Lending Club Case Study

Case Study : Overview

- Problem Statement :Identify patterns which indicate if a person is likely to default, while taking loans
- Identify risky loan applicants using EDA, and such loans can be reduced thereby cutting down the amount of credit loss.

Dataset Overview

- The dataset contains the 4 years of complete loan data 2007-2011
- Sample data

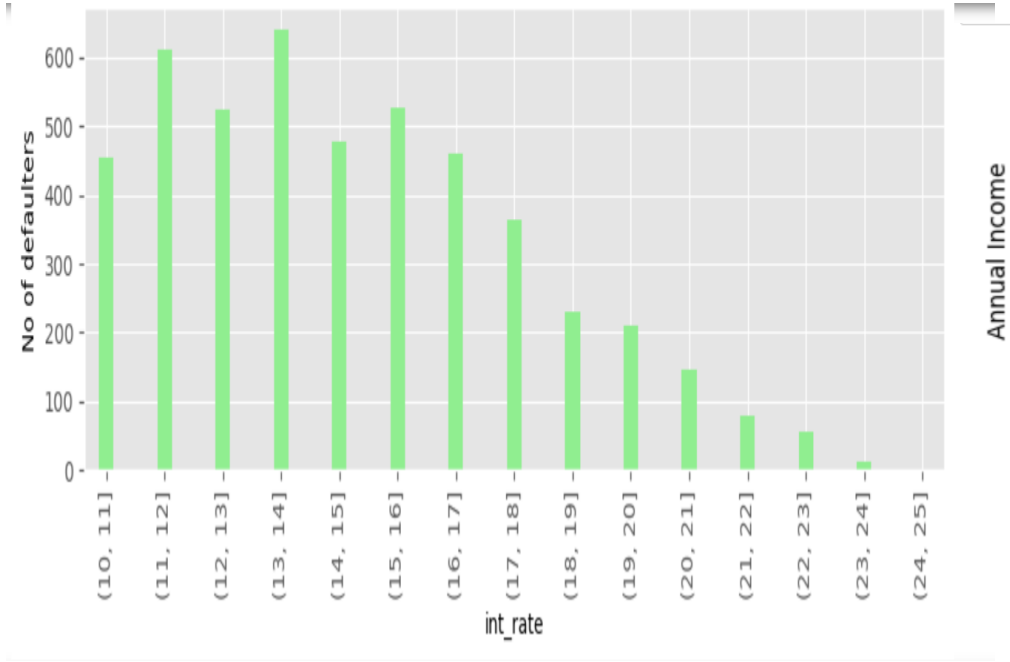
	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	emp_length	home_ownership	annual_inc	verification_s
24437	10000	10000	9796.198894	36 months	6.54%	306.68	A	A4	6 years	MORTGAGE	92040.0	Ve
7675	14000	14000	14000.000000	36 months	7.49%	435.43	A	A4	NaN	MORTGAGE	66000.0	Ve
24805	6000	6000	6000.000000	36 months	7.88%	187.69	A	A5	10+ years	MORTGAGE	35000.0	Not Ve
7783	5000	5000	5000.000000	36 months	7.49%	155.51	A	A4	4 years	RENT	75000.0	Not Ve
7798	7100	7100	7100.000000	36 months	6.99%	219.20	A	A3	2 years	RENT	50400.0	Source Ve

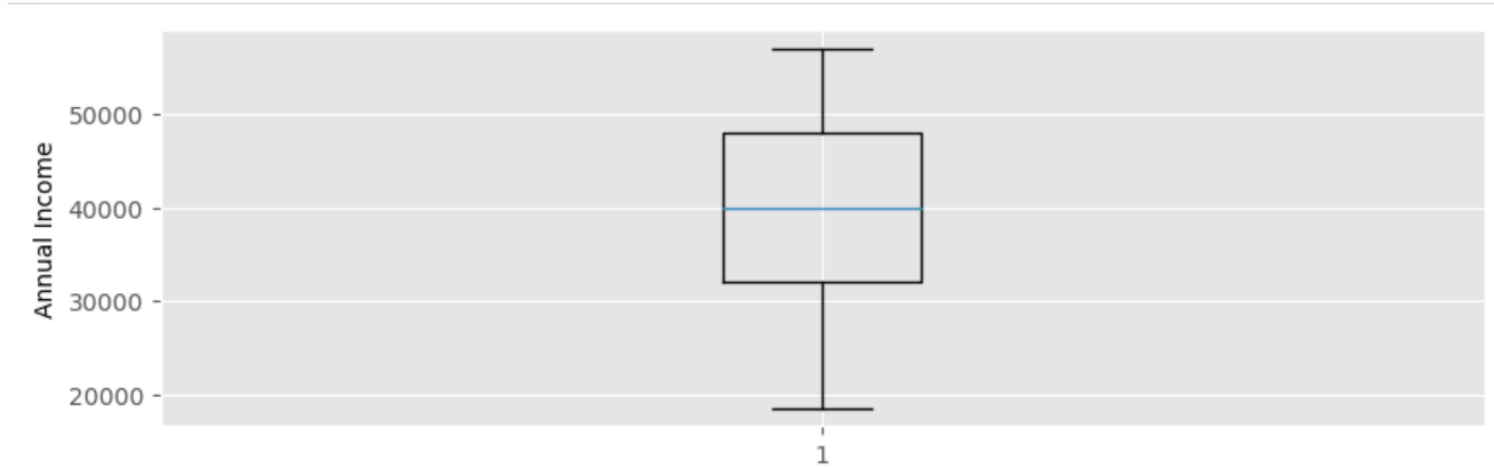
Analysis Approach

- Data Cleaning – Remove irrelevant columns
  - Remove the columns with all null values
  - Remove duplicates
  - Remove columns which are not relevant for analysis e.g. id, member id, etc
- Univariate Analysis
- Bivariate Analysis

Univariate Analysis

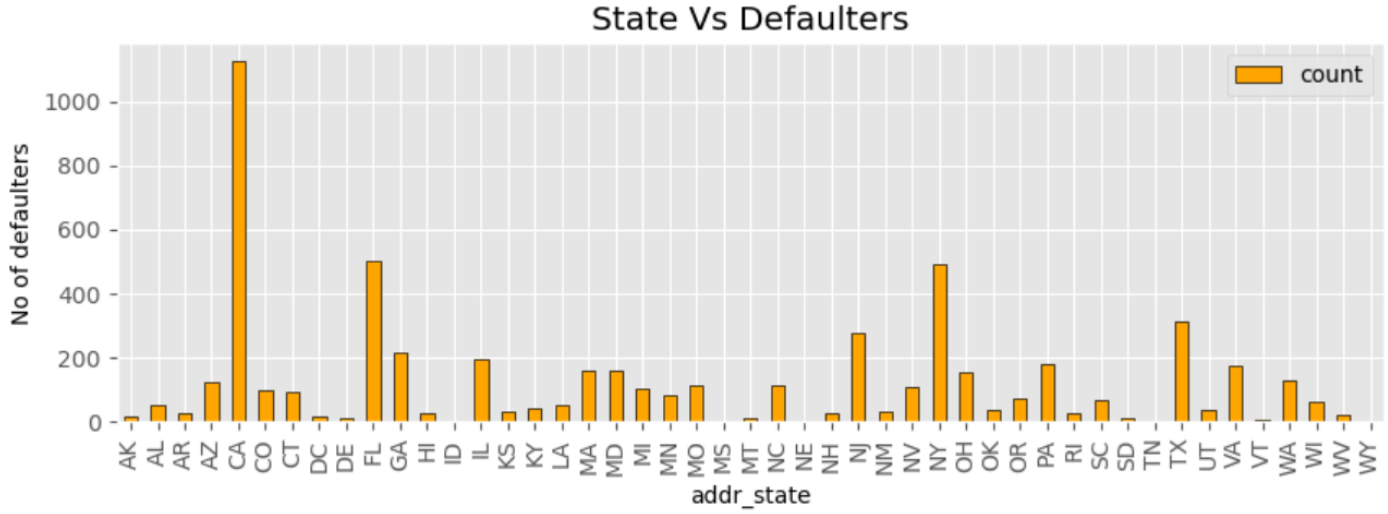
- Highest defaulters falls in rate of interest range 10 to 15%
- Most defaulters Annual income range from 10,000 to 60,000.

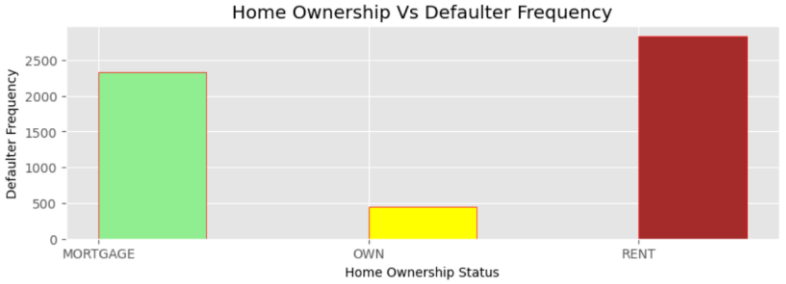




Univariate Analysis

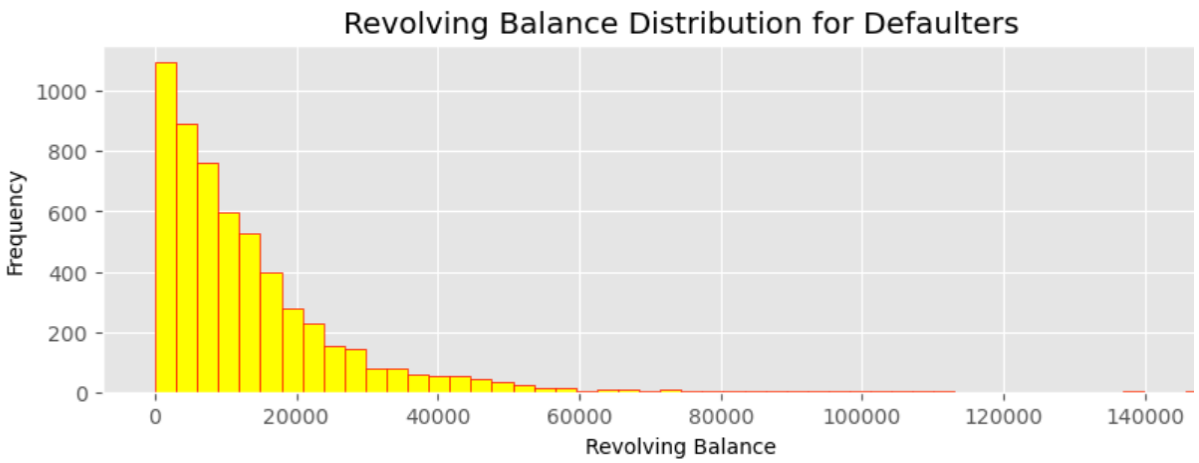
- Defaulters are Highest in state CA.
- Defaulters in states FL, NY rank next
- Those who stay in rented house seems to be having more defaulter tendency compared to those who owns a home

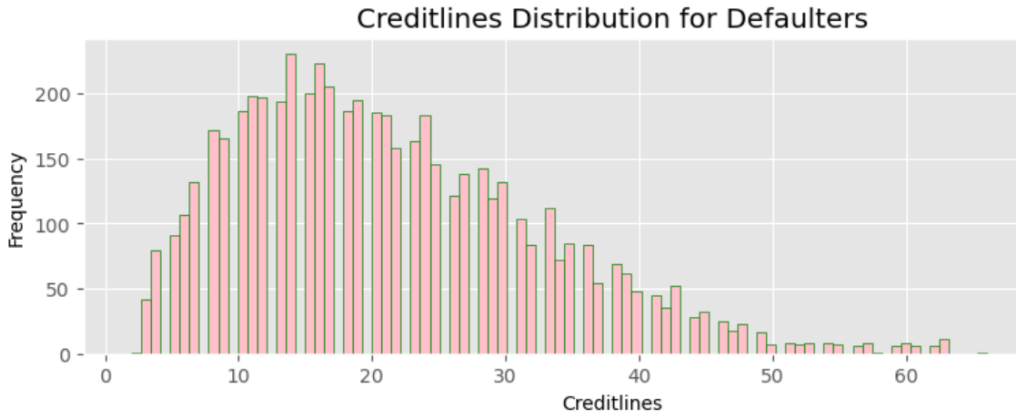




Distribution Curve Analysis

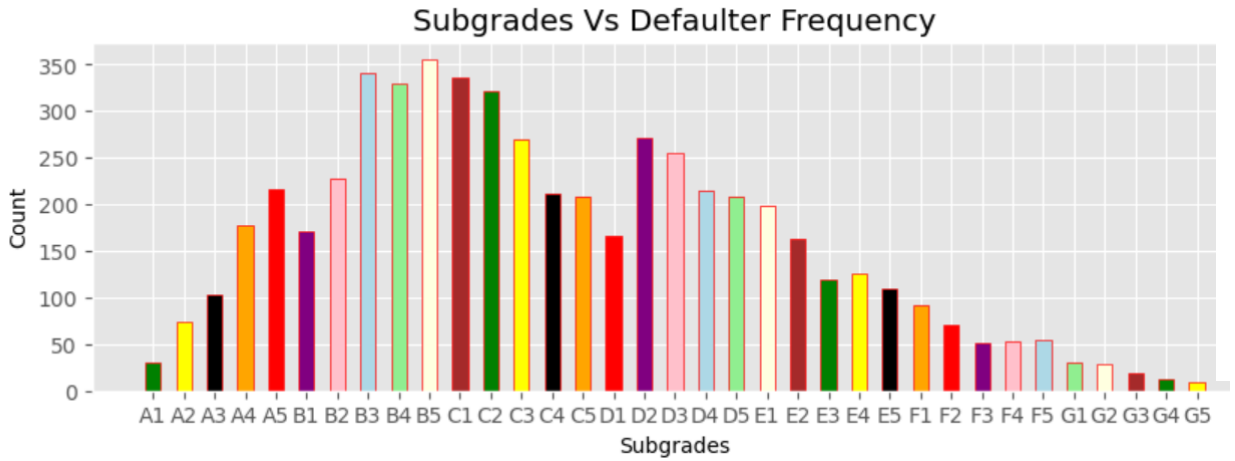
- Revolving balance of a consumer is the lowest, the higher the chance to be a defaulter
- Credit lines above 15 seems to be showing a lesser defaulter tendency

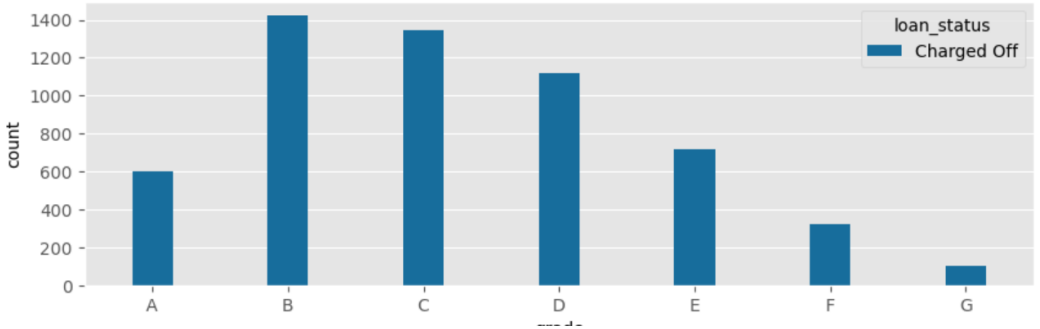




Grades and Subgrades

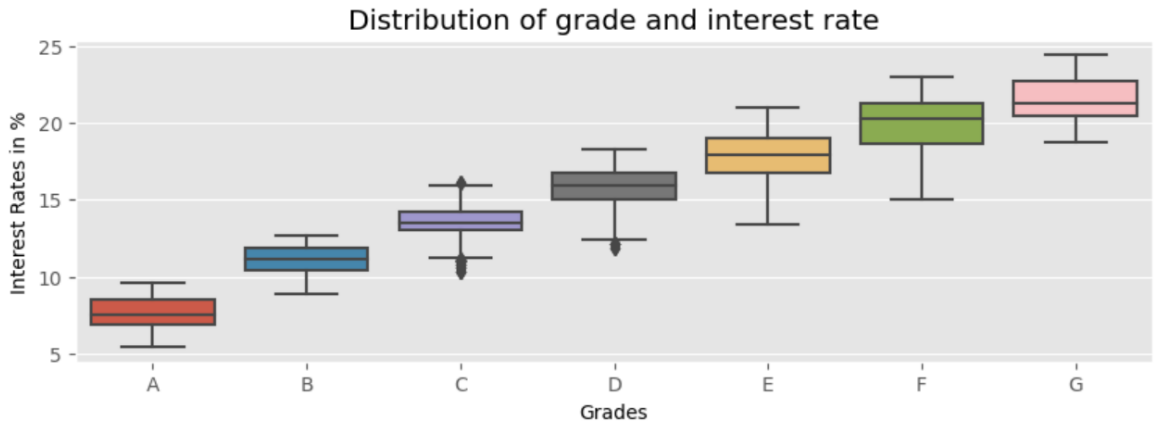
- The defaulters are highest for sub-grade B3 to C3, B5 being the highest defaulter frequency
- The defaulters are highest for grade B

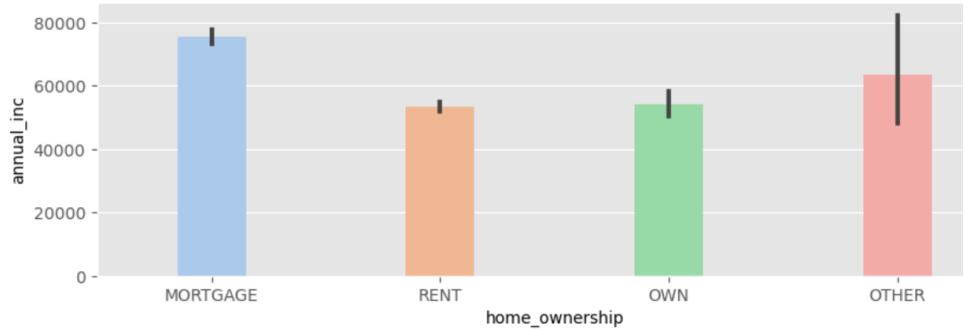




Bivariate Analysis

- Interest rates are increasing from Grades A to G
- Those who have high annual income are having mortgaged home owner
- Middle income range consumers are staying either in rented home or own home





Conclusions

- Following things can be concluded form above plots:
  - Most of loan applicants are either rented or mortgaged.
  - Maximum of loan application are for credit card payment and debt consolidation.
  - Home improvement and major purchases are other popular reasons of loan application.
  - Most applicants are from California state and then New York and Texas.
  - Majority of loan applicants are vastly experienced professional, i.e. more than 10 years.
  - Most people have multiple credit lines linked with them.