EDA/Gramener CASE STUDY

SUBMISSION

Group Members:
1.   Amit Shekhar
2.   Ankush K R
3.   Aparna Sastry
4.   Suman Kundu

Data set contains : 111 columns (comprising of fields related to customer profile as well as loan related information) of 39717 urban customers of a consumer finance company who have been granted loan between year 2007-2011 .

Problem Statement:

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 utilise this knowledge for its portfolio and risk assessment.

Approach   :

Exploratory Data Analysis (EDA) is an approach for analysing the data (based on the following steps) -

• **Data sourcing** - The given data set is already provided along with data dictionary for understanding the data fields.

• **Data cleaning and manipulation**

• **Data Analysis** –

       Univariate and segmented univariate analysis - to identify the important driver variables

       Bivariate analysis - to identify the important combinations of driver variables

Data cleaning & Manipulation

- Remove the columns which contain only NA values- These field columns will not be helpful for analysis.

- Remove % from the columns revol_utli and Int_rate so that these values can be read as numeric .

- Converting the following columns to numeric field variables using is.numeric function for further analysis.

  int_rate ,revol_util ,funded_amnt,installment ,annual_inc,Dti ,delinq_2yrs ,
 *inq_last_6mths_,open_acc,pub_rec ,revol_bal,total_acc*

- Employee length - <1 year to be replaced by 0 and 10+ year to be replaced by 10.

- For this analysis, only two loan status are considered i.e ' Fully Paid' &'Charged off'

Loan Status analysis

UpGrad

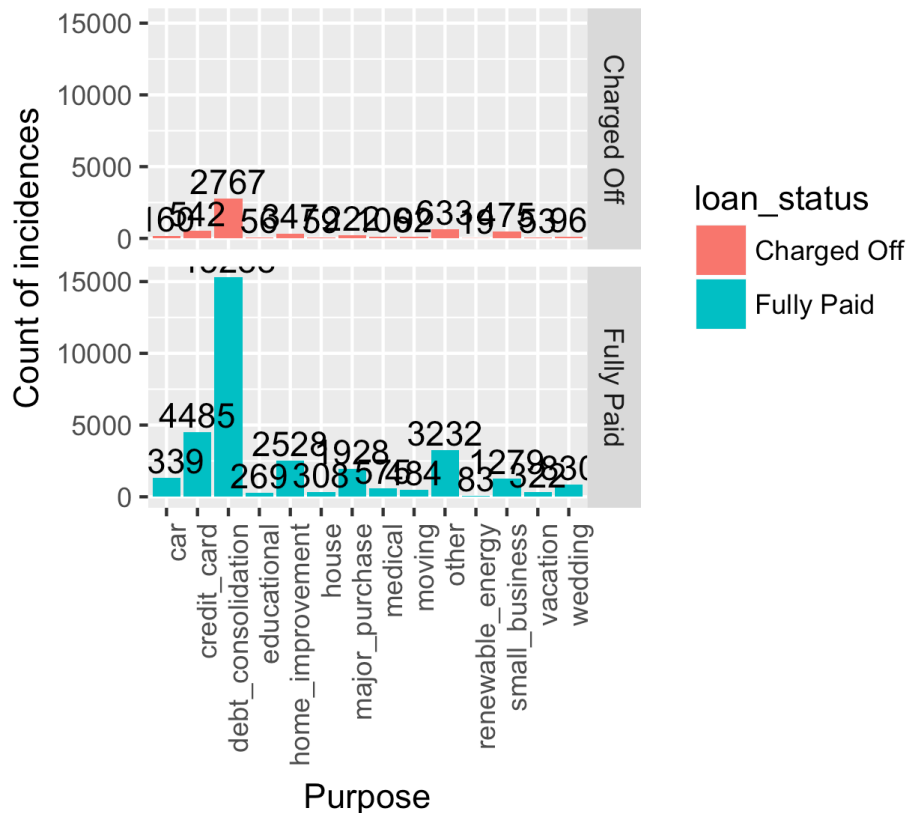Count of Charged Off & Fully Paid Loan Status
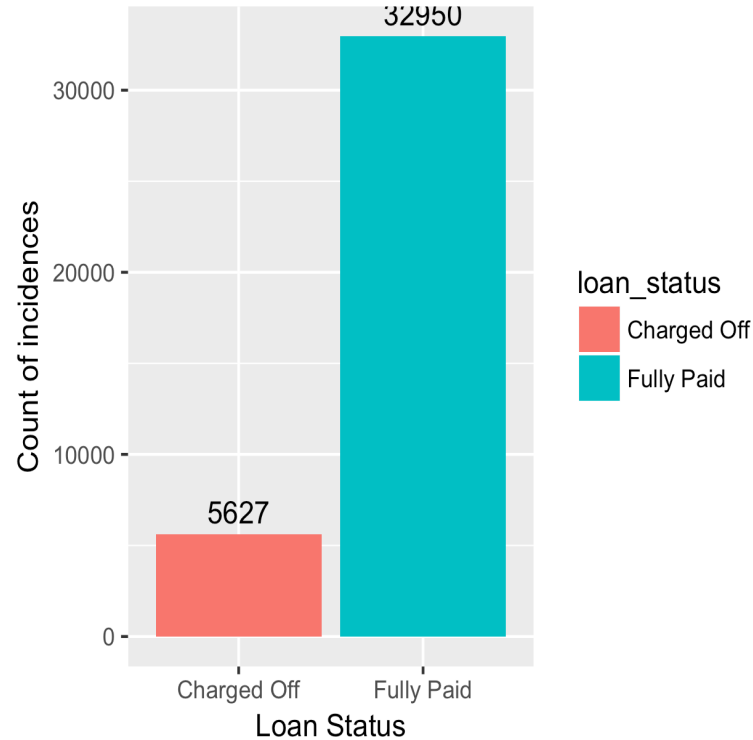
- The count of Charged off loans are around one-fifth the count of fully paid loans.
- It can be seen that the number of Charged offs are more in short term loans(i.e. 36 months) when compared with long term loans. The count of short term loans is little more than three times the count of long term loans.

Loan Purpose analysis
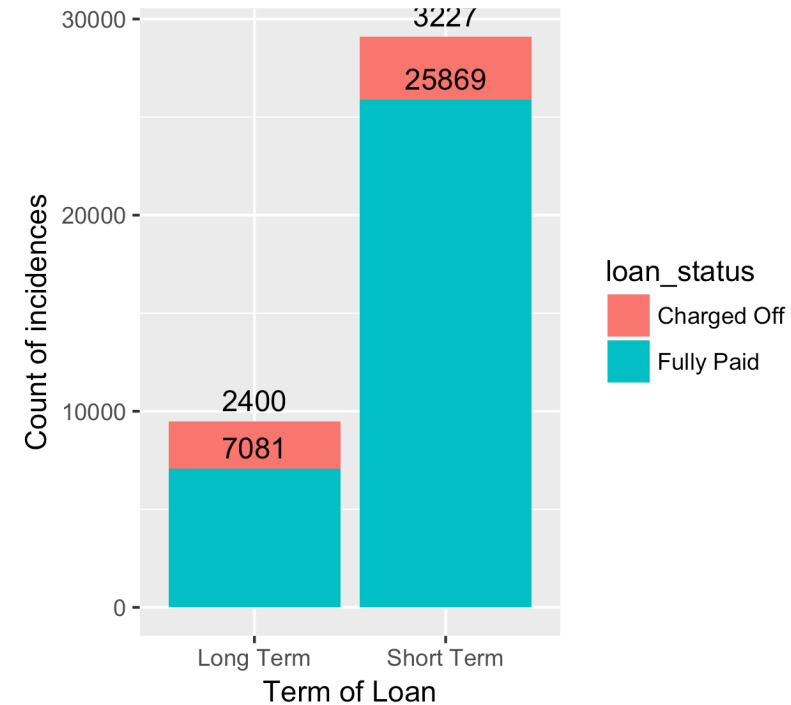
Count of various purposes w.r.t Loan Status

- For the purpose –'Debt consolidation, Credit Card and Small business – both Funded Amount and Instalment are higher leading to higher charged-offs.

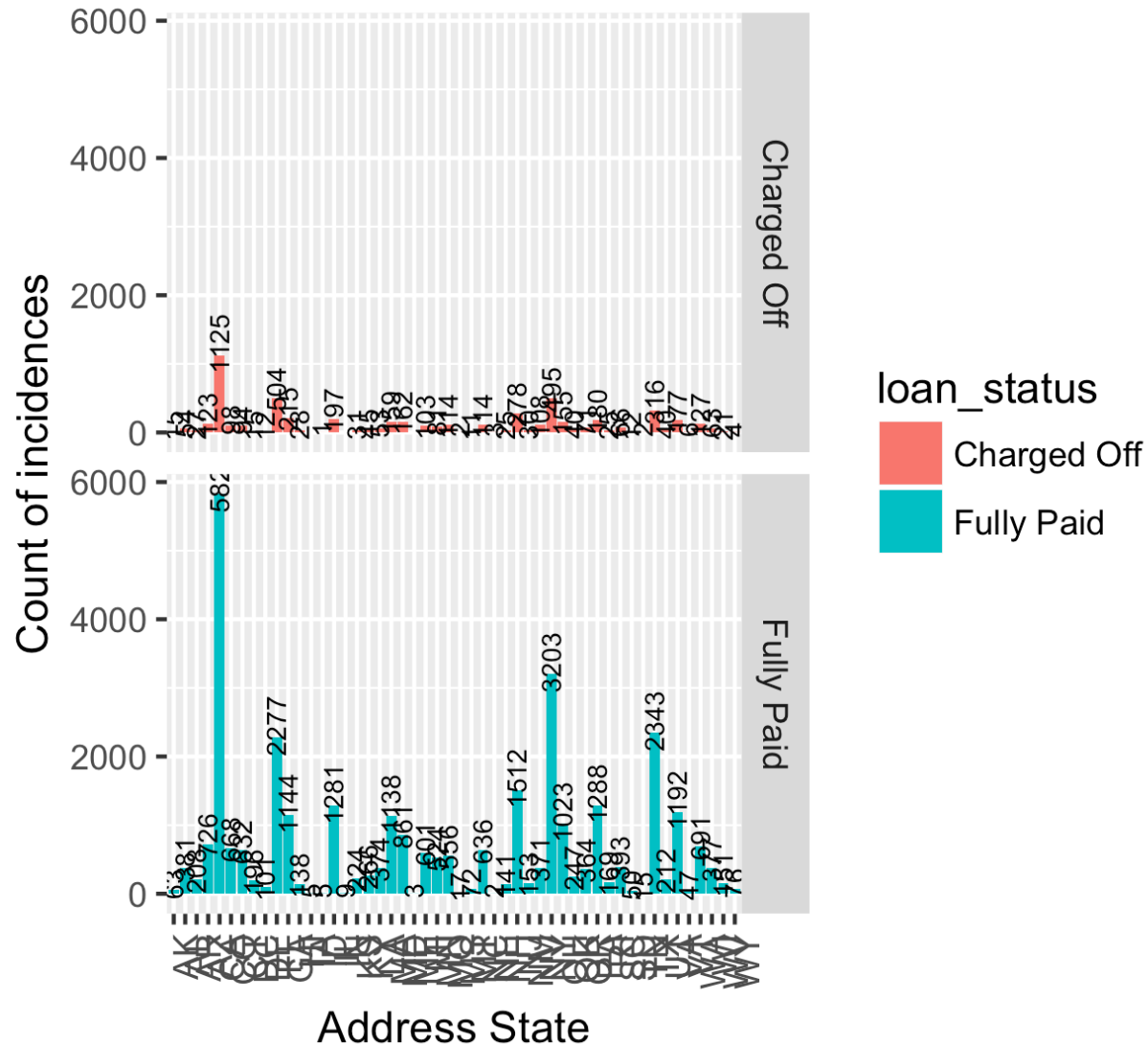### Funded amount vs installament by purpose

- Charged-off numbers are higher for the purpose - Debt consolidation, followed by Credit Card, home improvement, Other, small business etc. Number of Loans taken for debt consolidation purpose is very high compared with Loans taken for all other purpose.

# Loans Count for address states w.r.t. Loan s

- Top 5 States in terms of higher order of charged-off loans are CA(California),# NY(New York),TX(Texas),FL(Florida),NJ(New Jersey).
- Number of Loans taken from the state of California (CA) very high when# compared loans taken from all other address states.
- Number of charged off loans taken from the state of California (CA) is very high# when compared loans taken from all other address states

UpGrad

## Count of types of home ownership w.r.t. loa
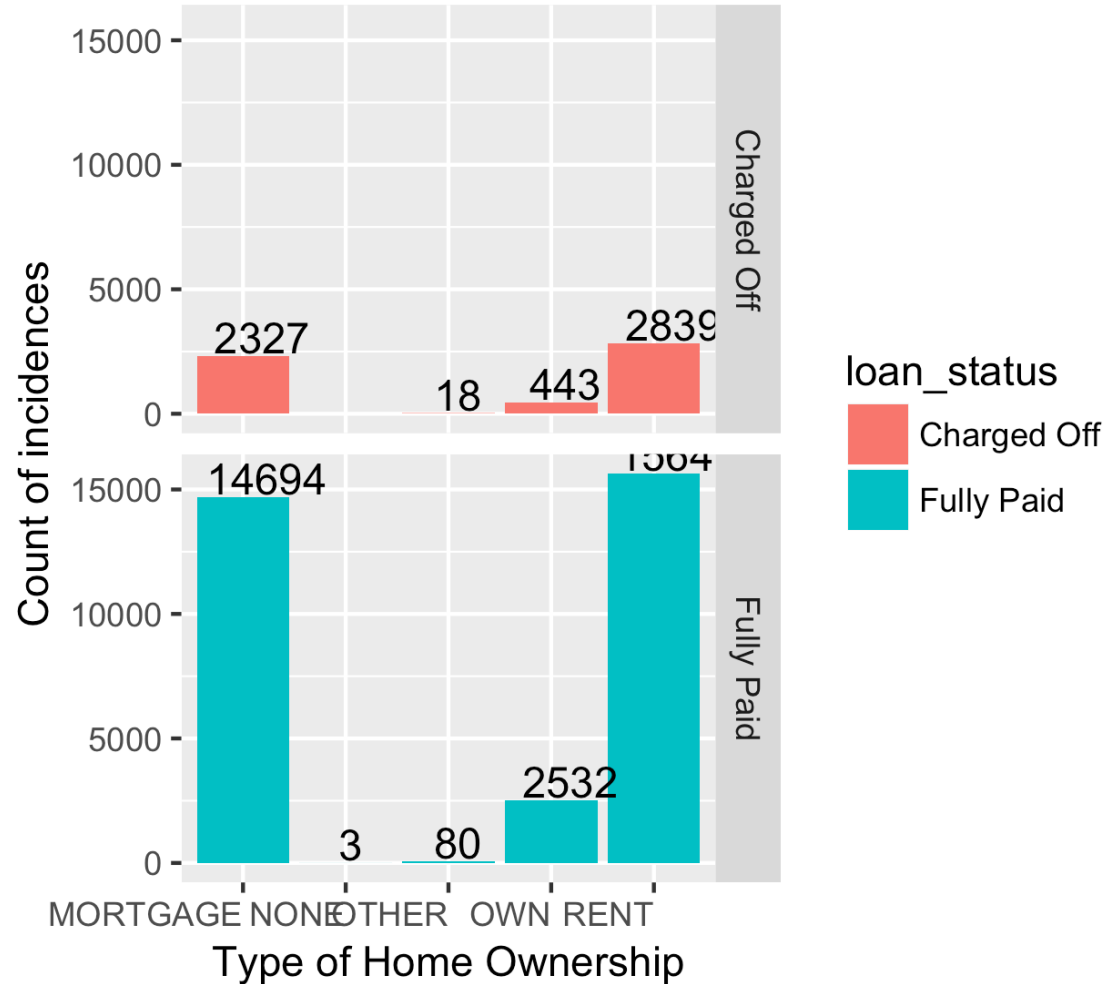
- Total number of Loans for applicants with home ownership as Rent is high followed by Mortgage.
- The count of charged off loans for applicants with home ownership as Rent is marginally high than Mortgage.
- The count of Fully Paid loans for applicants with home ownership as Rent is high.
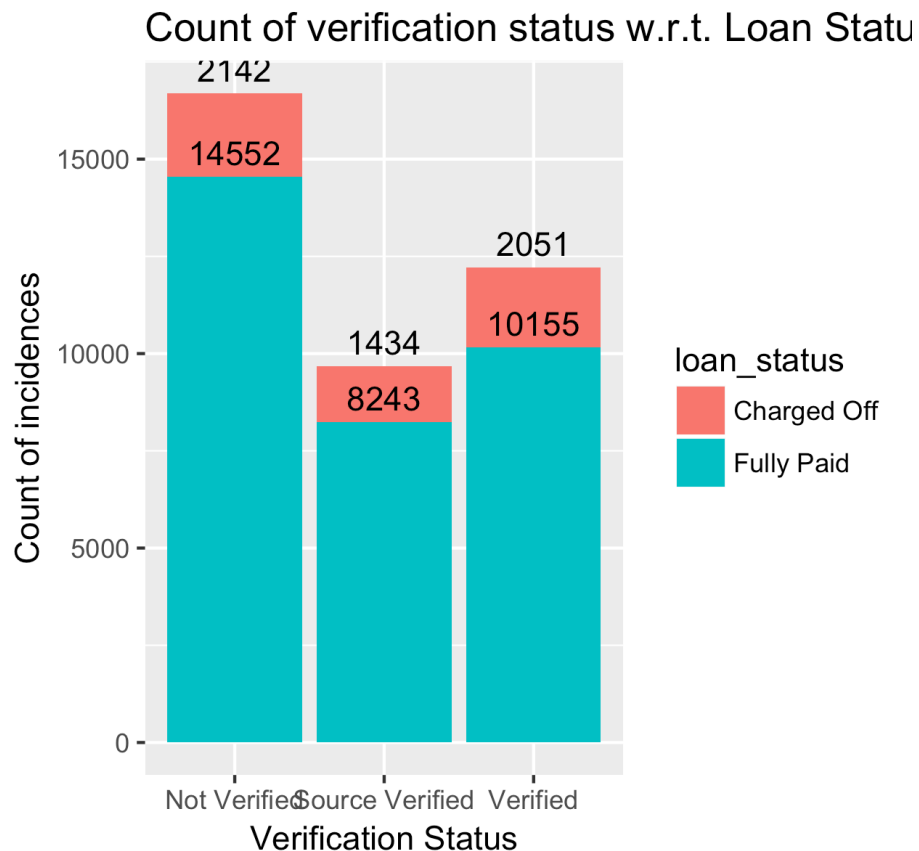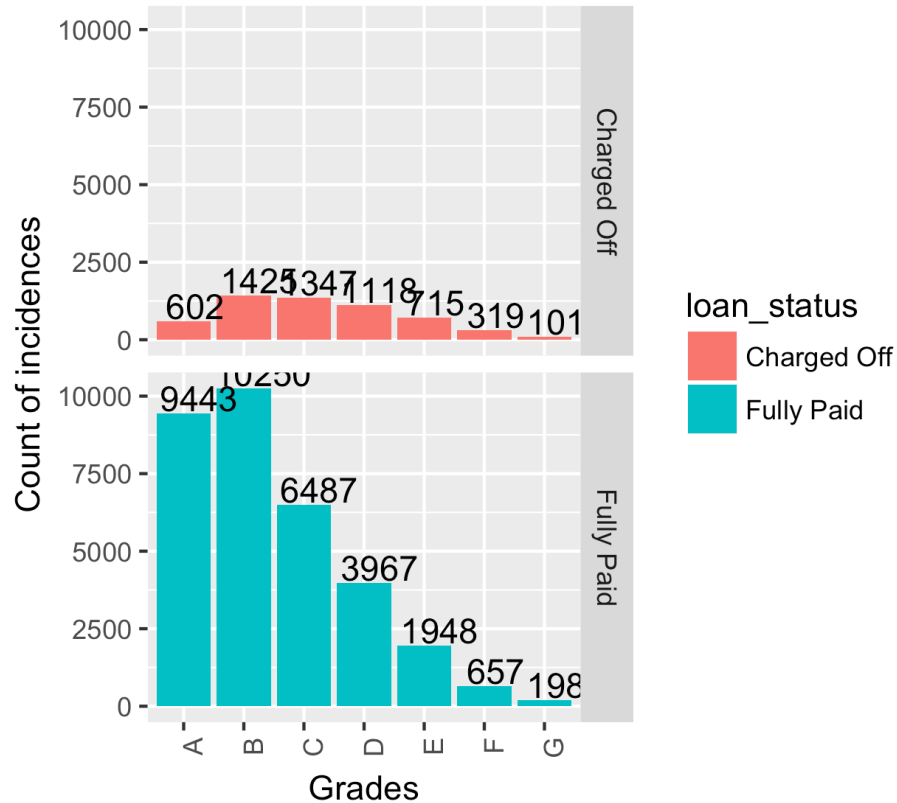
Verification status analysis

Count of verification status w.r.t. Loan Status

- Customers with purpose debt consolidation are higher in number even for lower DTI and funded amount.
- Customers with purpose – 'Credit card' or 'Small Business' & higher DTI – even with lower or medium Funded amount # are having higher proportion of charge-offs. Probably due to higher debt burden on them than their paying capacity.

Funded amount Vs Debt to Income Ratio w.r.t

- It is seen that there is a dip in the number charge off # when the loan applicant's income is source verified.
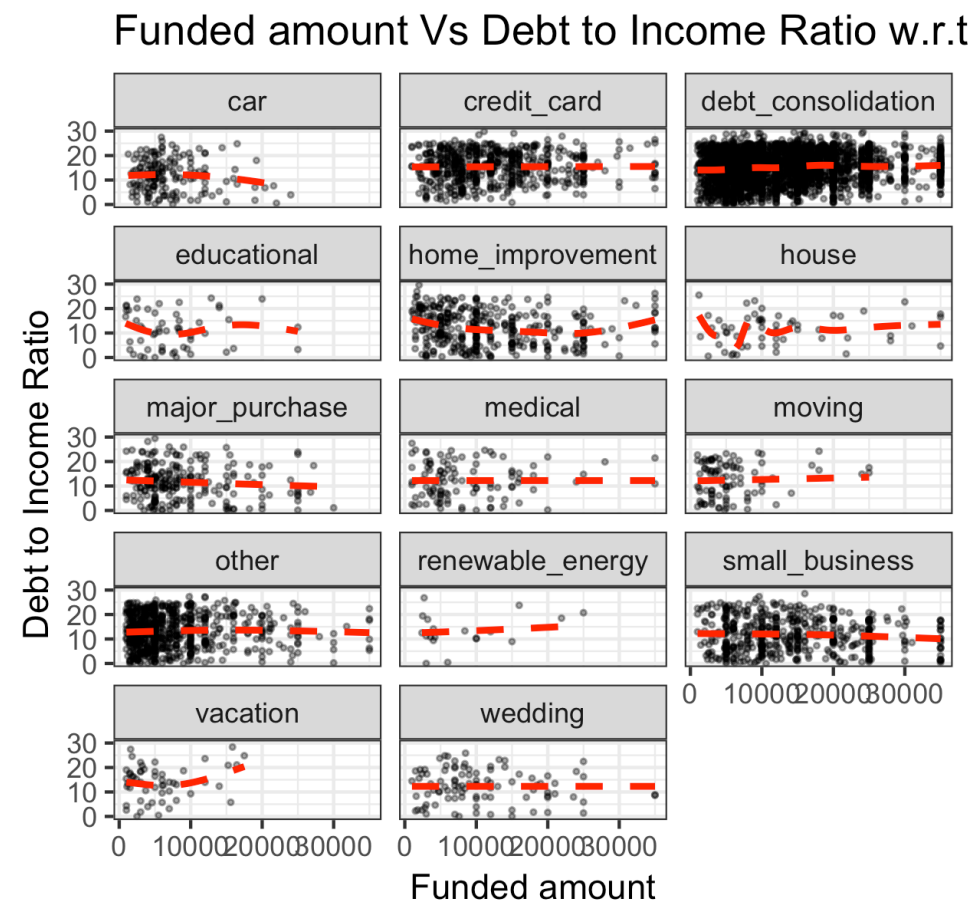
Count of loans with various grades w.r.t.loa

- Loan amount is higher for higher grades i.e. for Grades E, F, G.

Loan amount by grades across years

- Total number of Loans under Grade "B" is high when compared with loans under other grades.
- The count of charged off loans is maximum under Grade B, followed by Grade C & Grade D respectively.
- The count of Fully Paid loans is maximum under Grade B, followed by Grade A & Grade C respectively.

Employee length analysis
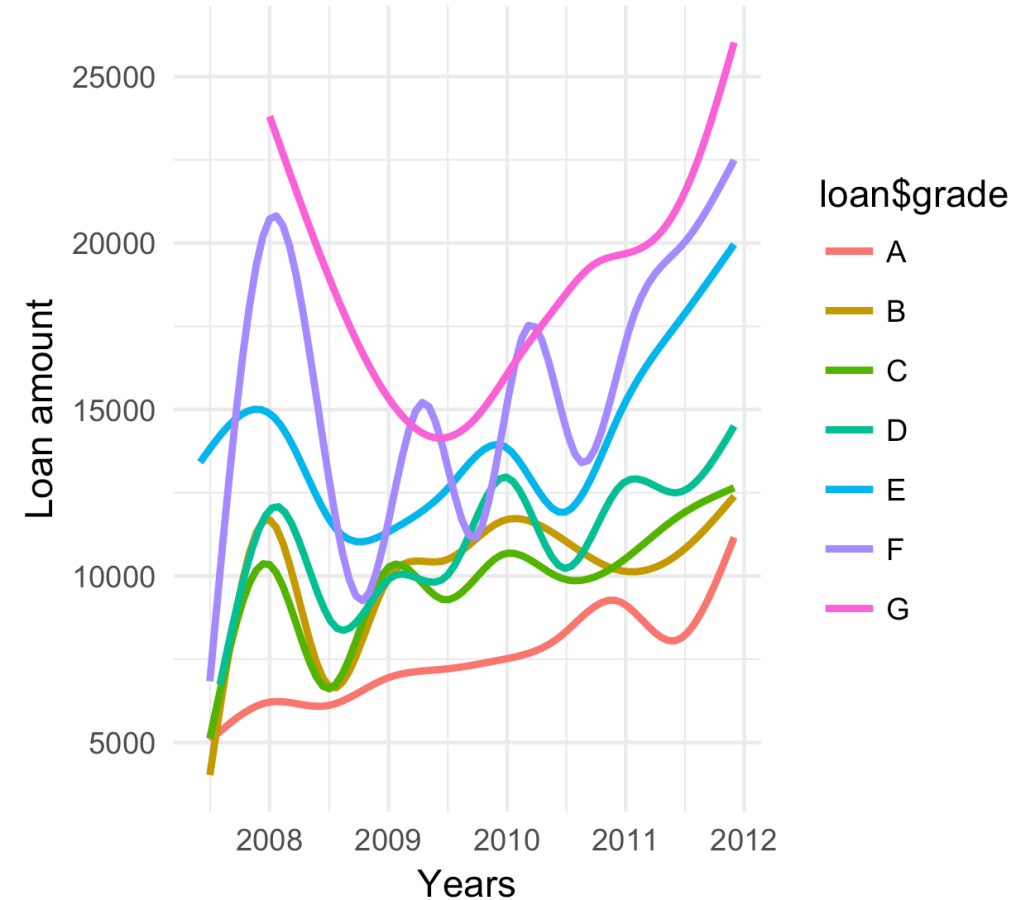
Range of emp length w.r.t.loan status
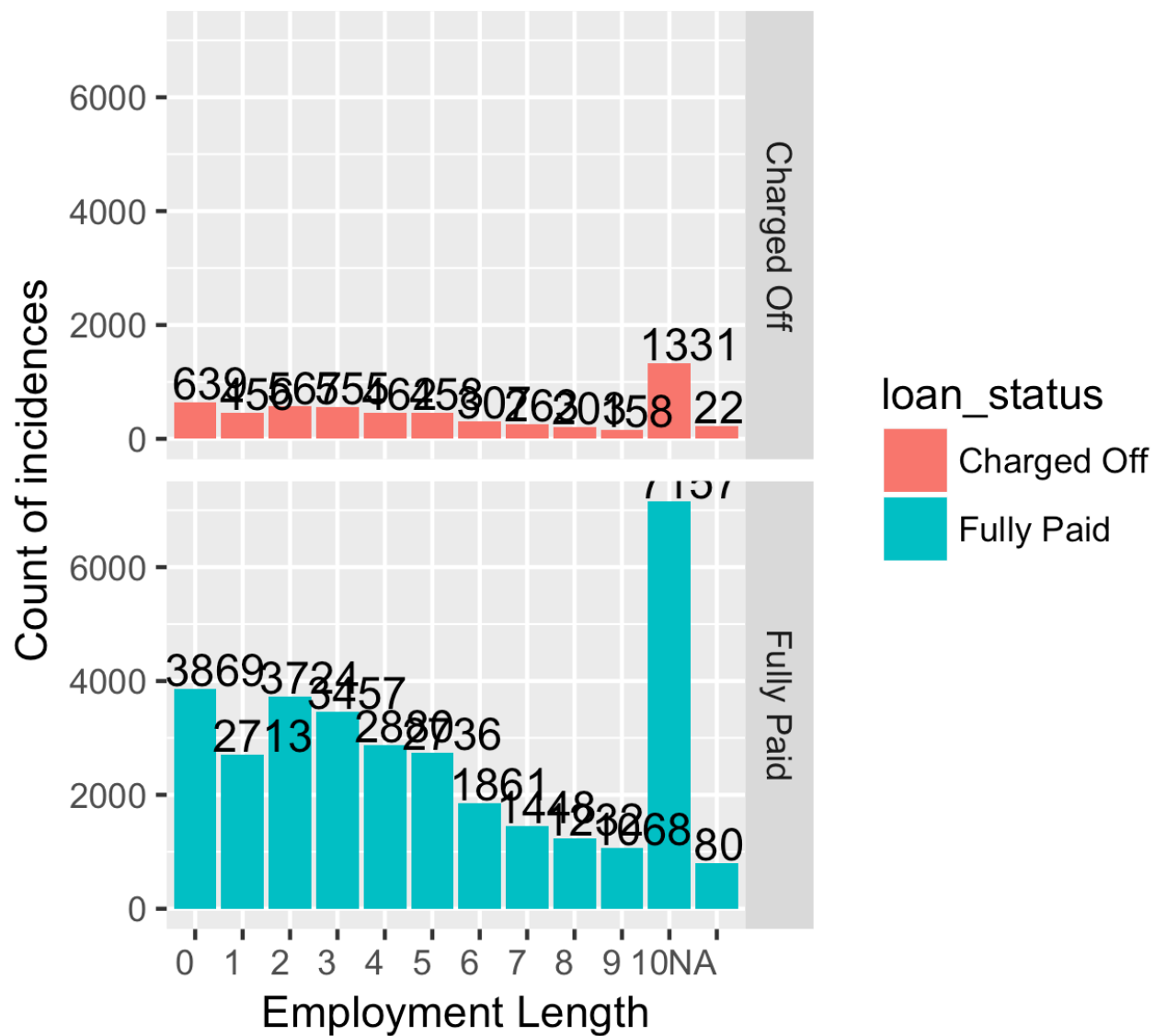
- Total number of Loans for applicants with employment length of 10 years or more is high.
- The count of charged off loans for applicants with employment length of 10 years or more.
- The count of Fully Paid loans for applicants with employment length of 10 years or more.
- There are around 1200 applicants whose employment length is not available.

Distribution of dti w.r.t. loan status
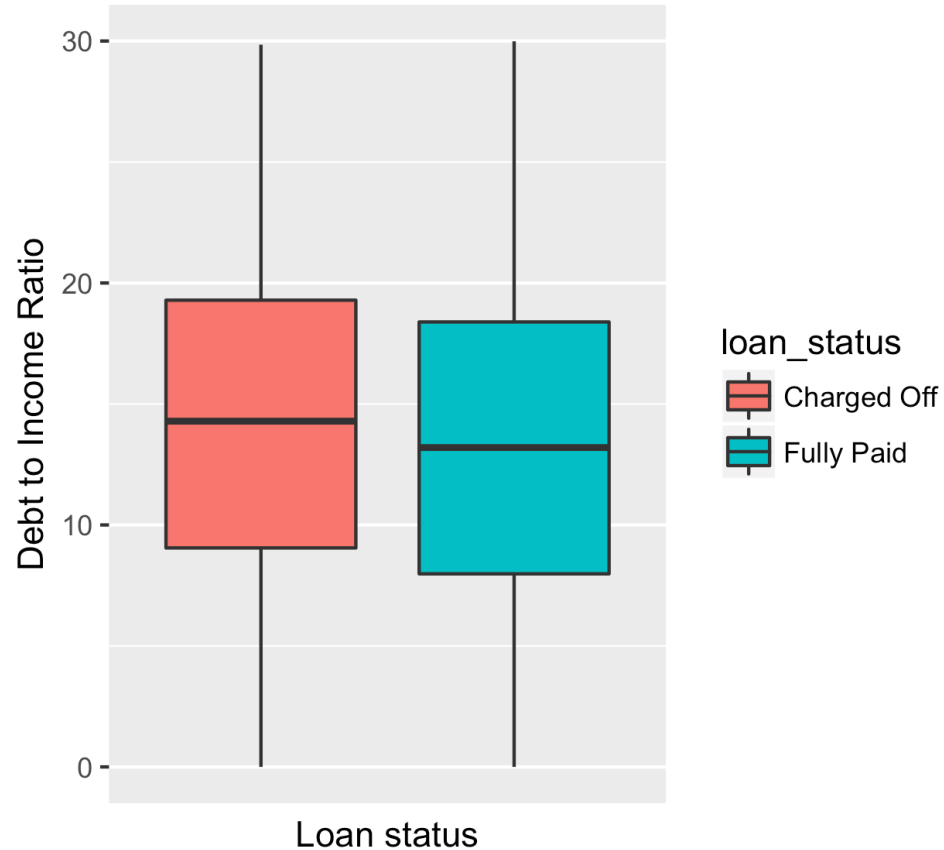
- Customers with purpose debt consolidation are higher in number even for lower DTI and funded amount .
- Customers with purpose – 'Credit card' or 'Small Business' & higher DTI – even with lower or medium Funded amount # are having higher proportion of charge-offs. Probably due to higher debt burden on them than their paying capacity.

Funded amount Vs DTI w.r.t. Purpose

- For charged off loans the median DTI is comparatively higher than 'Fully paid'.
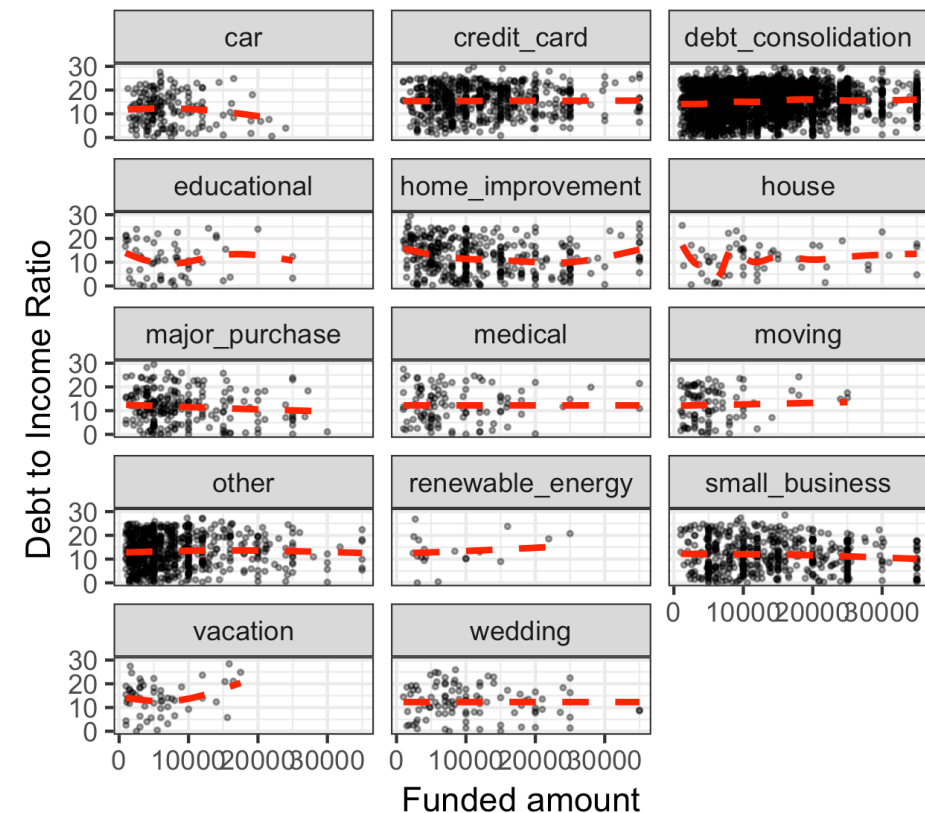
Revolving Utilization % analysis

## Distribution of revolving utilization % w.r.t. lc

loan_status
- Charged Off
- Fully Paid
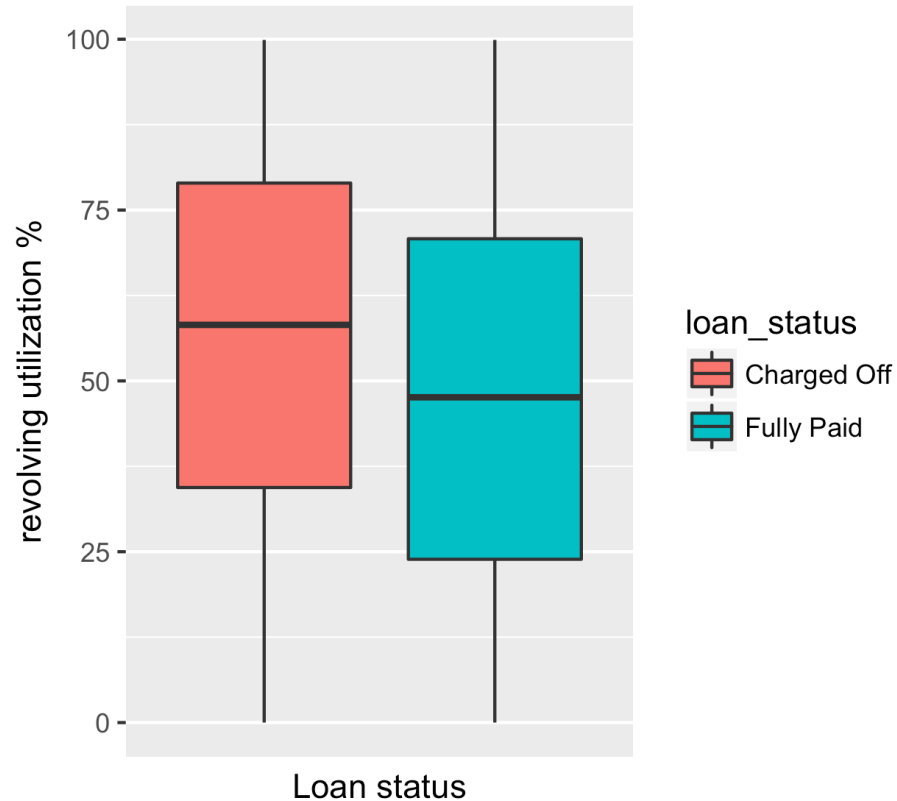
- Higher the resolving utilization, higher is the charged offs.

- Revolving Utilization analysis for top 5 states in terms of number of charge offs shows that it is higher for higher Annual income customers - typically for state Florida. However, still charge offs are concentrated towards lower income and Revolving Utilization.
- Most of the charged off cases are concentrated in lower annual_inc & lower revol_util baring a few cases where a fewer number of higher or medium annual_inc customers are also contributing to charged-offs.

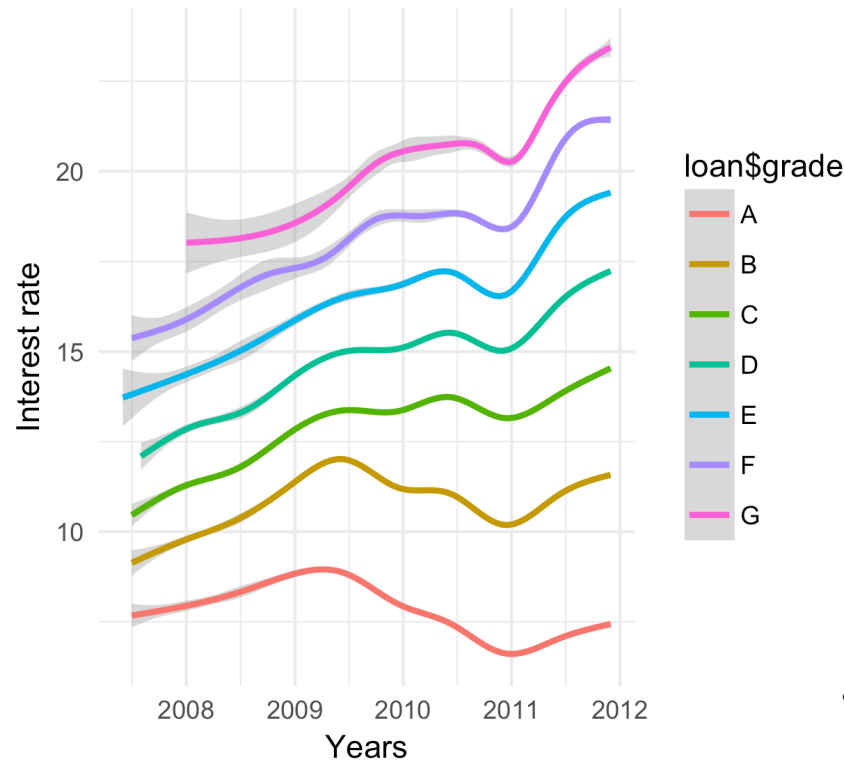## Annual Income vs Revoling Utilization %

Interest Rate analysis
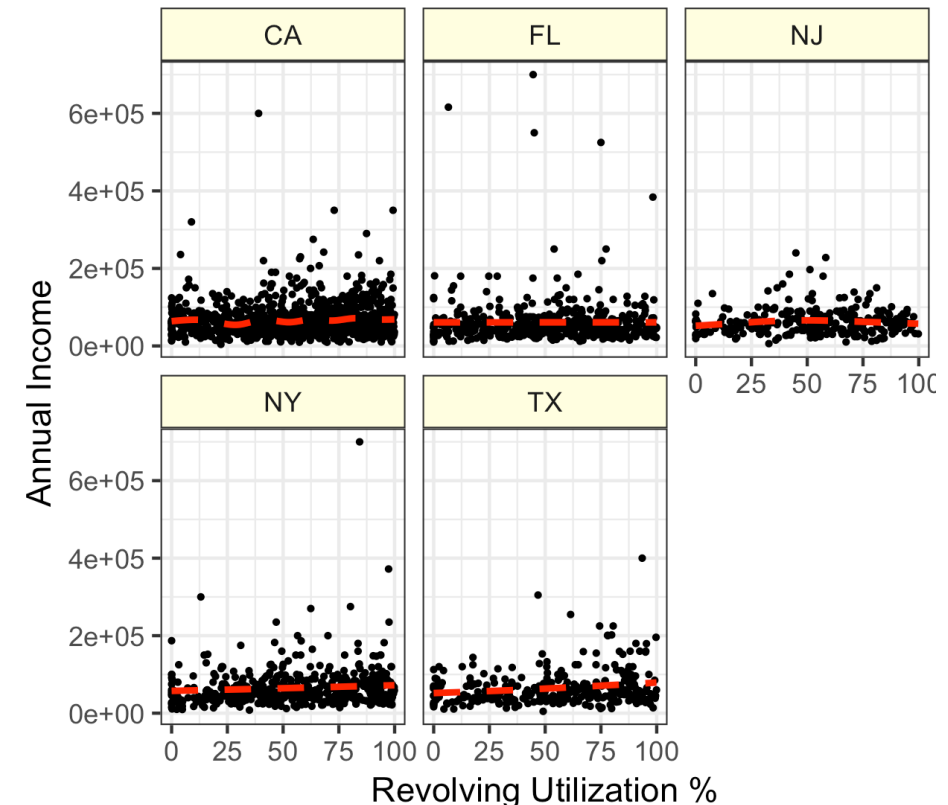
Distribution of interest rate w.r.t.loan status

Interest rate vs grade across years

Annual Income Vs Interest rate w.r.t.Term
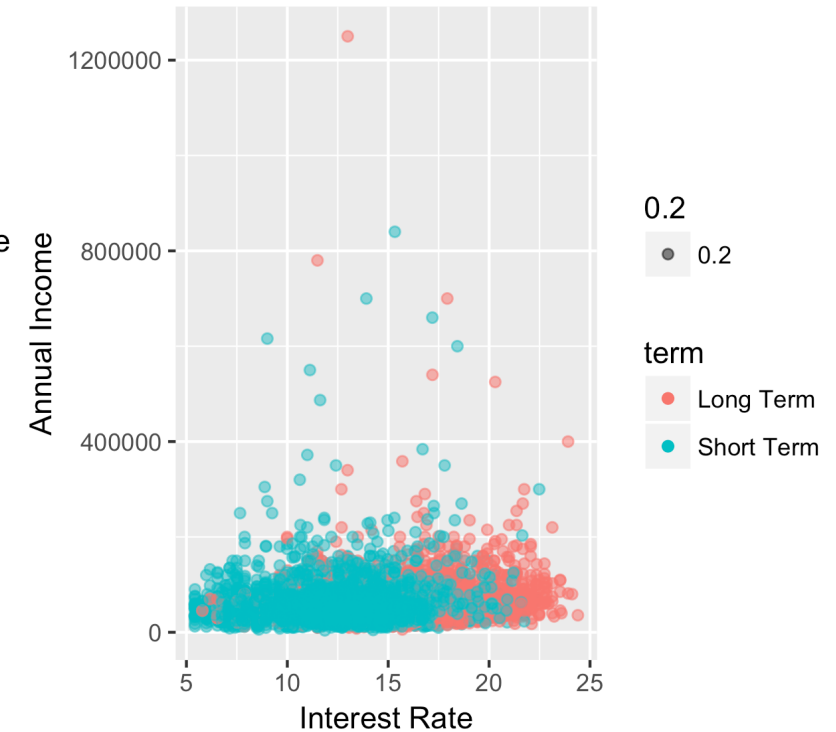
- The rate of interest for charged off loans is higher compared to Fully Paid loans

- Interest rate is higher for higher grades i.e. for E, F, G.

- Most of the long term loans(60 Months) are falling in higher interest rate for lower Annual income. This may be leading to higher instalments and may be contributing to charge-offs.
- A significant proportion of short term loans(36 Months) are under charge off category for lower annual income and medium & lower interest range. This may be due to improper income verification of the customers.

Funded amount analysis

Distribution of funded amount w.r.t.loan sta

Funded amount by term

Interest rate vs funded amount and term
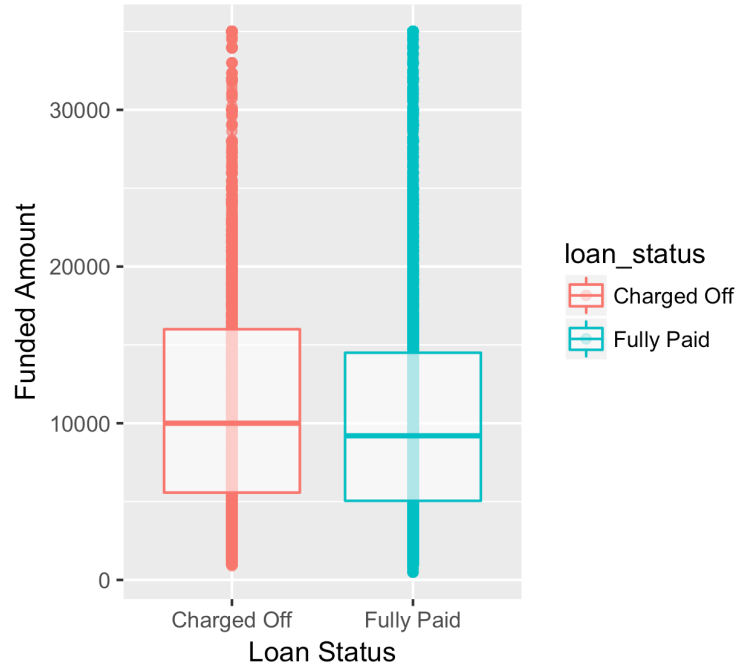
- Funded amount for charged off loans is slightly higher compared to Fully Paid loans.

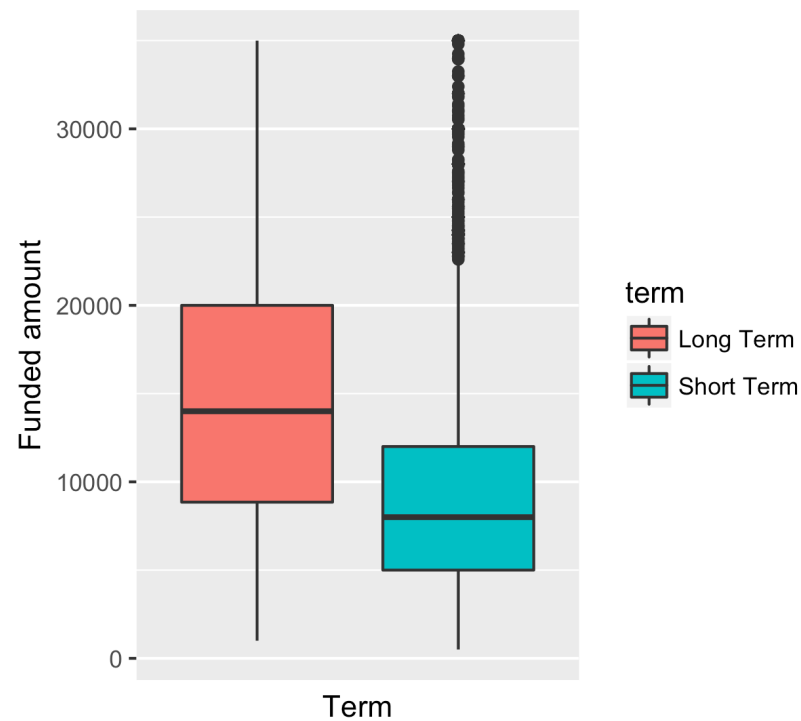- For 36 Months Term, median funded amount is generally lower than for 60 Months Term. There are some outliers for 36 months term
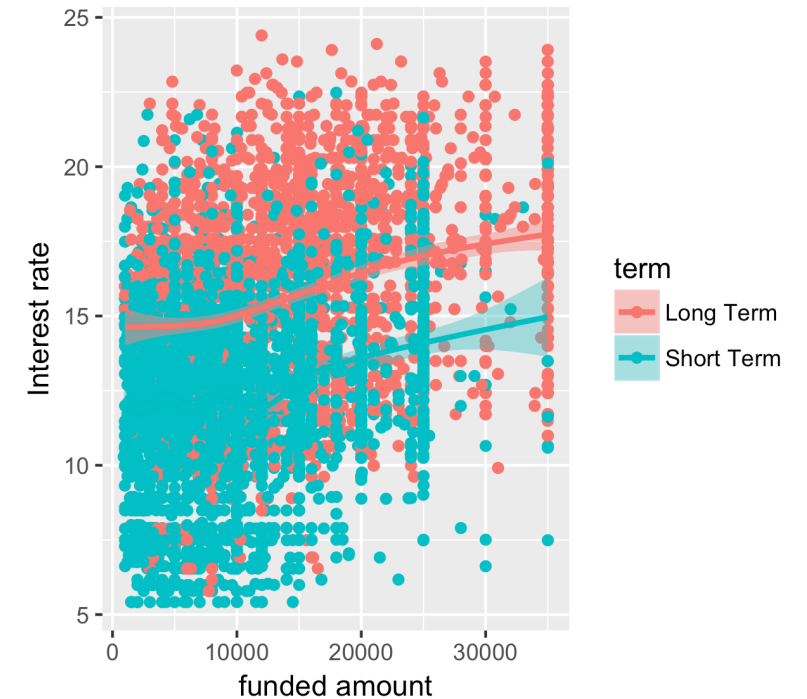
- Most of the long term loans(60 Months) are falling in higher interest rate even for lower funded amount. This may be leading to higher instalments and may be contributing to charge-offs.
- A significant proportion of short term loans(36 Months loans) are under charge off category for lower funded amounts and medium & lower interest range (up to 15%) – this may be due to improper income verification of the customers.

Conclusions

- **Debt consolidation & Credit card** are most popular reasons for borrowing and contribute to the highest number of loan charged off. Funded Amount and Installment together are causing higher charged-offs for **Debt consolidation, Credit Card and Small business**. Number of Loans taken for **Debt consolidation** purpose is very high in comparison to loans taken for all other purposes.

- CA(California), NY(New York), TX(Texas), FL(Florida) and NJ(New Jersey) are the **top 5 states** in terms of borrowing and are the highest contributors in terms of number of Charged offs. Number of Loans taken from the state of **California (CA)** is very high in comparison to loans taken from all other address states. Number of charged offs from the state of **California (CA)** is high in comparison to all other address states.

- Customers with **'RENT' & 'MORTGAGE' Homeownership** contribute higher in terms of number of charged off cases wherein those with homeownership as **RENT** are the highest. For **'Rented' Homeownership**, Interest **Grades B & C** are proportionately higher for loan charge-off in comparison to other grades. For **'Mortgaged' Homeownership**, Interest **Grades B,C,E** are proportionately higher for loan charge-off in comparison to other grades. The count of charged off loans is maximum under **Grade B**, followed by Grade C & Grade D respectively. Loan amount is higher for higher grades i.e. for Grades E, F, G.

- Surprisingly, the **proportion** of charge offs are **higher for 'verified' cases than not verified** . **Source verified** also contribute reasonably. Most of the charge off are between the medium interest range (approx. 9-17%) leading to suspicion that verification was not done properly . It is seen that there is a dip in the number charge off when the loan applicant's income is source verified.

Conclusions

- For 'Charged off' loans, the **median DTI** is comparatively higher than 'Fully paid' giving an insight the evaluation of loans' risk and debt management at Consumer finance company needs improvement. Customers with purpose debt consolidation are higher in number even for lower DTI and funded amount . Customers with purpose 'Credit card' or 'Small Business' & higher DTI are having higher proportion of charge-offs even with lower or medium Funded amount . Probably, due to higher debt burden on them than their paying capacity.

- A significant **proportion of lower term(36 Months loans)** are under 'Charge off' category for lower funded amounts, and medium & low interest range (up to 15%). This gives an insight that the Consumer finance company may be inefficient in pricing loans in accordance with associated risk. A significant proportion of short term loans(36 Months) are under charge off category for lower annual income and medium & lower interest range. This may be due to improper income verification of the customers. Number of "Charged off" loans are more in short term loans(i.e. 36 months) in comparison to long term loans(i.e. 60 months). Most of the long term loans(60 Months) are falling in higher interest rate for lower Annual income. This may be leading to higher instalments and may be contributing to charge-offs.

- Total number of Loans for applicants with **employment length of 10 years or more** is high. The count of charged off loans for applicants with employment length of 10 years or more

- **Revolving Utilization analysis for top 5 states** in terms of number of charge offs shows that it is higher for high Annual income customers; typically for state Florida. However, still the charge offs are concentrated towards lower income and Revolving Utilization. Most of the charged off cases are concentrated in lower annual_inc & lower revol_util baring a few cases where a fewer number of higher or medium annual_inc customers are also contributing to charged-offs.

- Consumer finance company should make verification (including source verification) stronger before issuing loans for across all annual income ranges .(not only for high income ranges) .

- In Top 5 states it is observed that borrowing is higher with higher number of charge-offs – Consumer finance company should focus on optimizing the pricing of loans (interest rates) – neither very high or nor very low .

- Consumer finance company should consider focus on verifying DTI of a customer before issuing loans- since consumer with high DTI are more prone to charge-off. Also, small business owners are proportionality more contributors to charge off – there DTI should be matched in particular.

- Since there is a significant proportion of charge-off for loans of shorter term(36 months) and lower funded amounts at lower interest rates – such loans should be more competitively and with better verification .

- Flexible payment plans should be provided to customers with high funded amount and instalment.