Lending Club Case Study

Data understanding

- This data if for loans of lending club.
- Loans are graded (A-G) and then sub graded (A1-A5 etc.)
- Loans do not seem to have any collateral.
- Data set is for about 40K records
- Data has got many variables and about 111 columns
- Most of about 59 columns have no values and need to be deleted.
- Columns that we have chosen for Univariate and bivariate analysis are grade,sub grade,interest rate,loan amount,dti,home ownership,tenure,verification status,employment data,state,months since last delinq,open accounts,public records,total accounts etc.

Data Cleaning and Manipulation

- Remove all the null columns
- Remove all the columns having all values either null ,0 or ‘’
- Add new column int_rate_float where from int_rate by removing % and typecaste as float.
- Binning of loan_amnt to loan_amnt_bins (segmentation)
- Binning of int_rate_float to int_rate_bin (segmentation)
- Binning of dti, mths_since_last_delinq, total_acc, open_acc(segmentation)
- Add new column emp_tenure_num

'< 1 year': 0, '1 year': 1, '2 years': 2,'3 years': 3, '4 years': 4, '5 years': 5, '6 years': 6, '7 years': 7, '8 years': 8, '9 years': 9, '10+ years': 10

Data analysis

- 5 key variables that Univariate Analysis reveals strong chance of default
    - Interest Rate
    - Grade
    - Sub Grade
    - Loan Tenure/Term
    - DTI

Data analysis

- Bivariate analysis indicate chance of default using the following combination

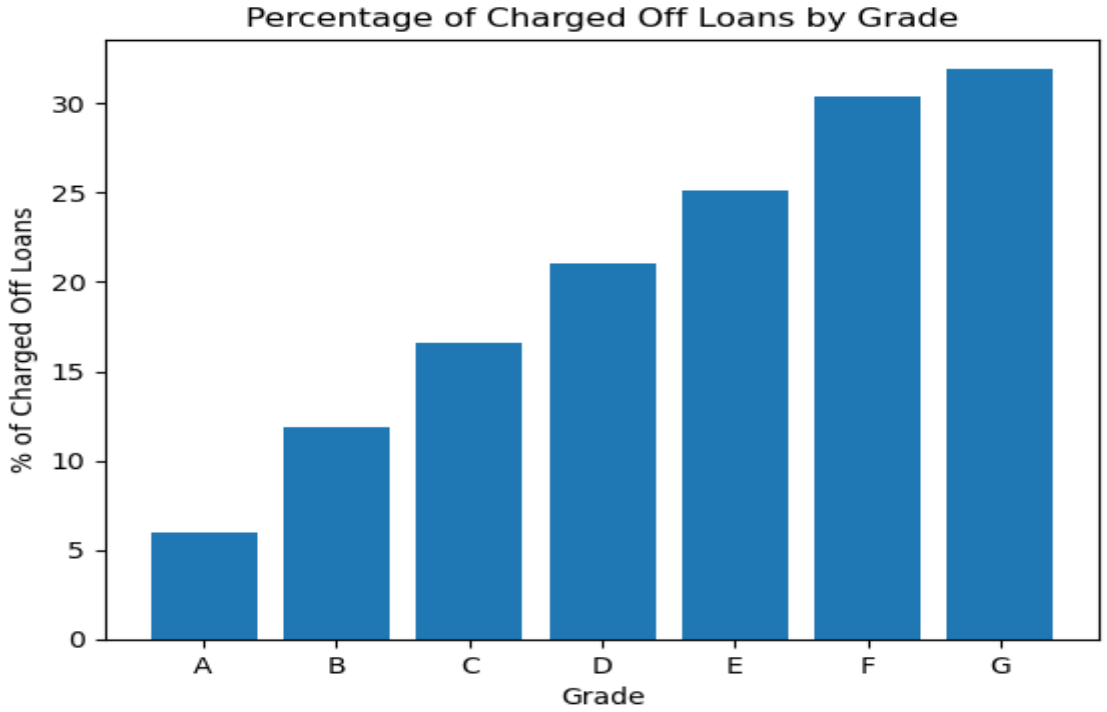
    Loan Amount and Interest Rate

    DTI and Interest Rate

    Grade and Interest Rate

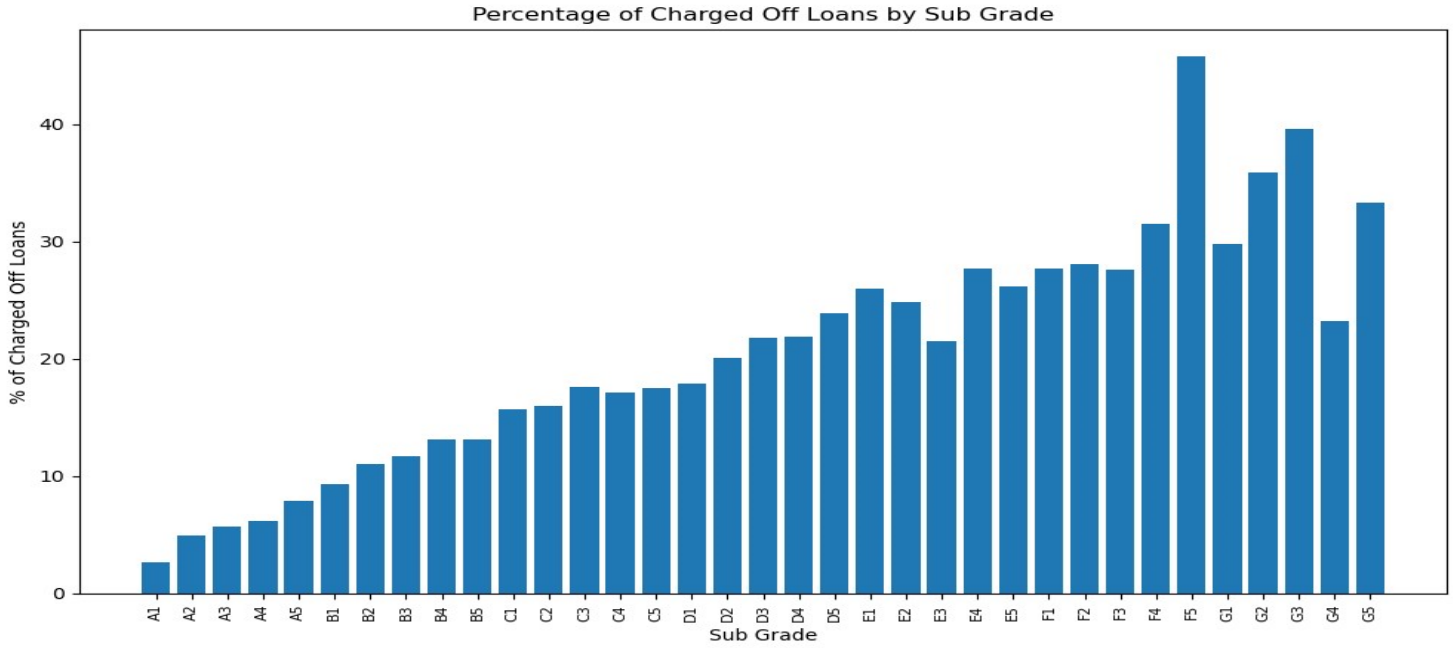
Univariate Analysis - Grade

- Grade vs % of loans in that grade that are charged off indicate as grade moves from A to G chances of default increase significantly.



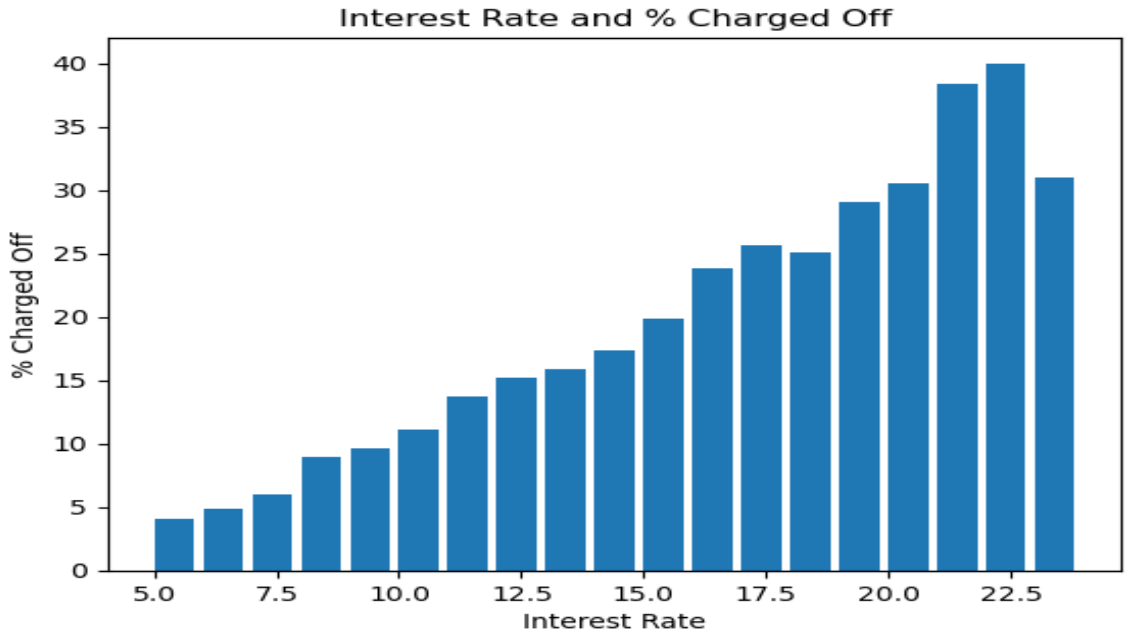
Univariate Analysis – Sub Grade

- Sub Grade vs % of loans in that sub grade that are charged off indicate as sub grade moves from A1 to A5 … chances of default increase significantly.



Univariate Analysis- Interest Rate

- Interest Rate vs % of loans in that Interest Rate range that are charged off indicate that as interest rate move higher chances of default increase significantly.

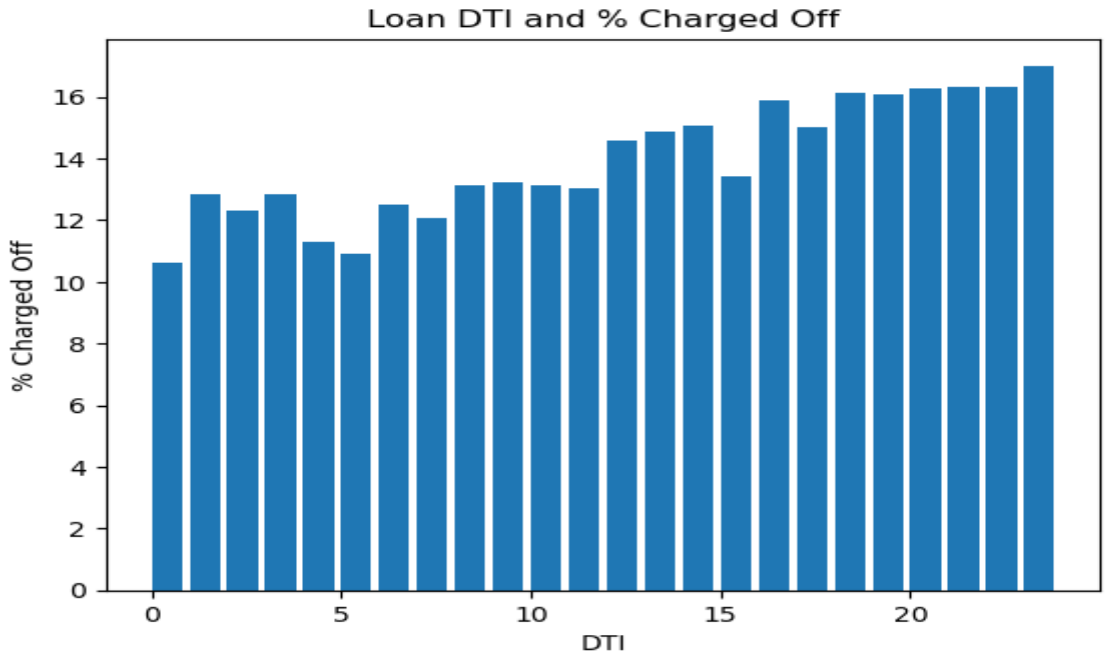


Recommendations

- Default is very highly dependent on Interest Rate.
- Grading and Sub grading of loans is effective in identifying default.
- Source verification is not effective
- Higher Interest Rate should be avoided for long tenure, low grade, low sub grade

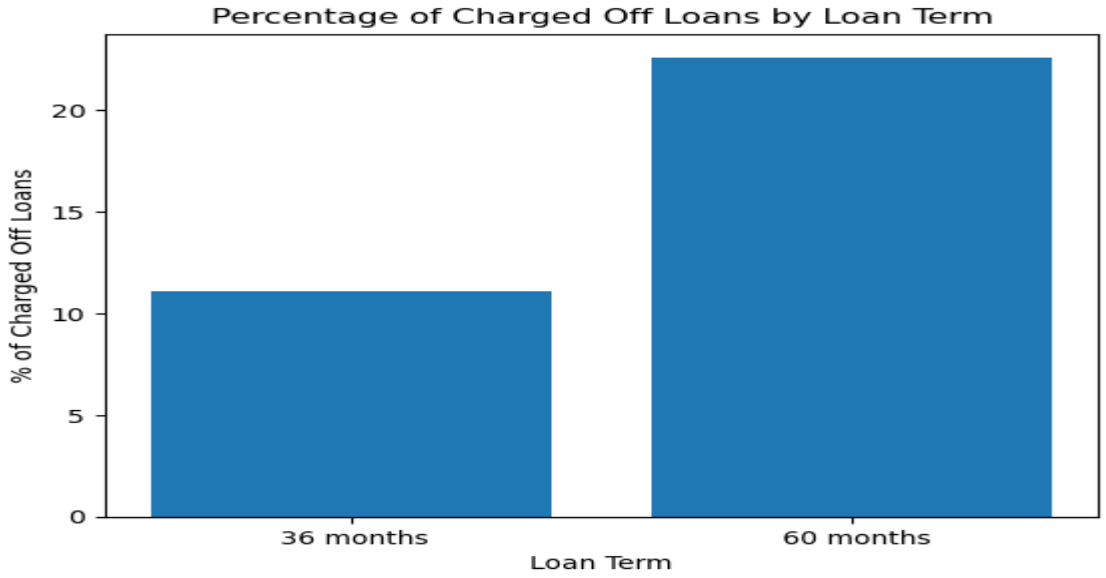
Univariate Analysis - DTI

- DTI vs % of loans in that DTI range that are charged off indicate that as DTI move higher chances of default increase gradually.



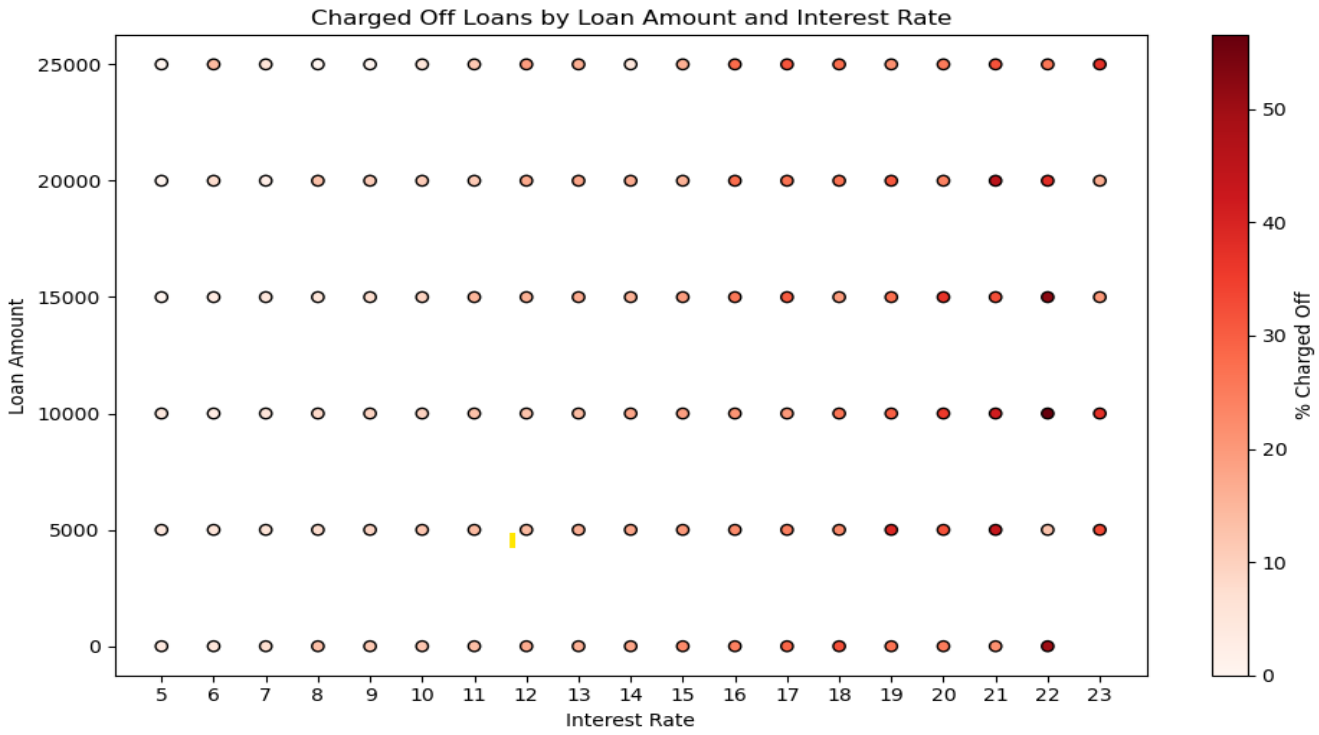
Univariate Analysis- Loan Tenure

- Loan Term/Tenure vs % of loans in that term that are charged off indicate that as loan term moves from 36 months to 60 months default % almost doubles.



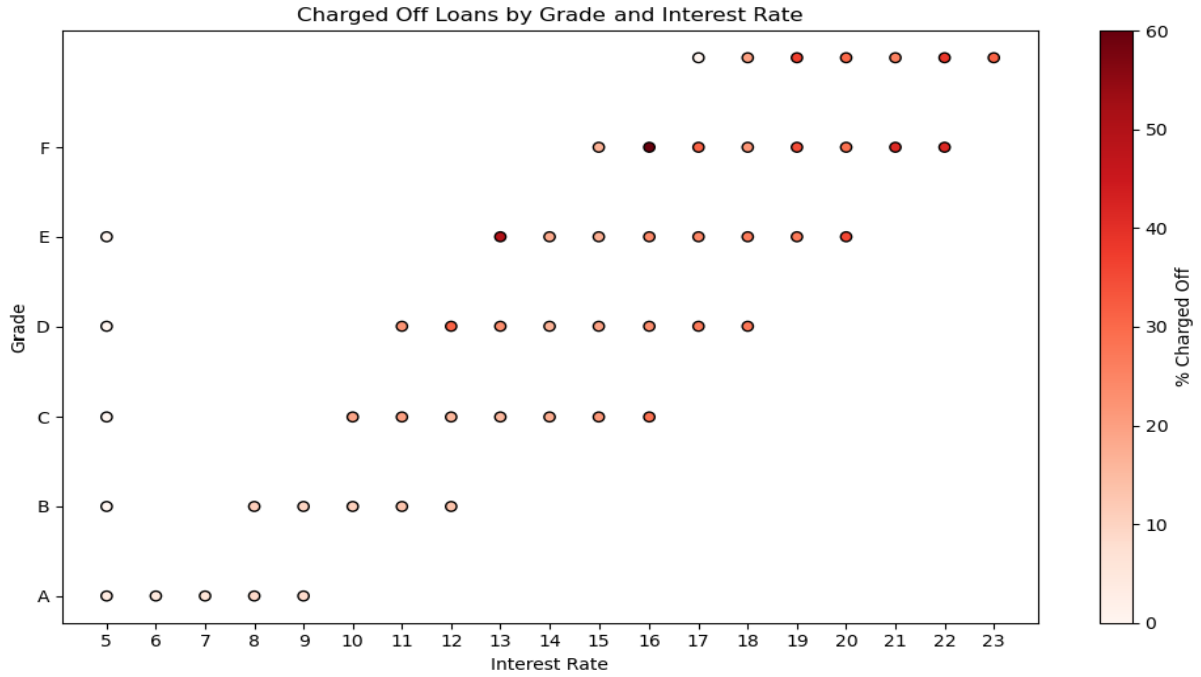
Bivariate (Loan Amount and Interest Rate)

- Loan default is linked more to interest rate than to loan amount.



Bivariate (Grade and Interest Rate)

- Loan default is linked both to grade and interest rate.



Bivariate (DTI and Interest Rate)

- Loan default is linked both to DTI and interest rate.