



EXPLORATORY DATA ANALYSIS

SUBMISSION

Group Name: Explorers

1. Kartik Porwal
2. Keerthika Nandagopal
3. Rahul Kumar
4. Viveka Manikonda



Business Understanding and Objectives

Business Understanding

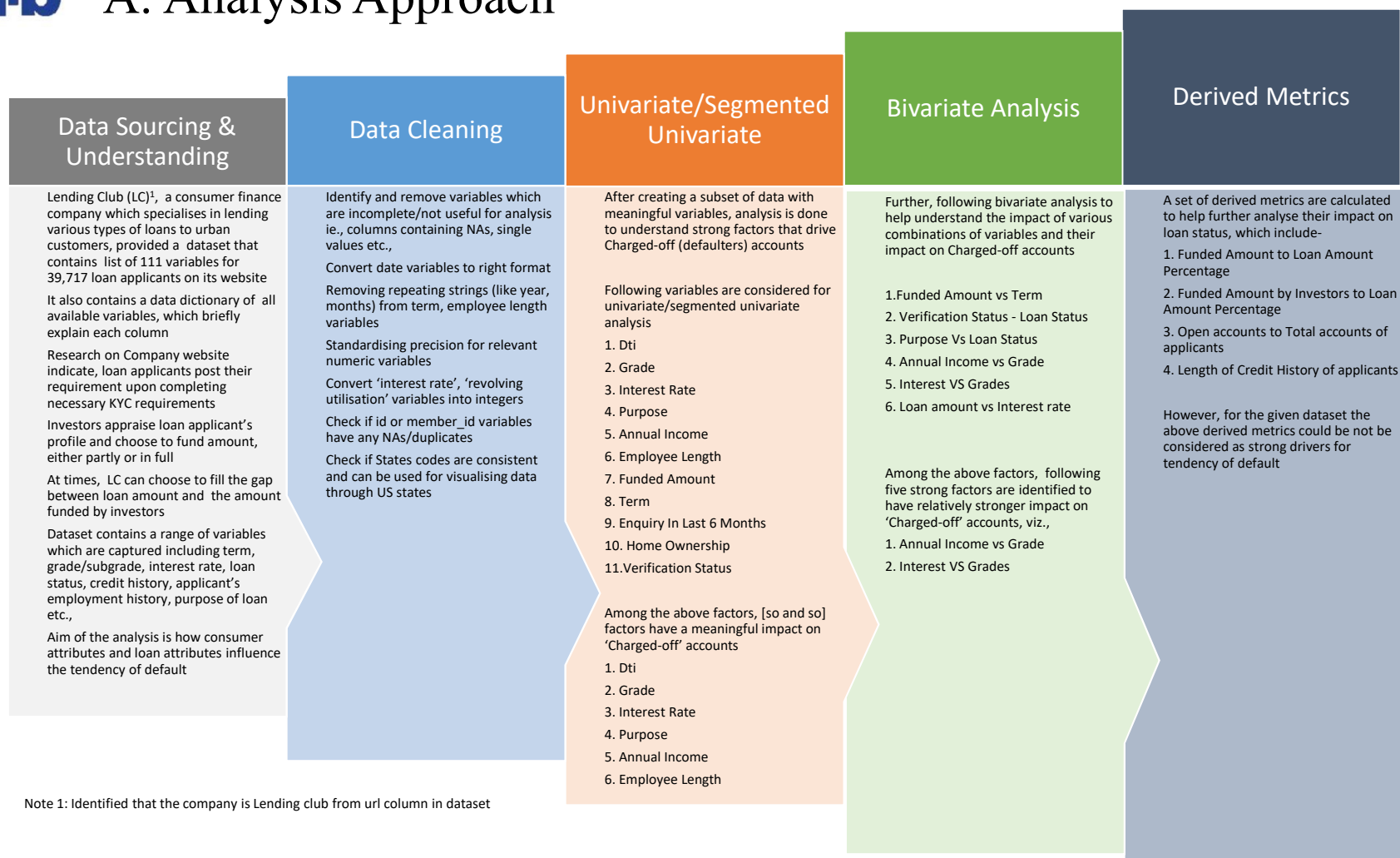
A consumer finance company facilitates personal loans, business loans, and financing of medical procedures. Borrowers who **default** cause the largest amount of loss to the company.

Business Objective

- To understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default.
- The Company can use this for identifying risky loan applicants and rejecting such loans in first place.

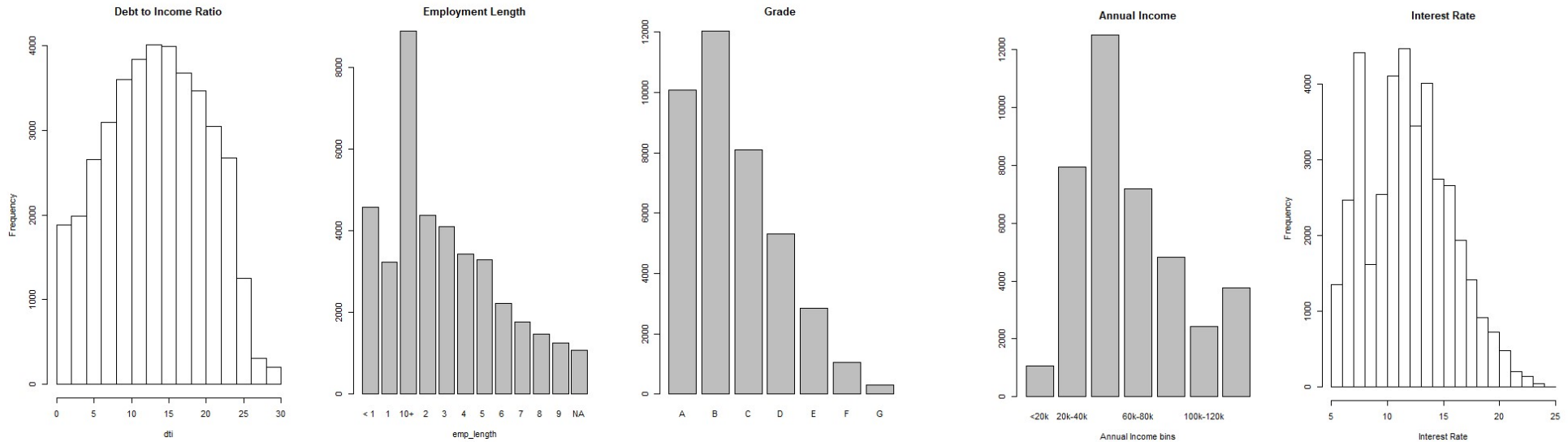
Data – Loan data for all loans issued through the time period 2007 to 2011

- 39717 sanctioned loans
- 5627 loans are **charged off (i.e defaulted)**



Note 1: Identified that the company is Lending club from url column in dataset

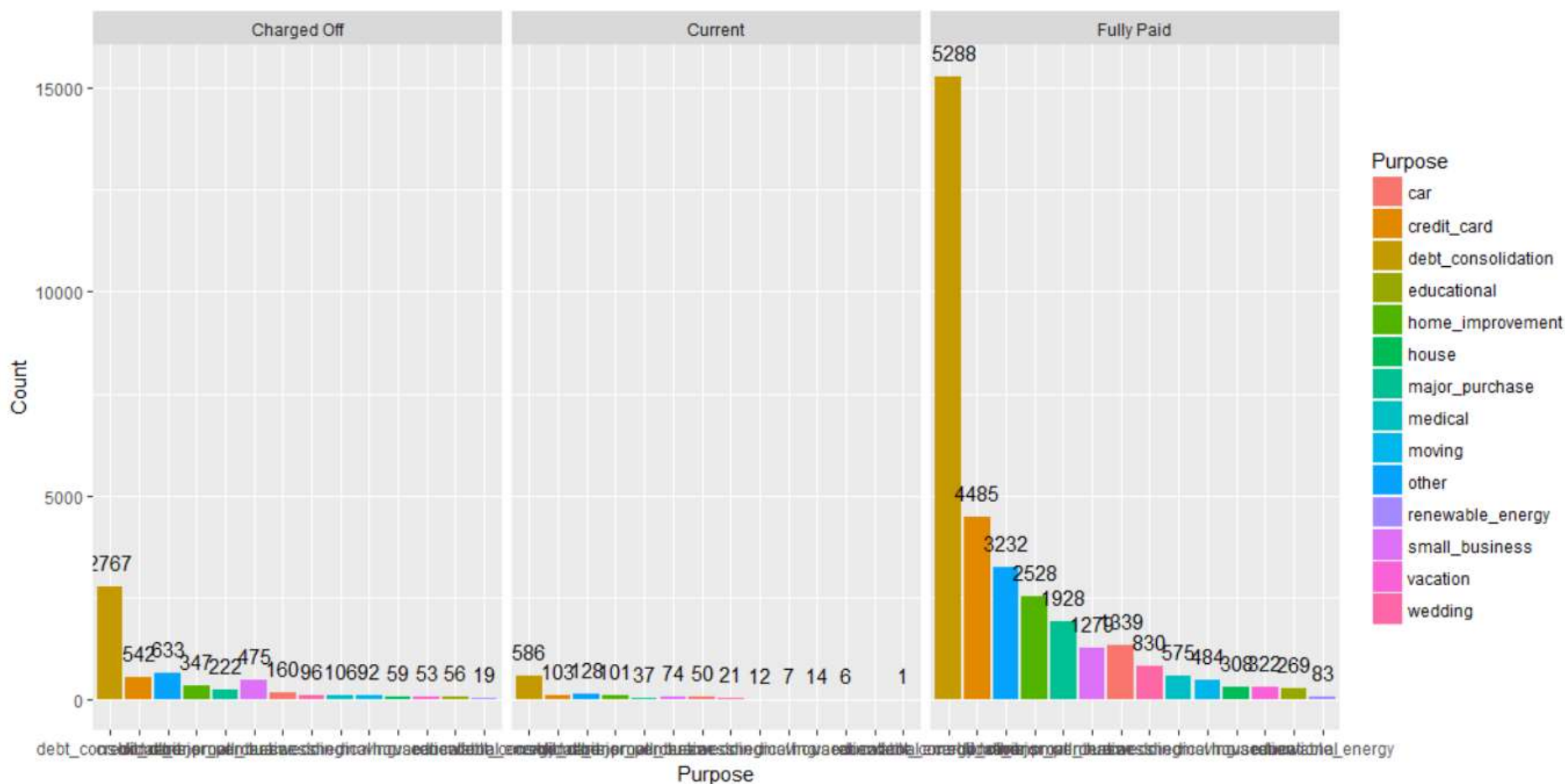
B. Overview of Key Drivers of Charged Off Accounts



- Debt Income Ratio – Ratio indicates amount of existing debt owed by applicant to his income. Higher the ratio, higher the risk associated with customer, however, there are no applicants with more than 30% Debt Income Ratio
- Employment length – Employment length of applicant could indicate his/her position in his career and indicate stability of his cash flows
- Grade – Dataset indicate it ranges from A (Best) to G (worst) and is indicative of applicant's credit history
- Annual Income – Represents annual income of applicant and higher the number, better is the applicant to meet his debt obligations
- Interest Rate – Interest Rate is reflective of applicant's credit history and borrowing purpose. Higher the interest rate, greater is the amount required paid towards interest

B. Overview of Key Drivers of Charged Off Accounts [Contd]

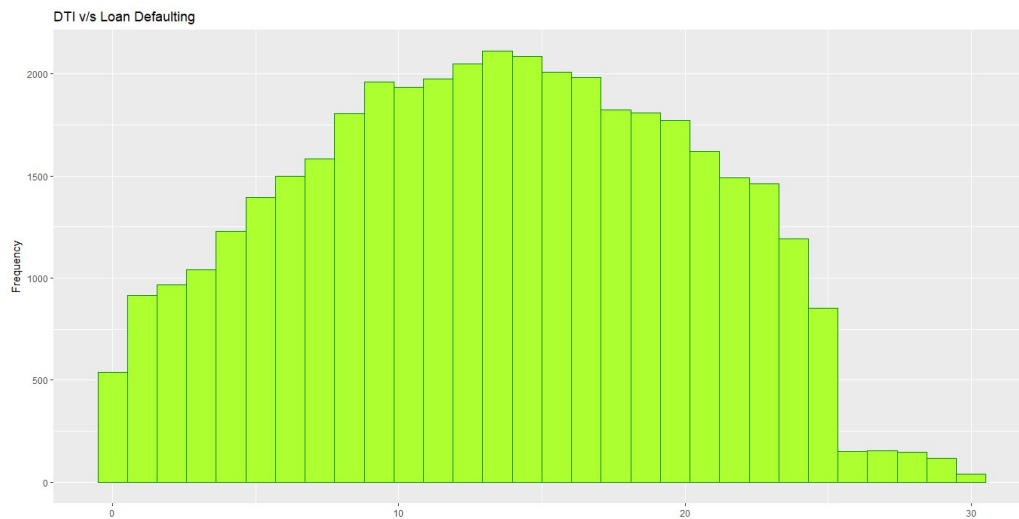
- Debt consolidation is most popular purpose amongst applicants





C. Detailed analysis on individual drivers

C1. Debt to Income Ratio



As per the DTI vs Charged off plot,

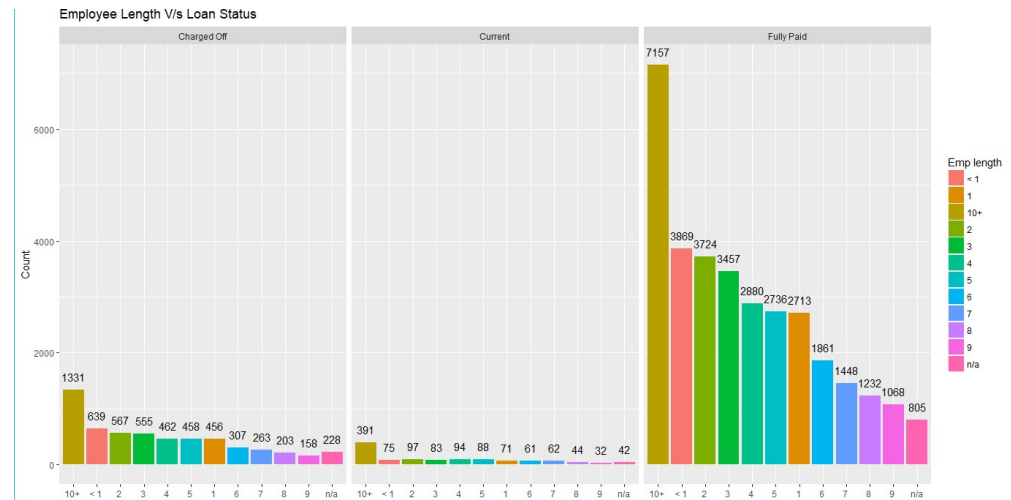
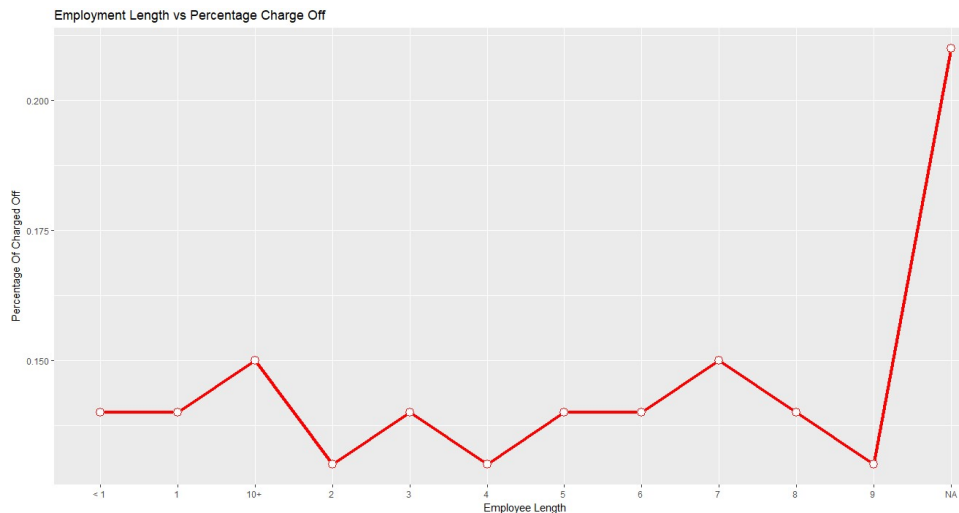
DTI is the ratio calculated using the borrower's total monthly debt payments on the total debt obligations, divided by the borrower's self-reported monthly income.

The charged off count increases gradually from 2 to 14 range of DTI is high from 14 to 22 then gradually decreases till # 25 after which is very low

Conclusion: DTI seems to be one of the factor for the customer to default the loan.

C. Detailed analysis on individual drivers

C2. Employment Length



In order to understand the relation between the trend of the Employee Length v/s Charged off Loans,

Quantitative Analysis : As per the plot on the right based on the count, the max charged off loans is for the customers with 10+ years followed by 1-5 years of experience.

Trend Analysis : Understanding the plot on the left, which shows the trend similar to that of the count plot, but has a high steep towards the right which is N/A's.

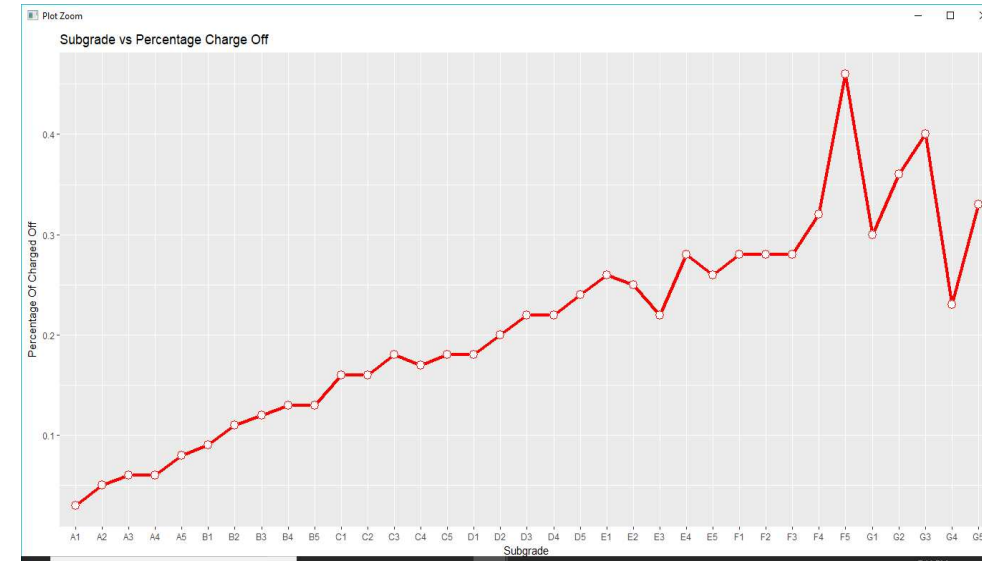
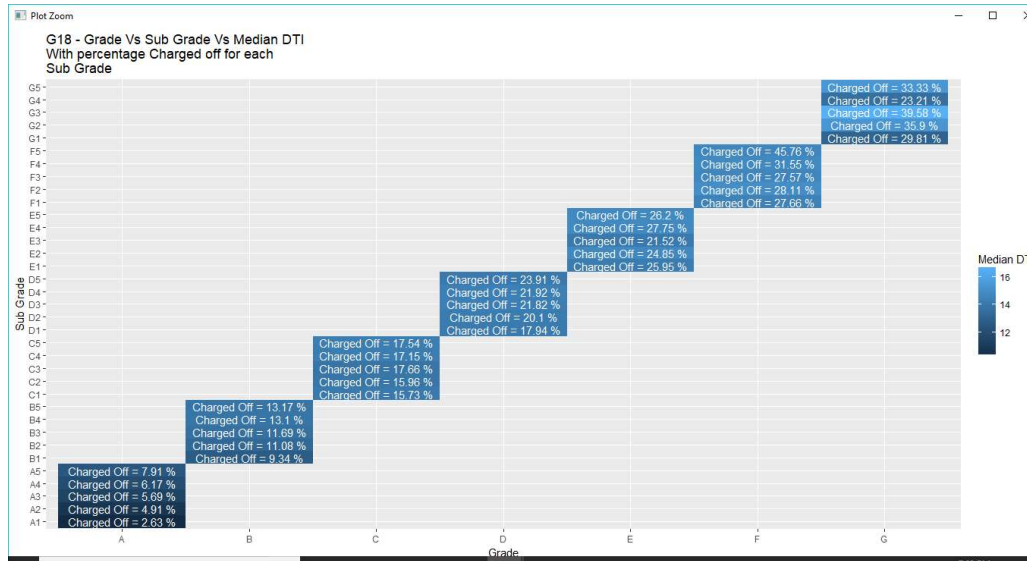
N/A's were not treated as part of this analysis, since it might affect the crisp of analysis if we change it to the median so Na's have been kept the way it is for this analysis.

Conclusion : The number of default loans are highest for the experience people with 10+ years, also there is a fine amount of loan defaults for people with experience 1-5 years



C. Detailed analysis on individual drivers

C3. Sub Grade



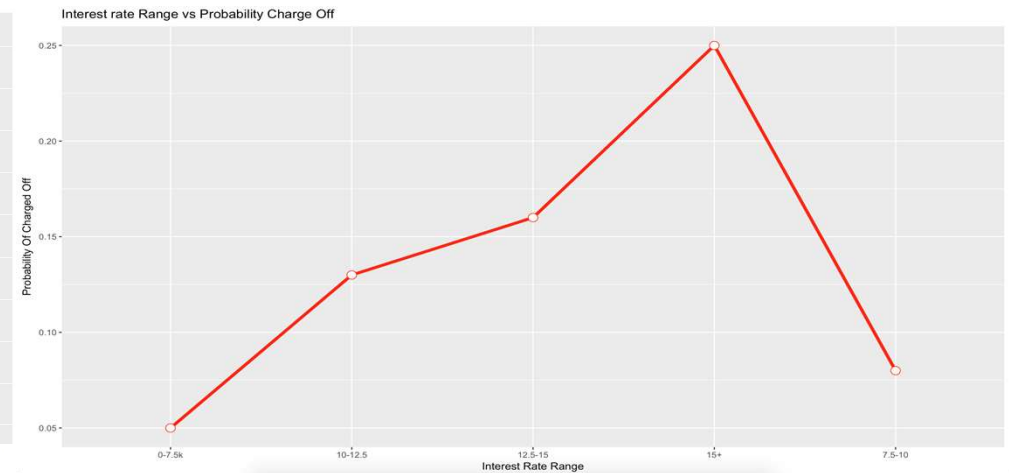
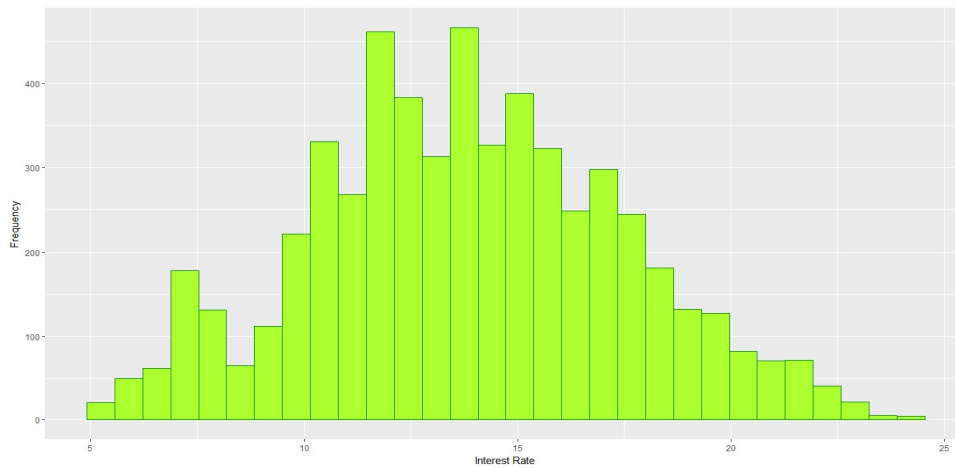
While analyzing the subgrade factor, the percentage of charged off loans increases with the increase in Grade/subgrade assigned to the customer.

For the plot on the left the Grade v/s Subgrade along with the DTI clearly shows the % of the charged of loans along with the trend which has been plotted on the graph on the right.

Conclusion : With the increase on Grade the probability of the loan to be charged of also increases.

C. Detailed analysis on individual drivers

C4. Interest Rate Analysis



By visualizing bar graph most charged off customers are in range of 10-20%.

After plotting line graph between bin interest rate and percentage of charged off we are able to see percentage of default person is increasing with increase in the interest rate with highest of 25% at 15%+.

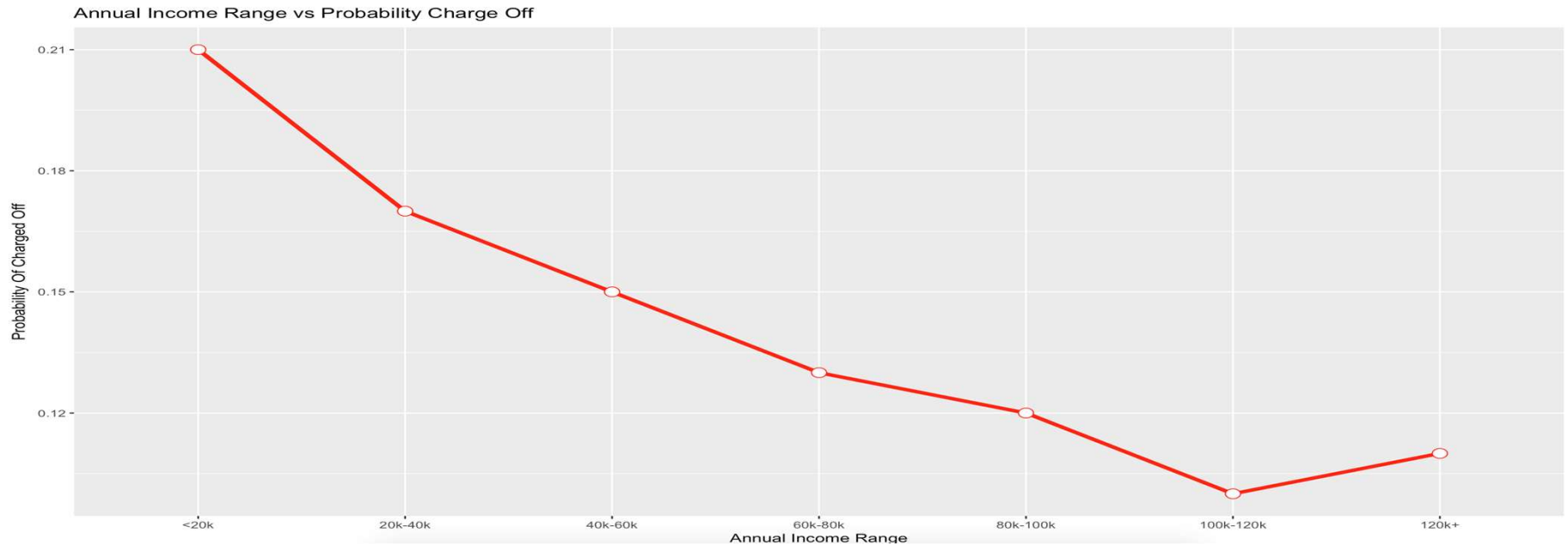
So, there is risk factor where bank have to careful before sanctioning loan for interest rate more then 15%.

Cocclusion:

Bank should decrease the interest rate or prefer to give loan to customers who have more annual income.

C. Detailed analysis on individual drivers

C5. Annual Income



As highest of 21% at (0 to 20000) salary bracket, the annual income decreases there are chances that person will go on default.

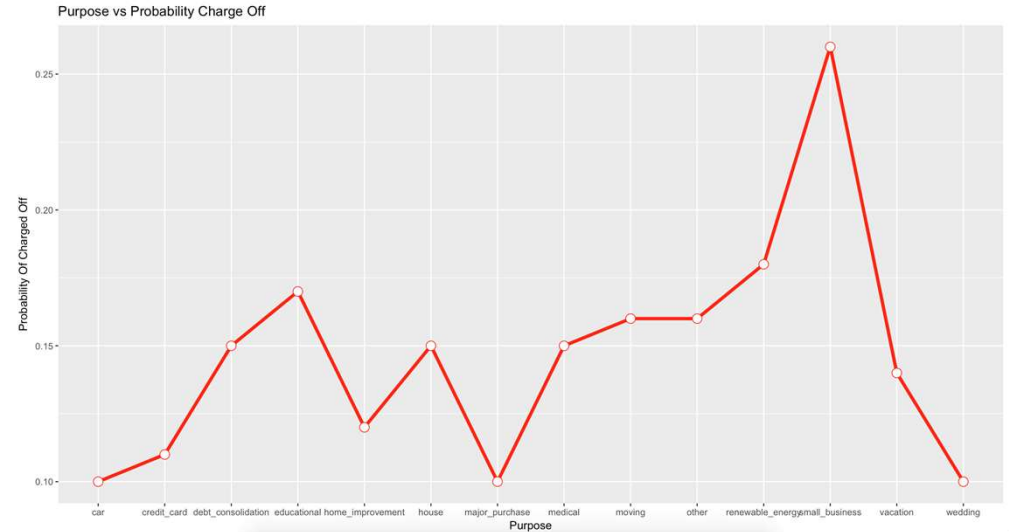
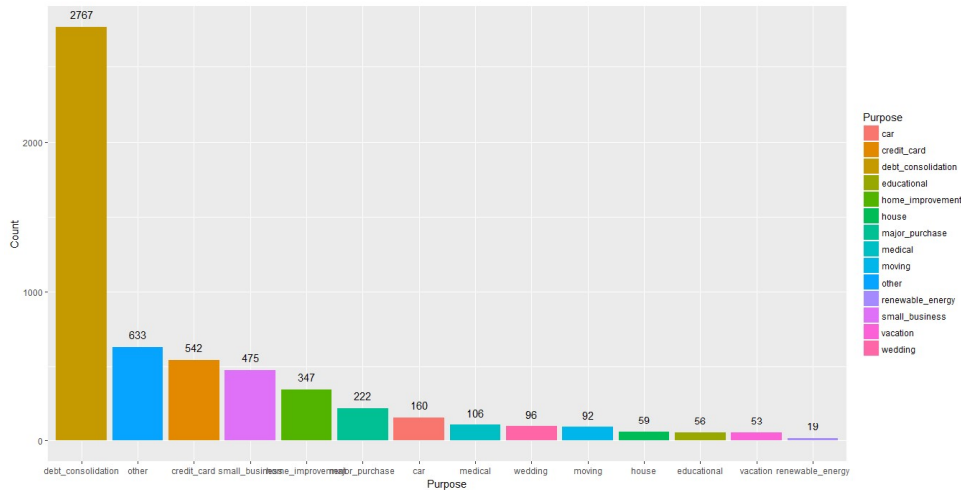
So, there is risk factor where bank have to careful before sanctioning loan if customer have annual income less than 20000.

Conclusion-

Loan amount is less for this category so, Bank always check customers purpose of loan for 20k or less annual income, why they are taking and how they will repay .

C. Detailed analysis on individual drivers

C5. Loan Purpose



Loan Purpose is risk factor where we are able to see customers count is more for "**debt consolidation**" then others which is 2767 more than total of all other purposes .

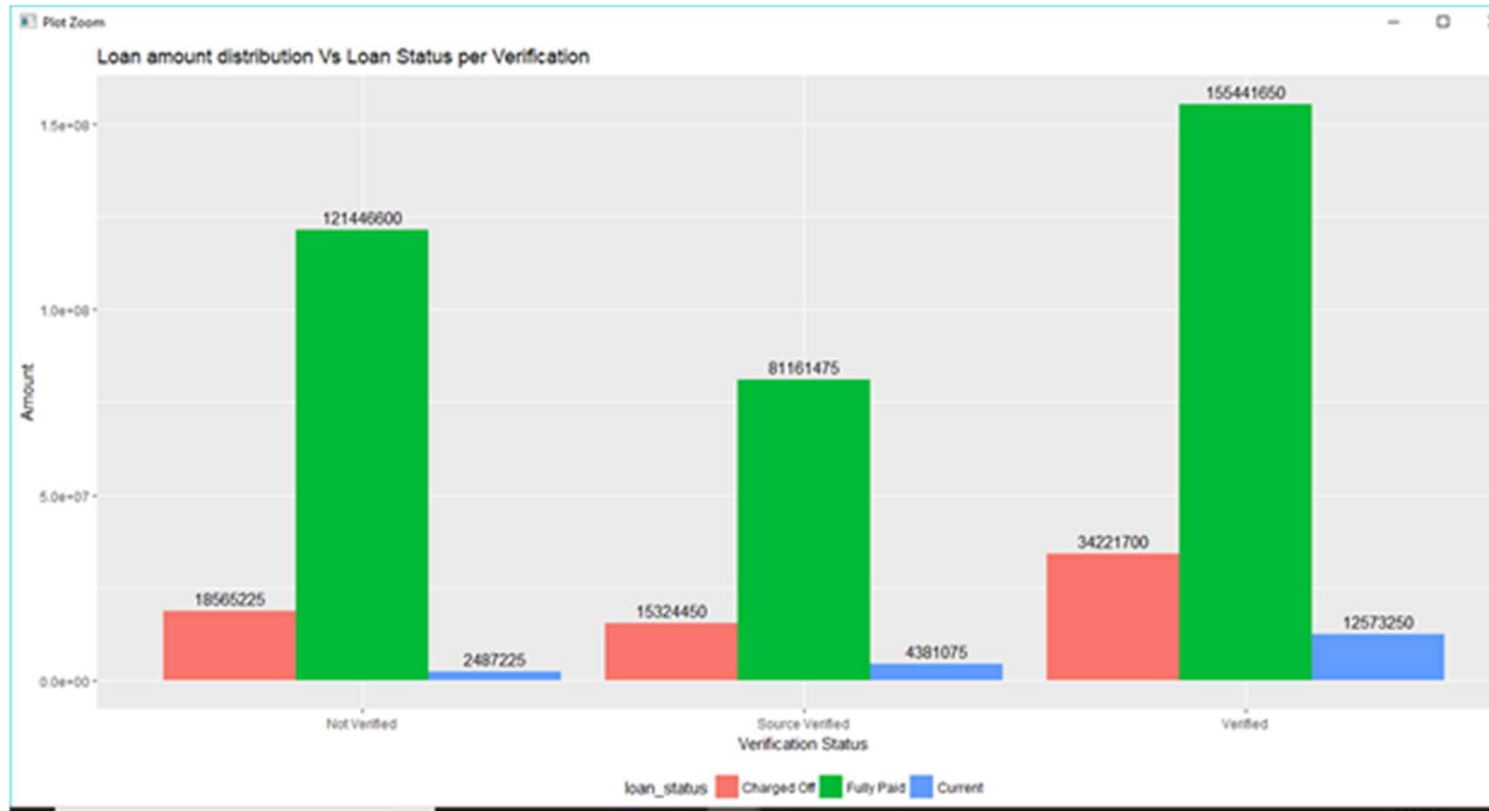
But when we check the percentage of customers are defaulter in line graph we are able to visualise applicants who has taken the Loan for '**small business**' has the highest percentage of charge off of 26.5%.

Conclusion

Bank should take extra caution like take some asset or guarantee while approving the loan for purpose of '**small business**'

D. Bi-variate Analysis

D1. Loan amount distribution by Verification and Loan Status



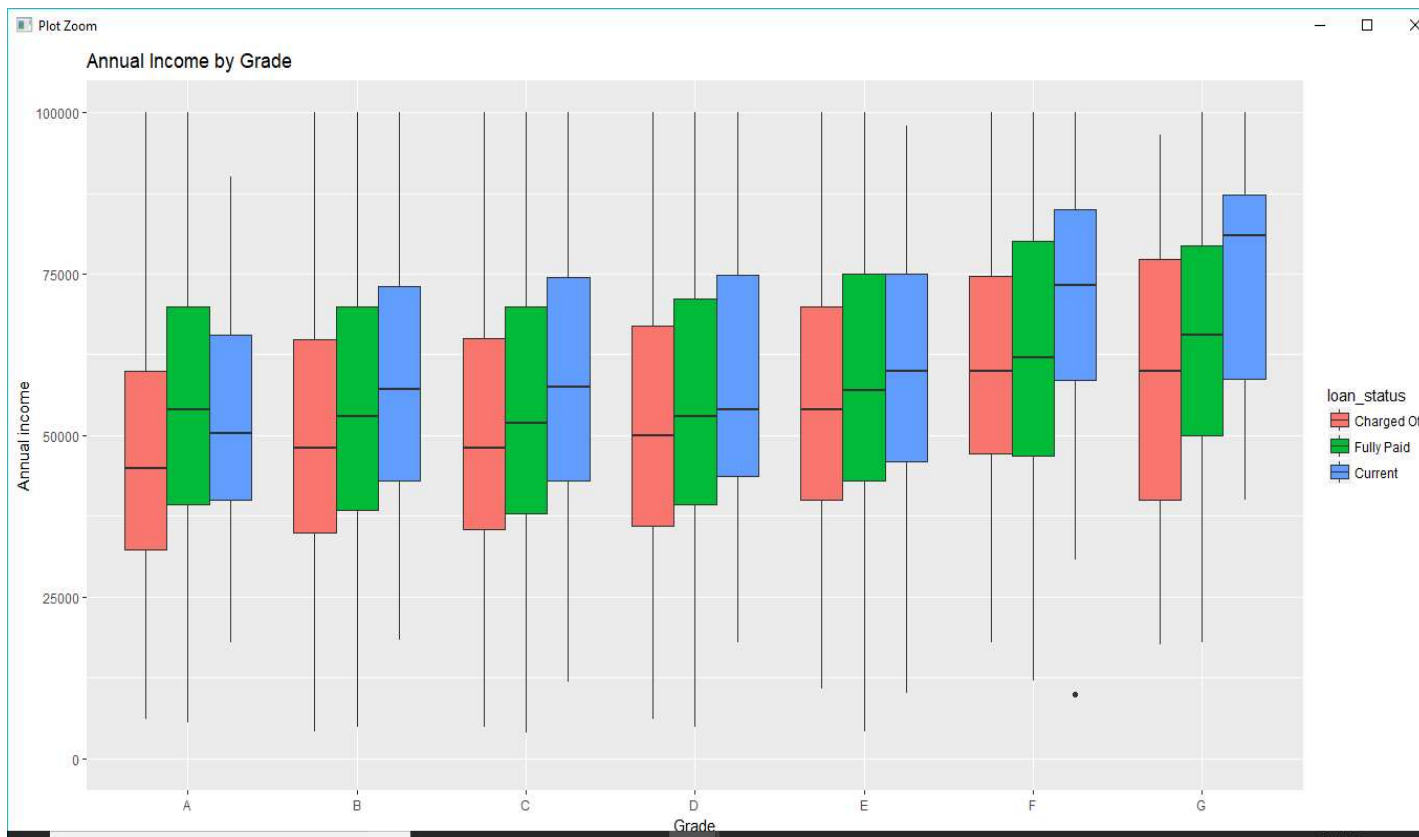
This chart indicate that Charged off accounts are highest in 'Verified' status, followed by 'Not Verified'

Conclusion - Verification Status cannot be confirmed as strong driver for loan defaults



D. Bi-variate Analysis

D2. Annual Income by Grade & Loan Status



On analysing the Annual income by grade for the loan default, even if the annual income is higher for the highest grade(G), the chances of loan being charged off is pretty high.

However, for the lowest grade(A) , the no of charged off with high annual income is low.



E. Conclusion

- Debt Income Ratio, Employment length, Grade, Annual Income, Interest Rate and Purpose are strong drivers and influence tendency of default
- Bivariate analysis indicate there could be other supporting factors like Funding Amount, enquiry in last 6 months etc., which play an important role and influence default tendency
- These factors could be used to identify customers who pose high risk of default and hence some preventive measures can be taken to reduce risk or curtail potential losses
- Further, analysis has to be updated as and when there is more data available, to confirm the existing key drivers or add any other variables