



Lending Club Case Study

Agenda

- Problem Statement
- Solution Approach
- Observation and Recommendation

Problem Statement

Lending Club, largest online loan marketplace, for its portfolio and risk assessment wants to understand the key driving factors behind the loan default which leads to credit loss.

For this problem statement, a case study is done by analysing the Loan dataset provided so as to find key driving features which identifies the defaulters (charged off).

Case Study - Approach

1. Understanding the data –

- Analyzed various features (at high level) present in the dataset with help of data dictionary and selected the features which would be used for further analysis

2. Data cleaning –

- Found the null values % and dropped the features with null% more than 50
- Removed the redundant or non-relevant columns for analysis (url, title, desc etc.)
- Used mode for imputation of categorical column [Employee length]
- Imputed Funded amount value in place of Funded amount Investor [when, Funded amount Investor = 0]
- Removed columns relating to Customer Behaviour, Customer Identifiers, features which have constant value (numeric or non-numeric)
- Found that Annual income had many outliers (1762 records)

Case Study - Approach

3. Feature Engineering –

- Issue date was split to year and month. Added attributes on amount fields to compare in between Loan, Funded and Funded investor amount.
- Created buckets for Annual income, Interest rate, Installment, DTI, Loan amount, Loan Status (0 as Fully Paid and 1 as Charged Off), Funded Amount Investor.

4. Data Analysis –

a. Univariate Analysis – Analysed various features individually and derived below findings :

- Loan status - we saw that the data is imbalance with 85% of data being Fully paid as compared against Charged-off
- Maximum number of loan was issued in year 2011
- Found outliers with respect to Charged off on Loan amount, Interest rate and Annual income, Annual income is skewed towards right – positive skewness
- Medium Annual income group [100000 - 150000] are the majority of the distribution
- Maximum loan requested was from 9600-15000 followed by 5300-9600
- 75% of loan requested was for 36 months term tenure
- Employee with 10 + years of experience have taken maximum loan
- Investor funded amount (final amount paid by Investor) never exceeded Funded amount (amount approved by Reviewer against the requested loan amount)
- Majority data are having verification status as 'not verified'
- Maximum loan is for debt consolidation
- California (CA) region has maximum loan entries
- Group1 with DTI less than 13.37 has the most loan

4. Data Analysis (continuation) –

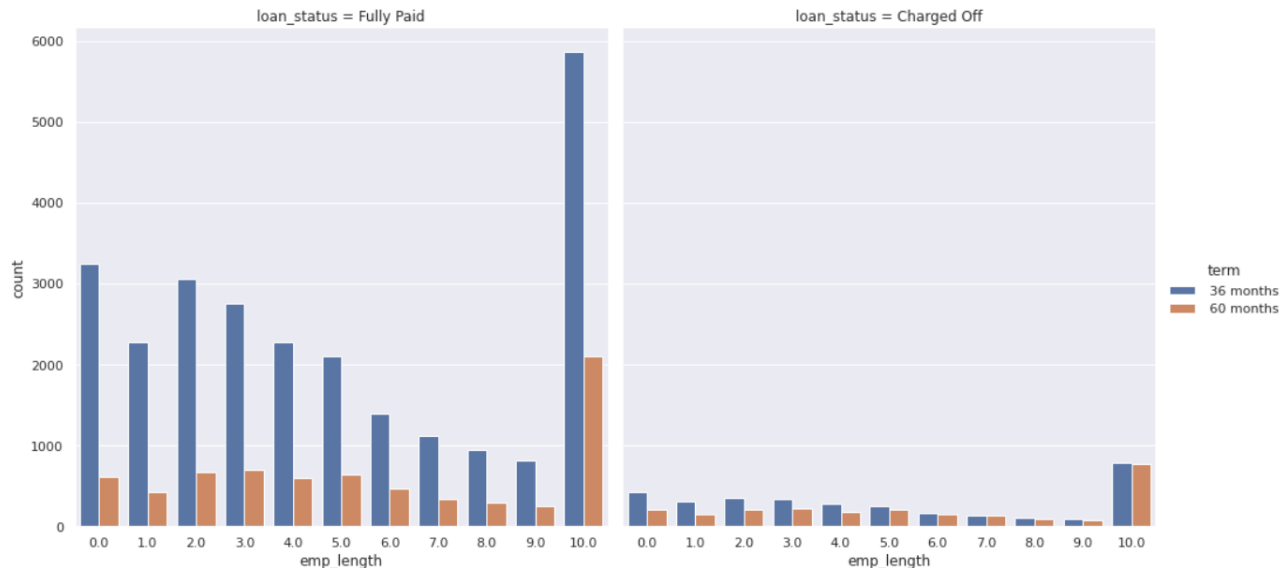
b. Segmented Univariate and Bivariate Analysis – Analysed together various features and derived below findings :

- Interest Rate group in the range of 20-25 has more than 35% defaulter
- Employee with tenure more than 10 year consists of 16% defaulter
- Annual income less than 50000 consists of 20.4% defaulter
- Purpose of loan with small business consists of 27% defaulter
- DTI more than 18.56 consists of 17% defaulter
- Home ownership as 'Other' consists of 18.3% defaulter
- Verification status as 'Verified' consists of 16% defaulter
- Grade G consists of majority % defaulter followed by F
- Term with 60 Months with respect to the total contribution has 25% more defaulter
- For Issue year, we see that 2011 has the maximum defaulter in numbers but if we compare it with Fully paid we can say that 2007 has highest default ratio
- California state has the maximum defaulter in numbers but if we compare it with Fully paid, we can say that Nebraska has highest default ratio
- If loan amount is greater than 15000, then it has higher possibility of not repaying the loan

Case Study - Final Observation & Recommendation

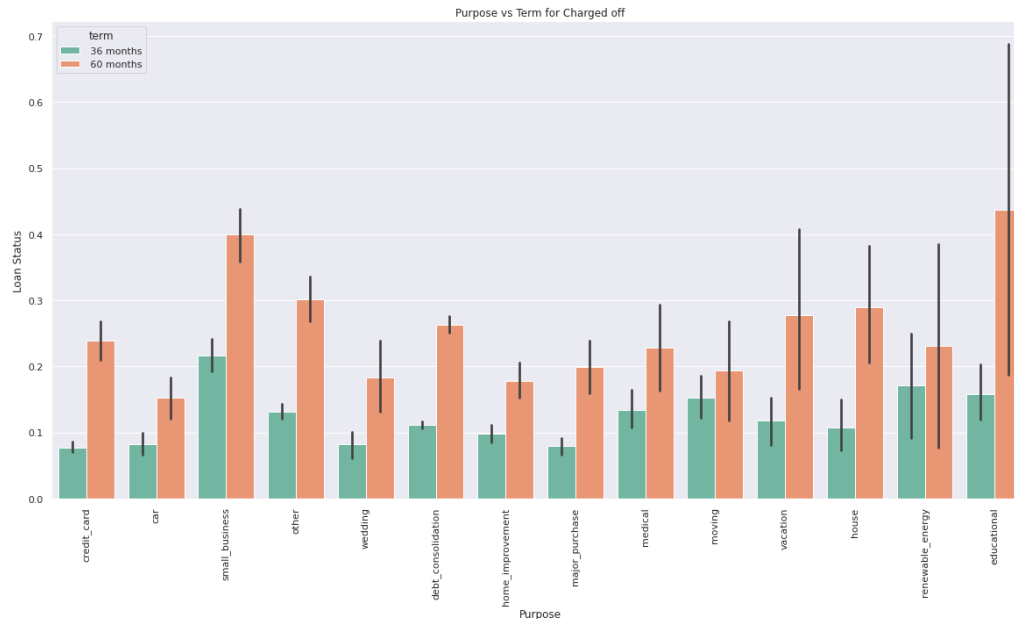
- Based on the feature analysis with respect to 'Charged off' –

- We have observed that employee with 10+ years of experience and with term period 60 months with higher defaulter ratio. And in terms of number, we observed that 10+ years with 36 months are more in



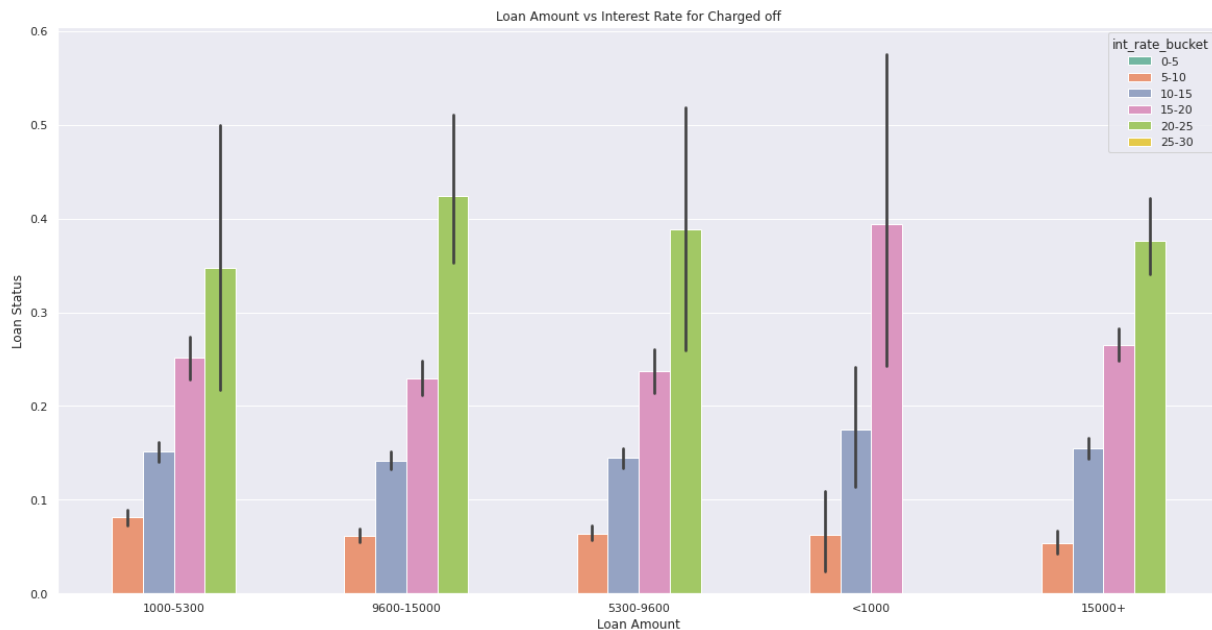
Case Study - Final Observation & Recommendation

- Maximum defaulters are those who have taken loan for Educational purpose with 60 months as term. And for Small Business, we observed that irrespective of term, the defaulter percentage is high



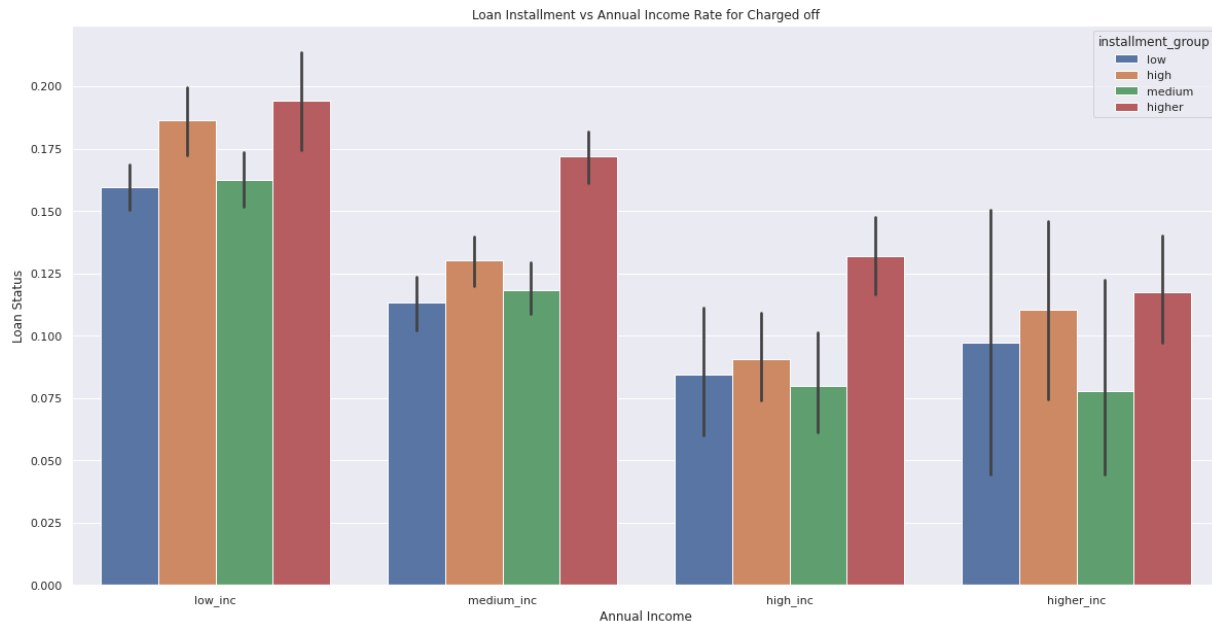
Case Study - Final Observation & Recommendation

3. Maximum defaulters are those who have taken loan with interest rate between 20-25 and for the analysis we can say that if the loan amount exceeds 15000, then the chances to pay the loan is low



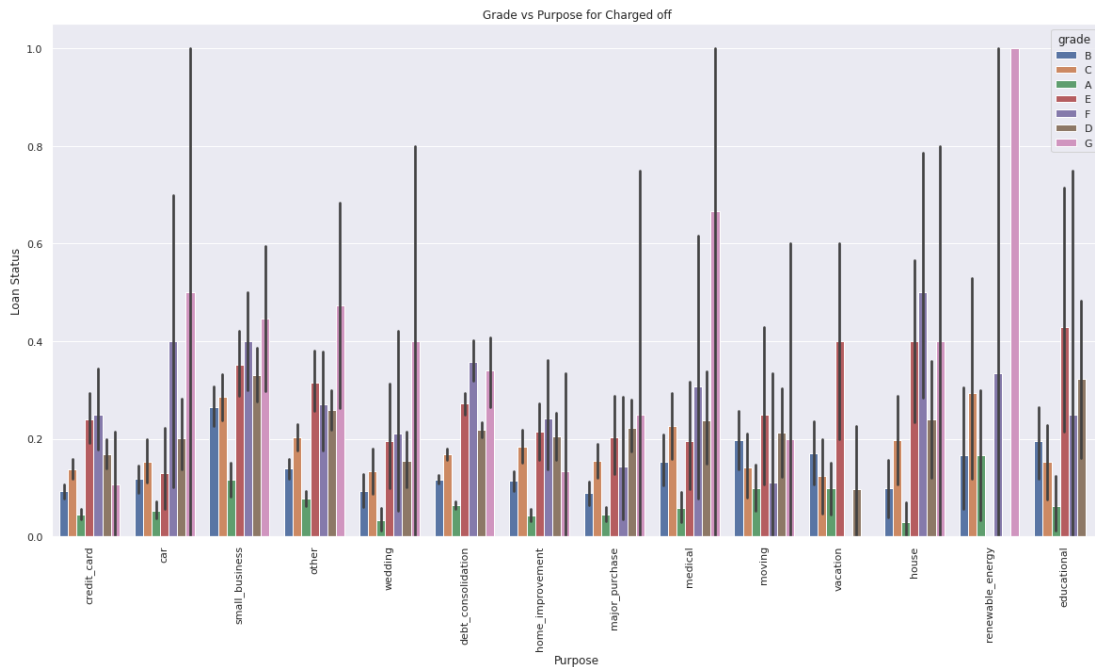
Case Study - Final Observation & Recommendation

- Maximum defaulters are those who have taken loan with lower annual income (less than 50000) irrespective of installment value



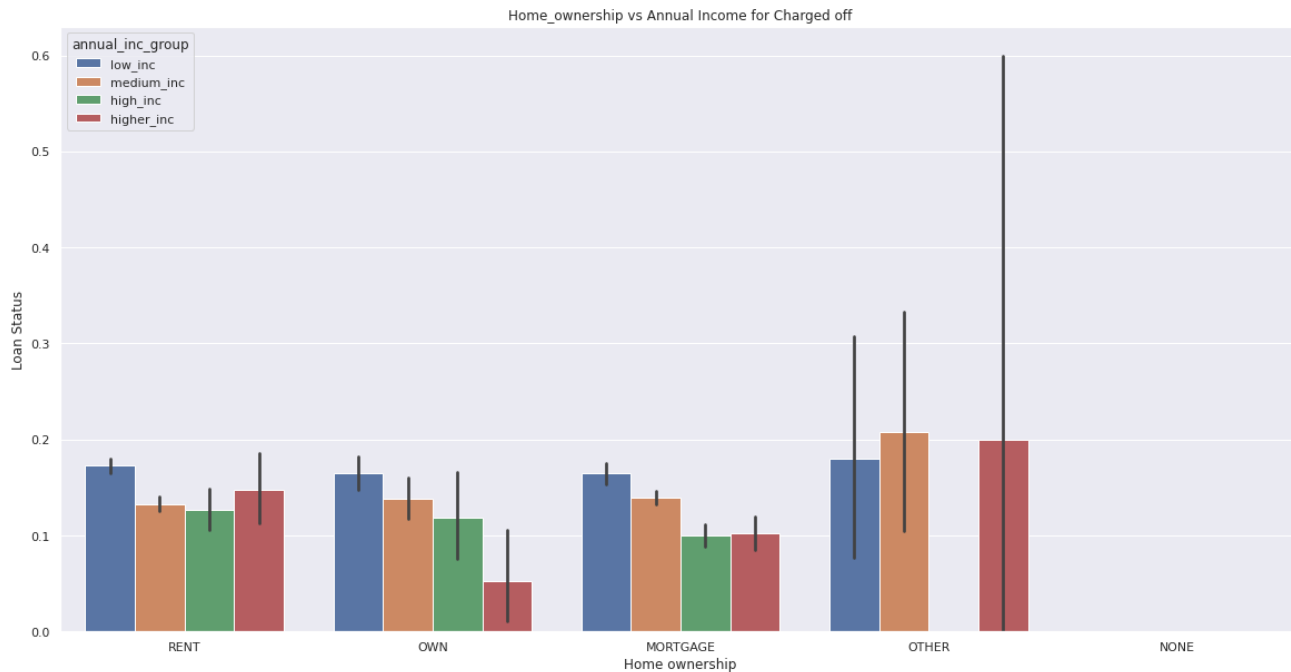
Case Study - Final Observation & Recommendation

5. Maximum defaulters are those who have taken loan with Grade G



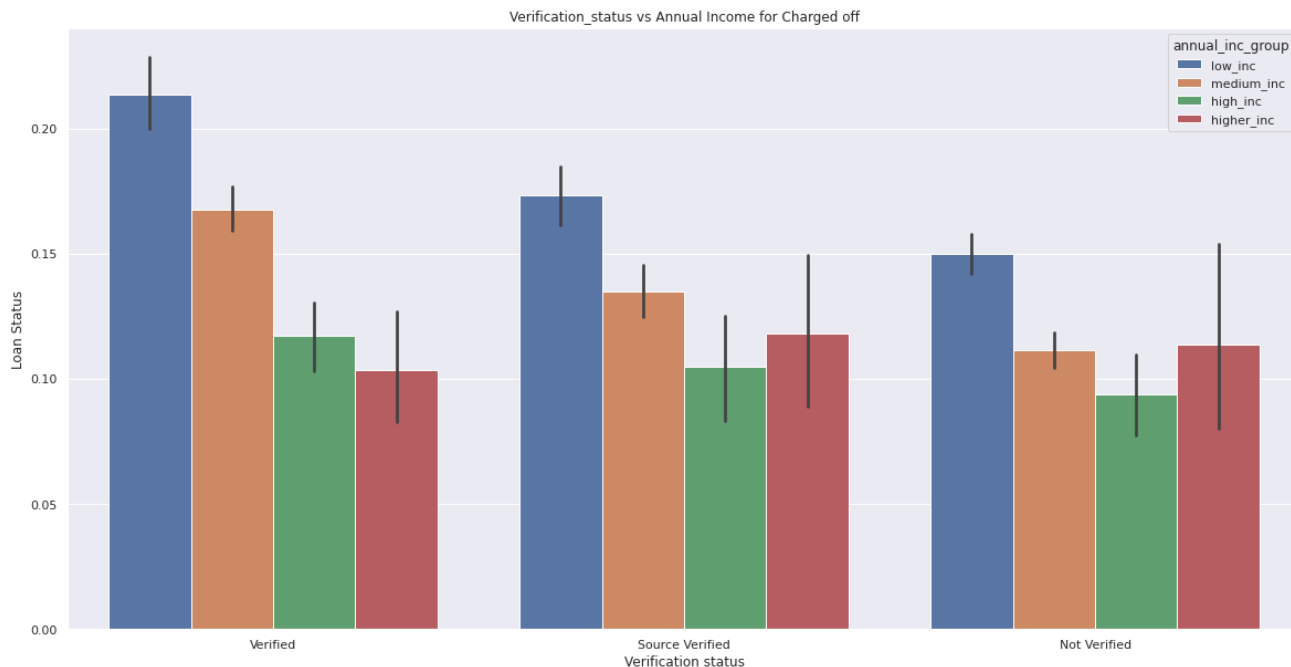
Case Study - Final Observation & Recommendation

6. Maximum defaulters are those who have rented properties



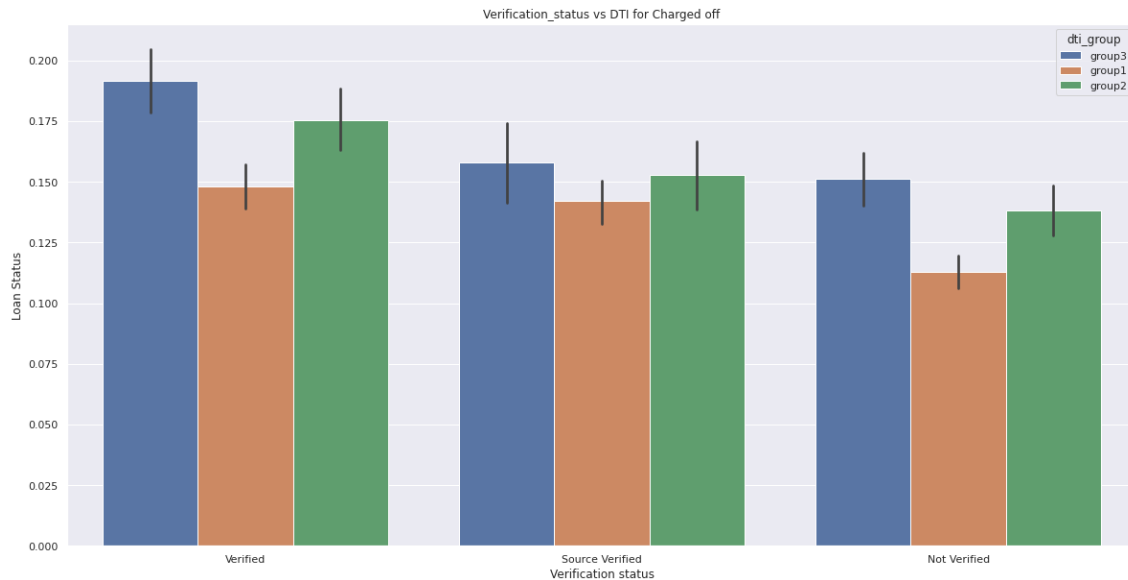
Case Study - Final Observation & Recommendation

7. Maximum defaulters are those who are in low income group



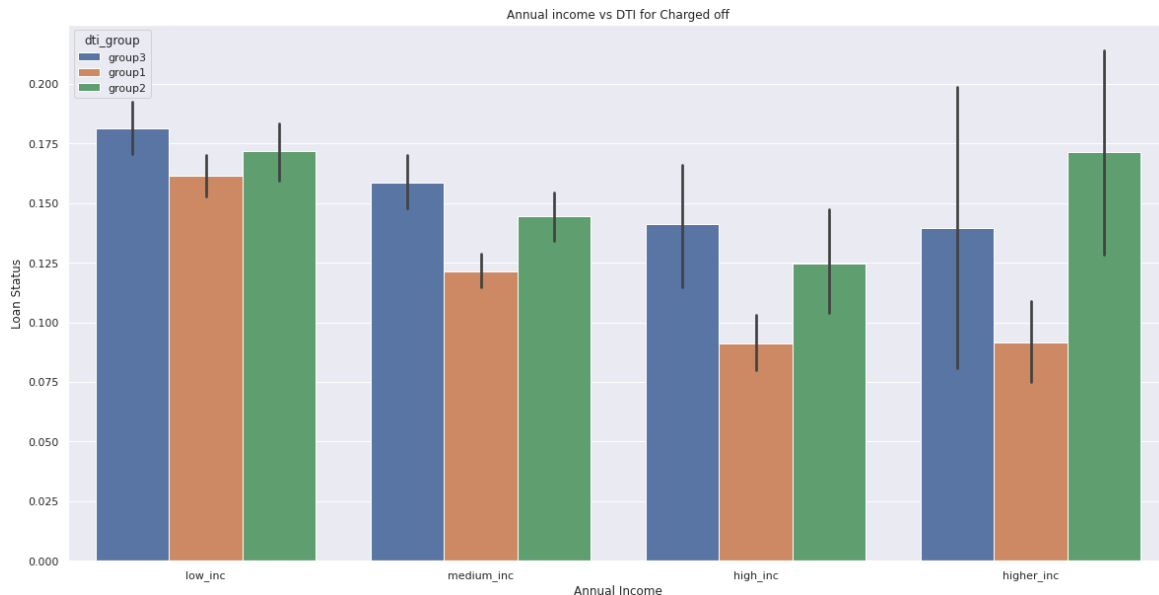
Case Study - Final Observation & Recommendation

- Irrespective of any verification status, with DTI more than 18.56 are the majority defaulters. When the Verification status is 'verified', it is observed that the majority defaulters are from this category which might infer that the verification process needs to be improved, so that defaulter ratio gets minimised



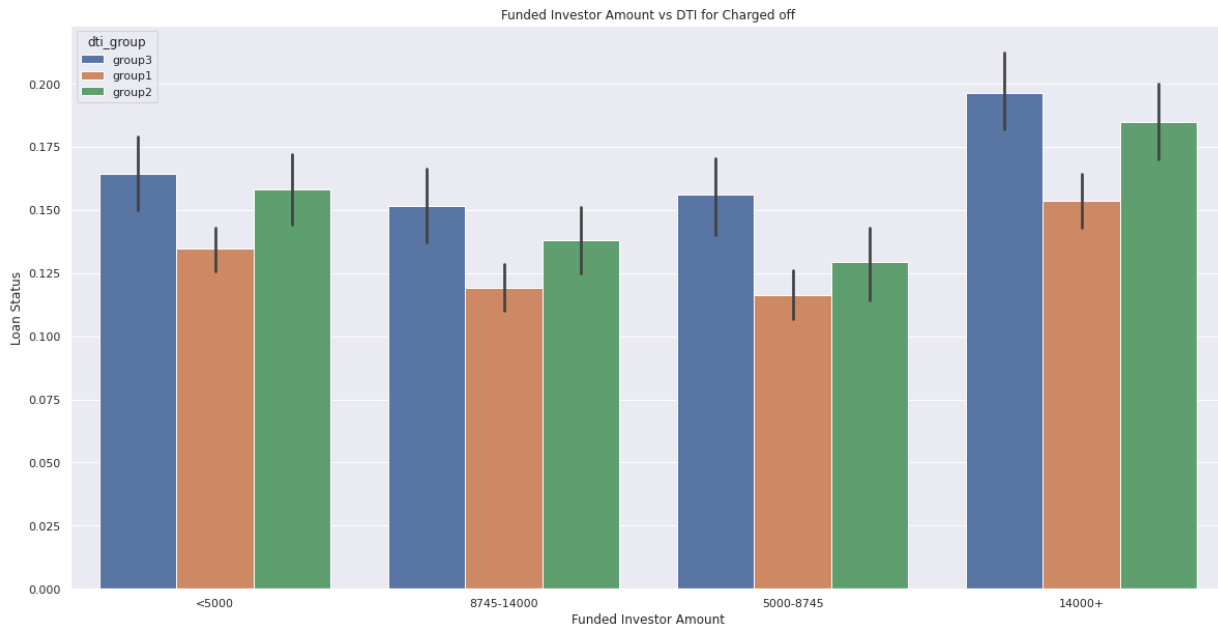
Case Study - Final Observation & Recommendation

9. Low income group are the majority defaulter for all DTI. We also see a trend, with DTI higher than 18.560000 are majority defaulter except for Higher income group where the DTI ranging from 13.370000 to 18.560000 are majority defaulters



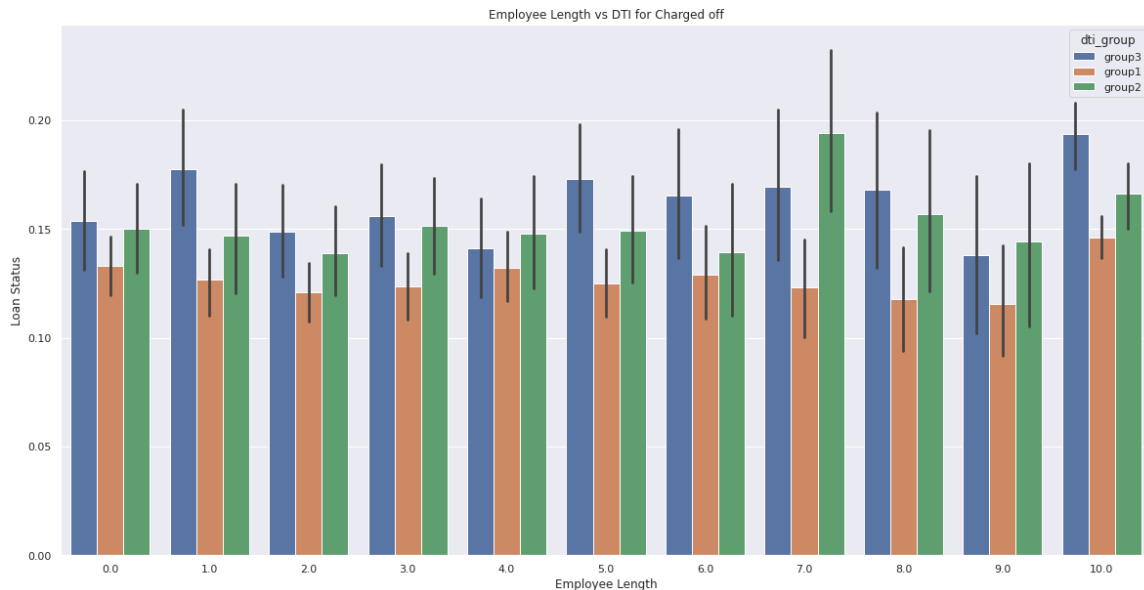
Case Study - Final Observation & Recommendation

10. Majority defaulters are those who have requested loan amount more than 14000 (for all DTIs). It even shows a trend on DTI with greater than 18.56 are the majority defaulter



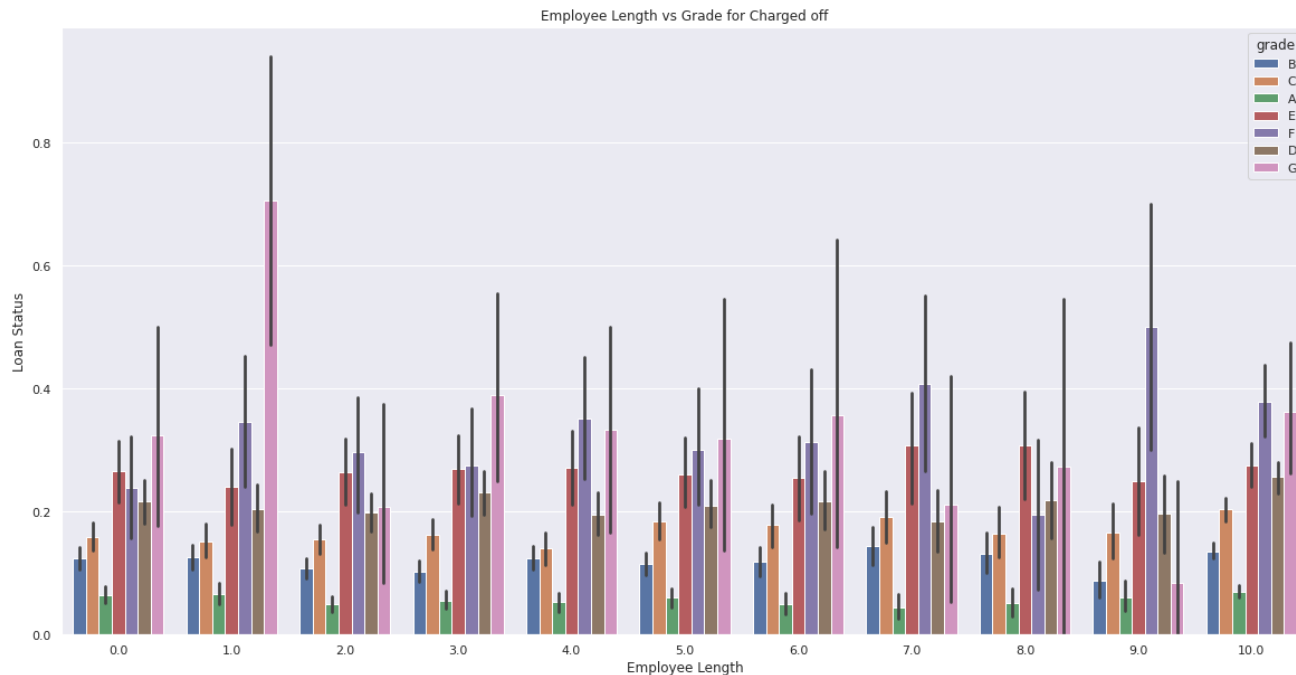
Case Study - Final Observation & Recommendation

11. Employee with 10+ years experience are the majority defaulters. In general, DTI greater than 18.56 are the majority defaulter except in cases where the employee years of experience are for 4, 7 and 9 years, where DTI between 13.37 to 18.56 are majority defaulters



Case Study - Final Observation & Recommendation

12. Majority of Defaulters are from Grade G followed by Grade F





Thank You!