LENDING CLUB CASE STUDY

Group Members:

1. Shrija Kale
2. Leon Dsouza

Problem Statement

Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

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 utilize this knowledge for its portfolio and risk assessment.

We need to identify important variables and derive conclusions from the dataset

Approach

Data pre-processing

- Data understanding
- Missing value treatment
- Outlier removal
- Redundant variables exclusion
- Variable standardisation and normalisation
- Logical variable selection

Univariate analysis

- Checking the type of variable (Categorical/Numeric/ Date object)
- Variable transformation
- Frequency plots, Pie charts and Line charts
- Box plots and dist plots to determine data distribution

Bivariate analysis

- Variable correlation with target variable (i.e. Loan default rate)
- Analysing Joint distributions by using:
  - A. Pivots
  - B. Stacked bar charts
  - C. Percentage distribution

Multivariate insights

- Analysis of significant variables with each other for loan defaulters
- Analysing patterns and trends which show high default rates

Outcome

- Recommendation of variables for further analysis
- Conclusions and insights from the recommended variables

Methodology

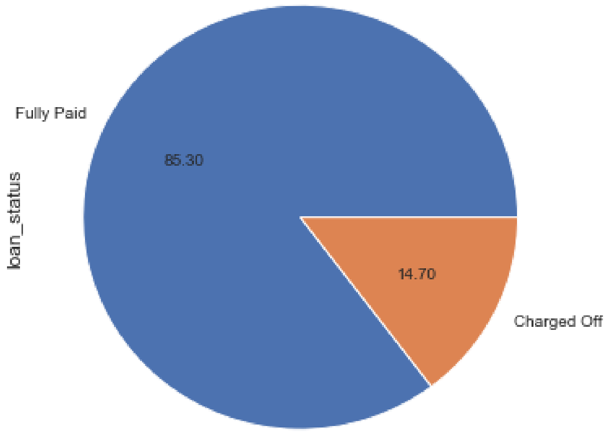
The approach is to understand the problem statement and identify variables that should be a part of analysis.

- Variables which capture customer behaviour post loan disbursals to be excluded.
- Inclusion of :
  - Customer demographics
  - Customer intent attributes
  - Loan attributes

Preliminary Analysis

1. Data size: 39.7K records with 111 variables
2. Data size after eliminating missing values and single value columns: 39.7K records and 48 columns
3. Assumption considered: Funded amount will always be >500, eliminated records below 500
4. Outlier removal for Annual income variable: Excluded values >95th Percentile
5. Final Data size: 36.3K records and 48 columns

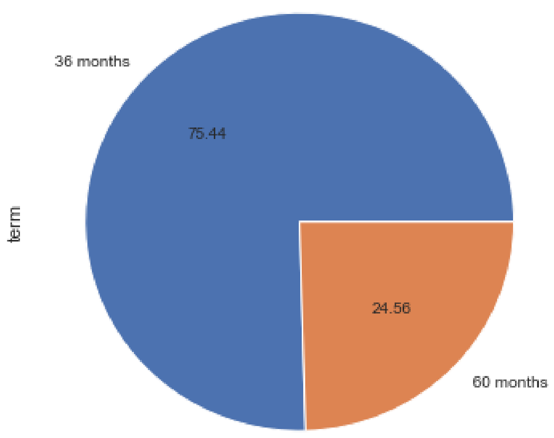
Target variable:
Charged off customers (Loan defaulters)



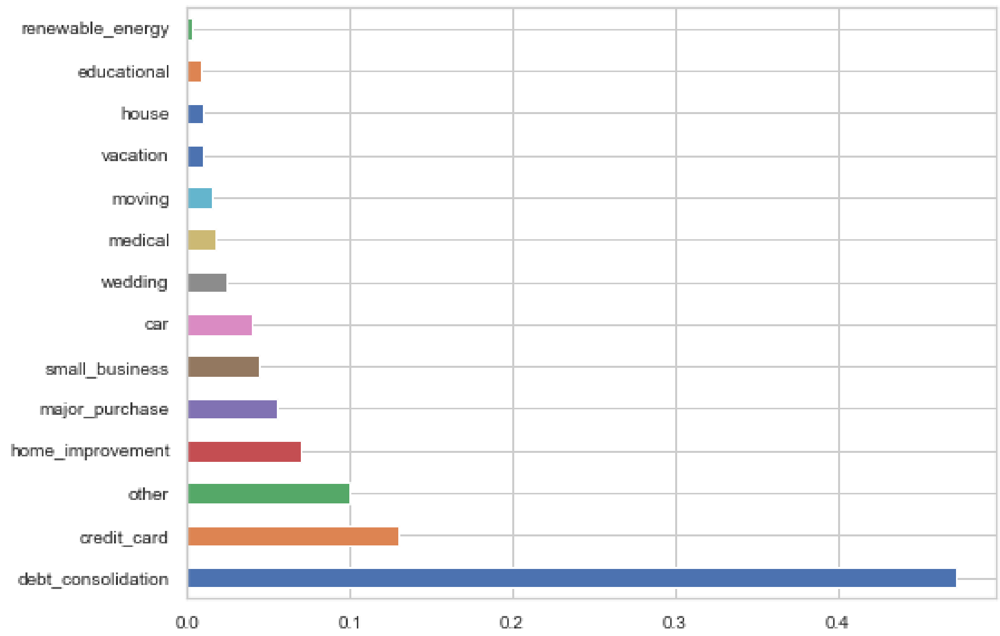
The dataset contains 14.7% of loan defaulters

Univariate Analysis

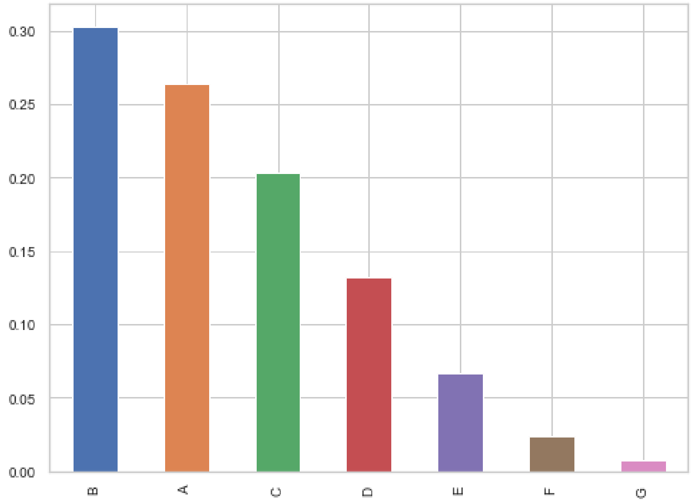
Loan Tenure



Loan Purpose



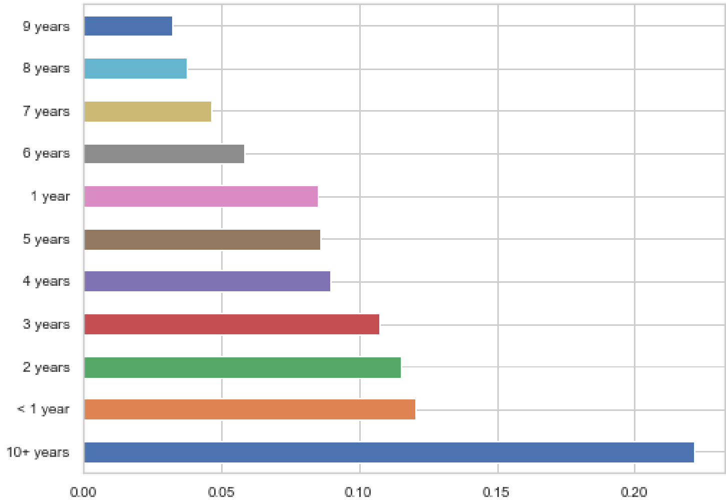
Grade



- The loan tenure is available in two installments 36 months and 60 months, of which ~75.5% customers opt for 36 months
- 76% of the total loan takers are from Grade A, B, or C
- ~60% of loan is taken for either debt consolidation (47.3%) or credit card settlements

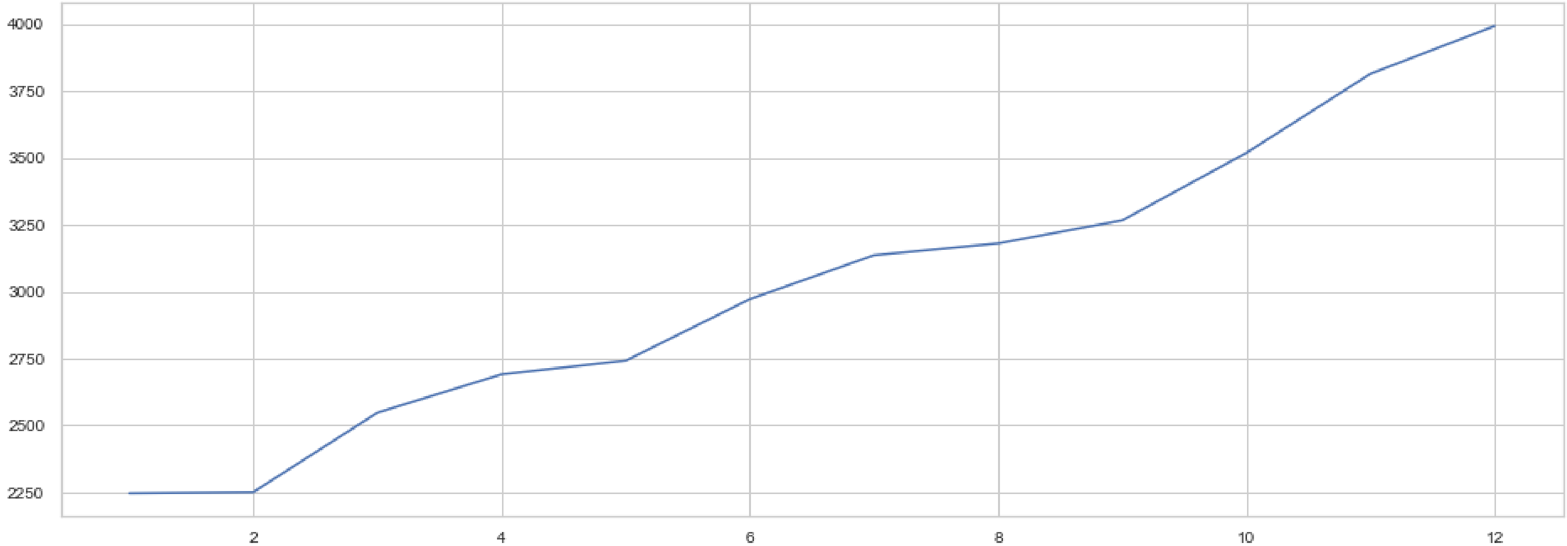
Univariate Analysis

Employment tenure



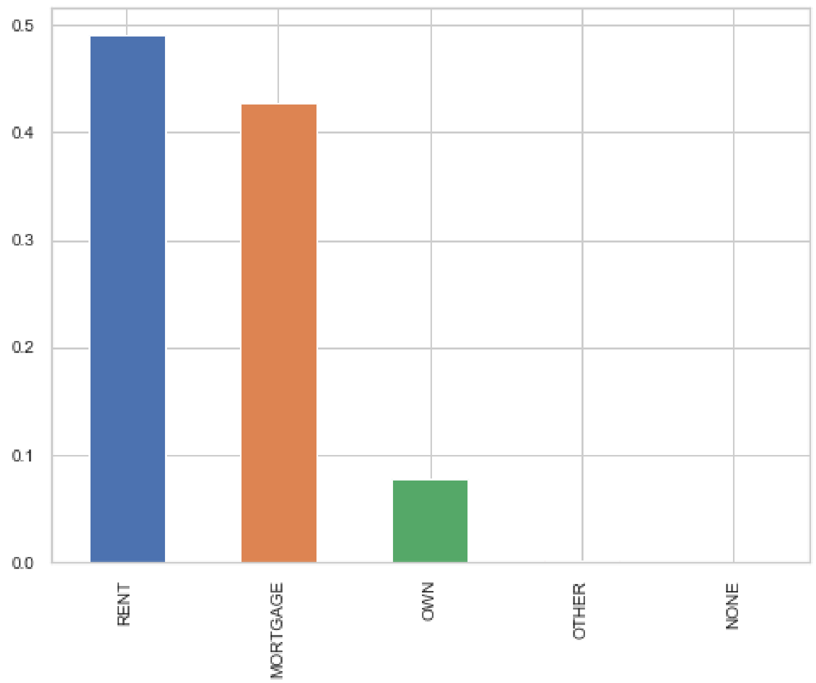
- 49% of the customers have employment tenure of 5 year or less, whereas 22% have 10 years or more
- More number of customers take a loan towards the end of the calendar year and that is why we see an increase in distribution towards the last quarter

Loan issued monthly trend



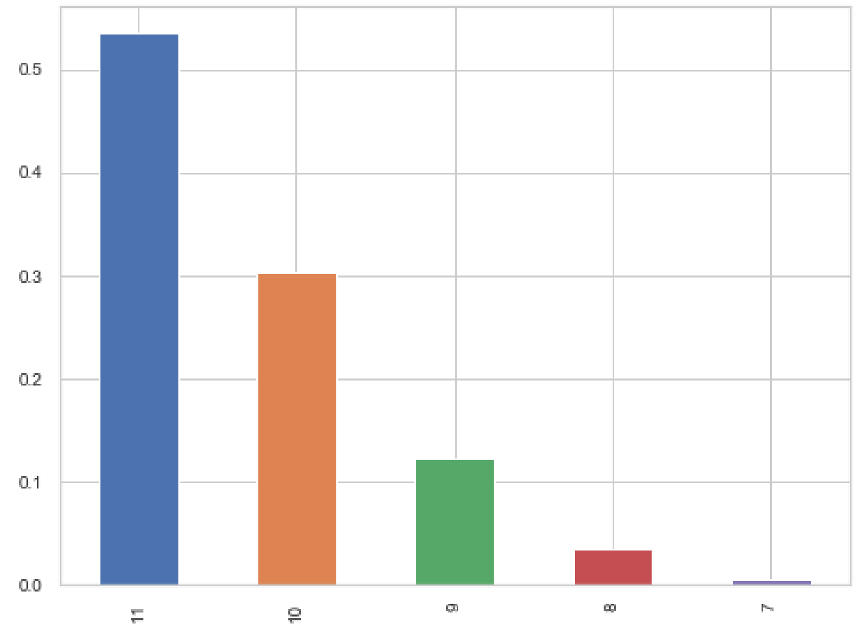
Univariate Analysis

Home Ownership



92% of the customers pay either rent or mortgage on their current residence

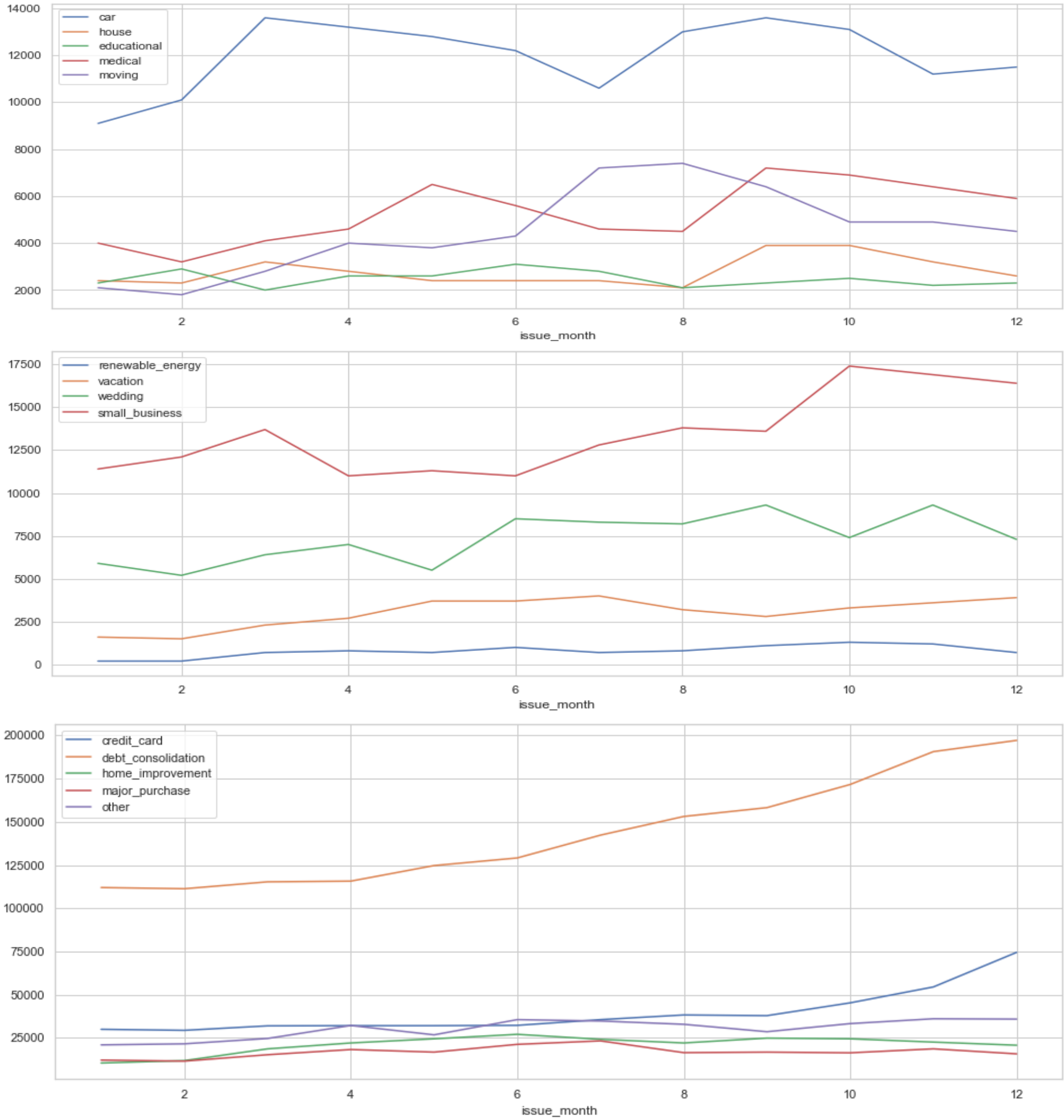
Yearly loan disbursals



Each year the number of customers taking loans is increasing

Bivariate Analysis

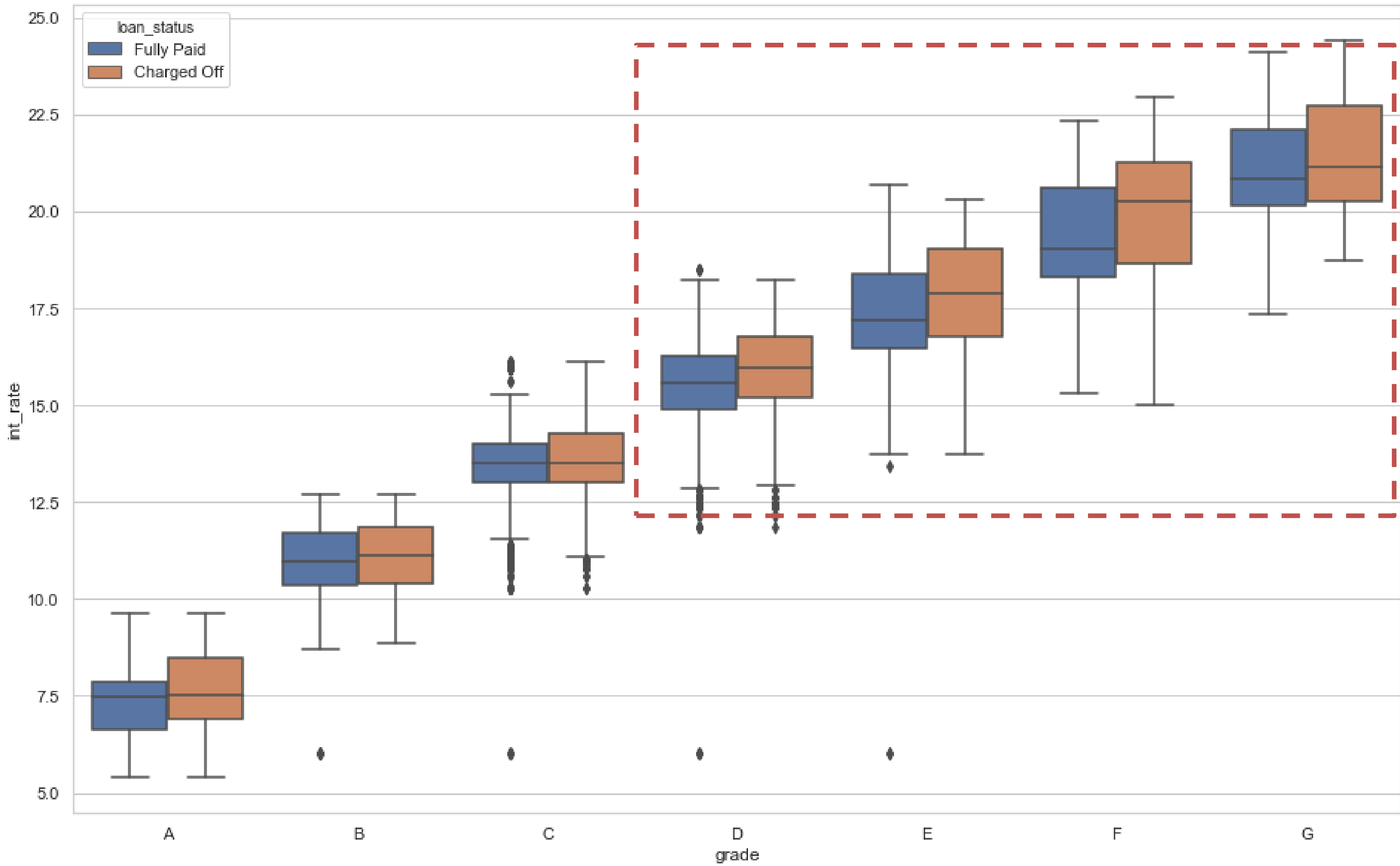
Purpose of Loan vs. Month



- We see a massive surge in loans for Credit card and debt consolidation towards the end of the year as compared to the rest of the year and this can be attributed to the idea of having a good credit score towards the end of the year
- On an overall basis there is a dip in loans taken May except for Vacation and Medical purposes

Bivariate Analysis

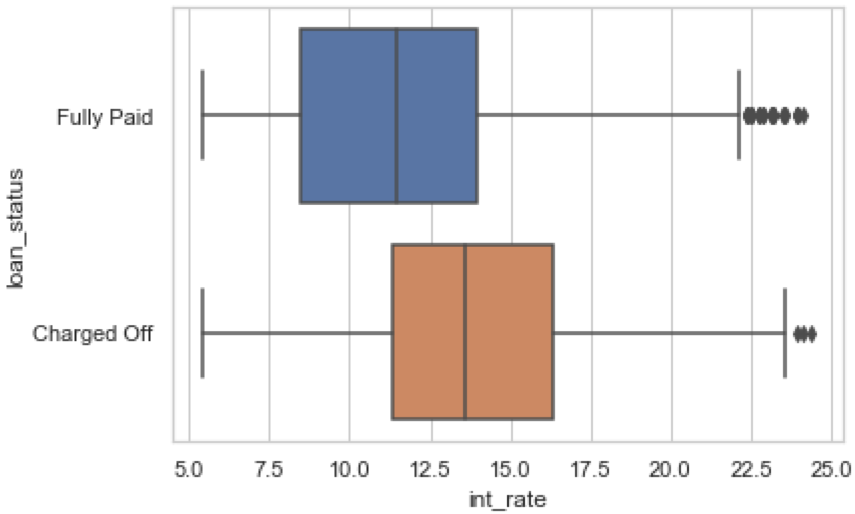
Grade Vs. Interest rates



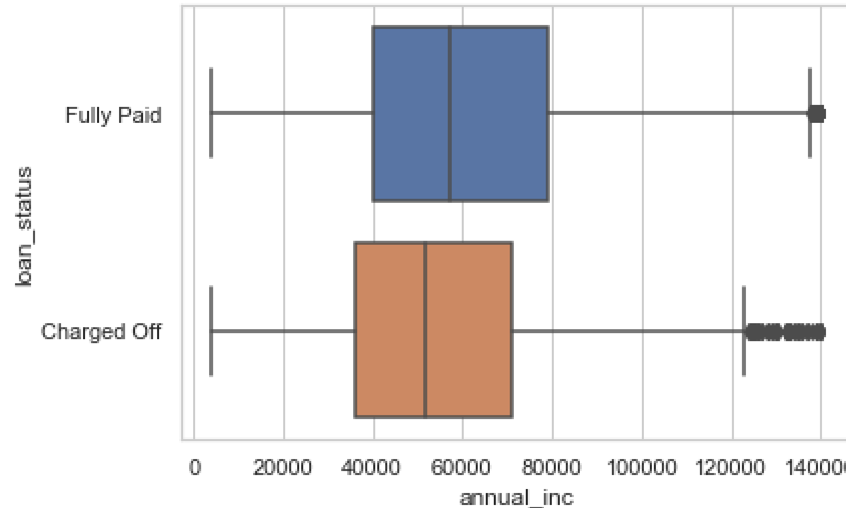
- The interest rate provided to a customer is directly proportional to grade
- The interest rates that the defaulters received was usually higher than the fully paid customers.
- For Grade A, B & C even though the interest rates were higher, they had similar median values within loan status.
- For Grades D-G, **Q1, Q3** and **median** are all higher for defaulters in comparison to fully paid

Bivariate Analysis

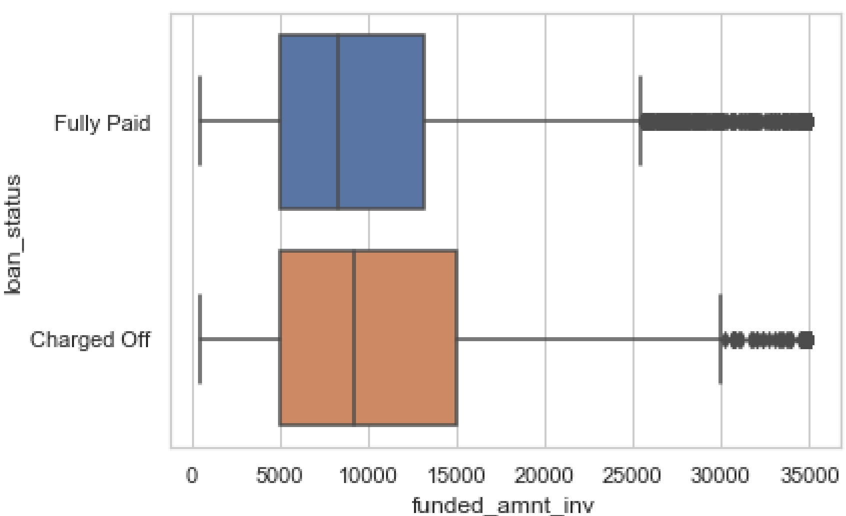
Loan status with Interest rates



Loan status with Annual Income



Loan status with Funded amount



- Defaulters are mostly the ones with higher interest rates as the median of the fully paid is Q1 of charged-off

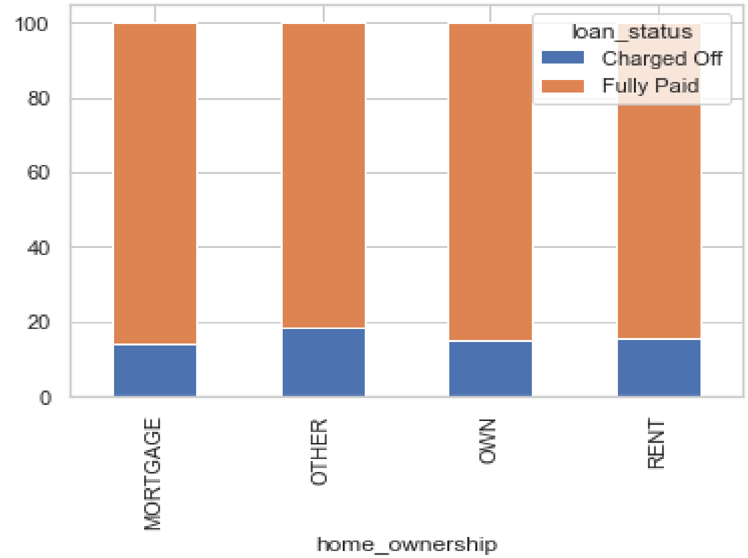
.

- Maximum annual income of charged-off for 75% of population Rs.10000 less than the fully paid, the median also lies on the lower end

- The gap between the upper whiskers of fully paid and charged off is huge along with the size of the box plot. It looks like people who take a higher loan are more likely to default

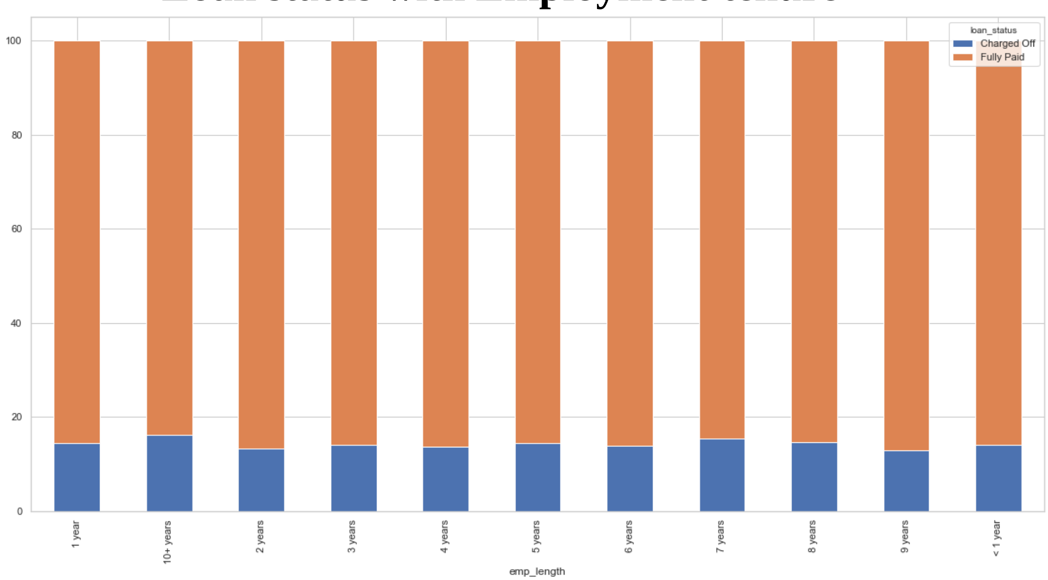
Bivariate Analysis

Loan status with Home Ownership



While the distribution of home owners remain almost the same across charged-off and Fully paid, we see higher number of customers who mention OTHER to be defaulters.

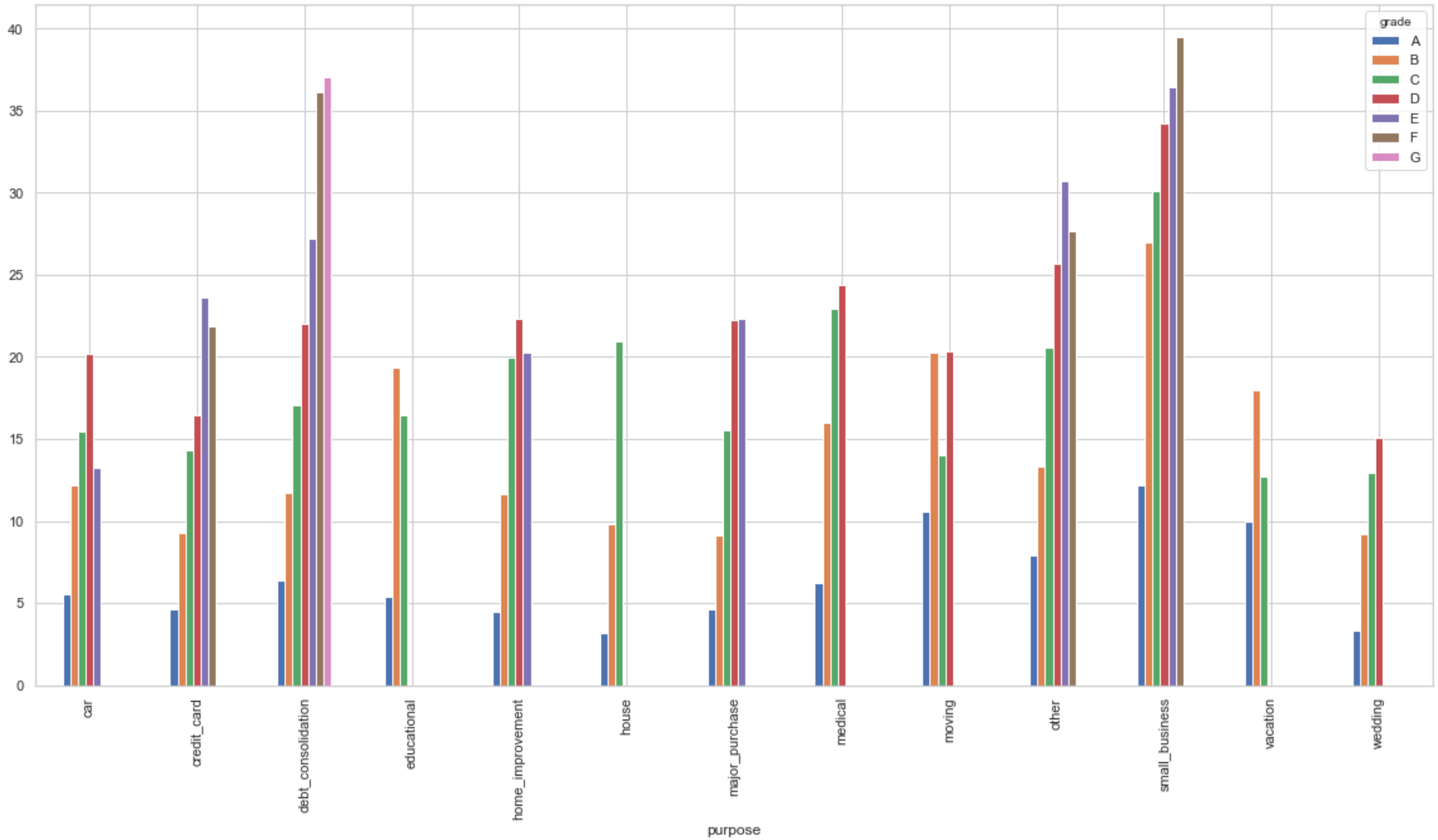
Loan status with Employment tenure



While the spread of completed and defaulters are pretty much even across all employment length, 16% defaulters in 10+ years being the highest and contributing 24% to the overall defaulter contribution.

Multivariate Insights

Purpose and Grade wise distribution of charged off customers

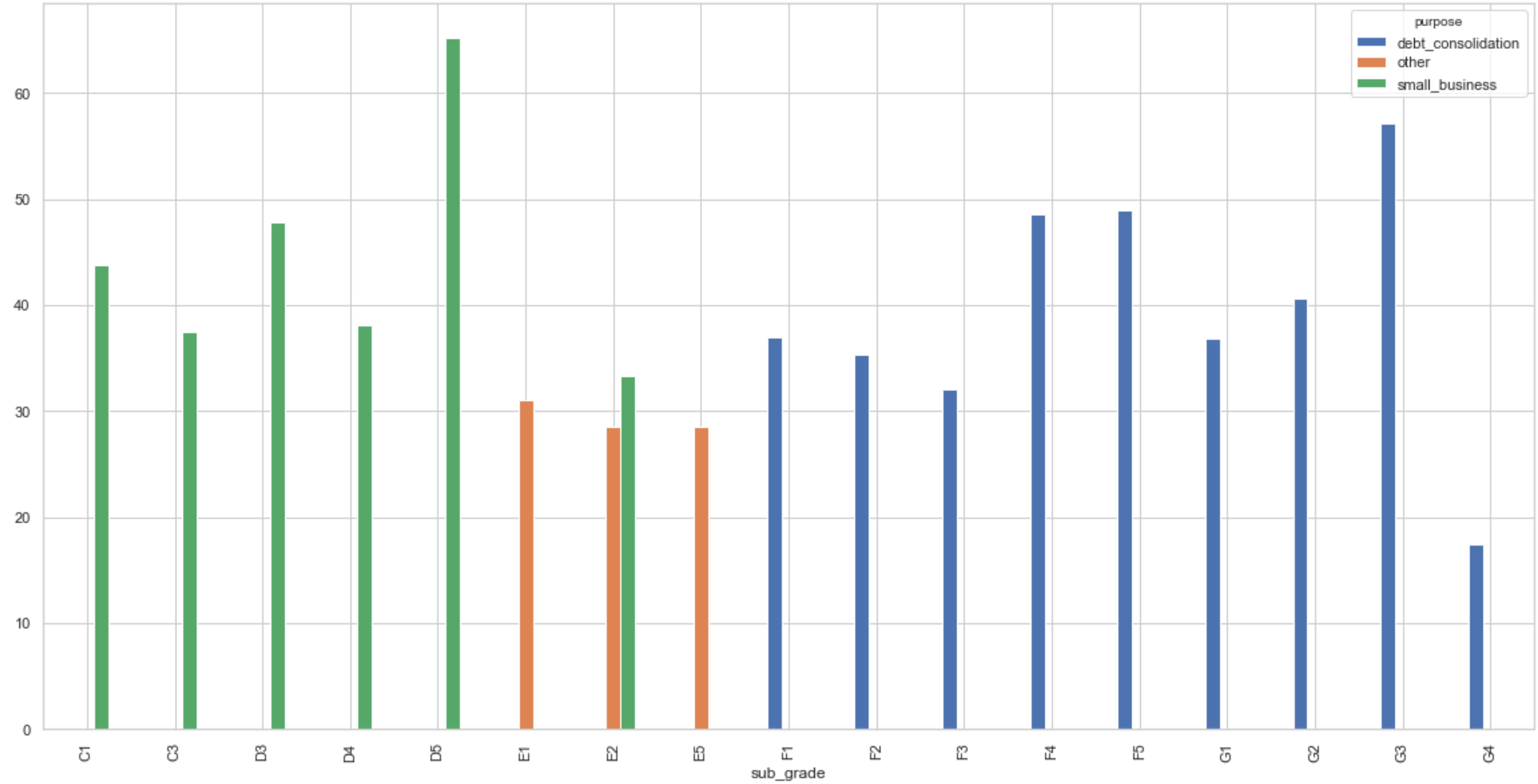


We are only considering records with default rate > 30% and at least 50 customers in a category

- Loans for small business taken by grade C, D, E and F are most likely to default, with highest default from grade F at 39.5%.
- Loans for debt consolidation taken by grade F, G are likely to default with 36.1% and 37% respectively.
- Loans taken for other expenses by grade E have a 30.7% chance to default.
- Medical, Home improvement, Wedding and Car have high defaulters in grade D in comparison to other grades (~20%).
- Grade A customers have lowest defaulters (<10%) across purpose category.

Multivariate Insights

Sub grade distribution of charged off customers in top 3 highest default purposes



Restricting the analysis to a minimum of 20 customers per category we observed

- Sub-grade D5 for small business has ~65% default rate, followed by G3 for debt consolidation at ~57% default rate.
- D3 for small business and F4, F5 for debt consolidation have very close to 50% default rates.

Outcome

Significant variables which can be considered in further analysis:

1. Loan purpose
2. Grade
3. Loan tenure
4. Employment tenure
5. Funded amount
6. Interest rates
7. Loan issue Month
8. Subgrade
9. Home Ownership

The above variables exhibit direct correlation with each other and the target variable i.e. charged off customers.

These variables can comfortably help us understand loan defaulter patterns and can be used in further model building to give us improved predictions

Conclusions & Recommendations

From the observations we have derived some conclusions as listed below:

1. Reduce the number of approvals for people who have taken a loan to set up small business, for the purpose of debt consolidation and CC settlements. They are likely to default.

2. Increase approval rates of A,B grades and avoid approving loans with E, F grades, they are likely to default.

3. Loans with very high interest rate and high funded amount are likely get charged off. Either try reducing the interest rate or don’t approve the loan.

4. Inspect the application who have other/rented houses.

5. Inspect customers with 10+ years of employment tenure

6. No. of loans approval is increasing exponentially for the Lending Club with each year, which indicates the club has a great growth.