

# GRAMENER CASE STUDY

## SUBMISSION

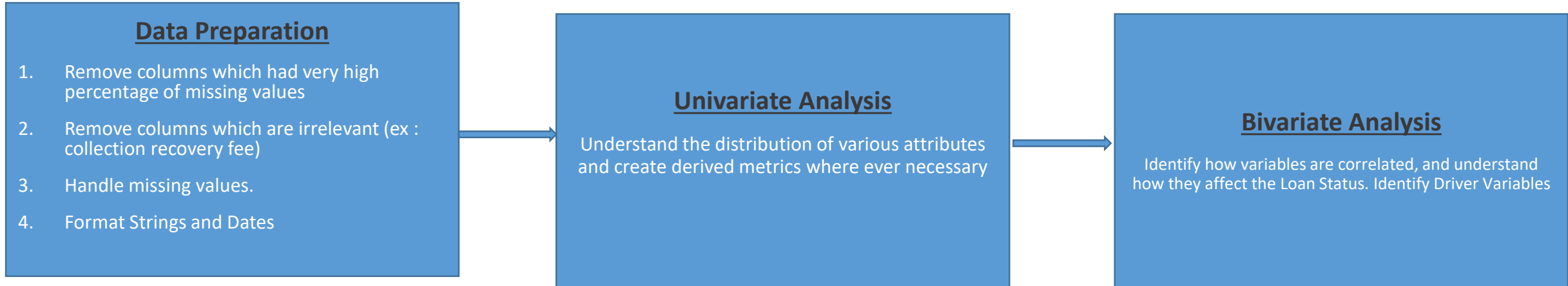
Group Name:

1. Aiswarya Ramachandran - APFE187000503

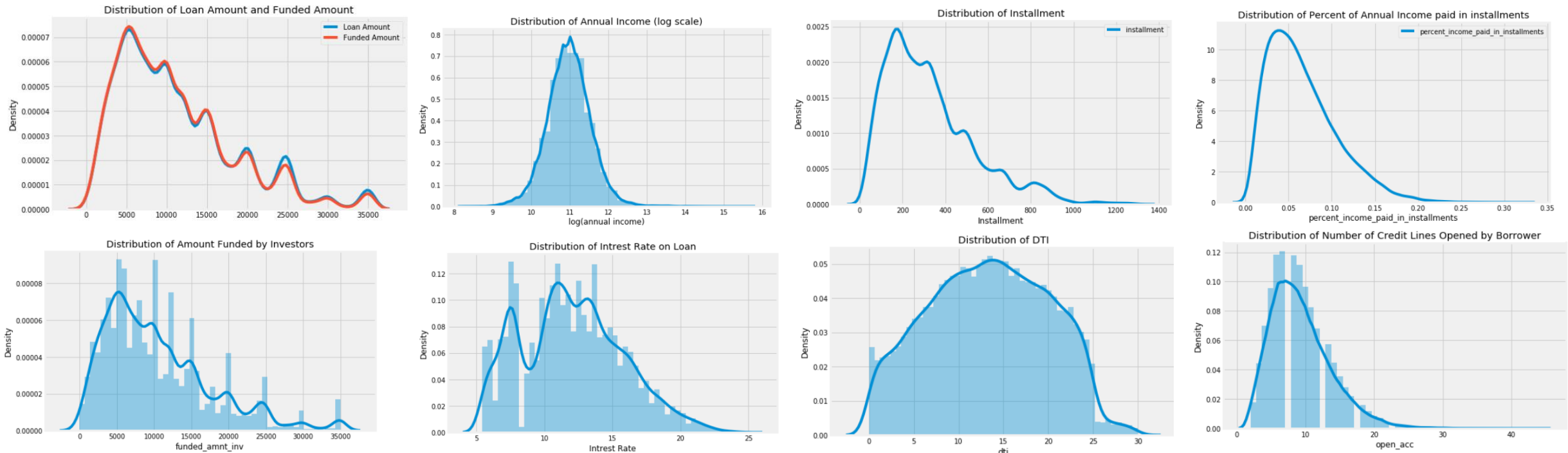
## Business Understanding

- A consumer finance company which specialises in lending various types of loans to urban customers, A consumer finance company wants to understand the consumer attributes and loan attributes influence the tendency of default.
- The data contains three loan status – the borrower has fully paid the loan, he is currently paying the loan or the loan has been Charged Off .If the loan has been Charged Off, it indicates that the borrower has defaulted on the loan

# Problem solving methodology

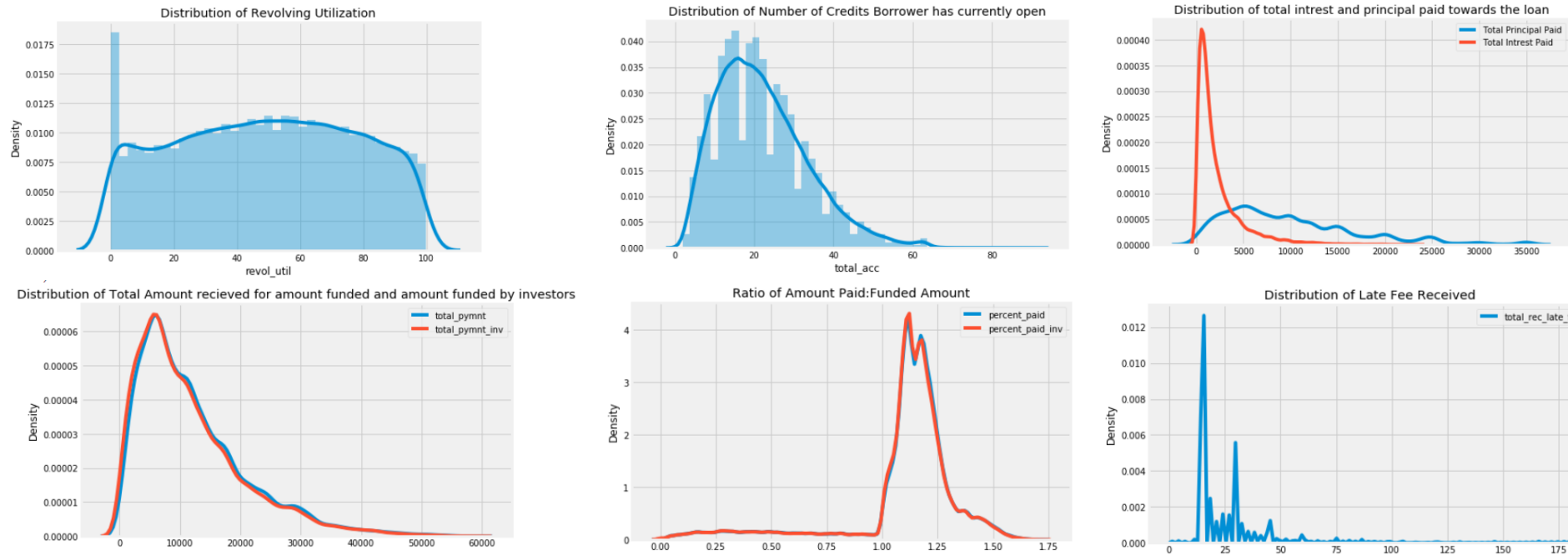


# Univariate Analysis – Continuous Variables



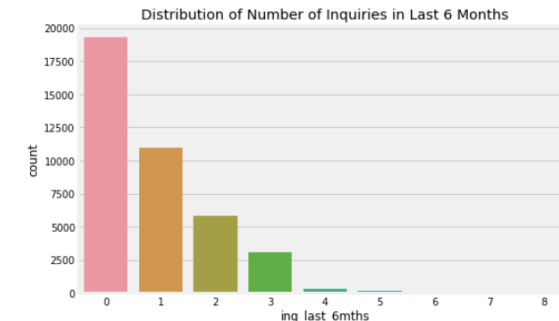
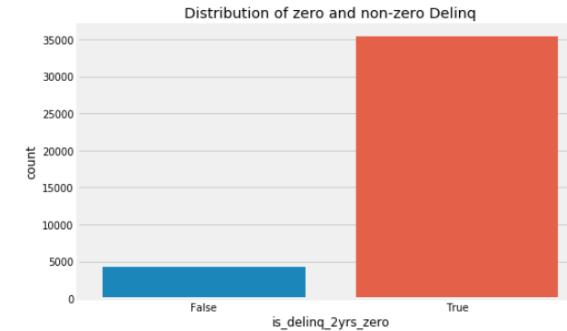
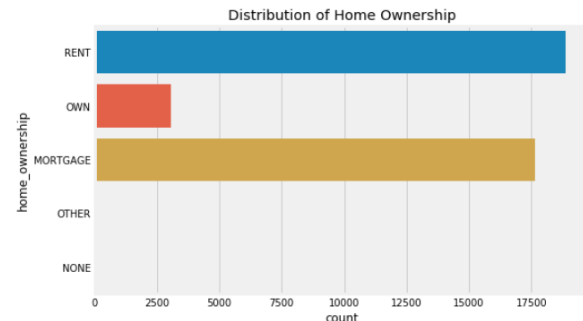
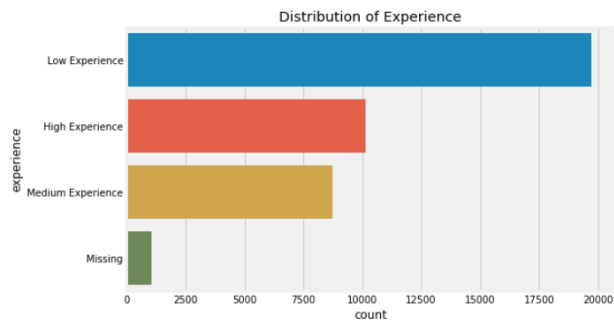
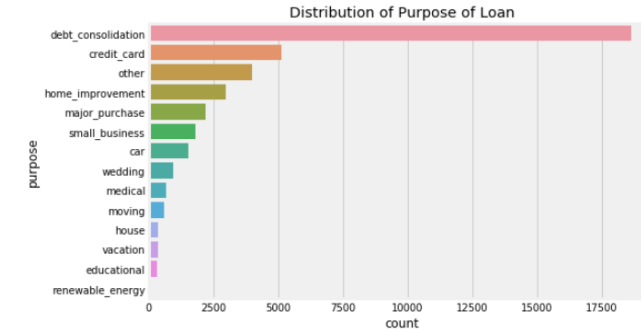
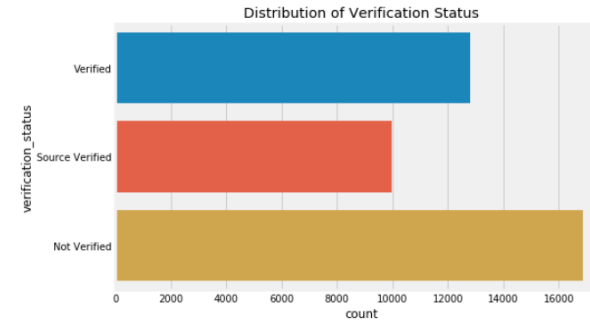
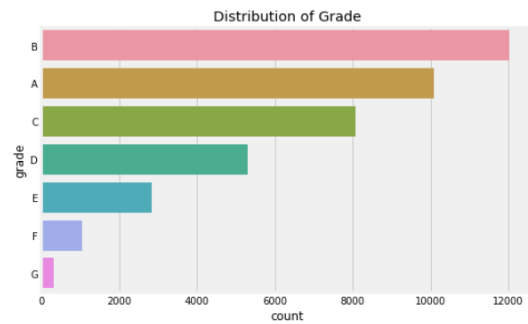
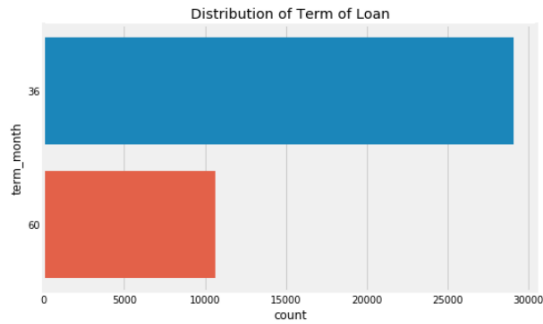
1. High correlation between Loan Amount, Funded Amount and Funded Amount By Investors.
2. Annual Income follows Zipf's law, hence to make it a normal distribution, log transformation was done
3. The distribution of amount paid towards monthly instalments is right skewed.
4. The Derived metric, percent of annual income paid in instalments is also slightly skewed

# Univariate Analysis – Continuous Variables



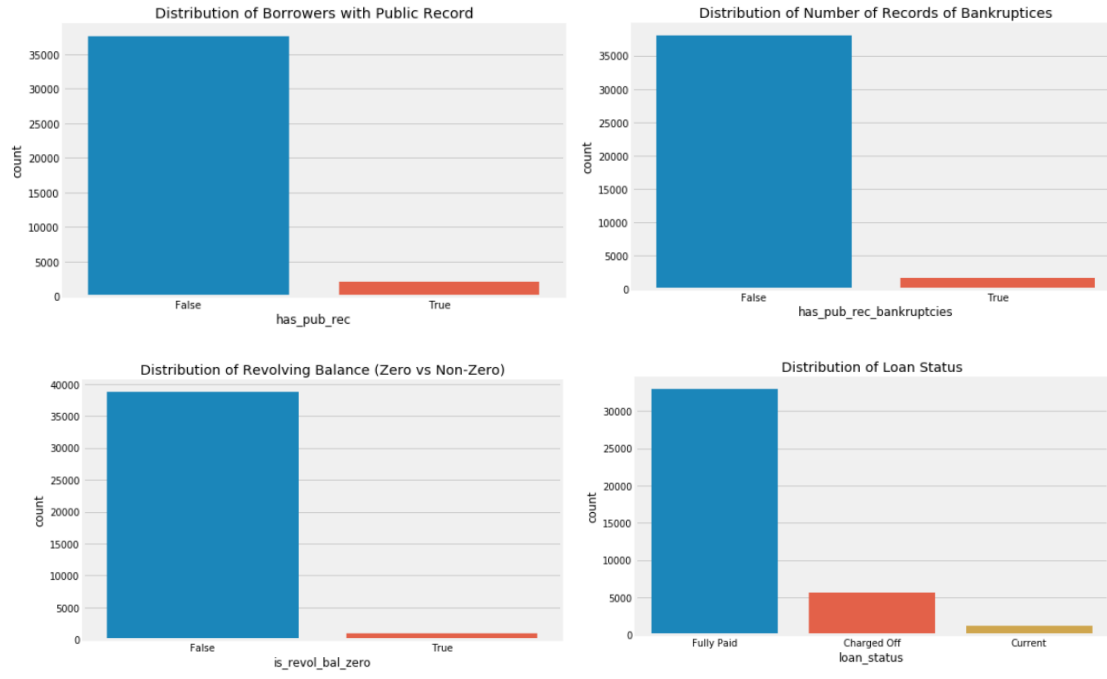
1. Distribution of Revolving Utilization, is similar to DTI
2. The distribution of Total amount paid towards the loan and total amount paid to investors are also highly correlated – created two metric, percent of income paid towards the funded amount and percent of income paid towards amount funded by investors
3. The distribution of Principal paid is right skewed.
4. 0.97 correlation was seen between total amount paid and total principal paid

# Univariate Analysis – Categorical Variables



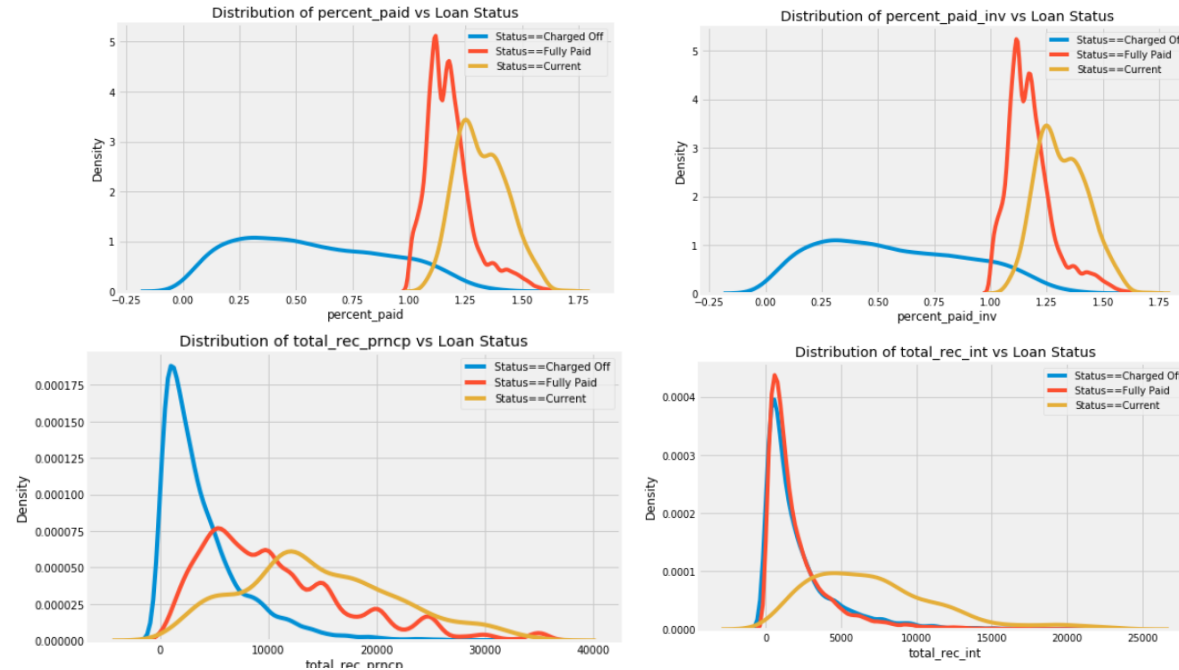
1. 73% of the loans have a term of 36 months
2. The duration borrower has been working has 12 levels, so we grouped them based on Experience. 50% of the borrowers have low experience
3. Most borrowers are in a RENTED house.
4. Around 47% of the borrowers have made only one inquiry in last 6 months
5. 89% of the times borrowers have not paid for 30+ days in last two years
6. Debt consolidation is the most important reason for taking a loan

# Univariate Analysis – Categorical Variables



1. 90% of the borrowers do not have any derogatory public record – this is a derived field from pub\_rec
2. 95% of the borrowers do not have any public record of bankruptcies – this is derived field from pub\_rec\_bankruptcies
3. 97% of the borrowers do not have revolving balance zero
4. Only 14% of the loans are Charged off. 82% of them are Fully paid

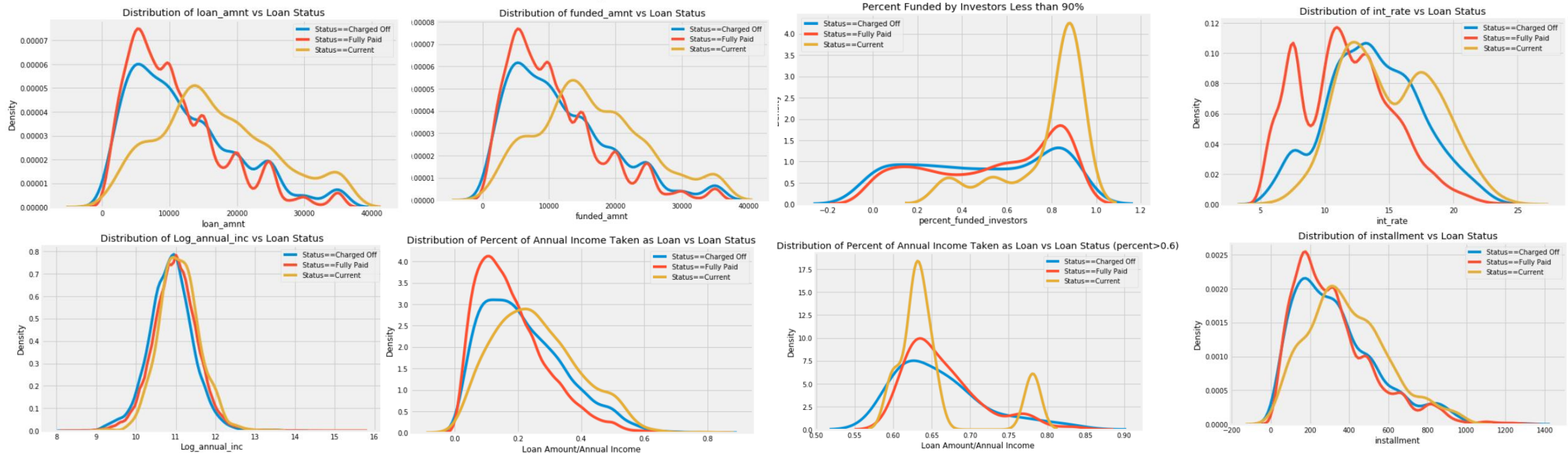
## Bivariate Analysis – Continuous Variables vs Loan Status



1. For the derived metric percent amount paid towards the amount funded (both total and amount funded by investors), if percent is lower high chances of default.
2. Lower principal amount paid, higher chances of default.
3. For Current Loans, the interest rate is higher. But, there is no significant difference between interest rate of Fully Paid and Charged Off Loans

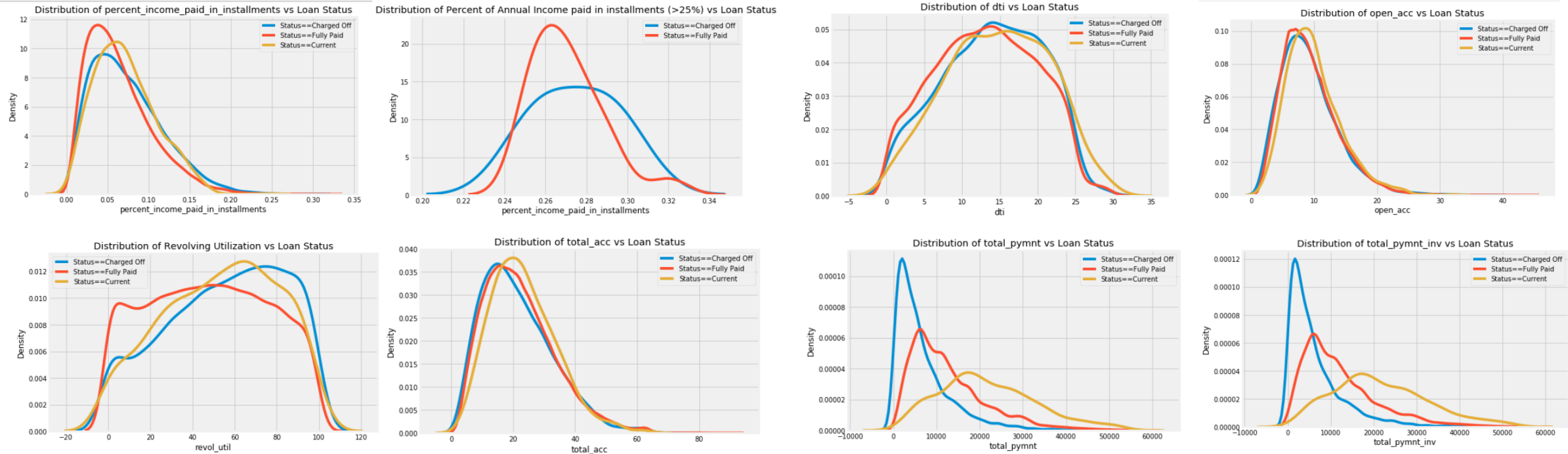


## Bivariate Analysis – Continuous Variables vs Loan Status



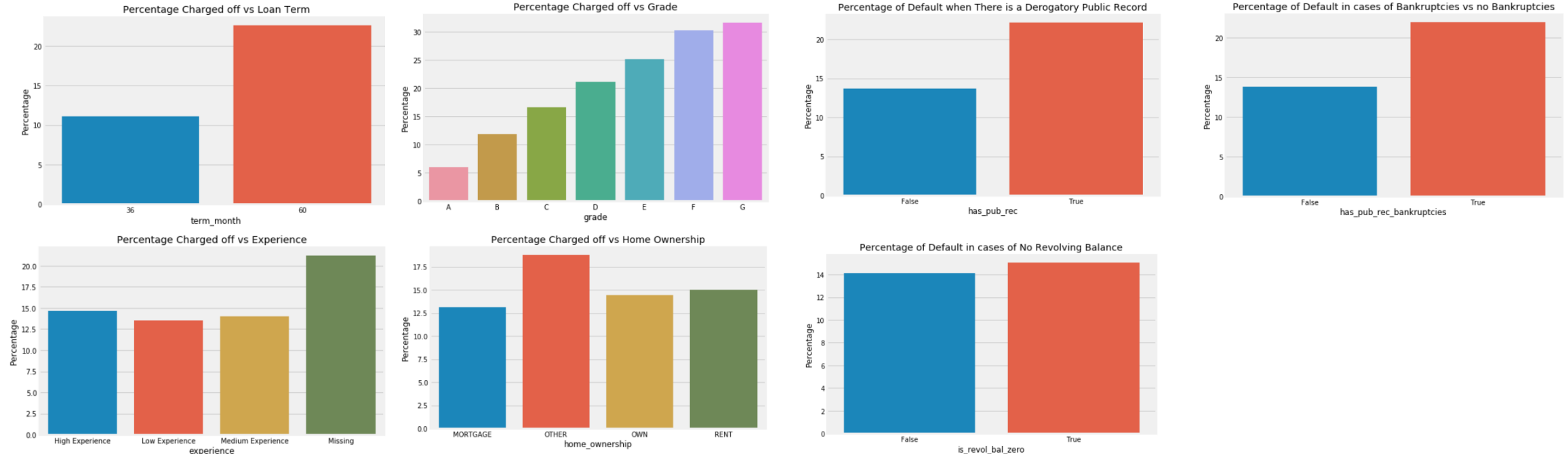
1. Higher loan amount or funded amount, higher chances of default. For Current loans, the mean loan amount or funded amount is higher.
2. Since, funded amount and funded amount by investors are correlated, a derived measure, percent funded by investors was created. If this value is below 50% there is a higher chance of default.
3. Lower the income higher chances of default.
4. Created a measure, percent of annual income taken an loan. For this measure, higher the percentage, higher chances of default. This is a much better measure than taking only loan amount or annual income
5. Installment amount higher for Charged off and Current Loans compared to Fully Paid loans.
6. Interest rate is higher for Charged Off Loans

## Bivariate Analysis – Continuous Variables vs Loan Status



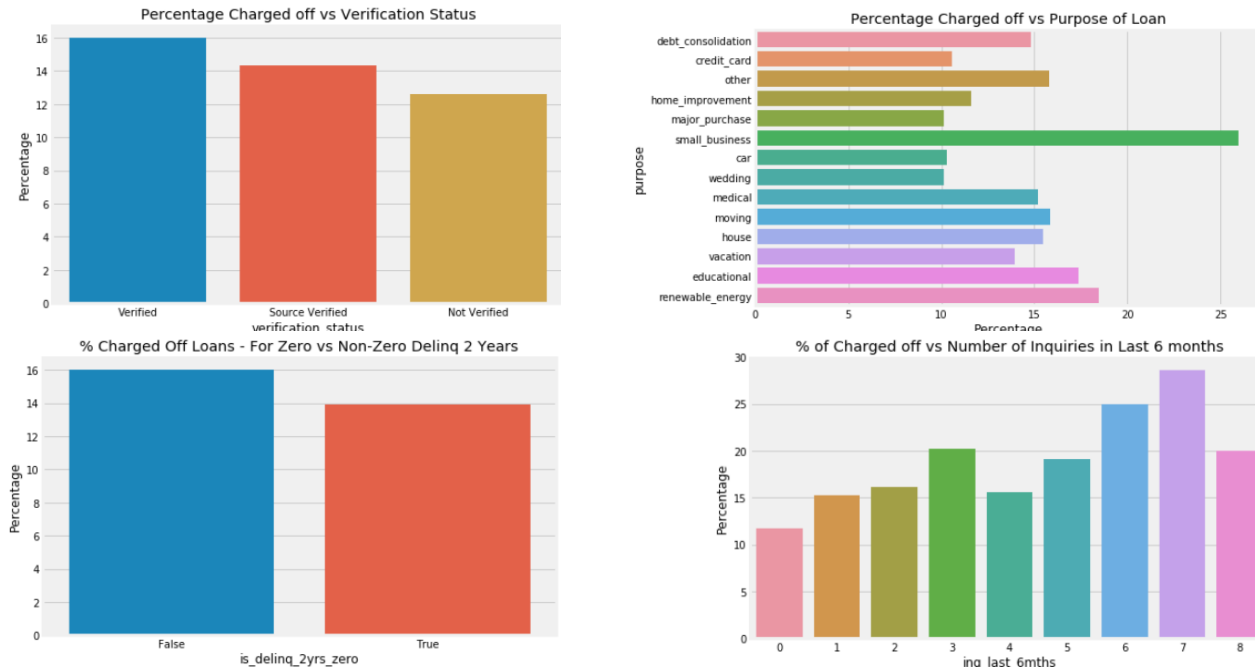
1. For the derived metric percent of income paid in instalments, if more than 25% is paid towards instalments there is a higher chance of default.
2. Higher DTI, more chances of default
3. Higher debt to credit ratio (revolving utilization), higher rate of default
4. Lower number of credit lines, higher default. This is not very significant
5. If total payment towards the amount funded is lower then higher chances of default

## Bivariate Analysis – Categorical Variables vs Loan Status



1. As grade goes from A to G, the default rate increases
2. 22% default rate when Term is 60 months, compared to 11% when term in 36 months
3. 21% default rate when there is no experience information available.
4. Home Ownership is “Other” 18% default rate and 15% default rate when home ownership is “RENT”
5. 22% default rate when there is a derogatory public record, compared to 13% when there is no record
6. 21% default rate when there is a public record of Bankruptcies, compared to 13.8% when there is no record
7. 15% of the loans which have revolving balance zero defaults , compared to 14% when revolving balance is non-zero

## Bivariate Analysis – Categorical Variables vs Loan Status



1. When income is “Not Verified” the default rate is lower (only 12%), compared to 15% when source is “Verified”
2. 25% default when purpose of loan is for “small business”
3. 28% default rate when there were 7 inquires in last 6 months
4. 15% default rate when there has been no late payments(30+ days) in last two years

# Conclusions

- Few Driver Variables for identifying default are as below:
  - **Interest rate of the loan** – higher the interest rate higher chances of default
  - **Amount funded** – Higher the amount funded more chances of default (This is very highly correlated with loan amount, taking either one into consideration will suffice)
  - **percent of annual income taken an loan** – if the percentage is higher more chances of default.
  - **Percent of annual income paid in instalment** – if borrower pays higher percentage of his annual income in instalments there is higher chances of default
  - **Revolving Utilization** – if higher debt to credit limit then more chances of default.
  - **Total Payment towards the loan** – if total payment is lower than higher chances of default. This is very highly correlated (0.97) with the total principal paid towards the loan, which also shows a similar trend
  - **DTI** - Higher DTI more chances of default. That is the ratio of borrowers total monthly debt payments to the total debt obligations is higher, then chances of default is higher
  - **Grade of loan** – as the grade increases higher chances of default
  - **Term of the loan** – higher the loan term higher default (60 months term has more default rate than 36 months)
  - **Work Experience of borrower** – If no experience is available, then 21% chances of default. Also, higher the experience, the default rate tends to increase.
  - **Is there any derogatory public record or any previous record of bankruptcies** – higher default rate if any such record is present
  - **Purpose of Loan**