• The rate of default increases as the number of Inquiries increases. The rate of default steadily increases from **12%** for zero loan applications, to **29%** for 7 loan applications

• This is particularly telling, when the loan term is 36 months. The rate of default steadily increases from **9%** to approximately **29%** for greater than 7 loan applications.

PURPOSE







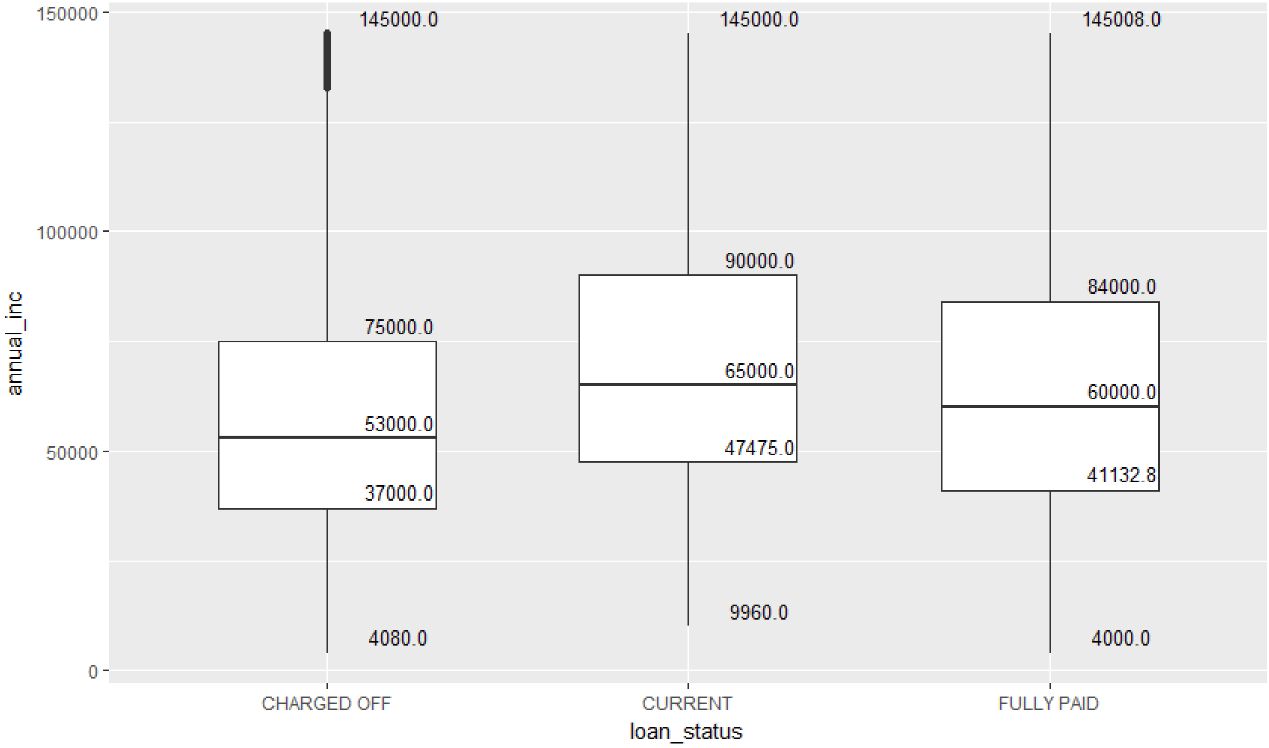
- The four categories '*debt consolidation*', '*credit card*' , '*other*' and '*small business*' account for **80%** of total default.

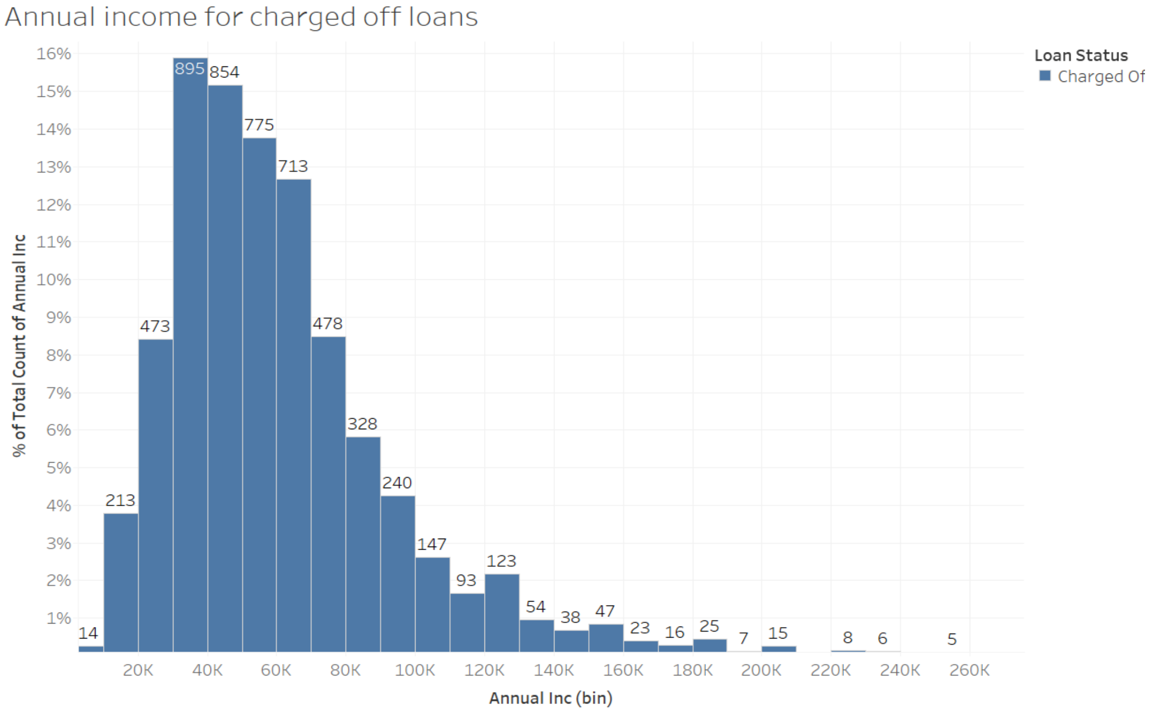
- When the term of loan is **60** months, the rate of default becomes higher in these categories as seen in the upper right hand image.

- The expected loss is highest for debt consolidation at a staggering **19** million.

- Although the number of 'small business' loans for 60 months are few, they produce a significant loss of **3.5** million, due to higher rate of default.

ANNUAL INCOME





• Those with lower annual incomes tend to default the most. For Annual income between 30-70k, the rate of default increases.

•As annual income increases, rate of default steadily decreases. As income increases above 70k, rate of default decreases significantly.

CONCLUSIONS

- From our Univariate analysis we know that ‘Rent’ and ‘Mortgage’ are the two most popular **home ownership** types. The **purpose** for which maximum number of loans are taken are for debt consolidation and credit cards. We also know that most of these applicants are from top **states** like‘CA’, ‘FL’, ‘NY’ or ‘TX. Finally, we understood that the number of loans taken increases steadily every **year**. When deciding loan status, a combination of bad levels from the above mentioned categories would result in a higher probability of default.

- Probability of default increases as the **interest rate** increases. While total customer default rate remains relatively low at **6%**of total population, at grade A, it increases steadily to a staggering **34%** of total borrowers by Grade G.

- Loans with **term** of 60 months have a higher rate of default. 60 month loans were introduced only in **May 2010** by the company, so the number of 60 month loans is much lesser than the 36 month loans, but, they display a higher relative rate of default.. Interestingly , all current loans in the dataset are of 60 months, which further validates this analysis. Since expected loss and default rate for 36 month loans is lower, it is encouraged to promote 36 month loan further.

- Default for loans that are ‘Verified’ and ‘Not Verified are almost the same. . This indicates that there is a need to improve the **verification** process. The best way is to verify the Source before providing the loan.

- The rate of default increases as the number of **loan applications** increase. This is significant when the loan term is 36 months. The rate of default steadily increases from **9%** to approximately **29%** for greater than 7 loan applications.

- ‘Debt consolidation’, ‘Credit card’, ‘Other’ and ‘Small Business’ account for nearly 80% of total defaults. Debt Consolidation causes highest expected loss. ‘Small business’ loans for 60 months, while although smaller, produce a significant loss of **3.5** million, due to higher rates of defaulters.