



Lending CLub Case Study

Lending Club is the largest online marketplace for availing personal, business and medical loans

- Sudarshan.S



Problem Statement

- Lending Club wants to identify the possible defaulters among the list of loan applicants
- Our job is to find the metrics that we can use to find risky loan applicants



Steps Undertaken To Pursue the Goal

1.Importing Libs and Loading Data Into DataFrame

2.Data Cleaning

3. Data Analysis

3.1 Univariate Analysis

3.2 Bivariate Analysis

3.3 Multivariate Analysis

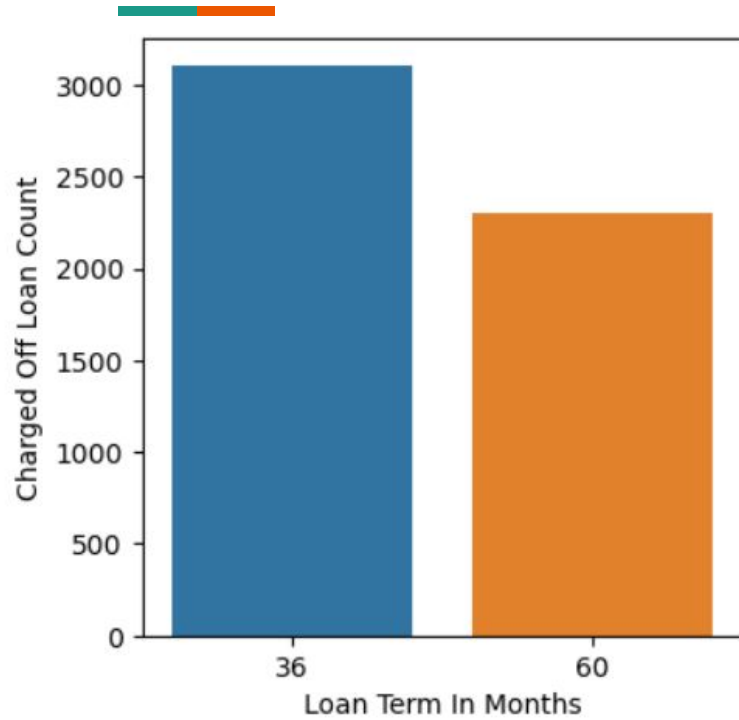
4 Consolidated Observations From Start

5 Important Observations In Business Terms

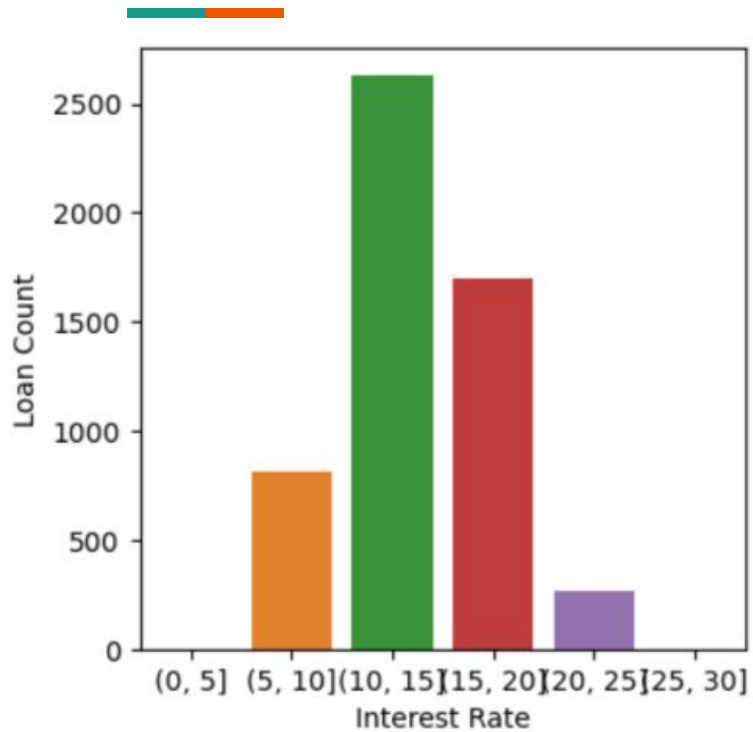


Observations Based On The Steps Taken

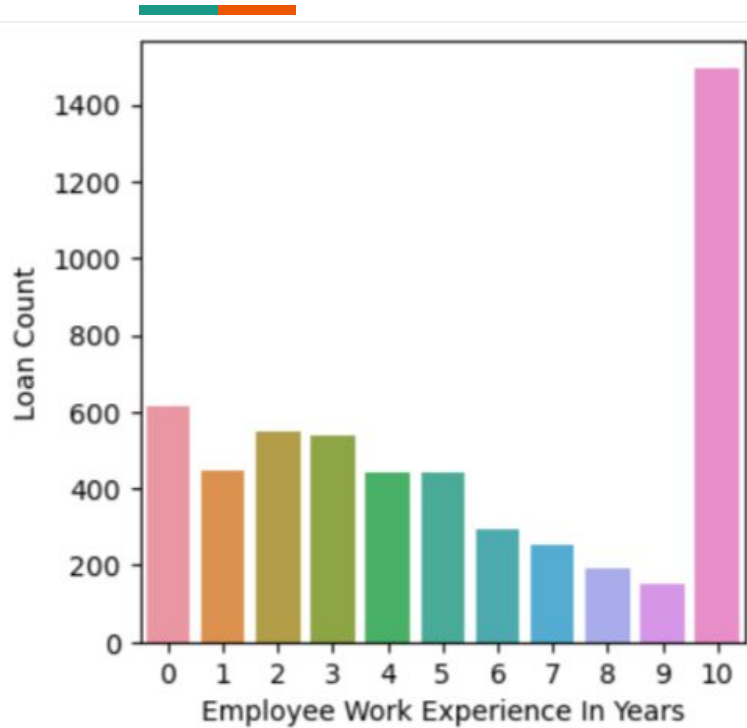
- We have observed 16 different metrics which we will use to come to conclusion on how some these can be used to detect risky loan applicants.
- Let's discuss them one by one using the graphs that helped us to come to this conclusions.



1 - The Lesser The Term The Greater than chance for default



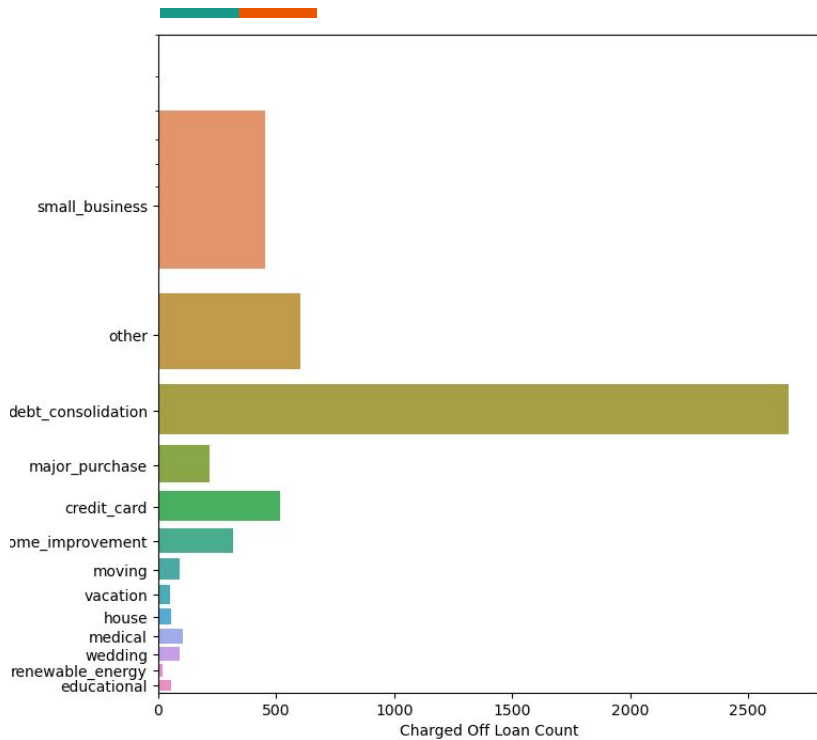
2 - Interest rate of 10% and above is when the default rates increase



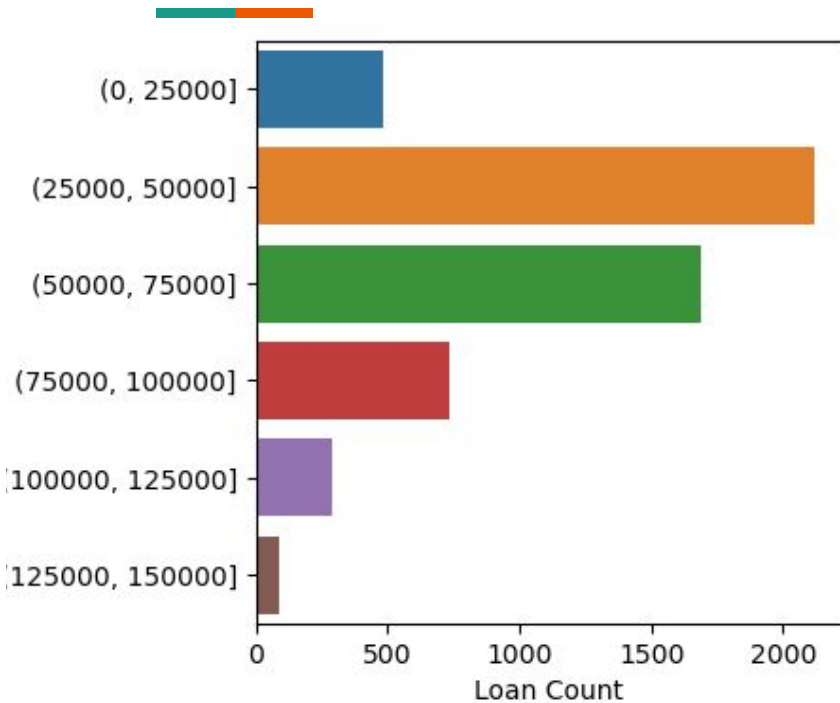
3 - Employees whose experience is 10 years are the ones with greater chance of default.



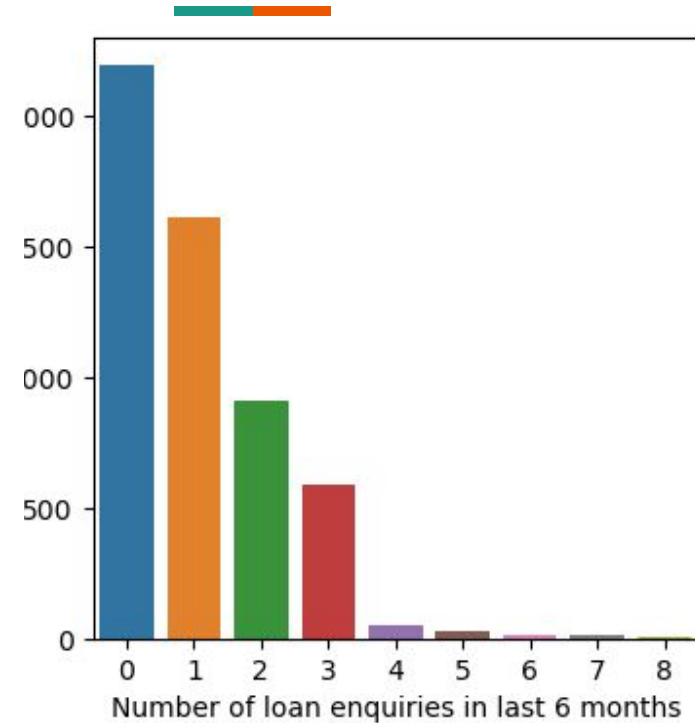
4 - People who are staying in a rental home has the highest chance to default



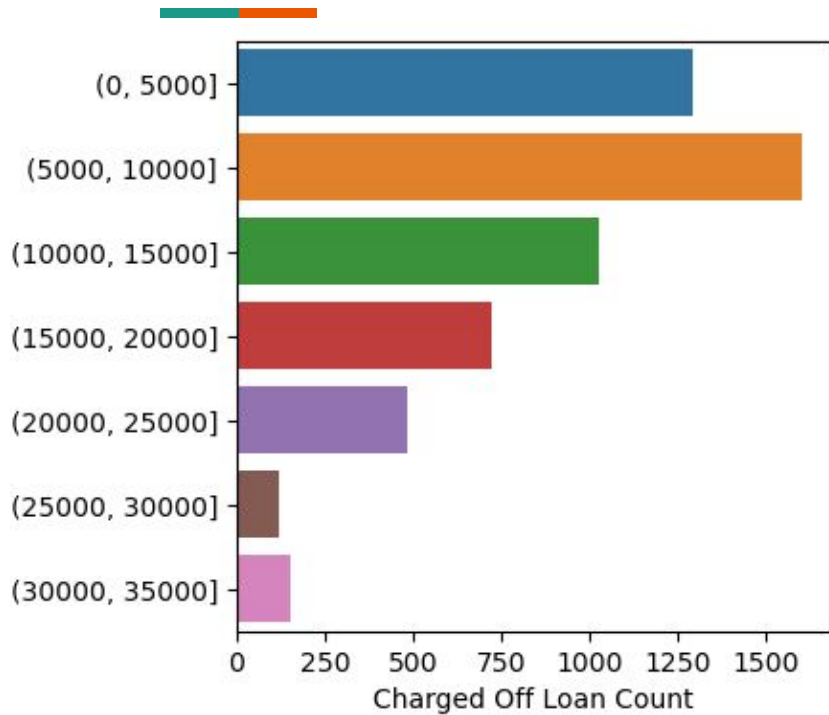
5 - People who are trying to debt consolidation or restructure the already existing loan are the number one defaulters



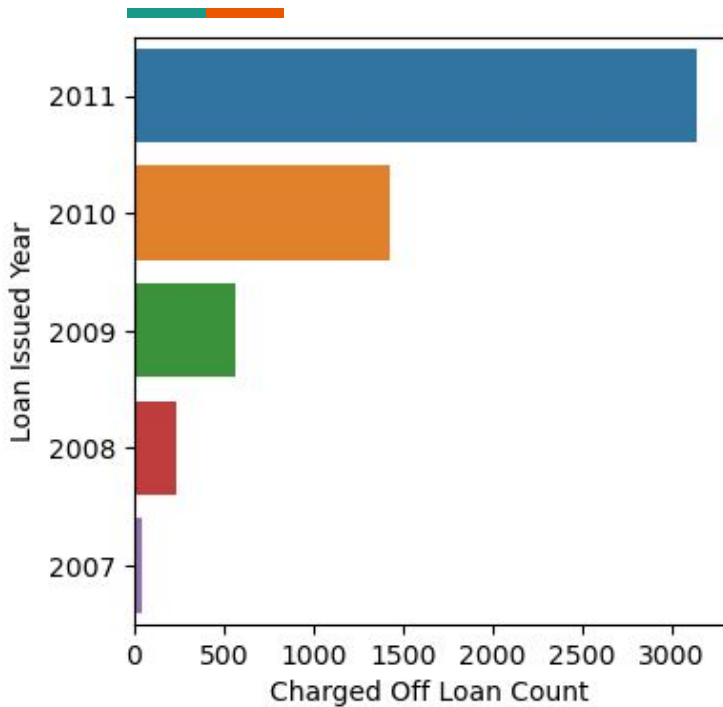
6 - People who are earning between 25k-50k are the top defaulters followed by 50k-75k



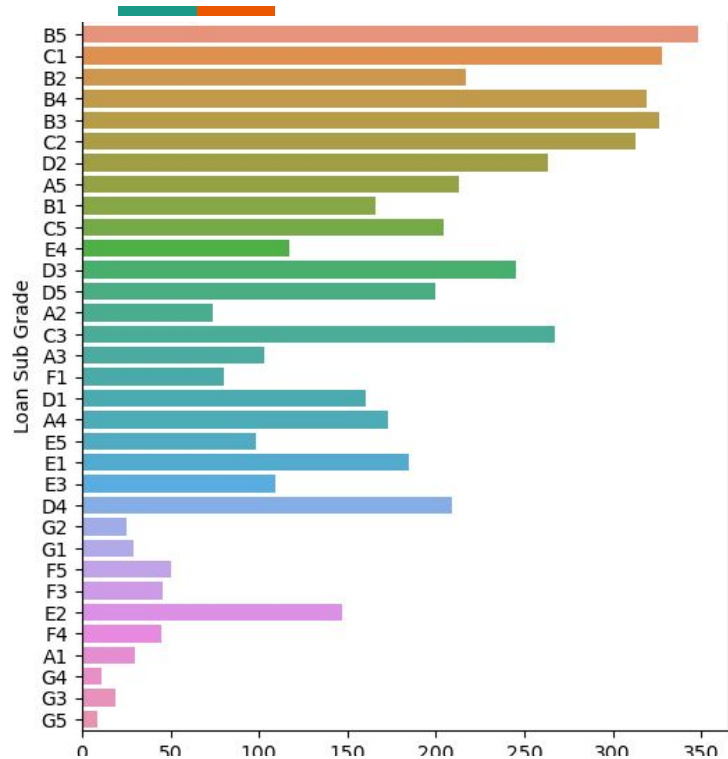
7 - The people with less number of enquiries are the ones with highest percentage chances for default. This is one of the most important metric . This could happen if Club is giving away risky loans to meet some target and the club employees who are tasked with verifying the applicants data are simply tagging them as verified on the first few requests



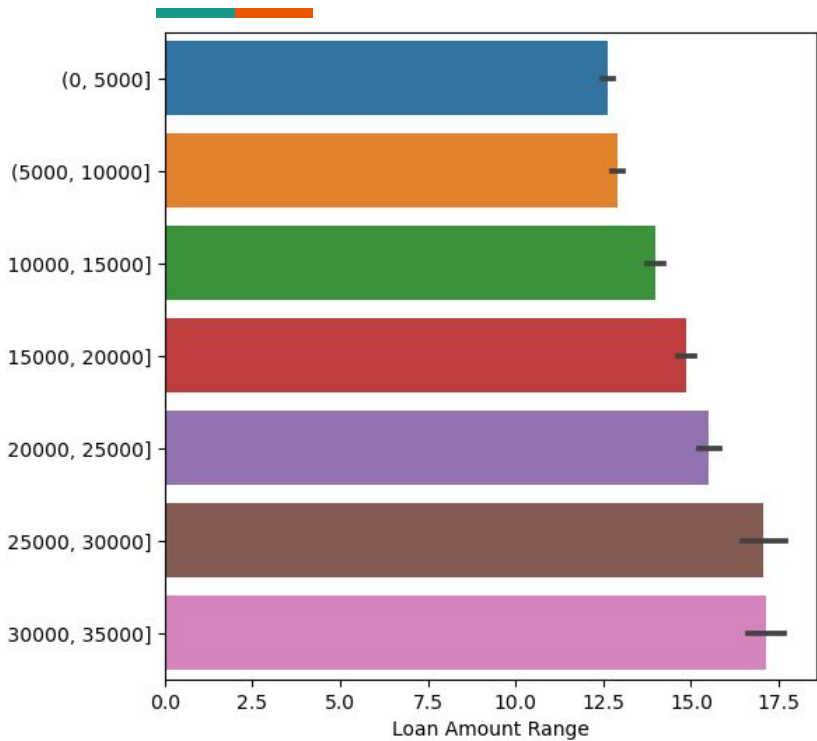
8 - The Loan range of 5k-10k as the most defaulted ones



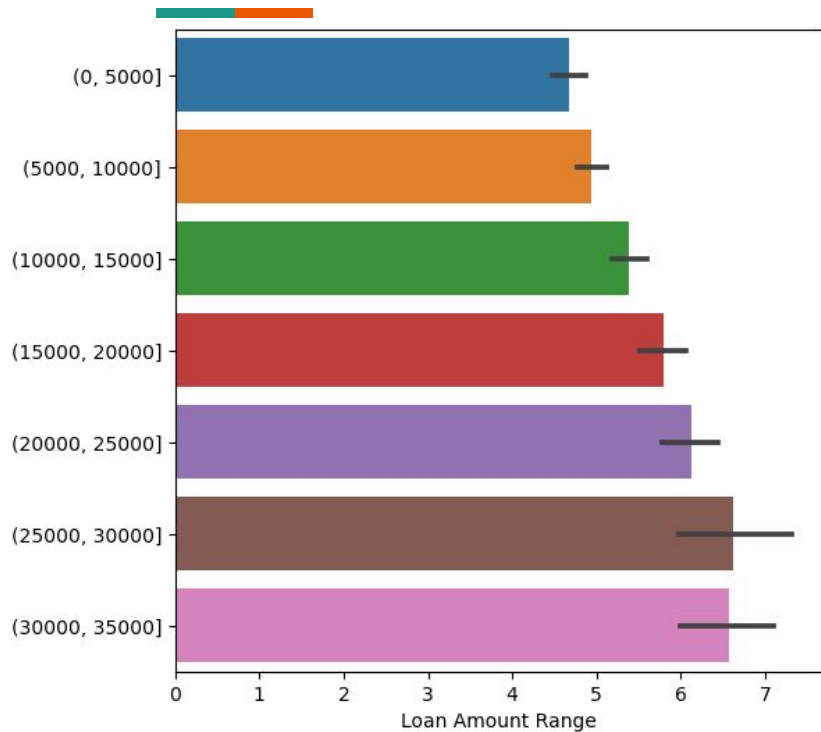
9 - Dec month and 2011 year where the maximum number of defaults have been made. We need to correlate this info with job layoff and economic downturn data to get more insights



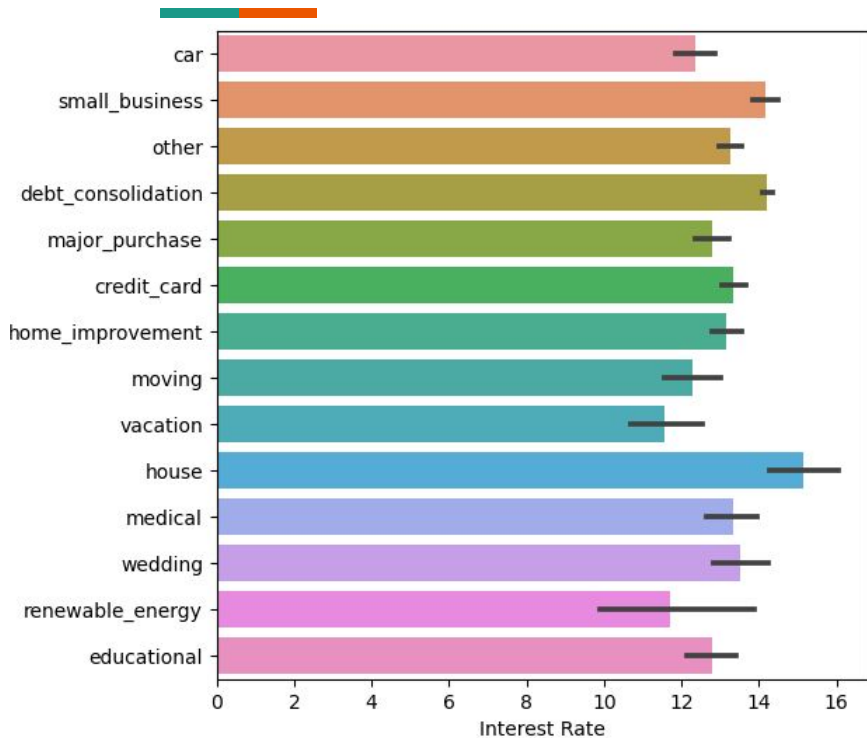
10 - B grade and B5 Subgrade loans have the highest number of defaults



11 - Higher Loan Amount rate means Higher Interest Rate



12 - Higher Loan Amount was given to employees with higher experience

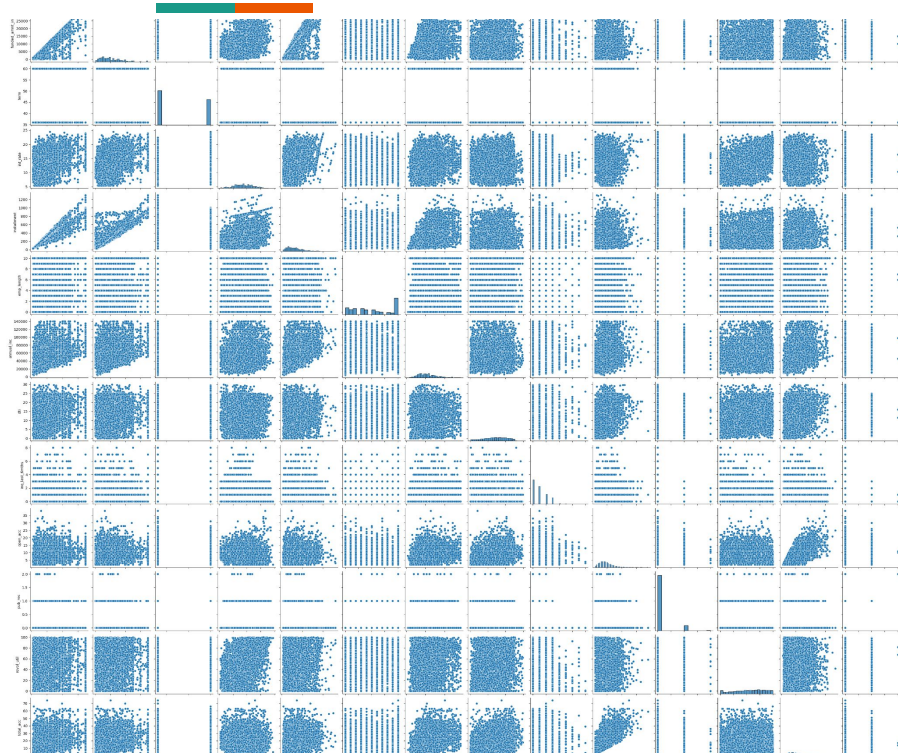


13 - For Defaulted Loans Interest rates are higher for all purposes



14 - Number of negative public records is correlated to number of public bankruptcies

15 - Funded Amount Invested, Loan Amount and Installment are highly correlated



16 - Interest rate is lower for short term loan and higher for long term loans



Important Observations In Business Terms

- 1 - Interest rates are higher for long term loans which is bad and is a primary contributor to default
- 2 - On Dec 2011 so many loan defaults are there . Whenever Economy has a downturn and layoff rates are high the club should be very careful about giving loans to highly experienced persons
- 3 - Whenever someone asks for a loan to restructure or transfer his/her existing loan the club should be extremely cautious as this is the no.1 reason for loan defaults
- 4 - The club should not pursue people to give loans. Rather it should wait for customers to reach the club for loan and if the loan is sanctioned on the first few requests , it is suggested that it is sent for further quality checks as loans given to people who dont ask and ask only a few times are the main defaulters
- 5 Club should be very careful in giving loan to people staying in rented out houses