LENDING CLUB

EXPLORATORY DATA ANALYSIS

|| Yash Mishra ||

PROBLEM

Company

Lending Club is a consumer finance company specialises in lending various types of loans to urban customers.

Objective

Driving factors behind loan default, i.e., the variables which are strong indicators of defaults.

Problem

Lending Club wishes to identify high risk clients and likely defaulters before sanctioning their loans.

ANALYSIS APPROACH

Clean

- Drop Cols with only null values
- Drop Cols with more than 20% nulls
- Cleaning and reformatting cols
- Drop cols with only one value
- Drop cols with one value having more than 80% proportion
- Drop cols which do not add value to analysis
- Create new cols from existing cols
- Create bins for numerical cols

Univariate Analysis

- Check relationship and find trends of target variable with each variable left after cleaning the data

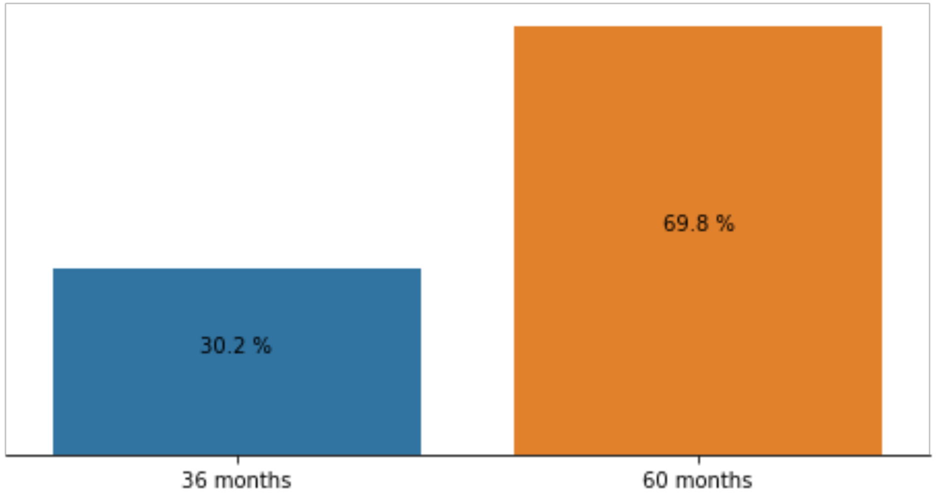
Bivariate Analysis

- Take combination of two cols at a time from the identified cols which have strong trend with target and observe their inter-dependency with each other and the target variable

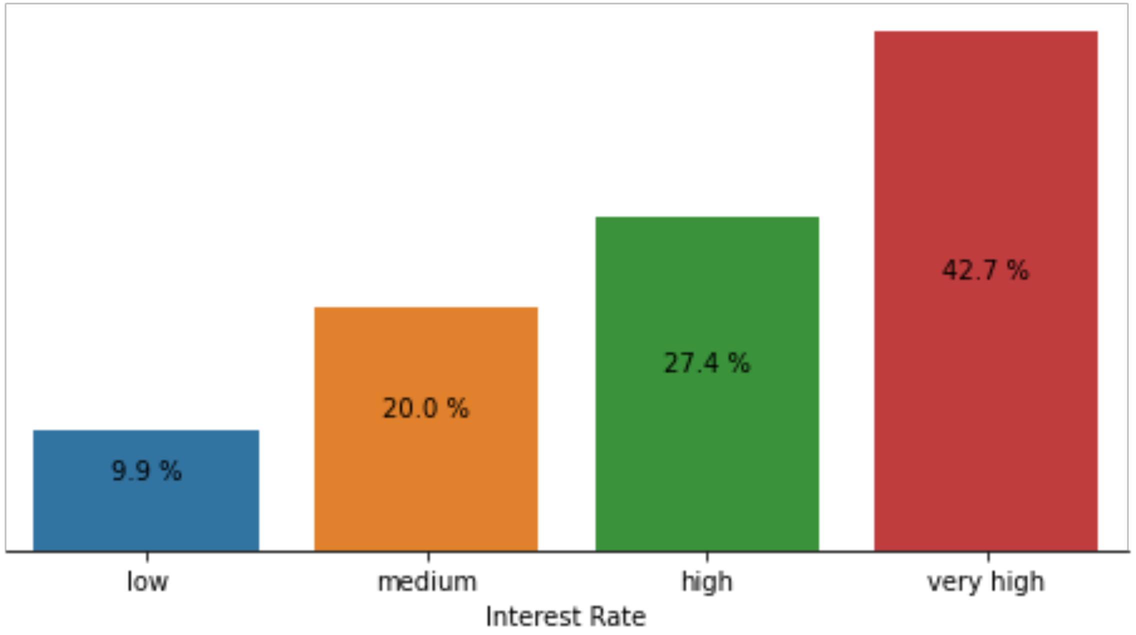
Multivariate Analysis

- Take more than two cols and see how they affect the target variable

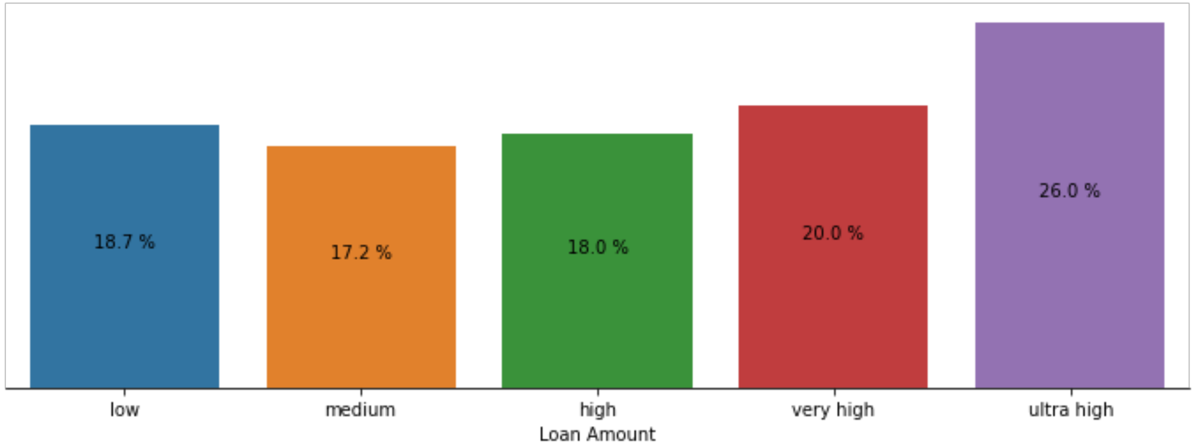
UNIVARIATE ANALYSIS



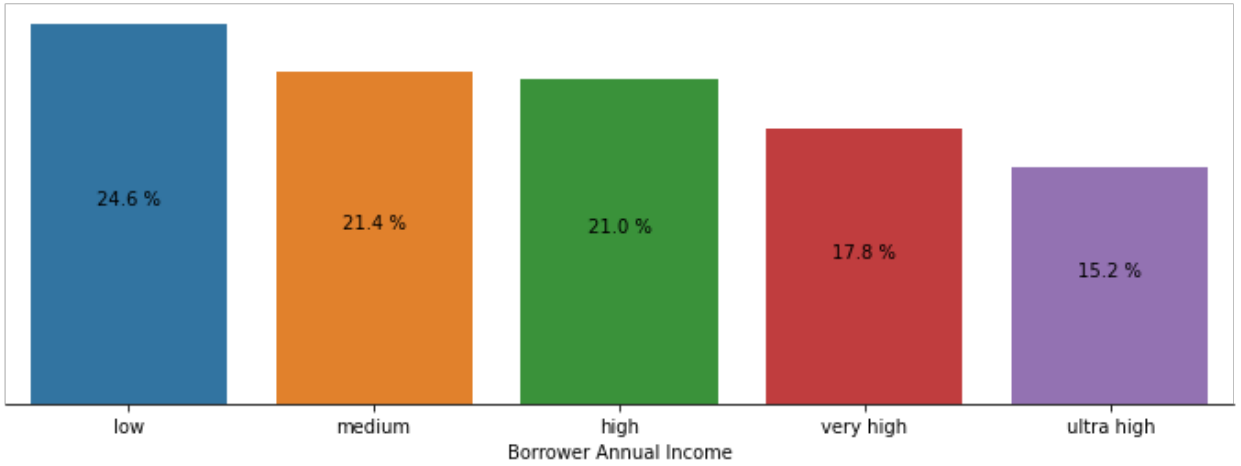
- People who apply for longer loan term of 60 months are more than twice likely to default. Longer Term Loans should be given more scrutiny.



- Higher the interest rate, higher the likelihood of defaulting. Borrowers with the high interest category ("very high" -> (14.42, 24.4]) tend to be worst at paying back their loans

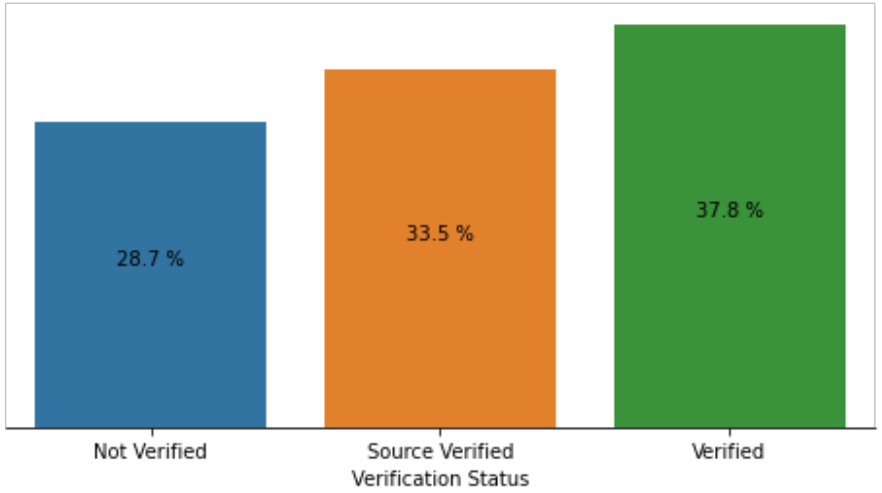


- Bank should also be careful of lending huge loan amount. People who borrow more than $15,000 (ultra high category) are most likely to default.

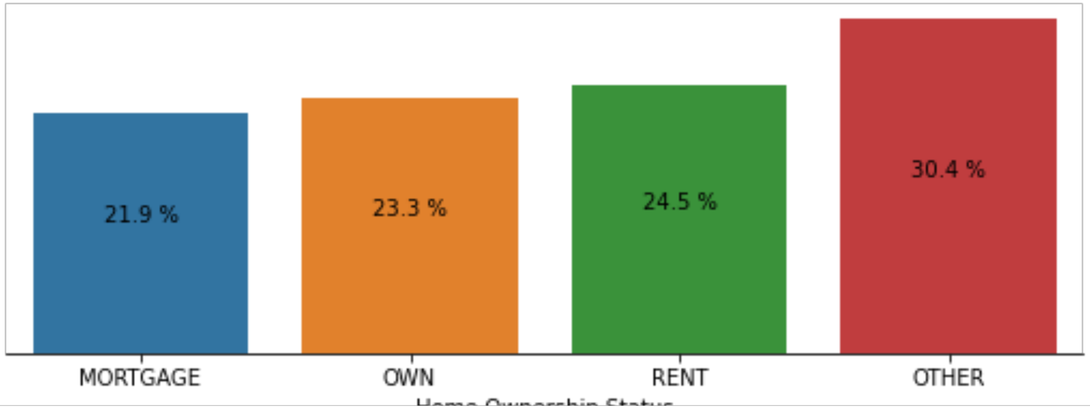


- Bank should be wary of customers with reported annual income of below $38,000 as they are most likely to default.

- State of a borrower cannot predict the likelihood of defaulting and bank should not bias their loan deciding factor on it.

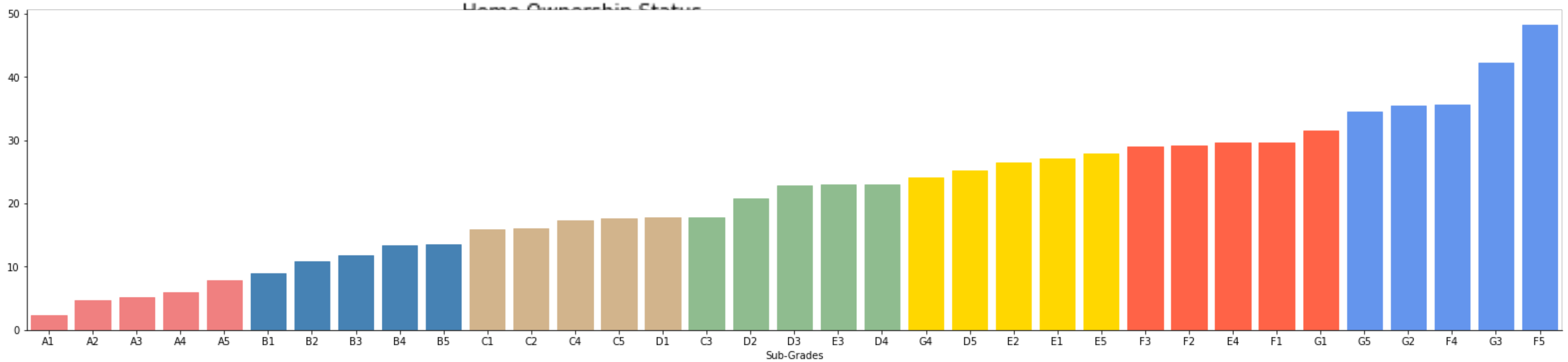


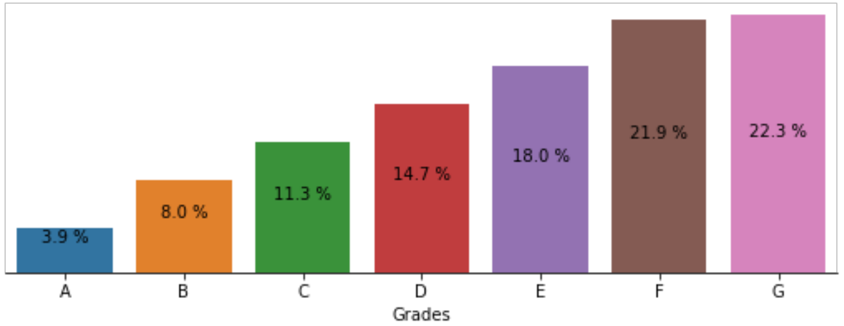
- Bank needs to investigate its verification process because ironically, “verified” borrowers tend to default the most.

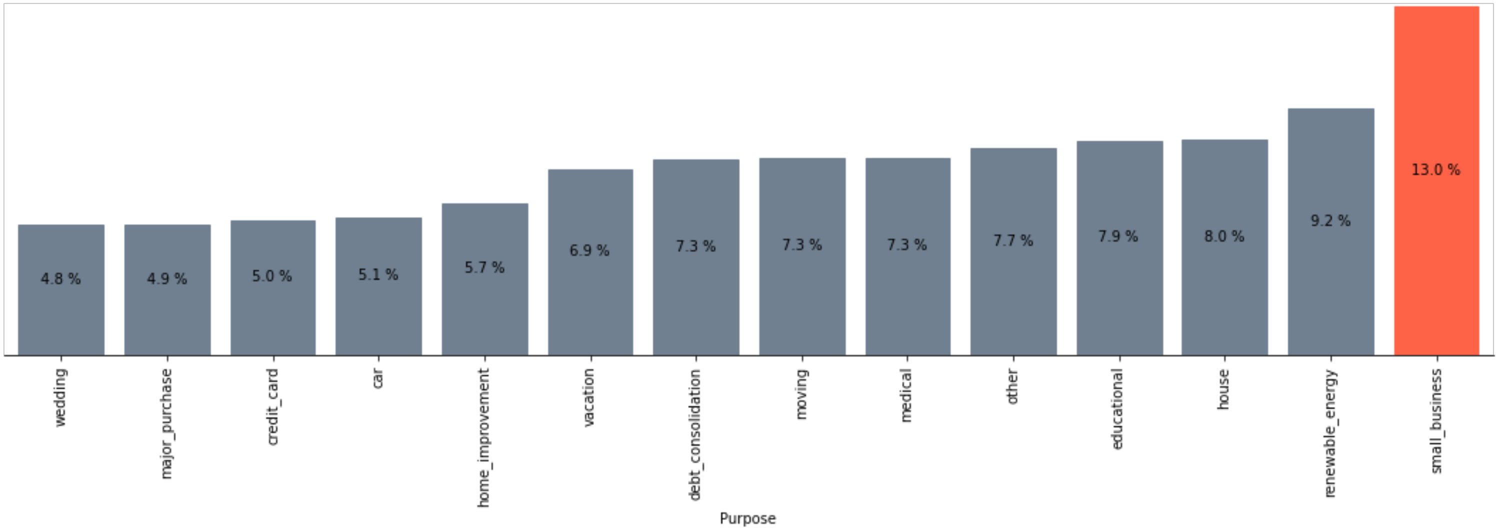


- Bank needs to scrutinize borrowers more if they don’t clearly mention their home ownership status clearly.

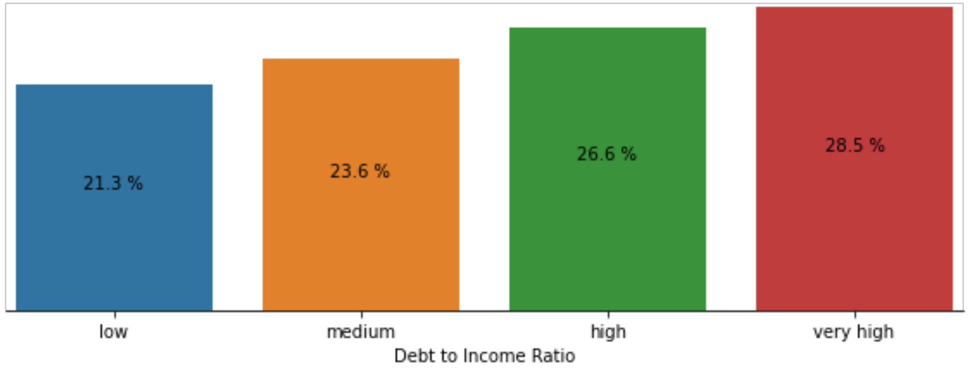
- It is also worth noting that the grading and sub-grading system by Lending Club is good at predicting defaulters



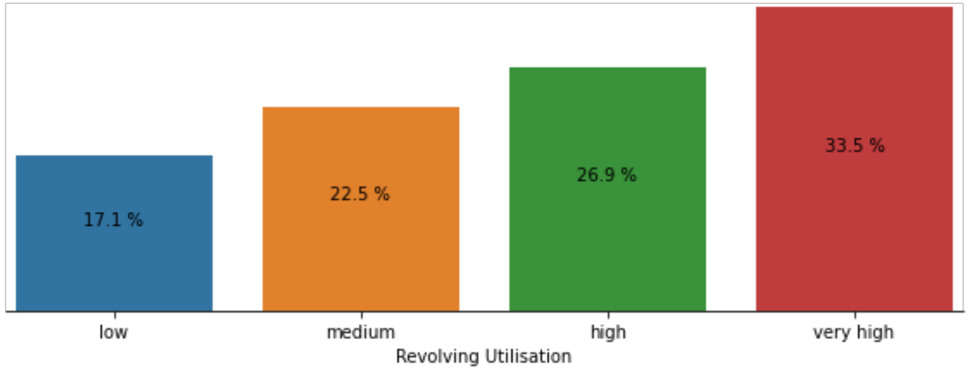




- Lending money for small business and renewable energy should raise a red flag and should be treated with caution

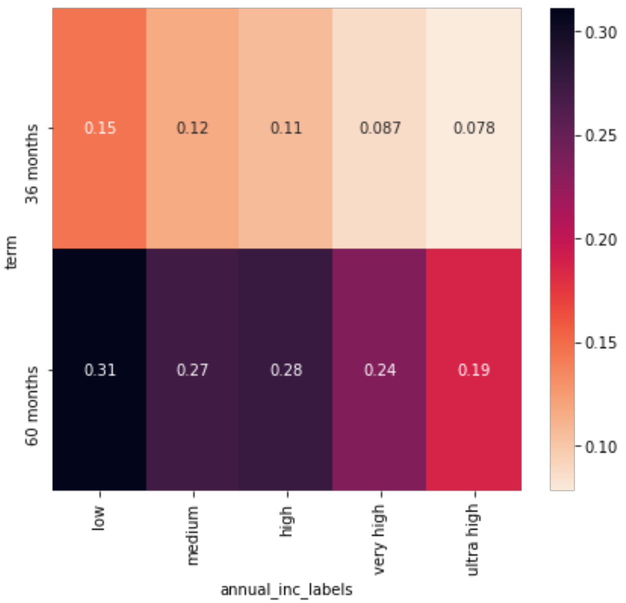


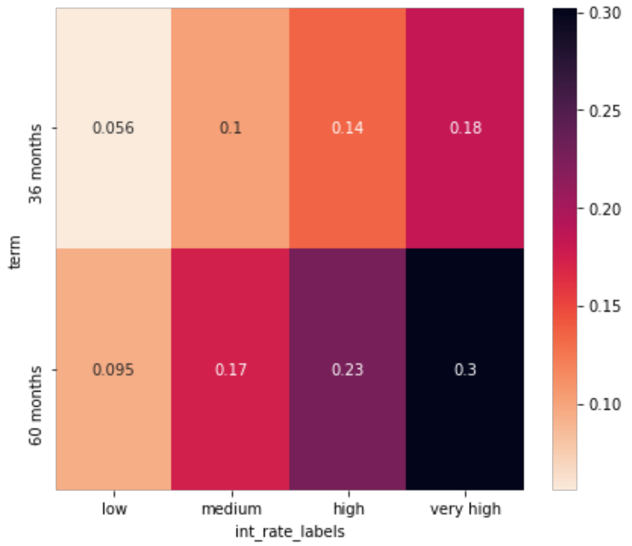
- Borrowers with 18+ DTI and 72+ Revolving Utilisation are more likely to default.

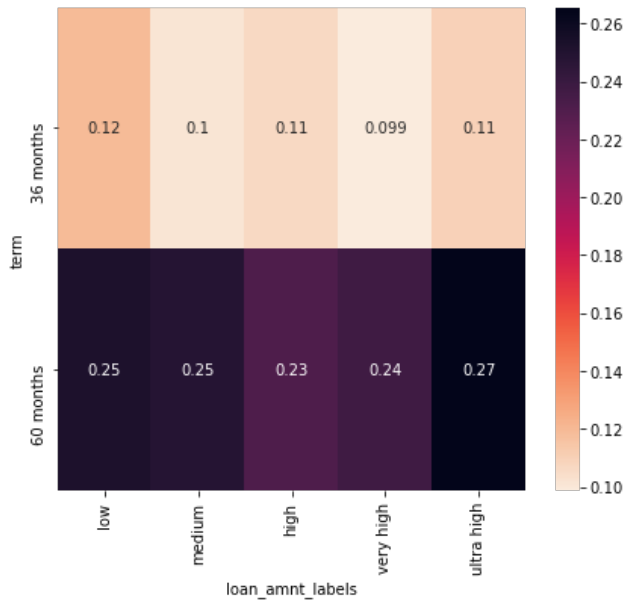


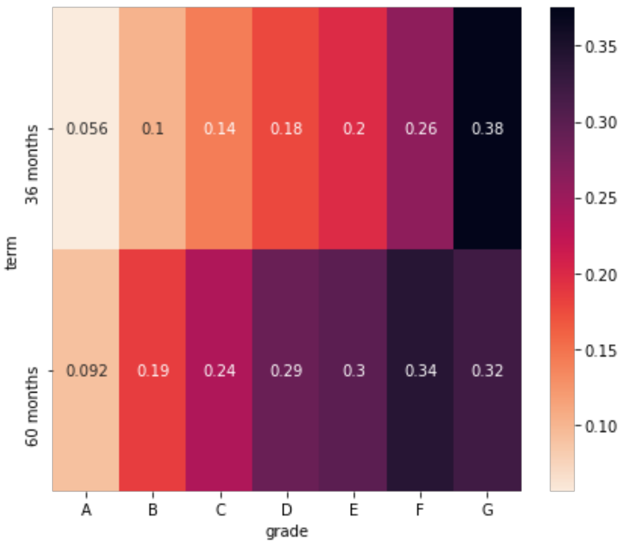
- Columns which are not strong indicators of a daulter: Borrow’s Employment Length, Date of Issuing Loan, Earliest Credit Line, total number of accounts by the borrower.

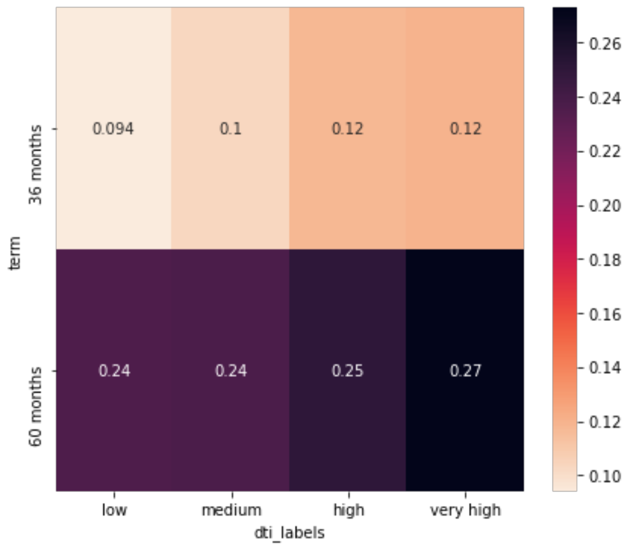
BIVARIATE ANALYSIS

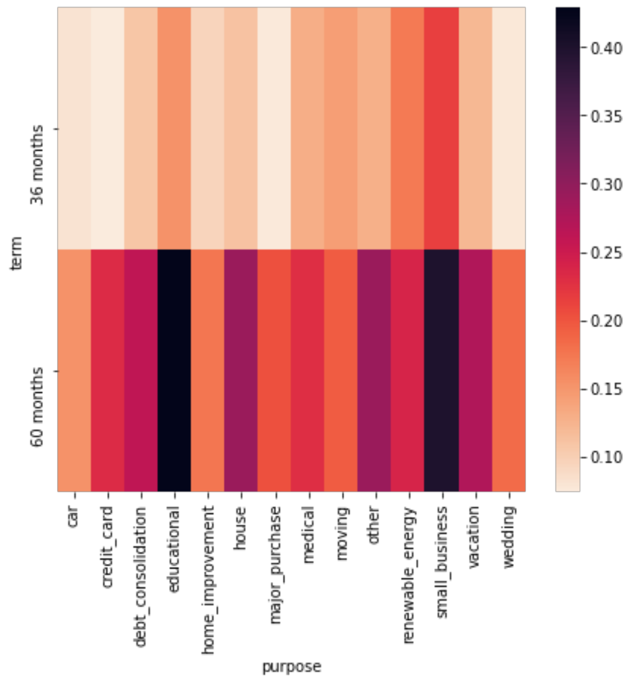






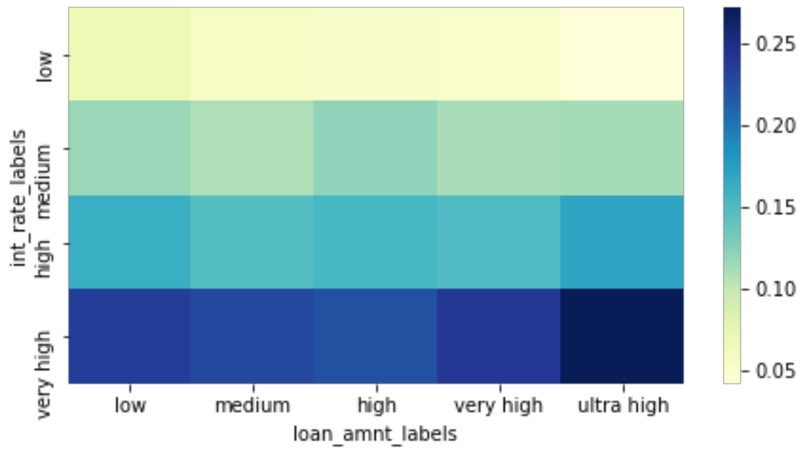


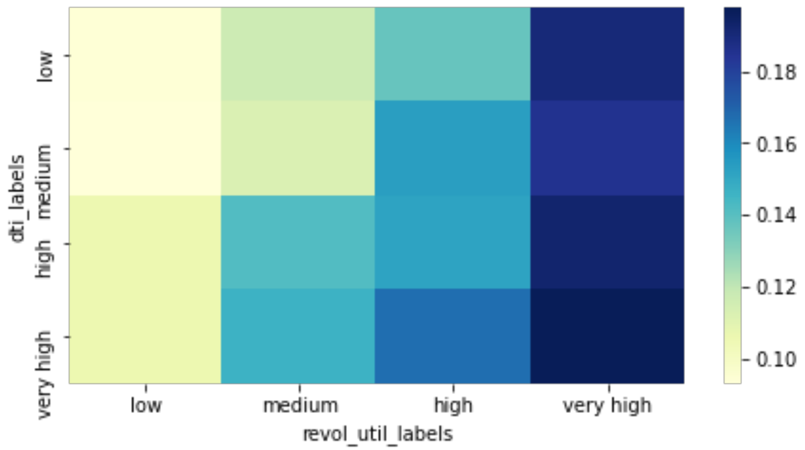


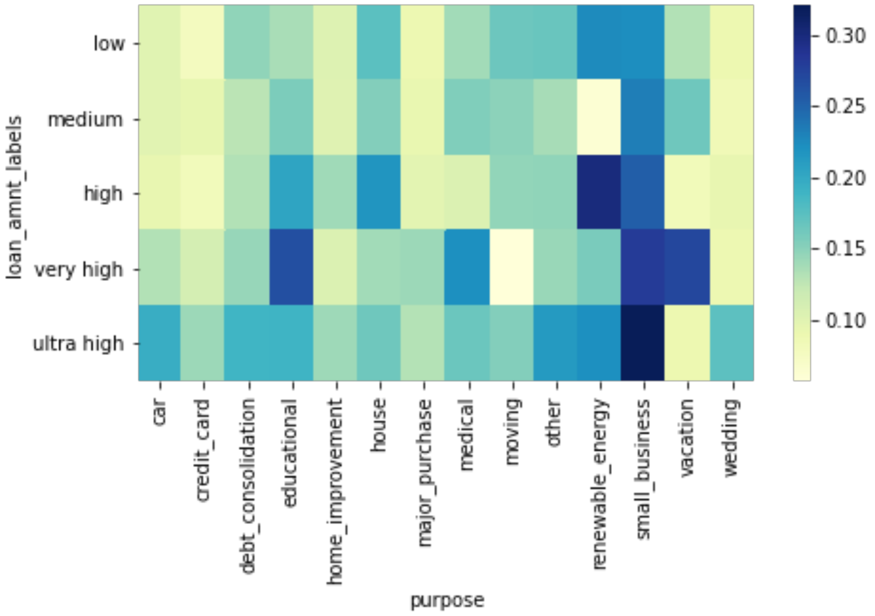


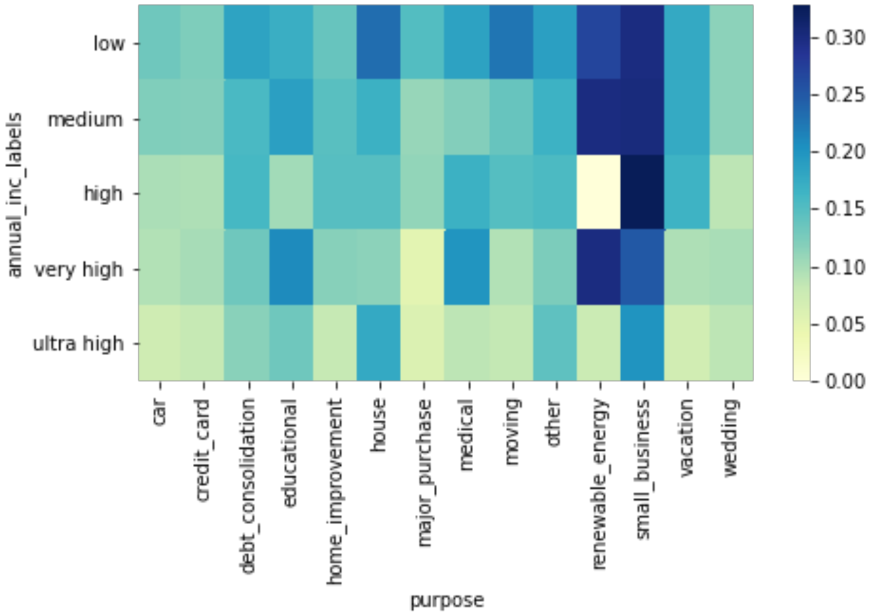
- People with low income who take 60months loan are at high risk of defaulting.
- Long period loans coupled with high interest rate lends should be avoided
- People, irrespective of their annual income, tend to do worse at returning money if they have borrowed for a longer duration (60 months)
- Grade G tend to do worse if they were sanctioned for loan for 36 months
- Individuals who take 60 months loan with low DTI ratio, should be scrutinized and lending them should be avoided
- Loans taken for "Educational" purpose and "Small Business" tend to be very risky, especially if, taken for 60 months.

BIVARIATE ANALYSIS



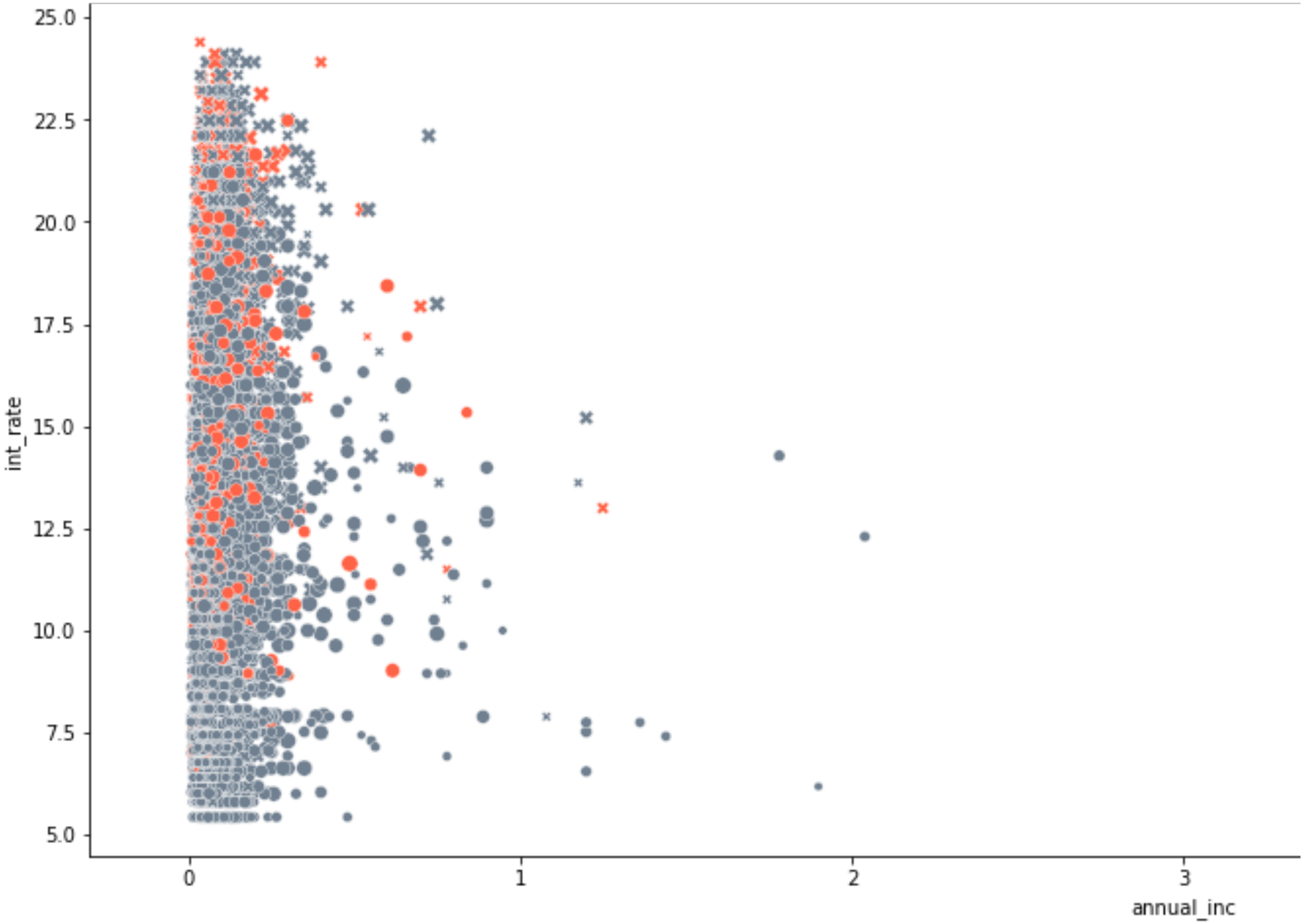






- If Loan is taken for huge amount at a very high interest rate, risk of defaulting is very high.
- Borrowers with low DTI ratio and very high revolving utilization lie in a high-risk category.
- High Risk Zone: Loans taken for Small business and renewable energy and especially if the loan amount is very high or ultra high
- Borrowers with low and medium annual income, borrowing for renewable energy and small business lie in a high-risk zone and they need to be approved sparingly

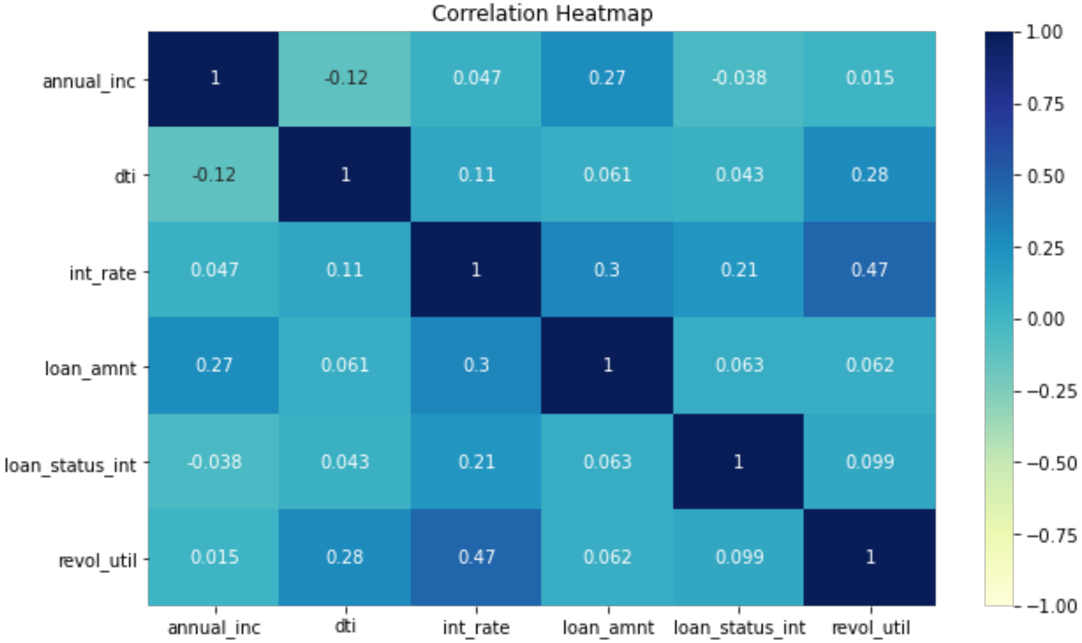
MULTIVARIATE ANALYSIS



- Defaulters start popping up at higher interest rates (7.5%+). If computer recommends high interest rate, bank must be alert about lending money.

- Defaults which have happened interest rates 7.5 to 10% seemingly happen for 36months term

- Among the top outliers, there is seemingly only one Defaulter who took a 60 months term



- Defaulters are highest correlated with interest rate. Higher the interest Rate, more likely the borrower will default