“GRAMENER CASE STUDY”

Group Name: The4DAz

1. **Prakash Babu**
2. **Asher Peter Babu**
3. **Jithin Jose Thomas**
4. **Ajay Joshy**

Objective

- Understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default.
- Risky loan applicants can be identified from these driving factors, this will help us to **reduce** the amount of **credit loss**.
- The company can utilize this knowledge for its portfolio and risk assessment.

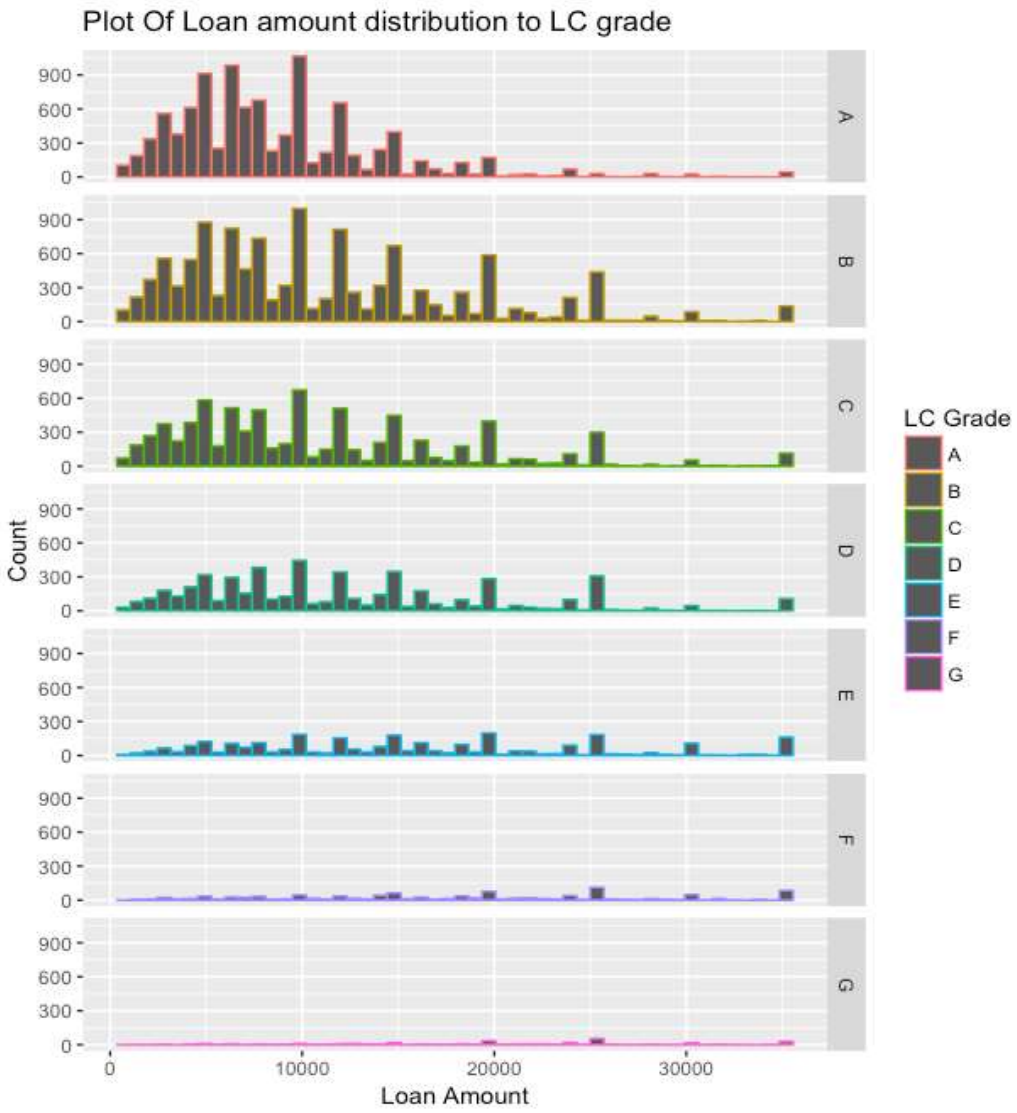
Data Set

- The dataset contains the complete loan data for all loans issued through the **time period 2007 to 2011**.
- Total of **39717 records** with **111 columns**.
- Details of the column variables can be accessed from the **Data Dictionary.**
- Highlighting only the **important column attributes:**
  - **id/member_id:** A unique LC assigned ID for the **loan listing/ borrower member**.
  - **loan_status :** Current status of the loan
    - **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.
  - **int_rate** : Interest Rate on the loan
  - **grade:** Lending Club assigned loan grade **(A,B,C,D,E,F,G)**
  - **purpose:** A category provided by the borrower for the loan request.
  - **annual_inc:** The self-reported annual income provided by the borrower during registration.
  - **dti:** A ratio calculated using the borrower’s total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower’s self-reported monthly income.
  - **revol_util :** Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
  - **pub_rec_bankruptcies:** Number of public record bankruptcies

Data Cleaning And Manipulation

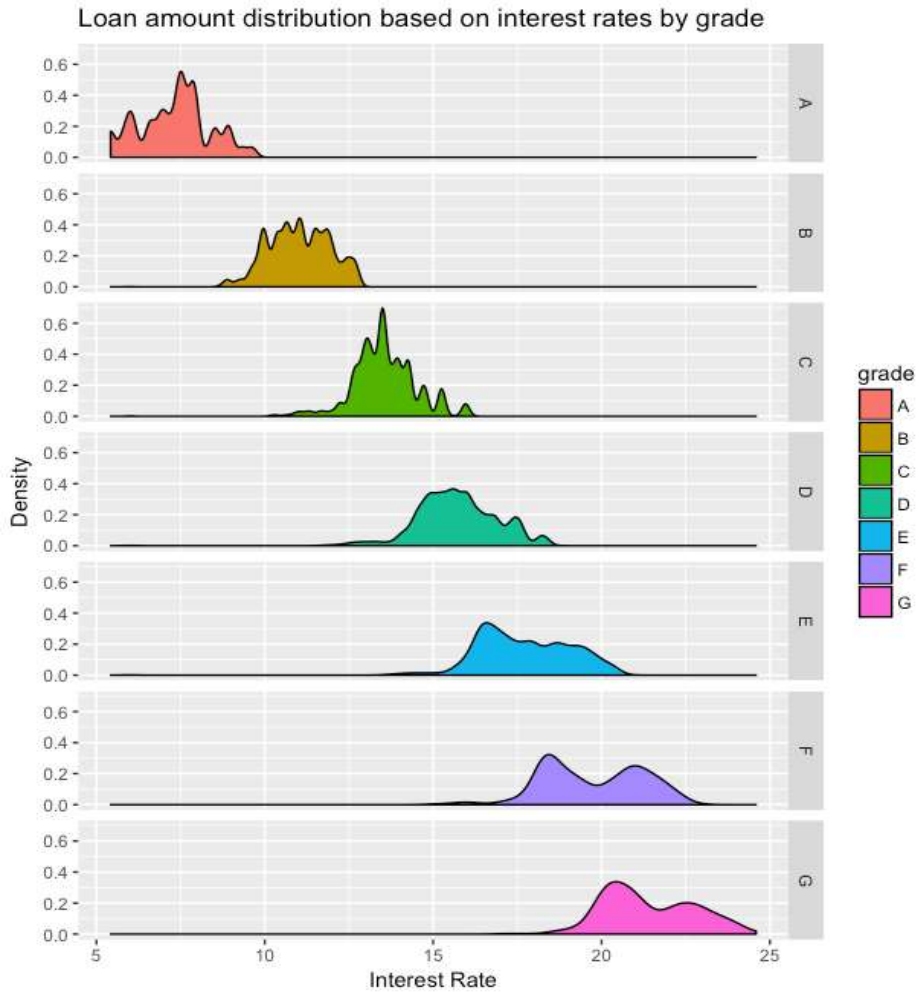
- **Possible Data Inconsistencies**
  - NA or missing values in columns of interest. **(If more than 50% missing, remove those columns)**
- **Other Issues**
  - Too many **unwanted column attributes**, which are irrelevant for the analysis.
  - **% sign** in the **interest rates** and **revolv_util**.
  - **+ sign employee length** column.
  - **Extract loan issued year** from **issue_d**.
  - **Impute** missing values with median values.
  - Treat **outliers** in **annual_inc** and **loan_amnt**.
  - Eliminate all records having **loan_status as "Current",** since those loans are still active

Loan amount distribution based on Grades



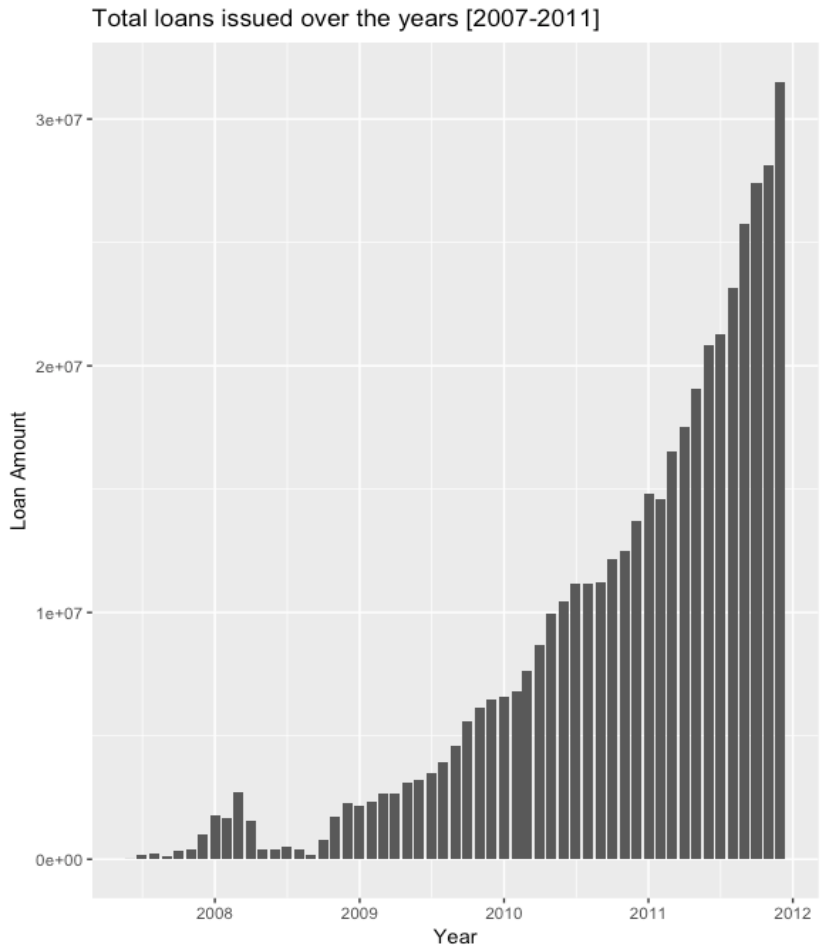
Those with **higher grades (A, B, C and D)** have received **more loans** compared to those with **lower grades (E, F and G)**

Loan amount distribution based on Interest Rates by Grade



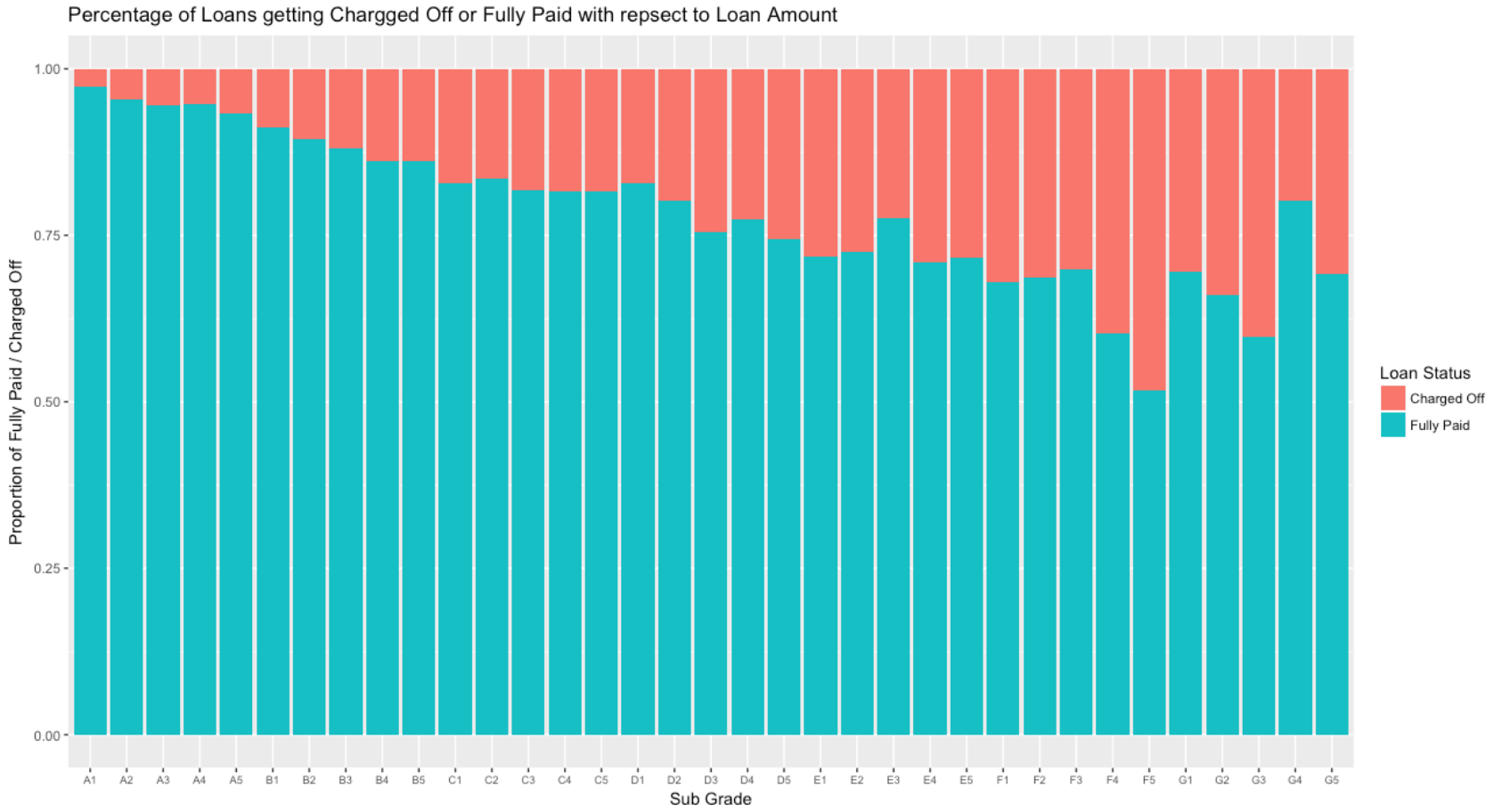
Grades are assigned based on risk, and **so interest rates go up as the risk goes up**.

Total loans issued over the years [2007-2011]



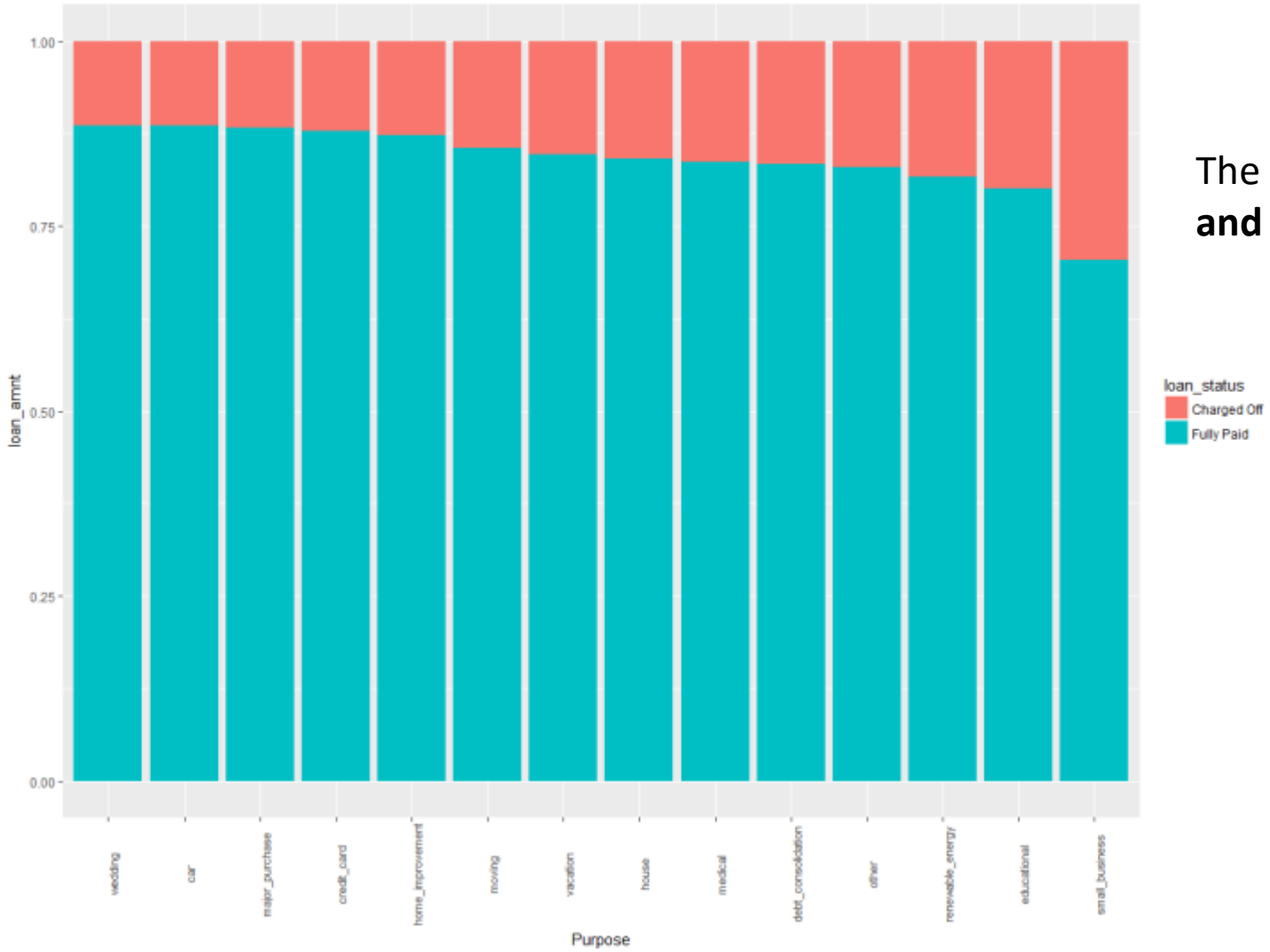
It is observed that as the years progress, the business for LC is increasing tremendously. **Hence the company is in dire need for finding out risky loan applicants**.

Proportion of paid vs default over the Grades



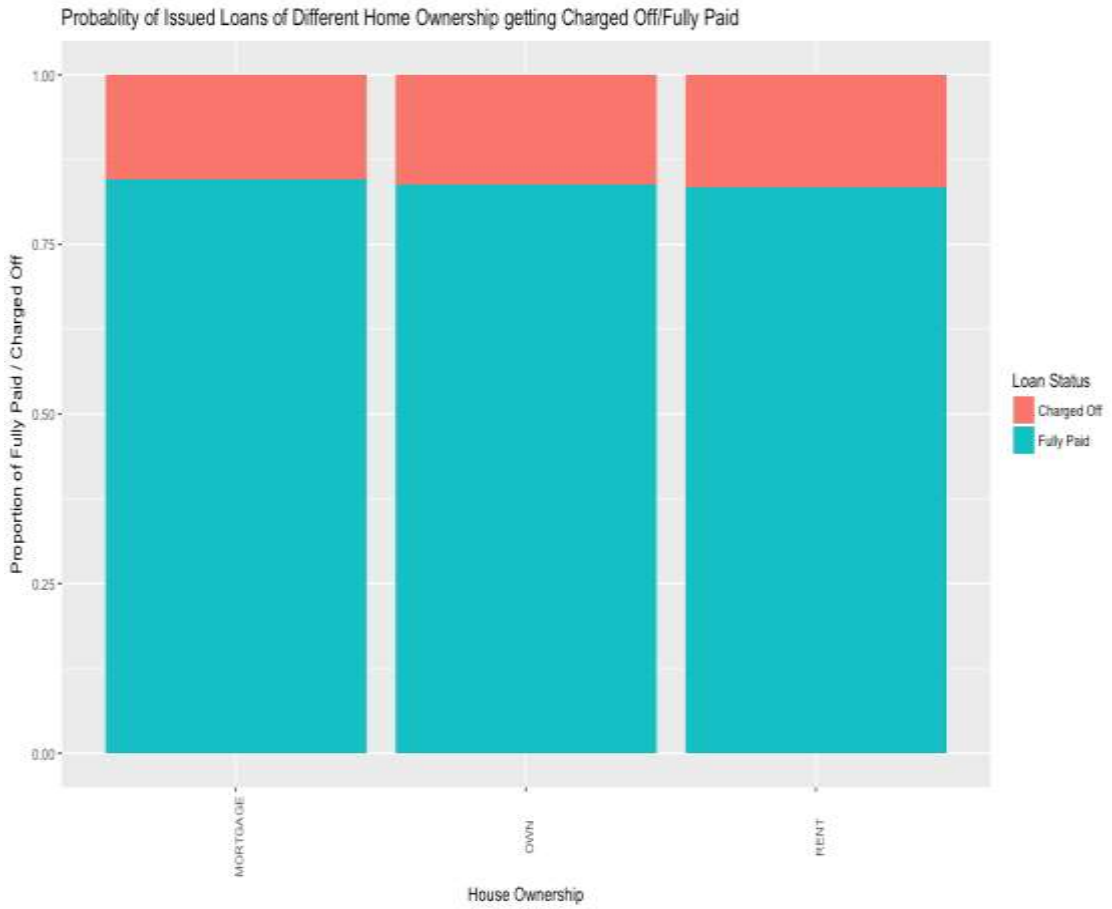
It is very obvious that as the **grade goes down**, the **proportion of the unpaid loan increases**

Proportion of Paid vs Default loan amount by different Purposes



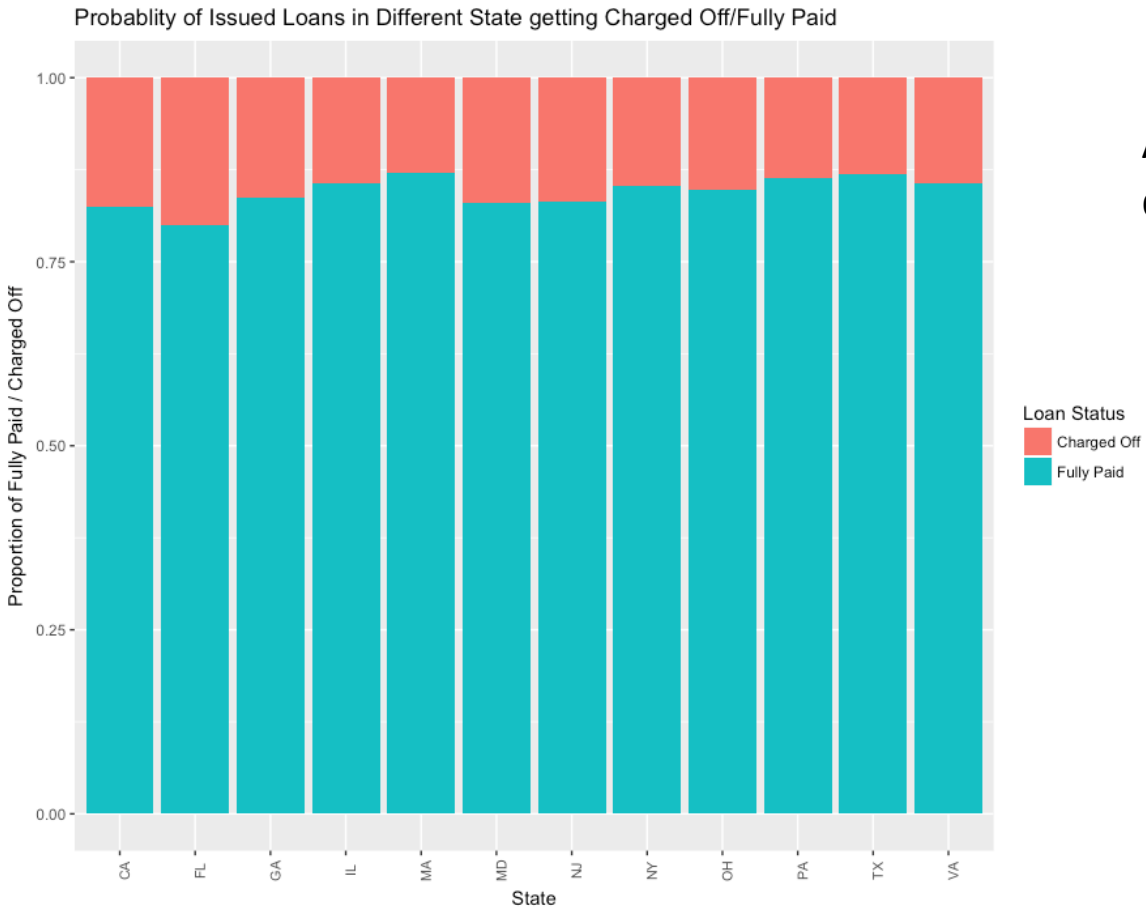
The probability of the loans for **educational and small business being unpaid is nearly 25%.**

Proportion of Paid vs Default loan amount by Home Ownership



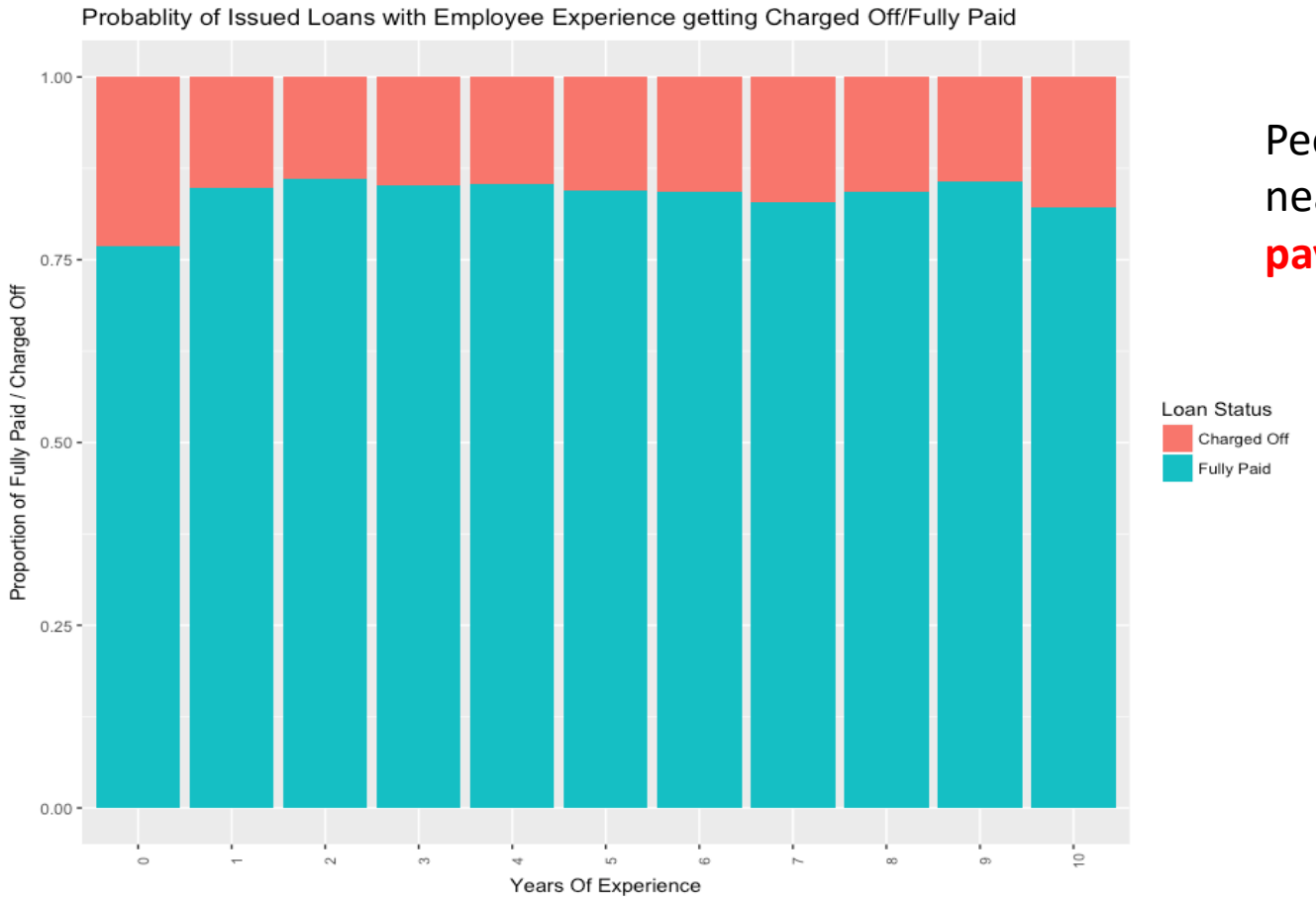
- All 3 categories of home owners have nearly **13-15% chance of defaulting.**

Proportion of Paid vs Default loan amount by State



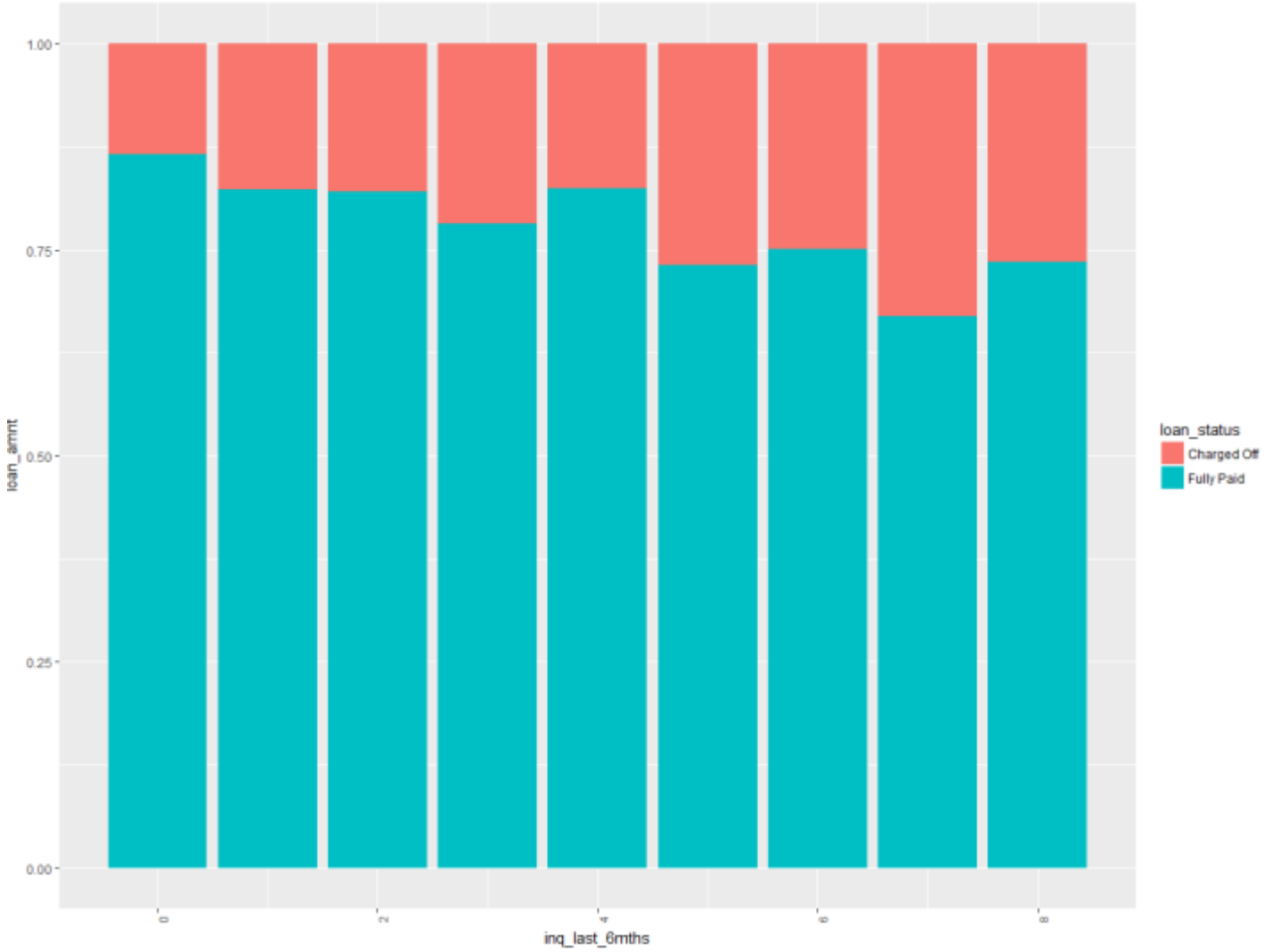
Almost all states (top 12) has **15-20%** chance to default

Proportion of Paid vs Default loan amount by Length of Employment



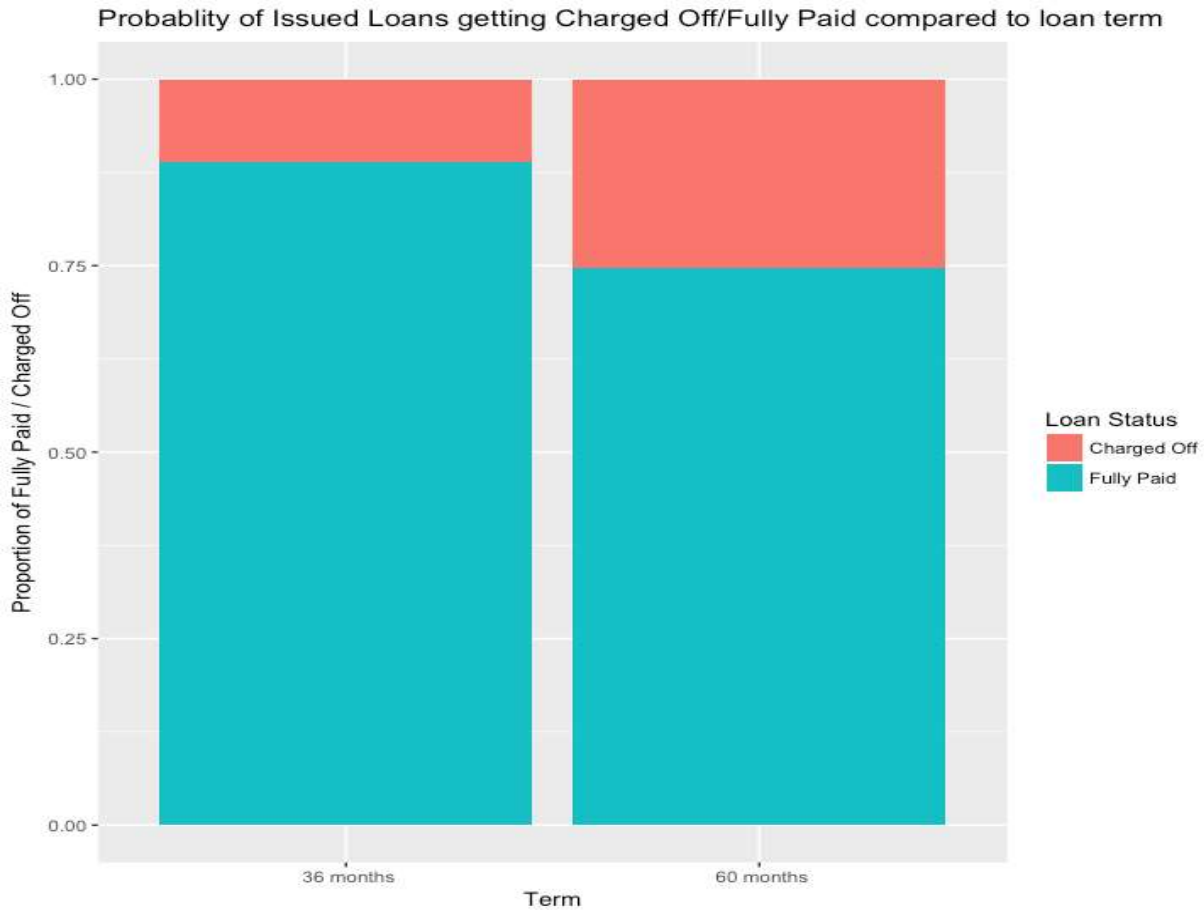
People with **less than 1 year** of experience have nearly **23% chance** of **defaulting on laon payments**.

Proportion of Paid vs Default loan amount by loan inquiries made in last 6 months



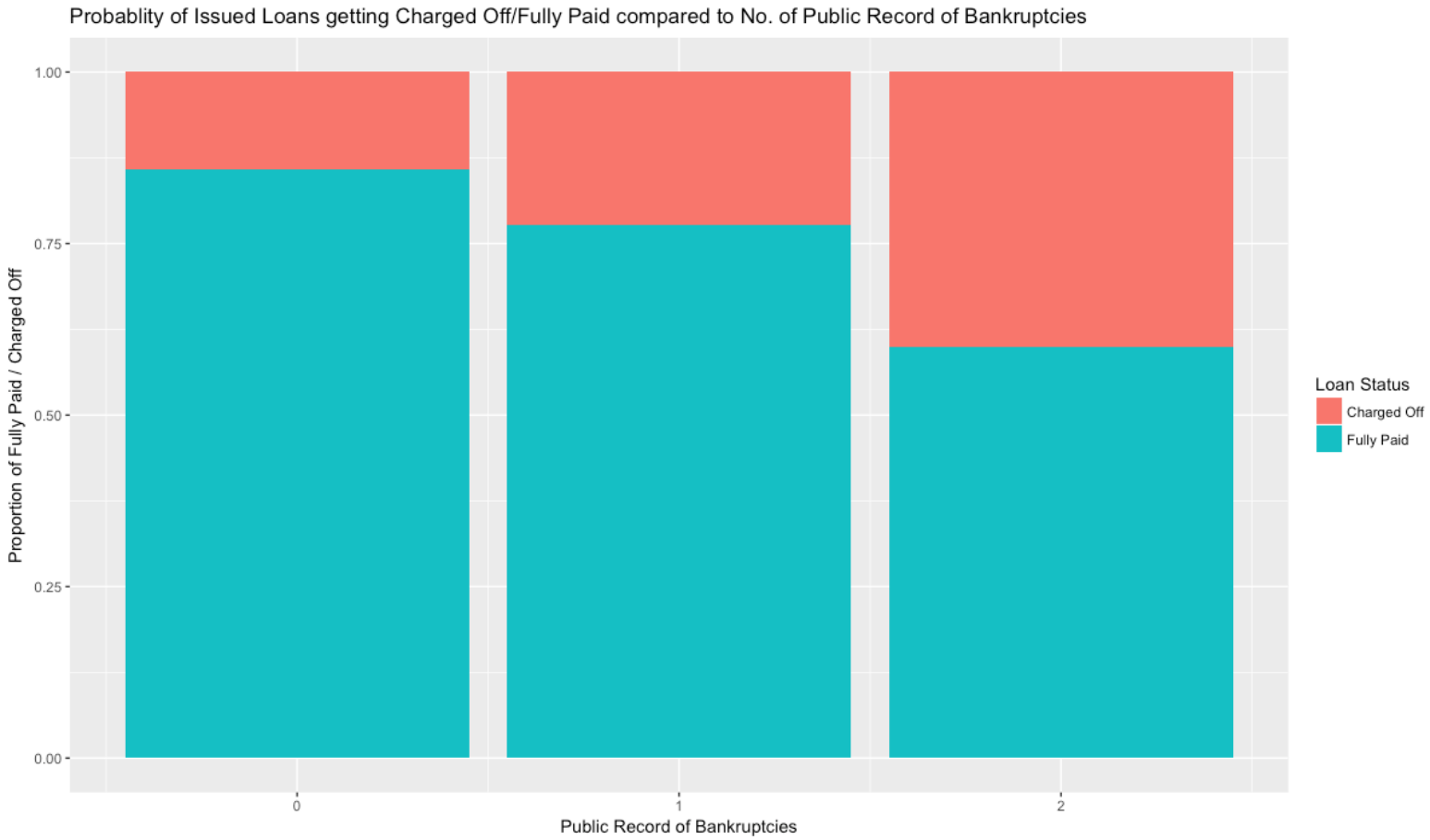
More number of loan inquires could indicate that the person has a habit of taking loans frequently or other banks might have made an inquiry and rejected a loan making the customer riskier. **Data shows this behavior is an indicator of loan defaults.**

Proportion of Paid vs Default loan amount by Loan Term



5 year loan term (60 months) have more tendency to get **charged off**

Proportion of Paid vs Default loan amount by Public Record of Bankruptcies



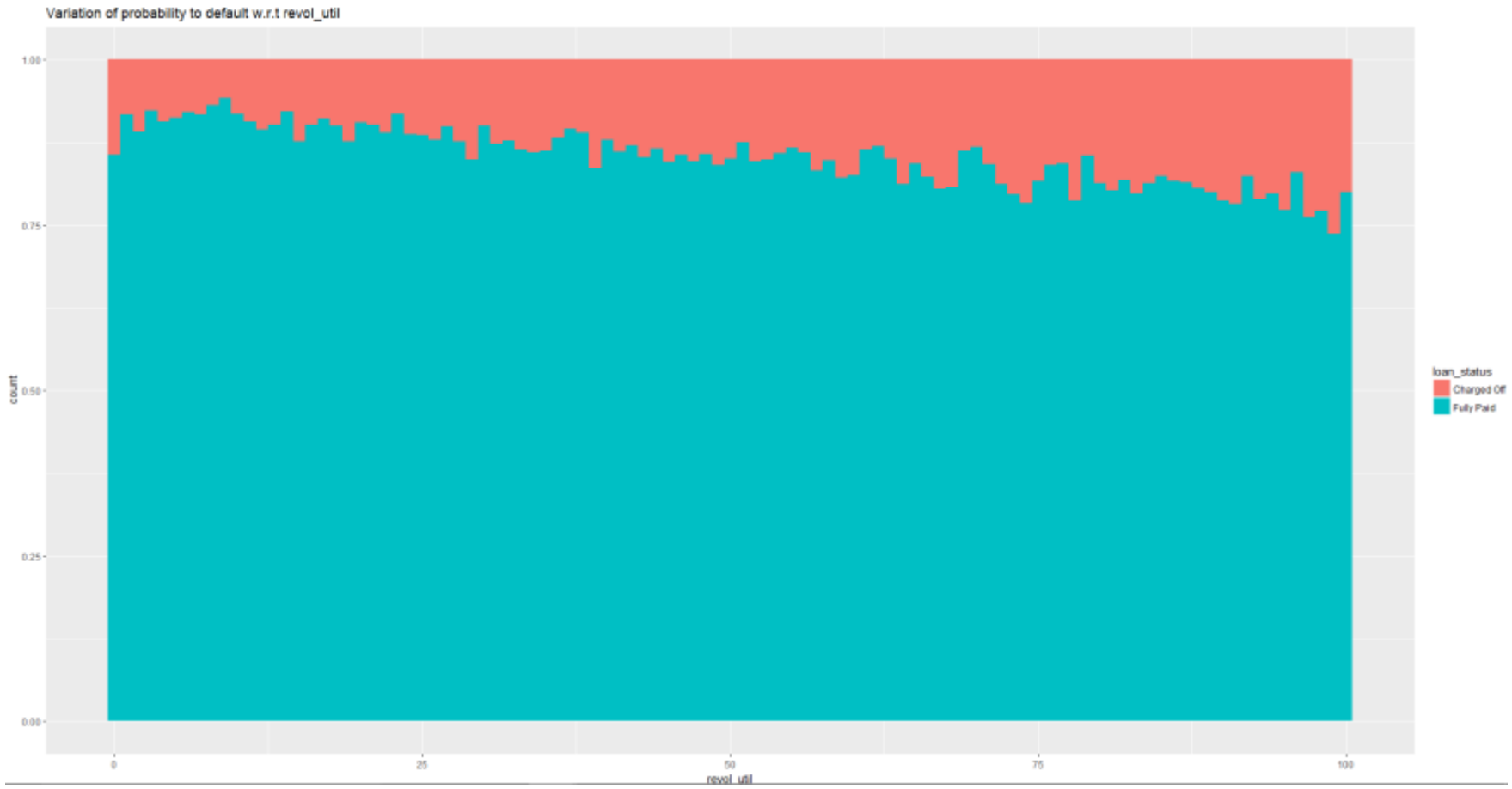
As **Public Record of Bankruptcies increases** the chances of **charged off also increases.**

Variation of probability to default vs dti

Dti is the monthly debt to income ratio. So more debt the person has greater is the value of dti index.

As dti increases there is an **increasing chance for his other loan payments to default**

Variation of probability to default vs revol_util

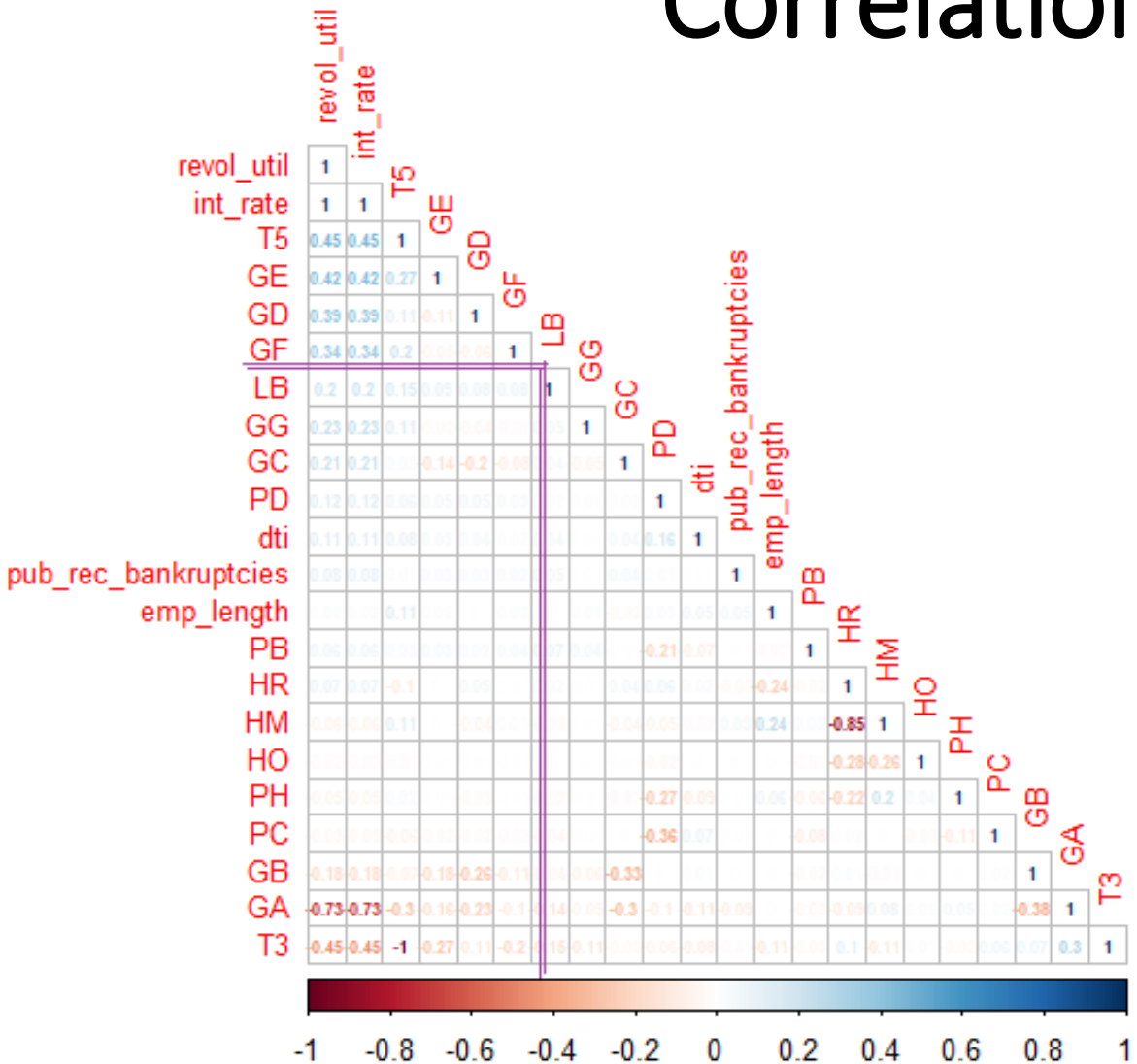


Revolving line utilization is the measure of amount of credit utilization by a customer.

Trend shows higher the revol_util, higher is the probability to default on loan payments.

This is indicative of a person who is nearly maxing out on his credit limit and is prone to default on payments

Correlation Matrix



Derived Variable Dictionary

LB	Bad Loan
T3	Term 3 yrs
T5	Term 5 yrs
Purpose	
PC	credit_card
PB	small_business
PD	debt_consolidation
PH	home_improvement
Grade	
GA .. GG	Grade A .. Grade G
Home Ownership	
HR	RENT
HO	OWN
HM	MORTGAGE

Correlation Matrix Analysis

- The **univariate and bivariate analysis** is further strengthened by analyzing the **Correlation matrix plot.**

- The correlation matrix clearly indicates the **positive and negatively** correlated variables with respect to bad or default loan we are interested in.

- It can be seen that that bad loans mostly have a positive correlation with **revol_util, int_rate, loan term of 5 years, Grade D and below**.
    - **This means that as revol_util, int_rate increase the number of bad loans tend to increase**
    - **Similarly as terms of loans increase to 5 years tendency to default increases**
    - **Also grade D and above tend to be more riskier**

- It can be seen that that bad loans also have a correlation with few variables like **emp_length, pub_rec_bankruptcies, Grade C and Above**
    - **It can be seen that as length of employee experience decreases tendency to default increases**
    - **Similarly as counts of pub_rec_bankruptcies decrease tendency to default decreases**
    - **Also grade C and below tend to be less riskier**

Conclusion

The following variables are strong indicators of loan re-payment default

1) **Grade** : Ranges from A to G. **Lower the grade, the greater the chances for loan default**
2) **Purpose** : The purpose for which the loan is taken is also a strong indicator of default
3) **Term** : **60 month loans** are more likely to default
4) **Public Record of Bankruptcies**: The more records there are the more the chances
5) **Loan inquiries made in the last 6 months**: The more inquiries a person has to his name the higher his risk.
6) **Debt to Income ratio** : The more debt a person has to, the more likely to default
7) **Revolving line utilization** : Rise in credit utilization is an indicator of likelihood to default