

LENDING CLUB CASE STUDY

Name: Abhishek Priydarshi

Background of Lending Club Company

Lending club is the marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.

- Borrowers interested in loans apply by filling an application on online platform at [LendingClub.com](https://www.lendingclub.com)
- Based on the data investors decide if they want to lend money to them based the loan parameters like interest rate, term of loan, monthly instalment, purpose, annual income, etc.
- While selecting the loans, investors also consider their own risk tolerance, investment portfolio goals and time horizon.

Business Objective

To understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. This will help the company/investors in making the decision to either accept /reject the loan.

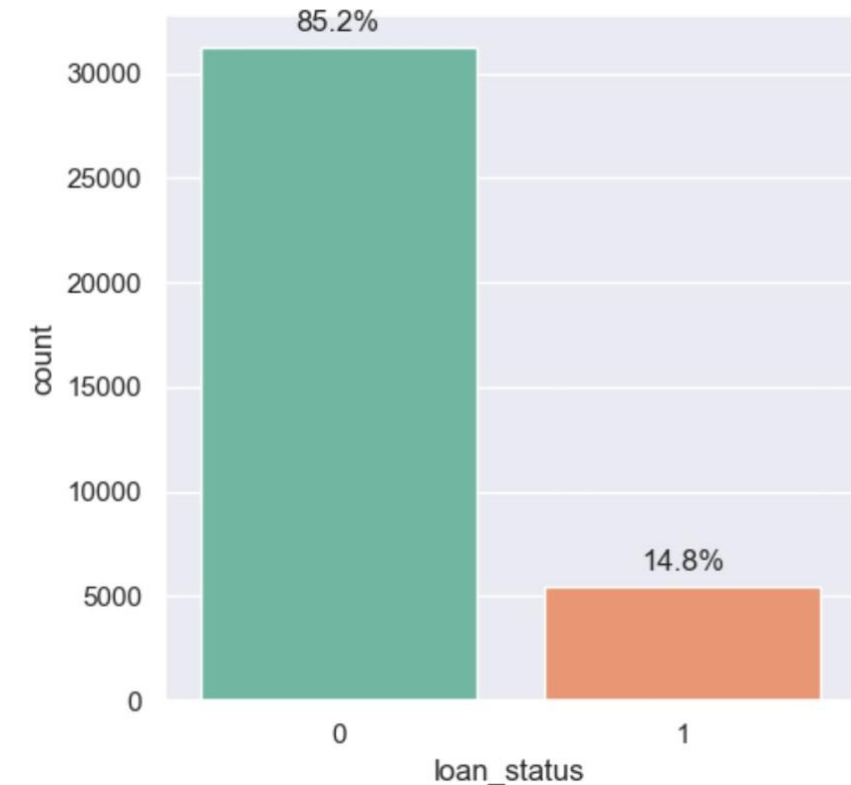
- > Data Dictionary has the detailed description about each columns in loan.csv file
- > There are mainly three different types of data in the data file such as
 - > Demographic data - such as Annual Income, Employment, State, etc..
 - > Loan Characteristics data - such as Interest rate, Grade, Loan Amount, Term, etc..
 - > Customer Behaviour data - such as past-dues in 2yrs, Revolving Balance, derogatory records, etc
- > Since loan status column has the values whether the loan is Fully Paid or Charged-Off, we can consider this as the target variable.
- > Rows with loan_status is “current” doesn’t required for the analysis, hence removing those rows (1140 rows)

Target Variable : loan_status

- Overall Default Rate is 14.8% as per provided dataset

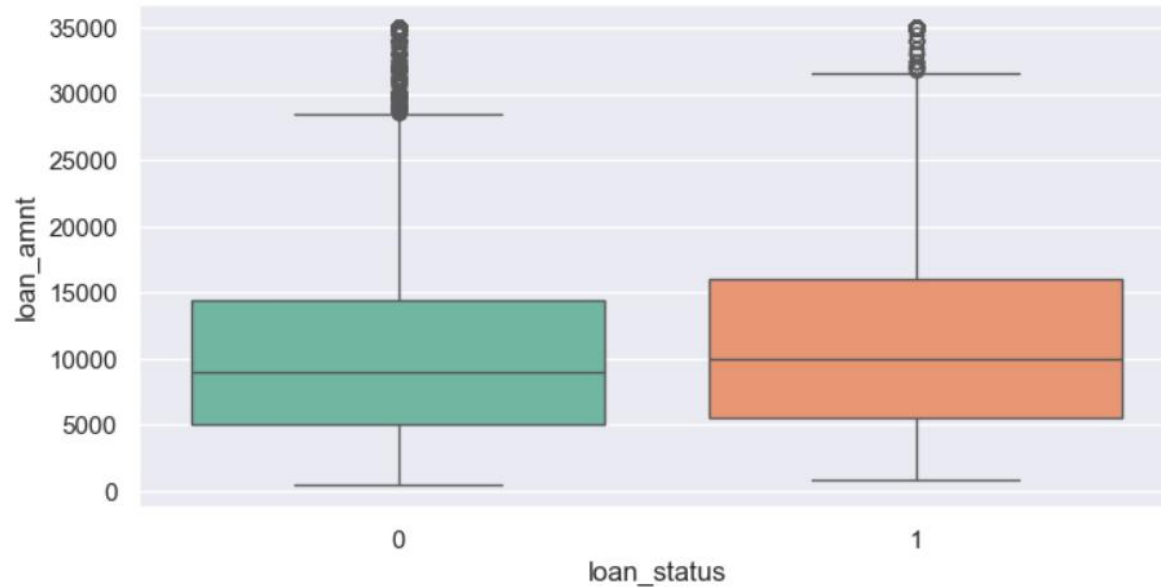
Assumptions:

- All the Active loans are not considered in Analysis
- Fully Paid Loans are represented as Code Zero(0) : **Non-Defaulters**
- Charged-Off Loans are represented as Code One (1) : **Defaulters**

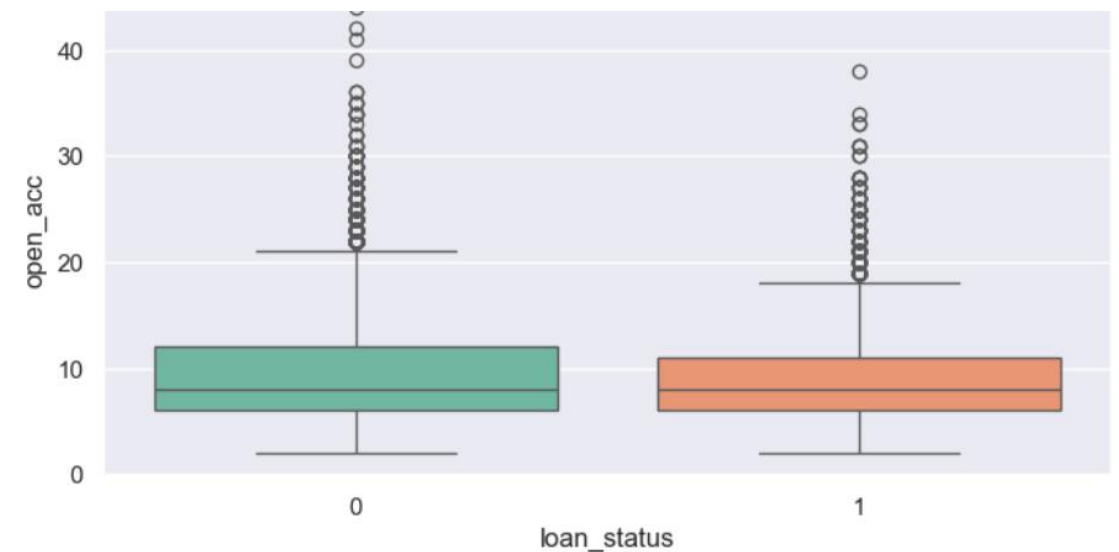
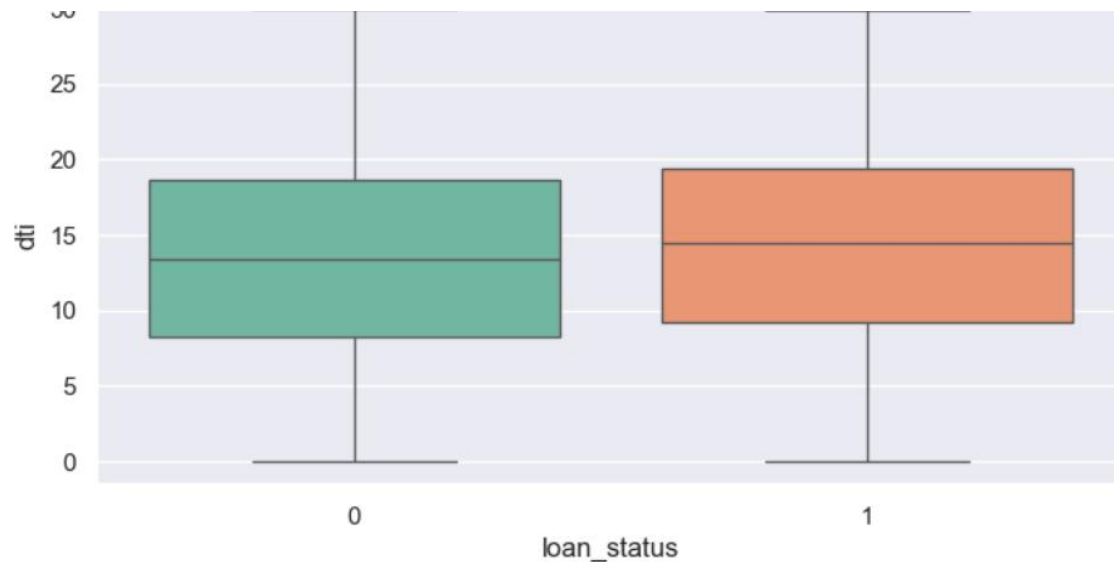
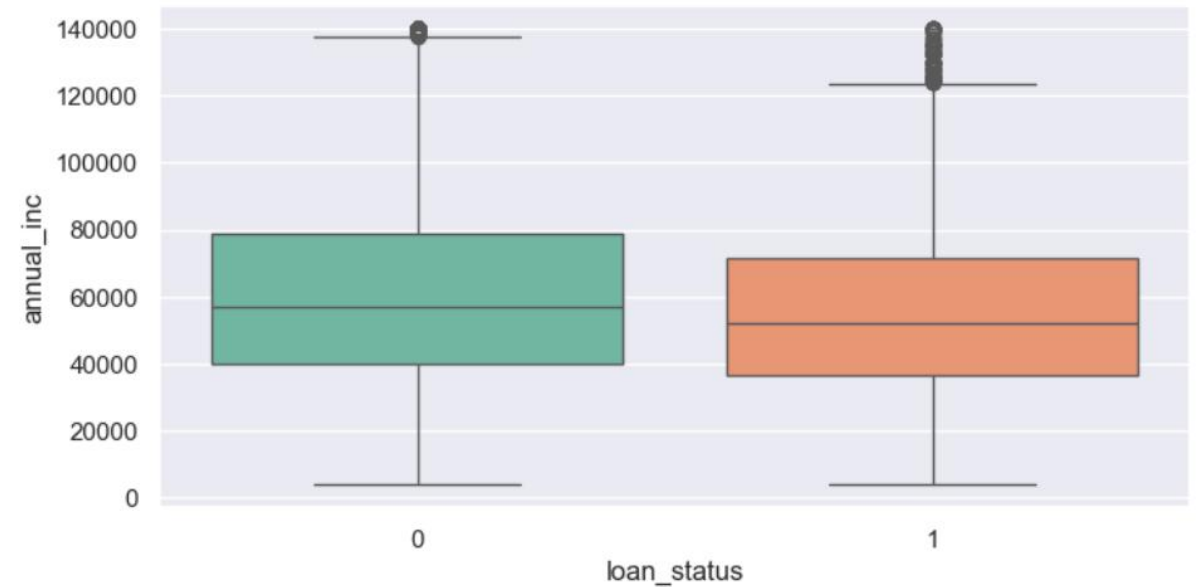


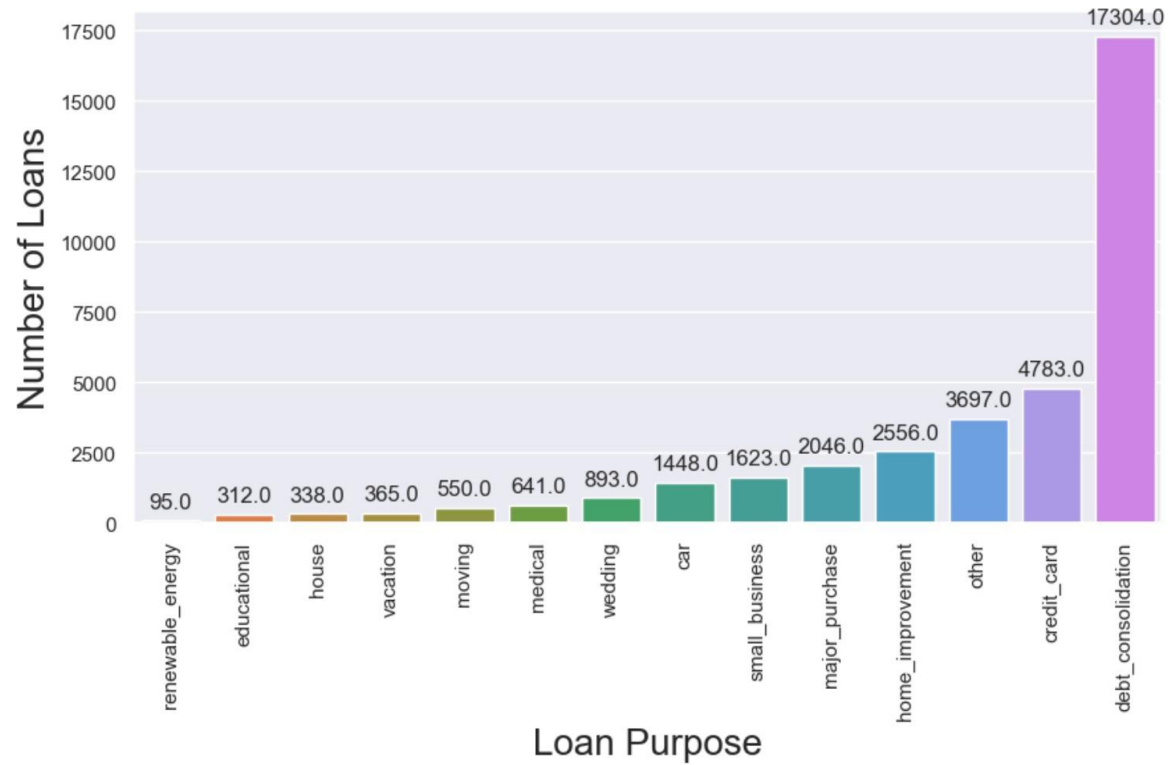
- > Dataset contains 111 columns and ~39K rows
- > There are 56 columns which has missing values more than 90%, lets remove those columns
- > Number of values in pub_rec is greater than pub_rec_bankruptcies, Hence we can drop column pub_rec_bankruptcies.
- > Since there are no rows which have more than 5 missing values (around 10% of total column), not dropping any rows
- > The columns like 'desc' and 'mths_since_last_deling' are not useful inferring target variable, hence removing it
- > There are anther 25 columns which doesn't contribute to the determination of target variable, hence dropping them.
- > Create derived metrics, month and year from Datetime column
- > Handling Outliers in Annual Income - remove the records above 95%

Box Plot



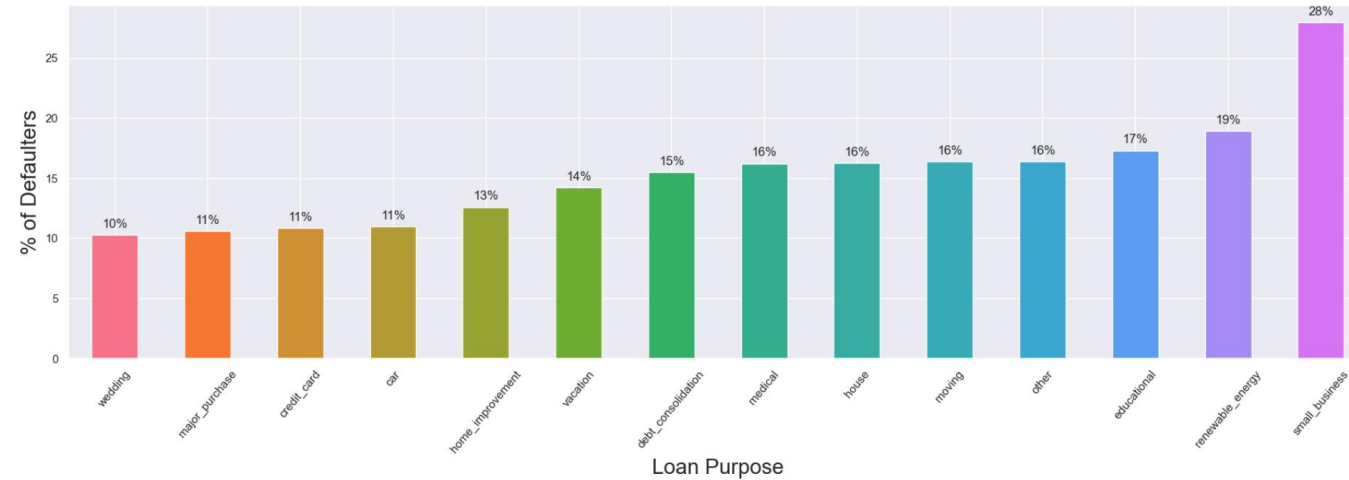
Box Plot





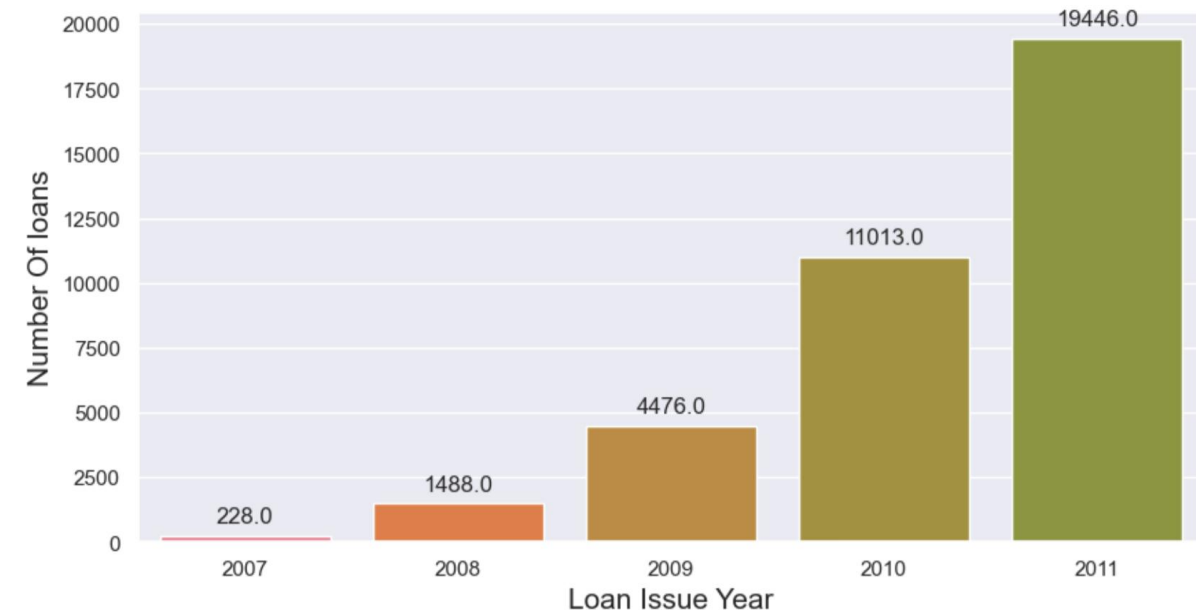
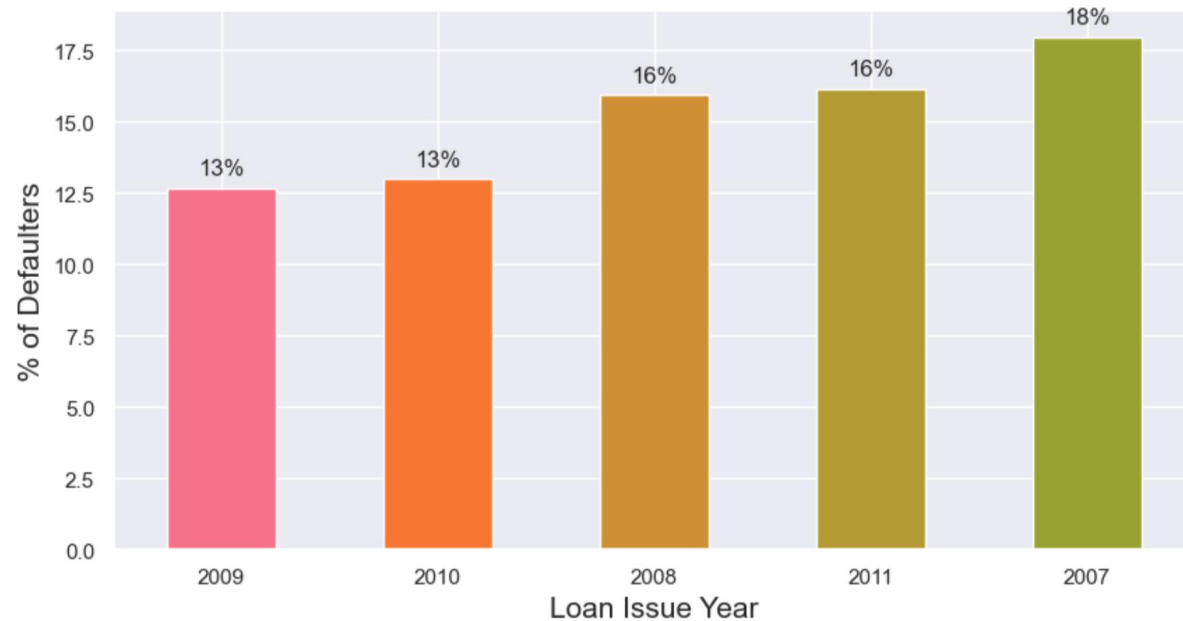
Popular loan purpose:

- Debt Consolidation
- Credit Card
- Other
- Home Improvement
- Major Purchase
- Small Business



Default Rate : Popular Loan purpose

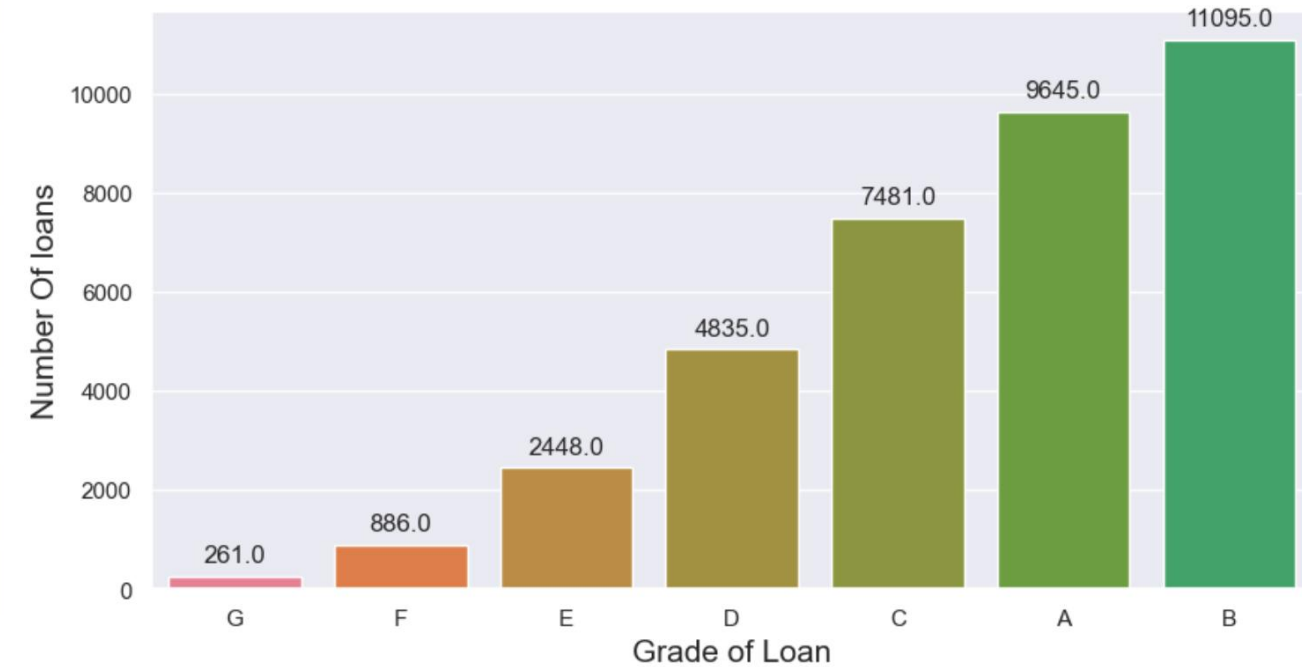
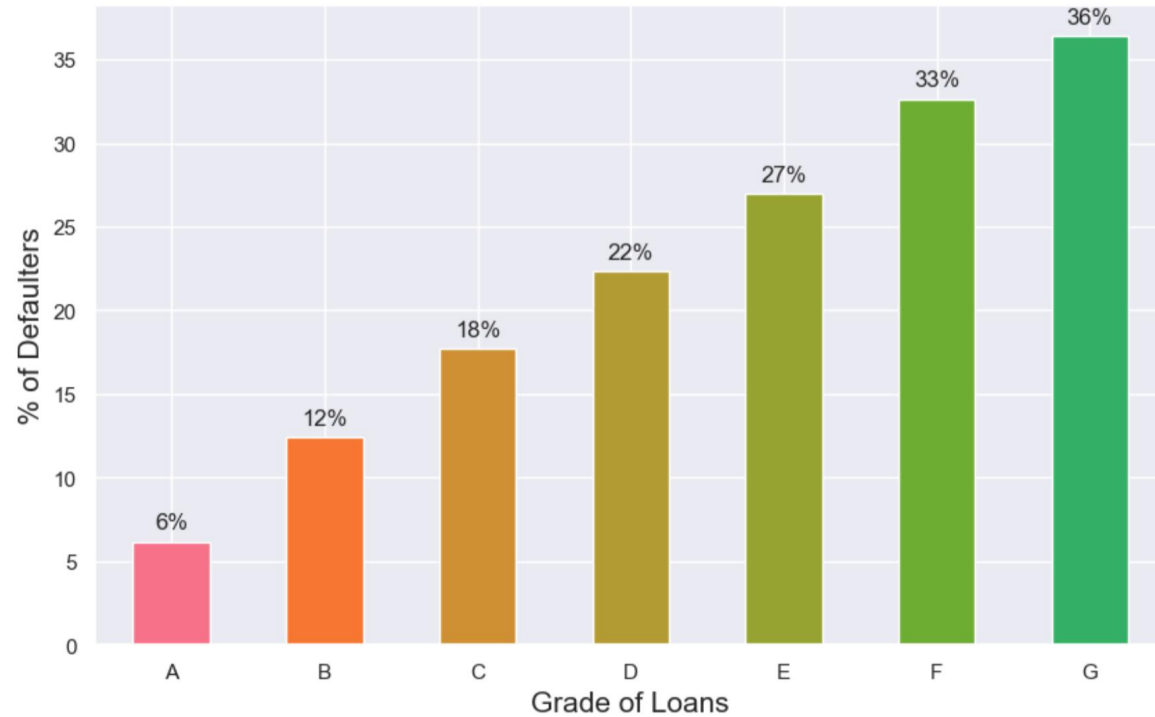
- Small Business 28%
- Debt Consolidation 16%
- Credit Card 11%
- Other 16%
- Home Improvement 13%
- Major Purchase 11%



- Default rate decreasing from 2008 to 2010, sudden increase in 2011.
- Highest in 2007 followed by 2011.

- As per the trend, loan applications are increasing from 2007 to 2011
- Maximum loans were issued in 2011.

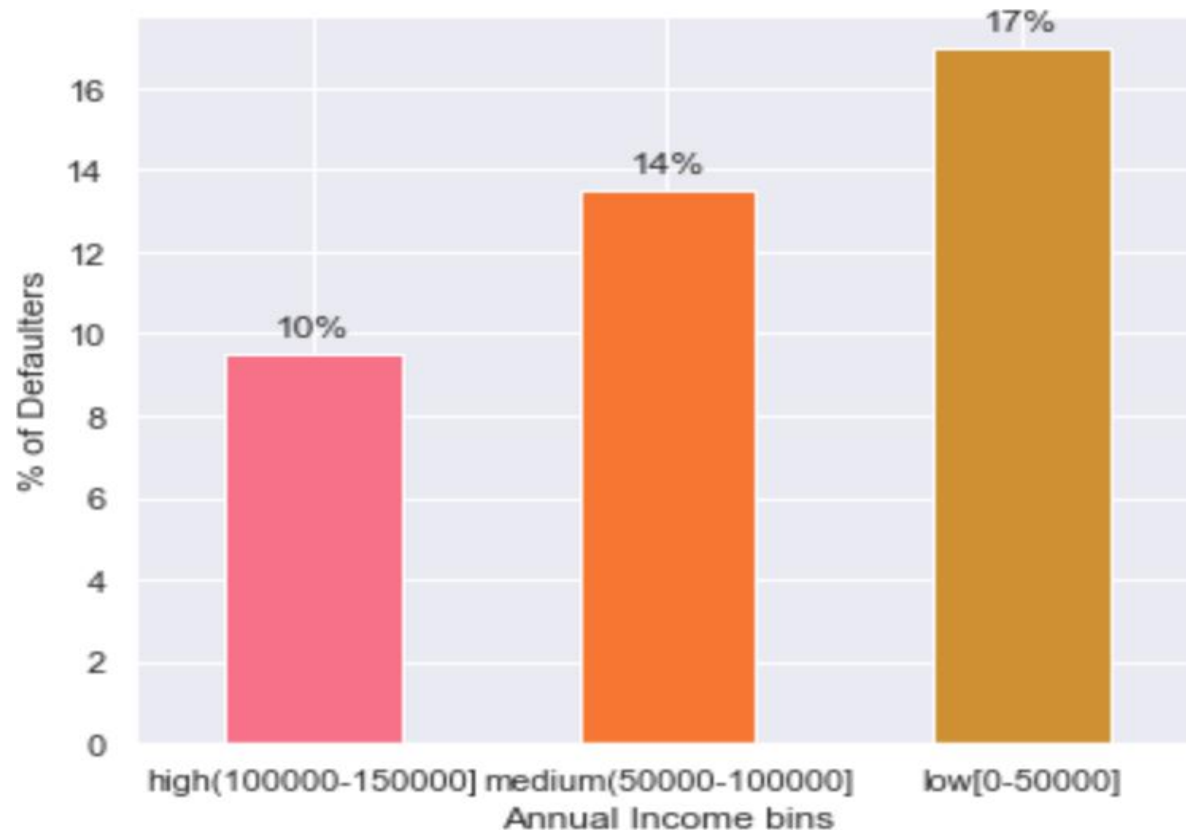
Grade of Loan – By Frequency distribution & default rate



- As the grade increases from A to G the default rate increases

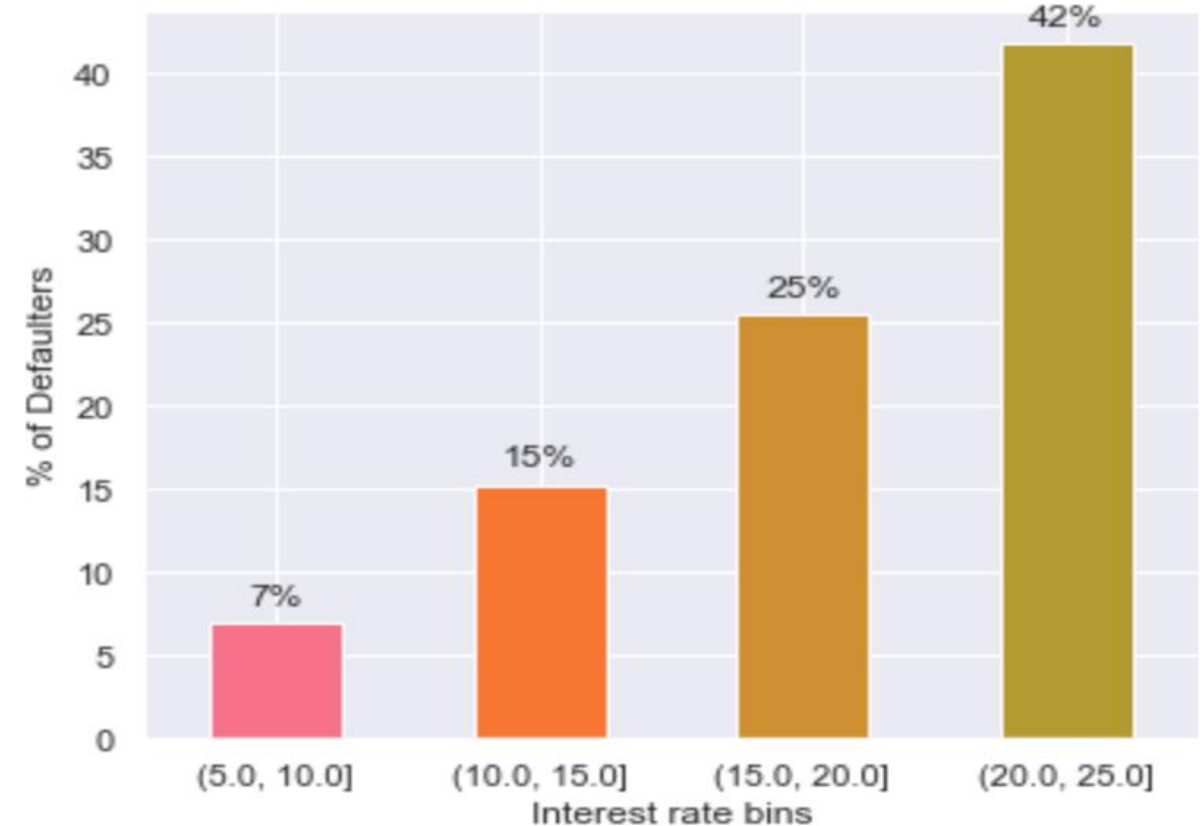
- Number of loans availed is less towards lower grade

% of loan defaulters on basis of binned annual inc



- Low annual income leads to high default rate

% of loan defaulters on basis of binned interest rate



- Default rate increases with interest rate
- More the interest rates , higher are the default rates
- Highest for 20-25% interest rate bin

% of loan defaulters on basis of annual income

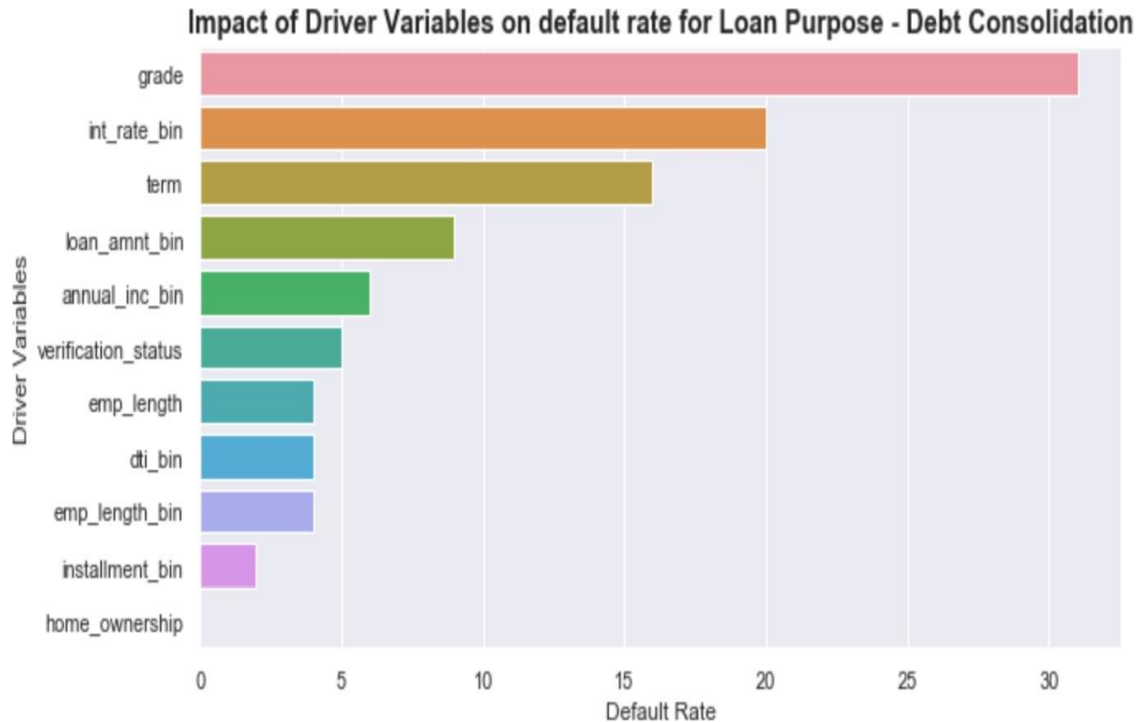


- Higher the loan amount , higher is the default rate.
- Default rate is highest for loan amount between 30000-35000 USD.



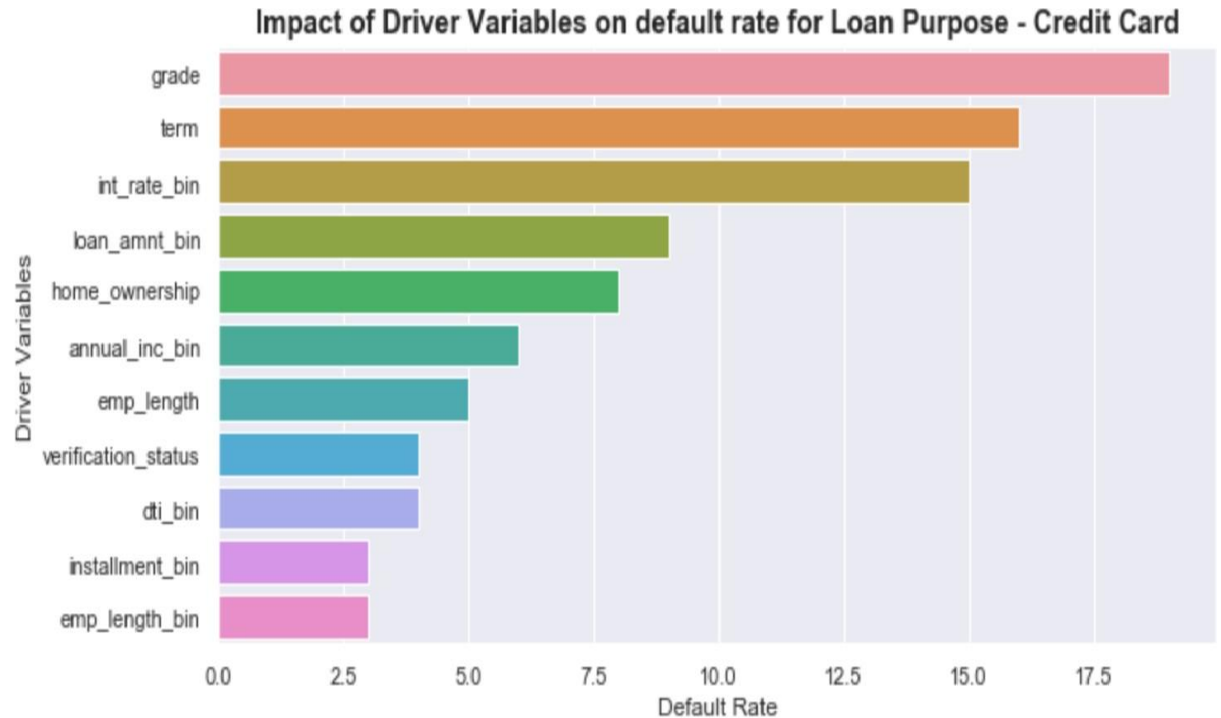
Top-5 Driver Variables for Loan Purpose - Debt Consolidation & UpGrad Credit Card

Debt Consolidation



1. Grade
2. Int Rate
3. Term
4. Loan Amount
5. Annual Income

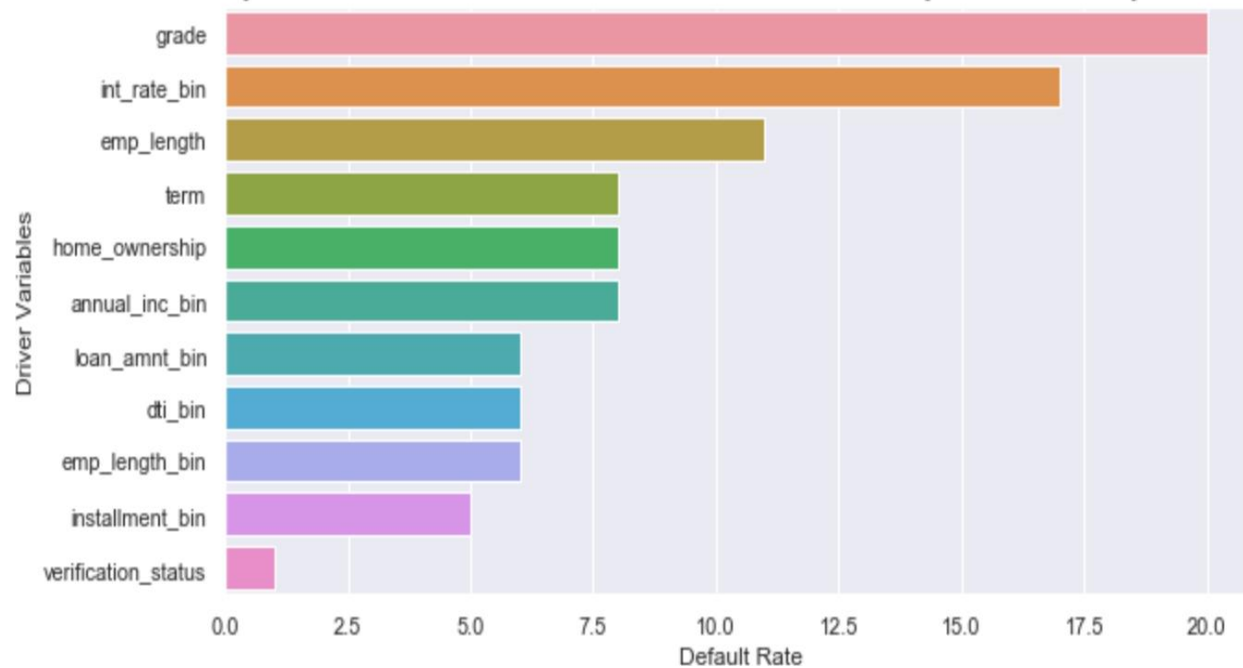
Credit Card



1. Grade
2. Term
3. Int Rate
4. Loan Amount
5. Home ownership

Home Improvement

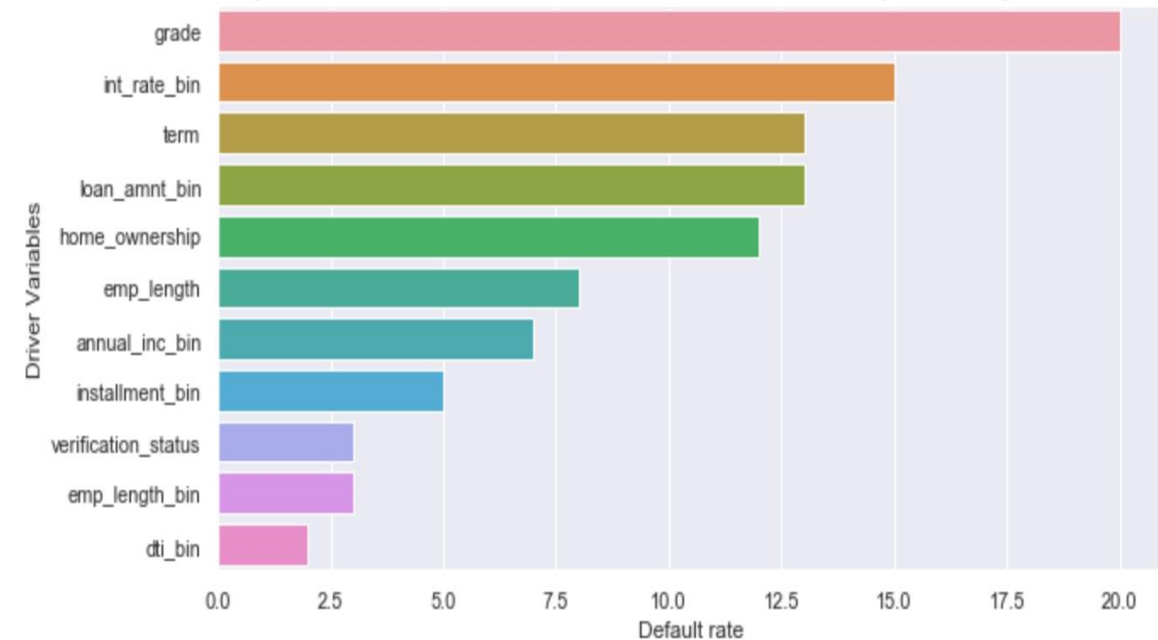
Impact of Driver Variables on default rate for Loan Purpose - Home Improvement



1. Grade
2. Int Rate
3. Emp Length
4. Term
5. Home Ownership

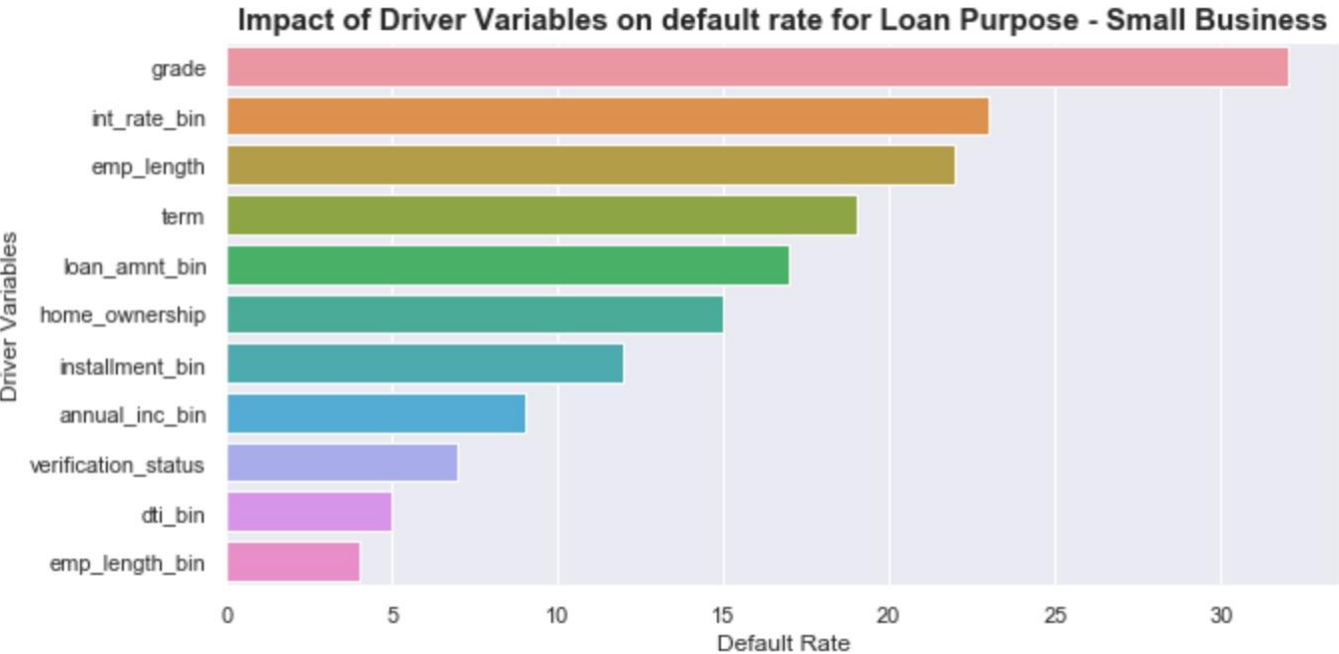
Major Purchase

Impact of Driver Variables on default rate for Loan Purpose - Major Purchase



1. Grade
2. Int Rate
3. Term
4. Loan Amount
5. Home ownership

Top-5 Driver Variables for Loan Purpose - Small Business



1. Grade
2. Int Rate
3. Emp Length
4. Term
5. Loan Amount

Driving features to be considered before approving loan :

1. **Purpose of Loan** – loan purpose has to be carefully considered, for e.g; giving loan to “small_business” very risky as it has the highest % of default rate.
2. **Annual Income** - Lower the borrower’s annual income, Higher are the chances of default.
3. **Grade&Interest Rate** – both of them are positively correlated and as grade/int rate increases, the loan default rate also increases. For e.g default rate is highest for grade “G” which also has the highest interest rate.
4. **Loan Amount** – Higher the loan amount , more is the loan default rate.
5. **Term** – Loans granted for 60-month term have higher default rate when compared to loans with 36-month term.
6. **Loan Issue Year** - There could be some unpredictable factors which leads to default. For e.g higher percentage of defaulters are those who took loan on 2007. One of the reason could be because the year after that faced the worst recession.