# Lending Club Case Study

Group:

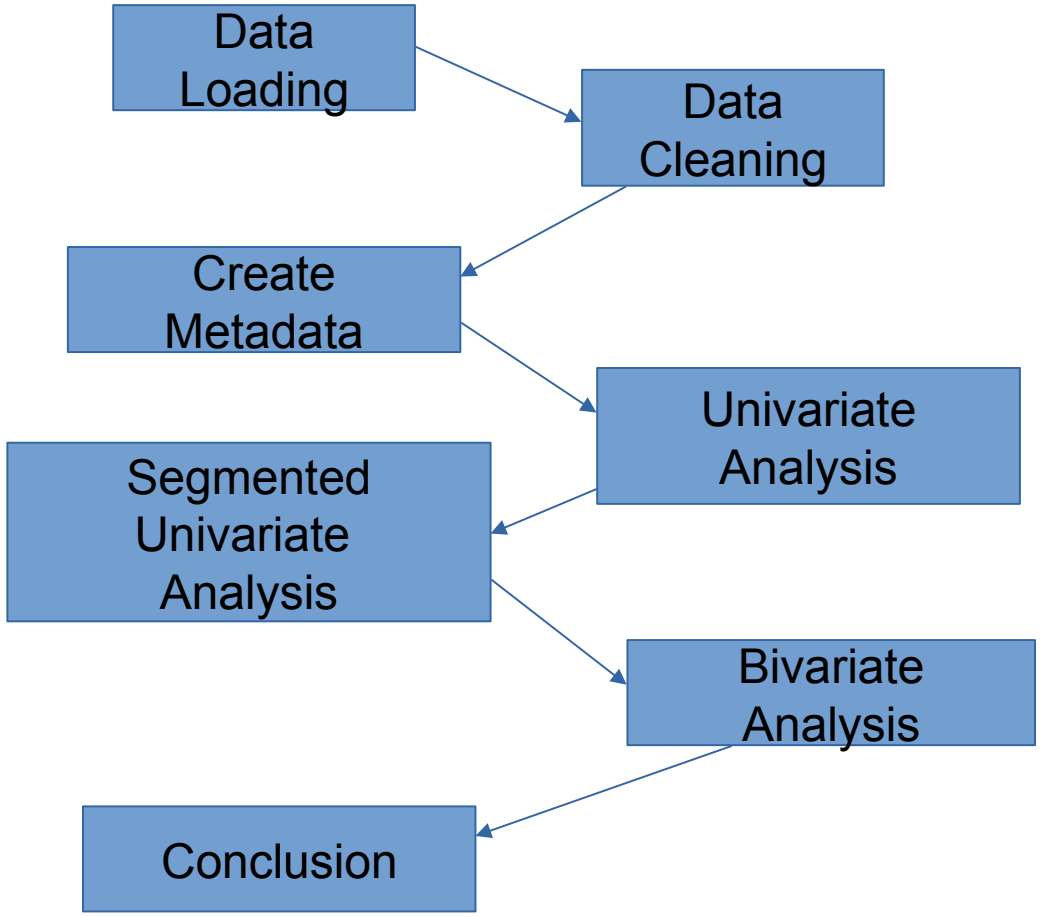
Mandheer Singh

And Kailash Mirani

# Abstract

- Exploratory data Analysis
  - We have the data of loan given to urban customers
  - We need to find patterns if a person is likely to default or not based on consumer or loan attributes
  - While lending money banks have two types of risks i.e.
    - Loss of business-not approving loan to repaying customers
    - Financial Loss- Approving loan to the customers not repaying
  - Objective of this study is to find risky applicants and reduce loans for lower financial loss
  - Find variables which are strong indicators of default

# Problem Solving methodology




- Data loading and understanding
- Data cleaning and preparation
- Creating metadata
- Univariate analysis
- Segmented univariate analysis
- Bivariate Analysis
- Conclusions

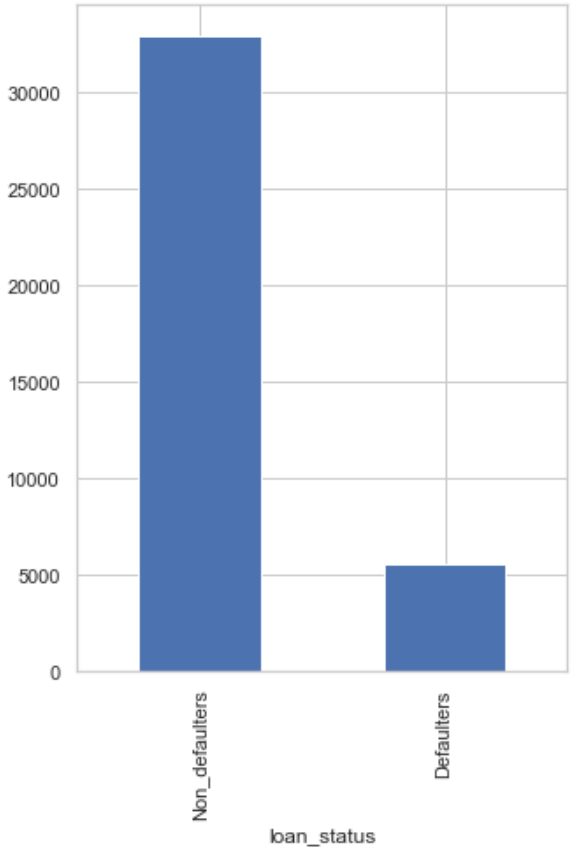
# Data Loading and understanding

- After Loading the data we found that the shape of the data is (39717, 111)
- Data about the loans issued has loan attributes and member attributes
- id and member_id columns could be used to uniquely identify records
- Data is available for loan issued from 2007 to 2011
- There is loan_status column having {‘Fully Paid’,’Current’,’Charged Off’}
- Loans are divided into grades,subgrades
- There are many columns which do not have any records.
- Many columns have same entry for all records
- Many columns are irrelevant to this study

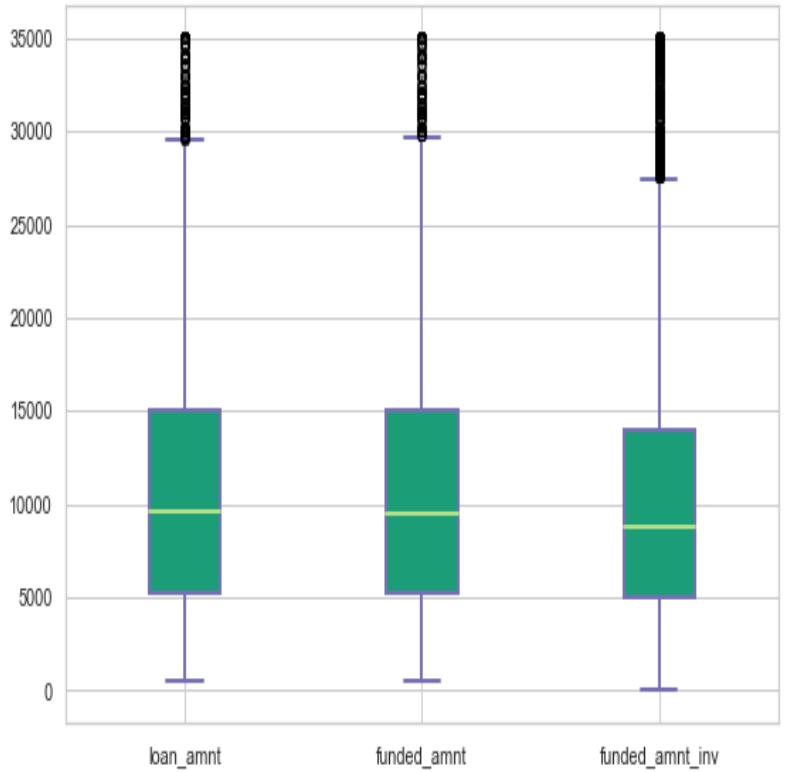
# Data Cleaning

- We found that there is no completely empty row
- We found there are 54 completely empty columns, so we dropped those
- Then there are irrelevant or unuseful columns or columns with same values for all, we dropped such columns
- We have found columns with lot of missing values like ['mths_since_last_record','mths_since_last_delinq','desc'] so we dropped such columns
- Then we changed some data types from int to object for columns [id,memeber_id]
- We also changed data type of dates from object to datetime
- We converted annual_income to monthly_income and dropped annual_income column
- Int_rate converted from object to int64
- We imputed missing values of some categorical columns using “blank” word
- After Cleaning we got a dataframe of (38577,21)

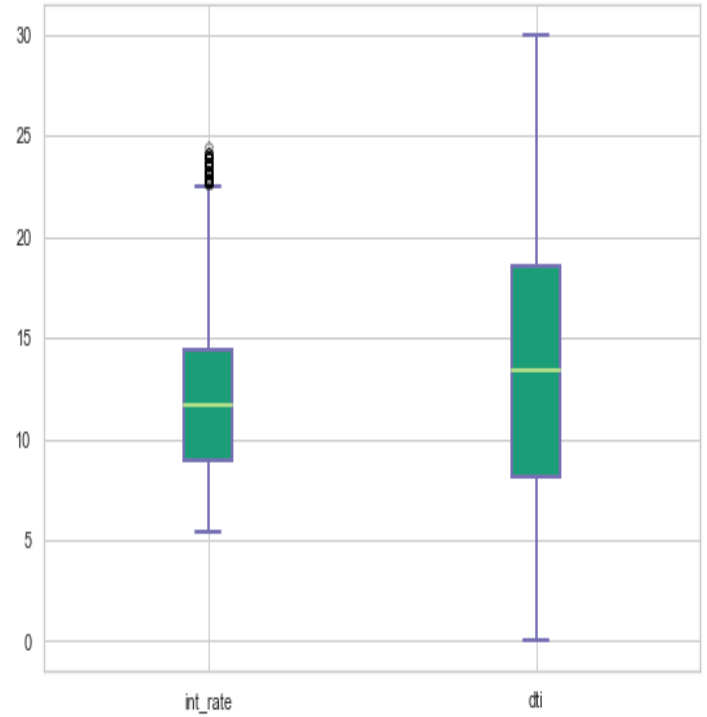
# Univariate Analysis




- Above plot shows the frequency of defaulters and non defaulters
- There are around 35000 non defaulters and around 5200 defaulters

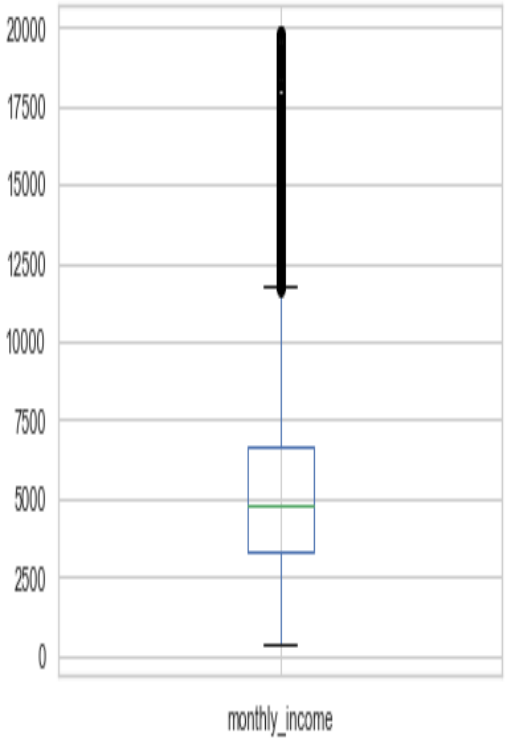



- Presented variables are in similar range
- Against funded amount inv, loan amount is littile more
- Investers are not lending full amount
- They are investing lesser than recommended by LC in most of the cases



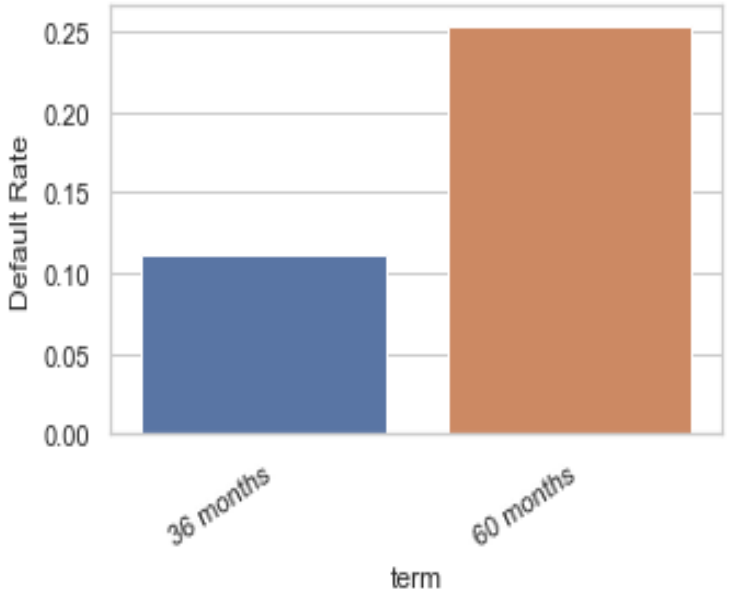

- We can see ranges of int_rate and dti
- Median for int_rate is around 12 and for dti it is around 14

# Univariate Analysis



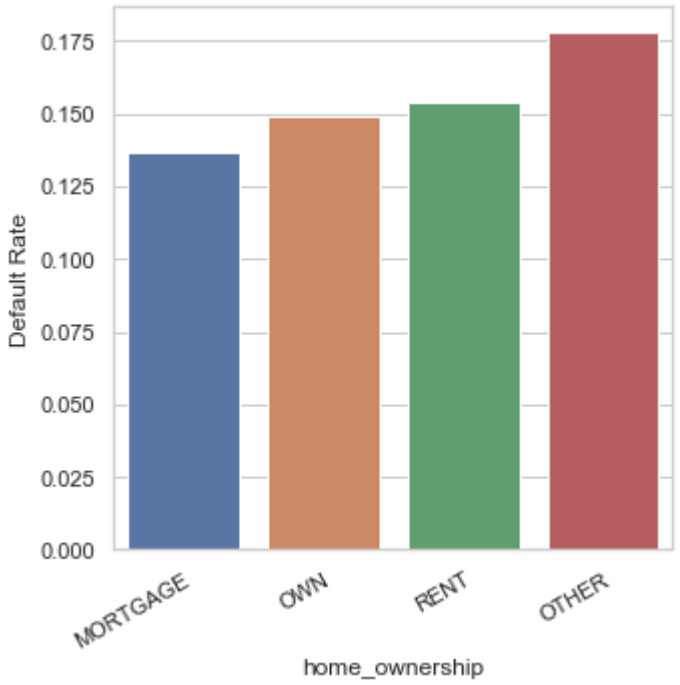

median income is around 5000 per month


- We have plotted loans wrt issue date
- Loans have been issued from 2007 to 2011
- With the timeline No. of loans are increasing

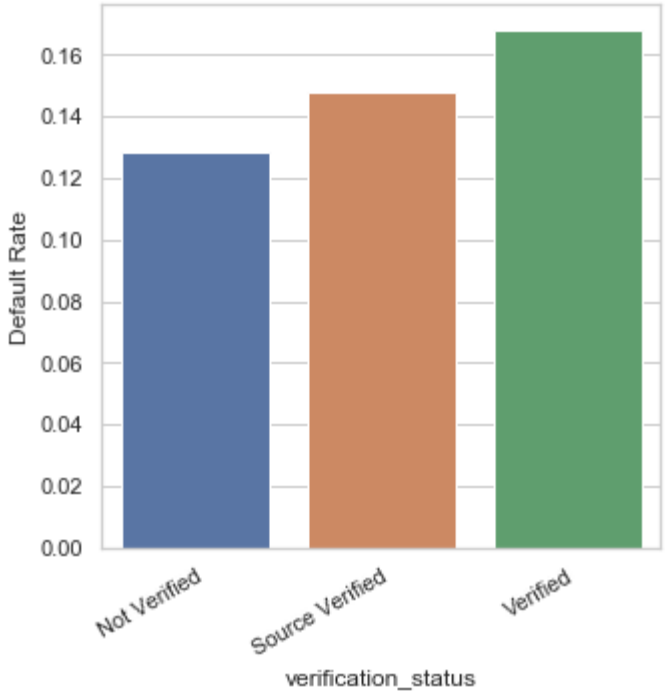



- We can see that there is approx 25% default in 60 months loan while only 10% default in 36 months loan
- Thus It is a strong indicator of default

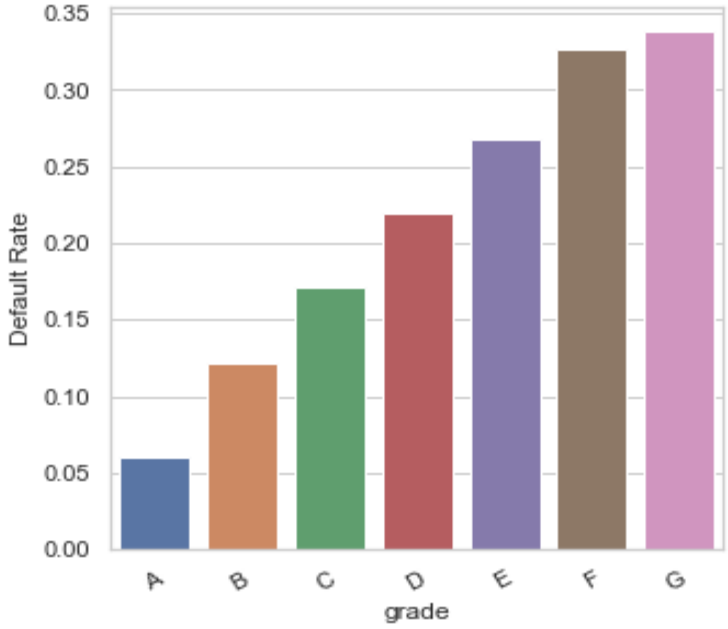
# Segmented Univariate Analysis




- From the above plot it is clear that people with Other category have highest default rate
- Thus If an applicant has other as ownership of home we need to be extra careful

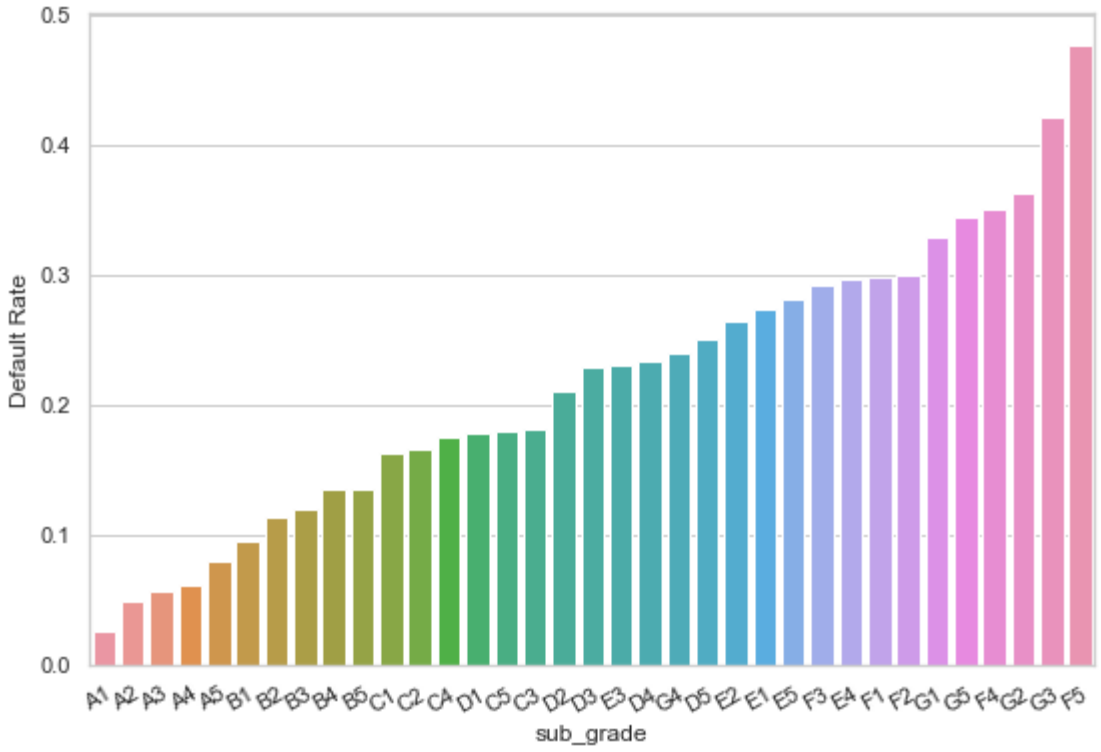


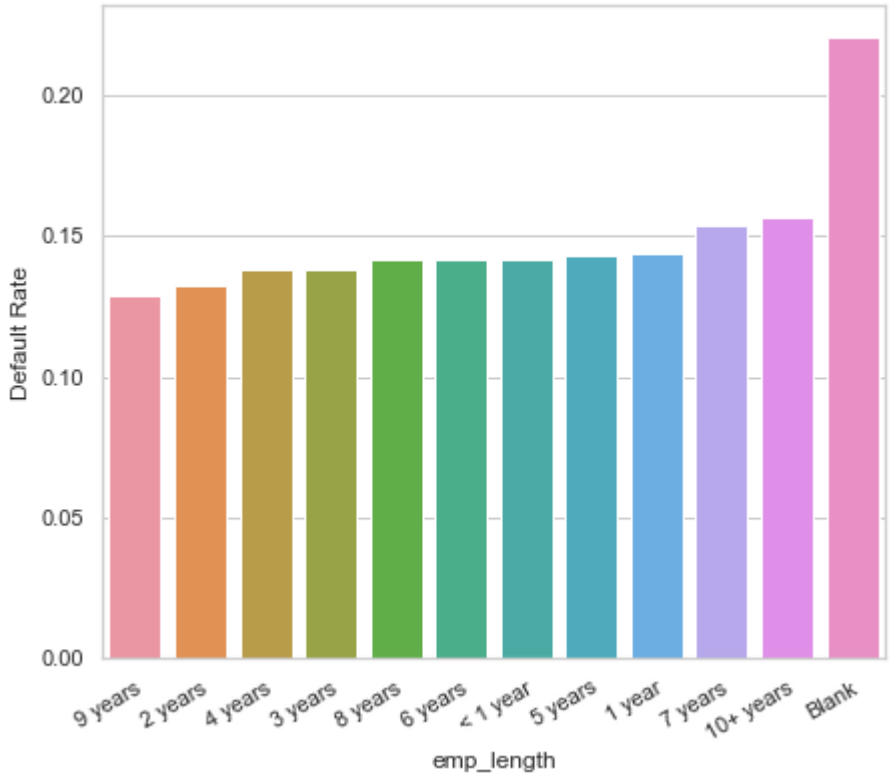

- We need to notice that Verified applicants has more default than not verified
- It shows that we need to check our verification mechanism
- May be there is some wrong practice in that




- Grade is strong indicator of defaulter
- Grade A B C are safer loans
- While Other grades are a bit riskier

# Segmented Univariate Analysis

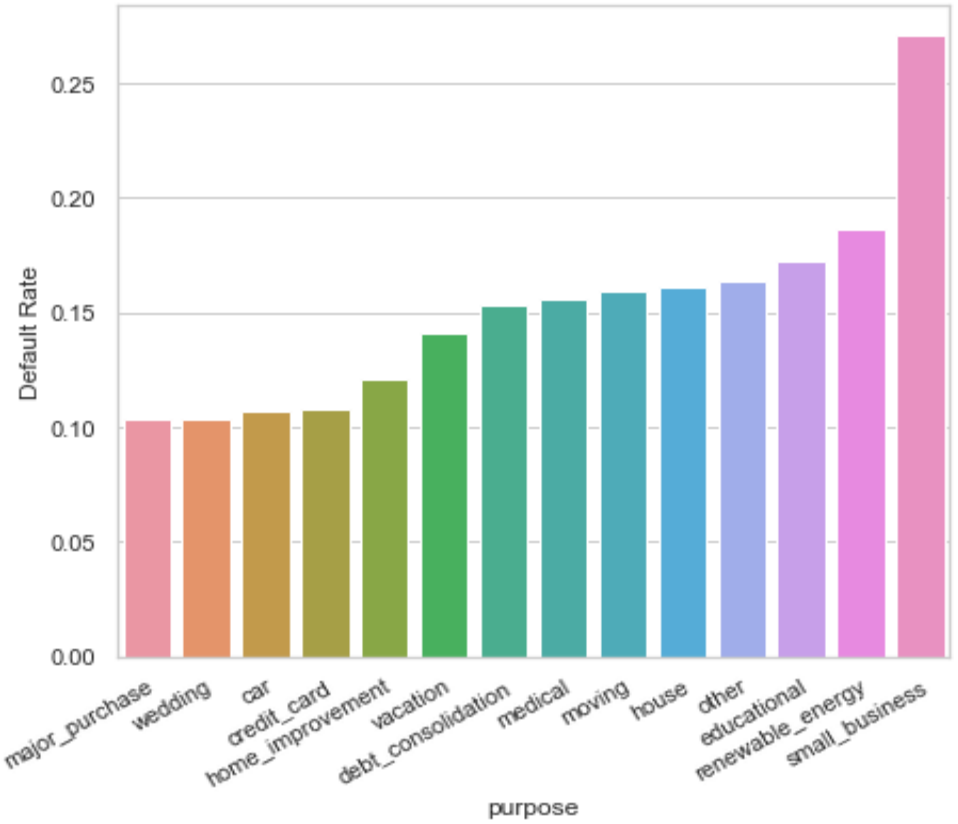


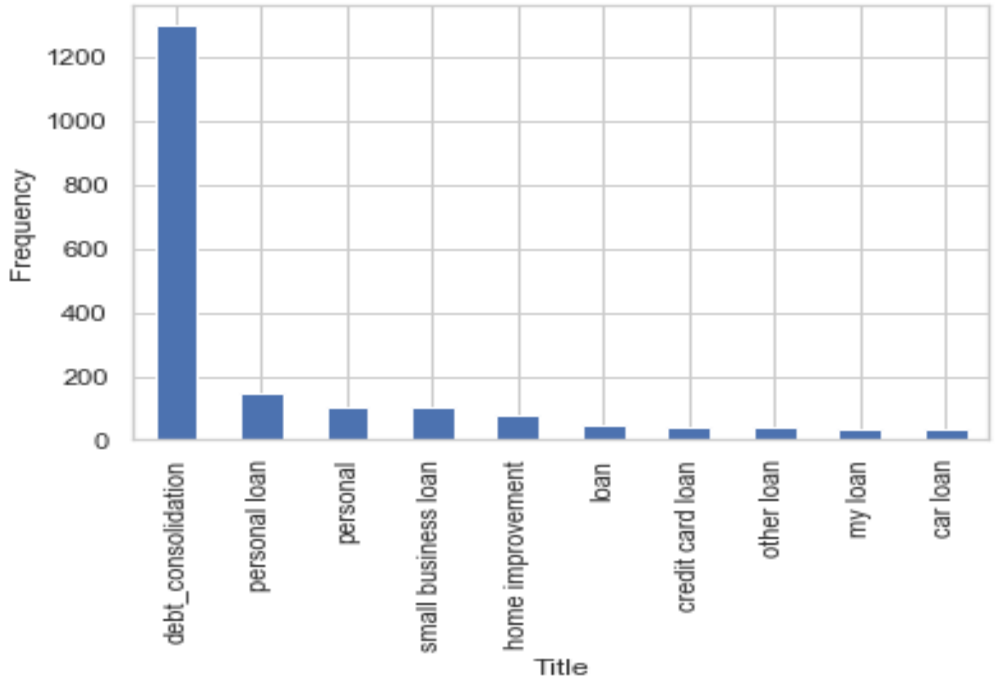

- Smilar like Grade, Subgrade is also strong indicator of defaulter
- F5, G3, G2 and so on are indicating high default

- We find out that people who do not mention there employement length are defaulting more
- May be they are not working or not getting regular income

# Segmented Univariate Analysis

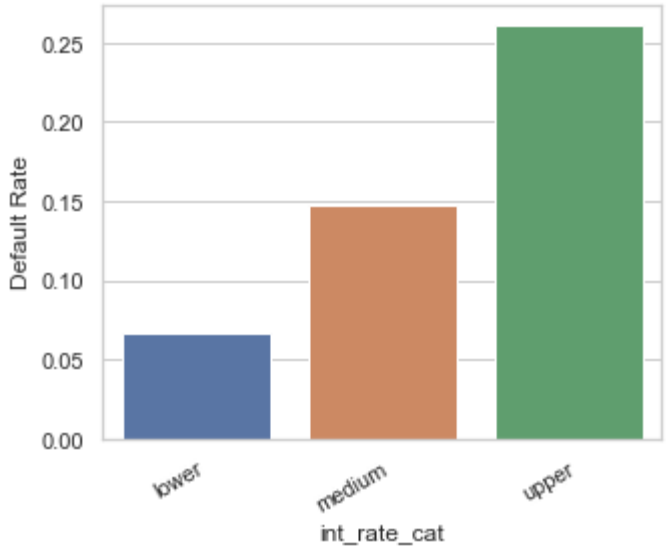



- We found that there is more default in ‘small_business’,’renewable_energy’ ‘education’ and so on.
- Intrestingly we also find if we compare frequencies ‘debt_cosolidation’ has most in default
- We can say this is also a strong indicator of default

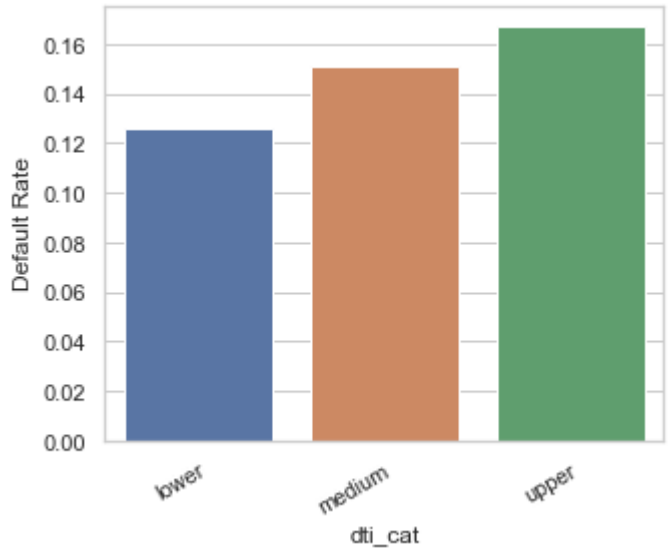



- There is a huge default for debt consolidation
- It is also a strong indicator of default

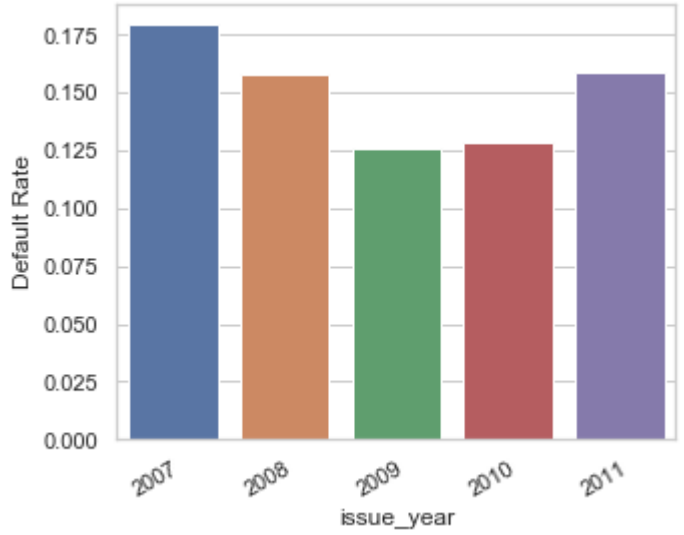
# Segmented Univariate Analysis




- There is highest default rate for interest rate above 15%
- As interest rate increases default rate also increasing
- This is also a good indicator of default

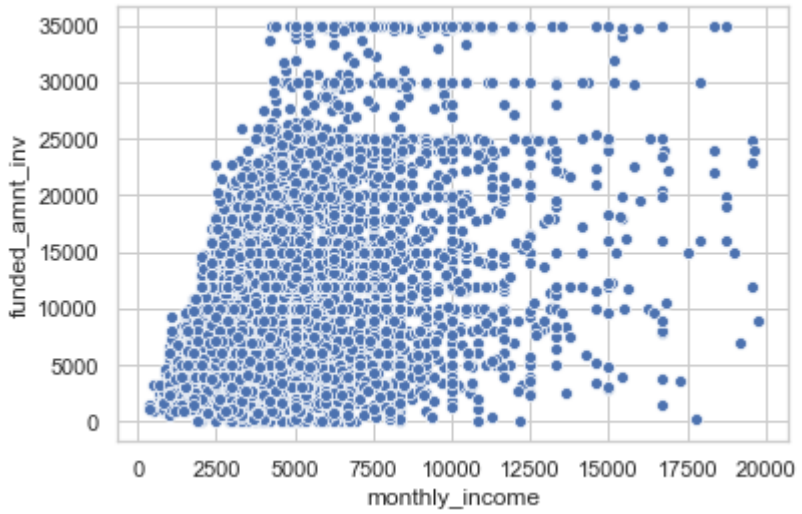



- Default rate is increasing with dti
- Showing that more debt more default or less saving more default

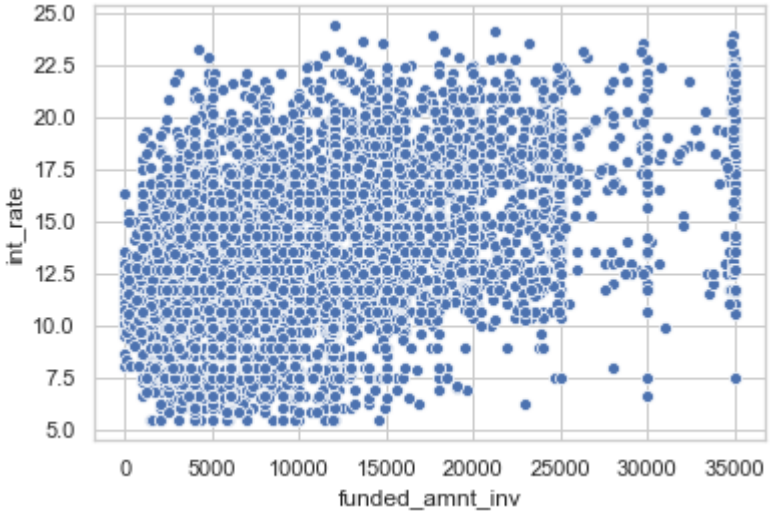



Highest default rate for loans issued in year 2007

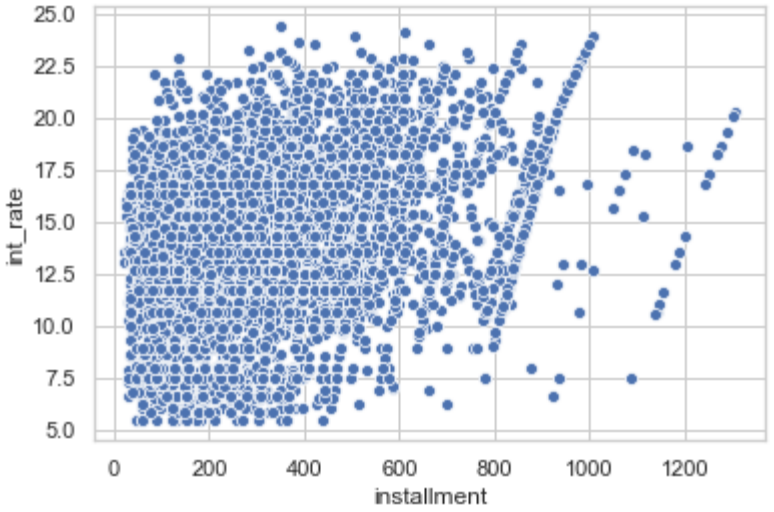
# Bivariate Analysis



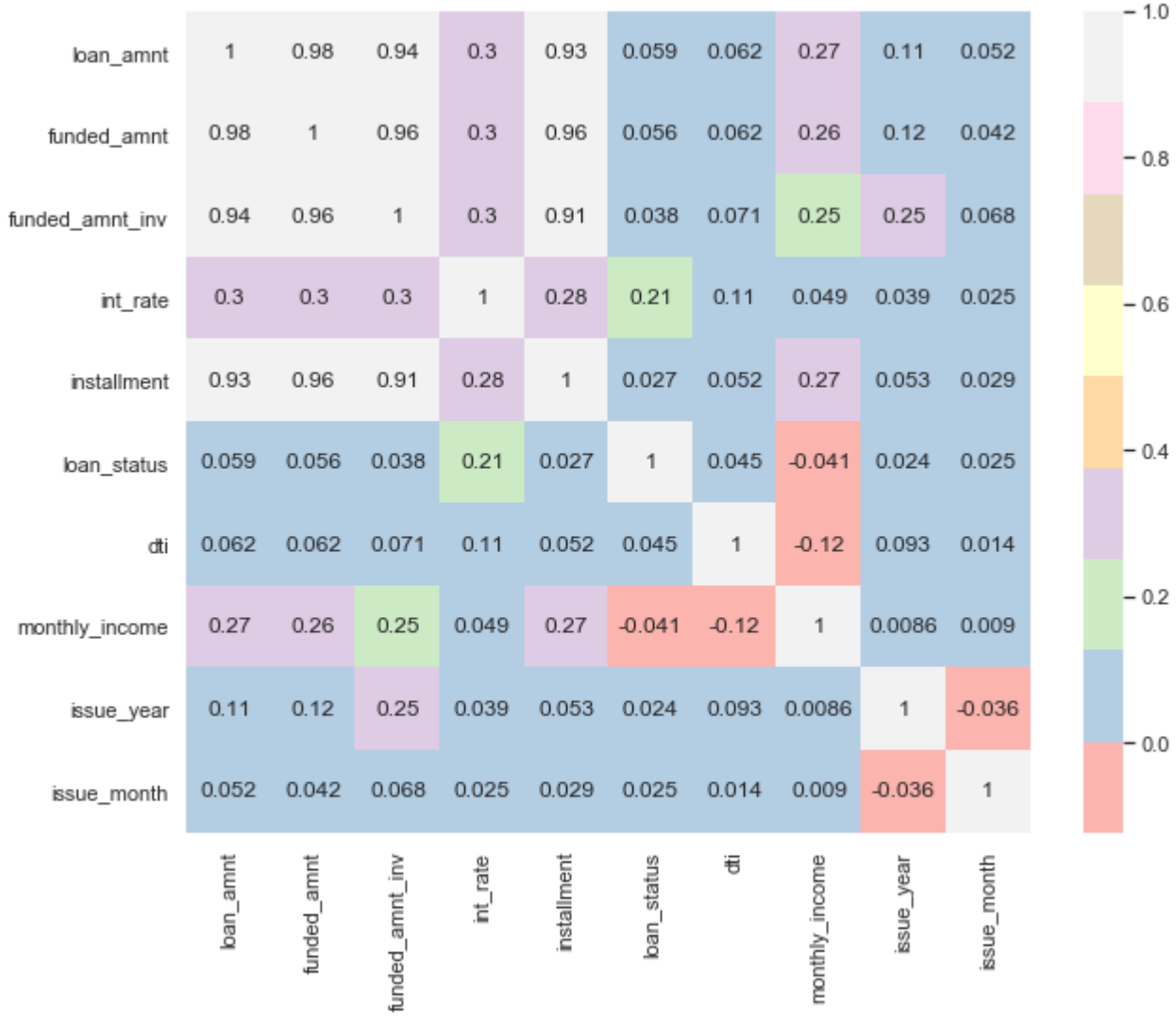

- We have plotted for defaulters only
- We can see at the top of the graph
- We have given 35000 loan to the people with less income too




- We have plotted for defaulters only
- We can see at the right corner
- We have given 35000 loan from interest rate 7.5 to 23 %



# Bivariate Analysis




- High correlation b/w loan_amnt and funded_amnt and funded_amnt_inv which is obvious
- Monthly income and dti has a negative correlation means if we increase monthly income dti will decrease
- Loan status has a good correlation with int_rate hence it is a good indicator
- Loan status has a small but negative correlation with monthly income
- More monthly income means more funded_amnt

# Conclusion

- We need to check verification schemes
- People more likely to default
    - loan for debt_consolidation,small_business
    - Who does not mention employment length
    - With Term 60 months
    - Higher interest rate
    - With Grades C,D,E,F and G
    - Subgrades F5,G2,G3 and so on
    - Home ownership as ‘other’

- Term,Grade , subgrade, purpose, title, int_rate, home_ownership are good indicators of default