

GRAMENER CASE STUDY

SUBMISSION

Group Members:

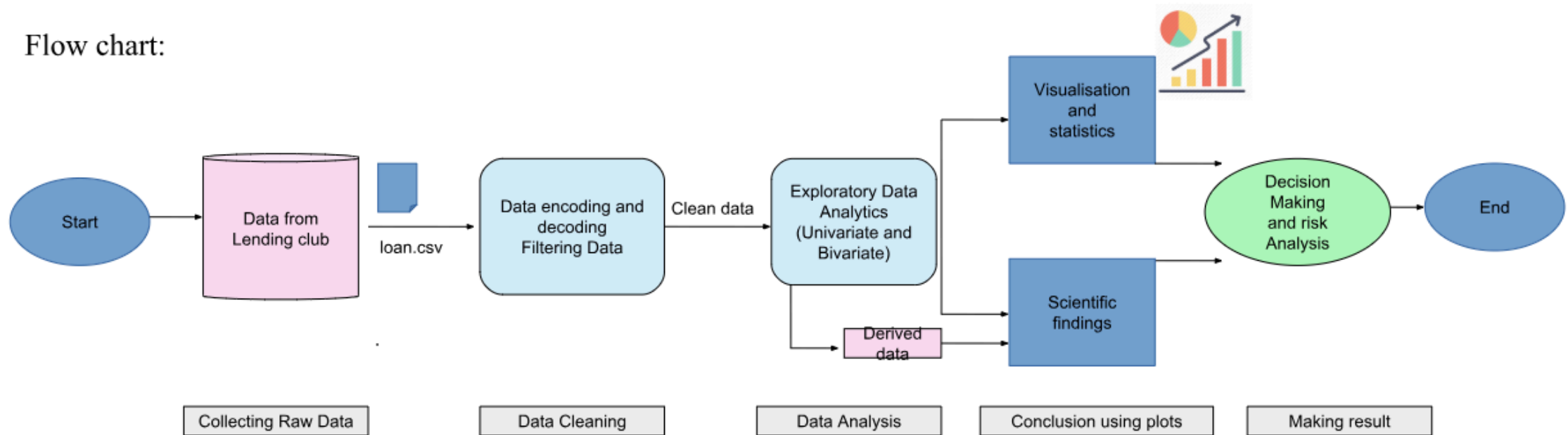
1. Sudhanshu Singh
2. Shweta Singh
3. Megha Banchhor
4. Joelbright Jacobmuthukumar

ABSTRACT

This case study is about a consumer finance company which is one of the largest online loan marketplace, facilitating business loans, personal loans and financing of medical procedures. Lending loan to borrowers leads company to make a decision. Company finds lending loans to 'risky' applicants as biggest source of financial loss (credit loss). In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'. The aim of this case study is to identify this risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Using EDA i.e. Exploratory Data Analytics we are applying the techniques of EDA and risk analysis to figure out the driving factors or driving variables behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment and hence can reduce the credit loss and improve performance.

PROBLEM SOLVING METHODOLOGY

Flow chart:

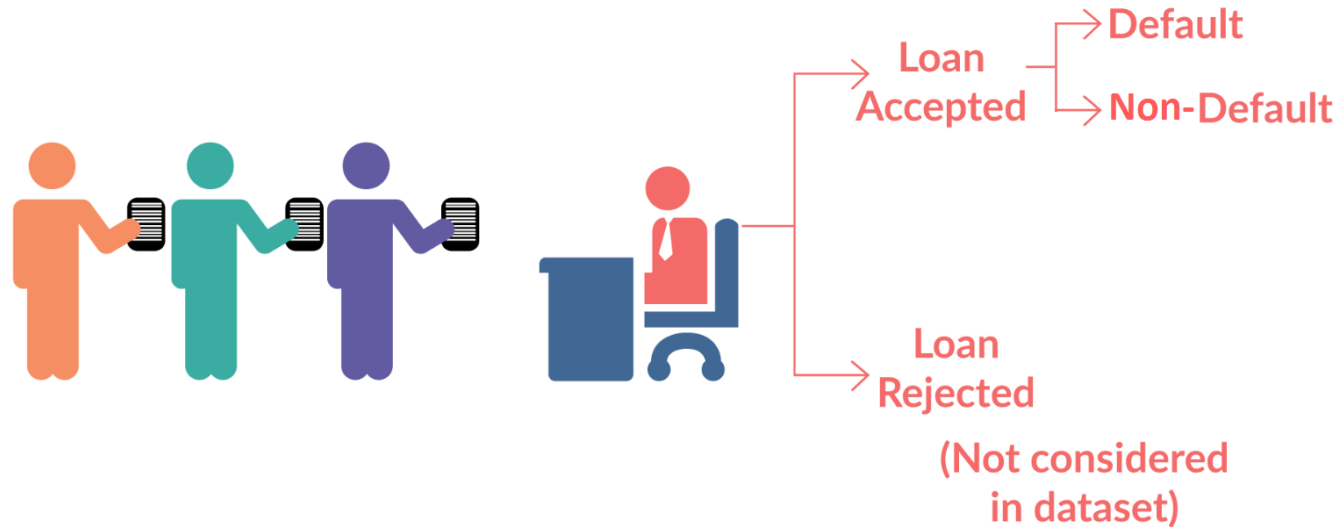


Flow Chart representing the steps involved in the case study for identifying risky borrowers for lending loan using Exploratory data Analytics.

OBJECTIVE

To identify the risky loan applicants or driving variables/factors behind loan default.

LOAN DATASET

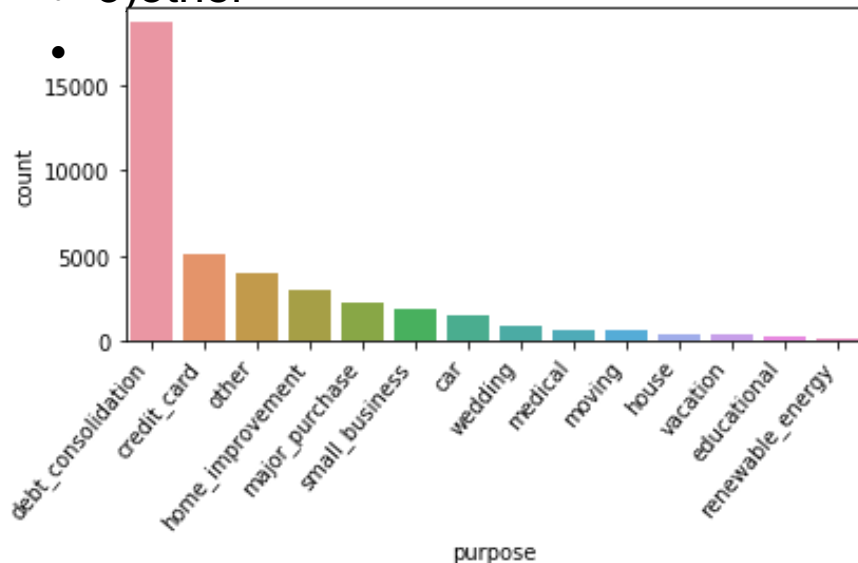


UNIVARIATE ANALYSIS

Total data size : 39717

On Categorical Data:

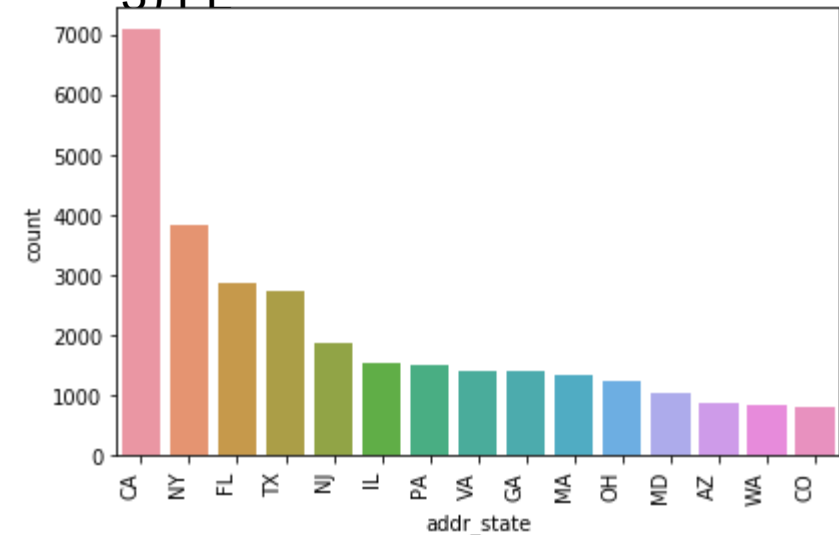
- Highest loan taken purpose wise
- 1) debt consolidation
- 2) credit card
- 3) other
-



This shows that "debt_consolidation" is the most popular reason for taking loan

Highest loan taken state wise

- 1) CA
- 2) NY
- 3) FL

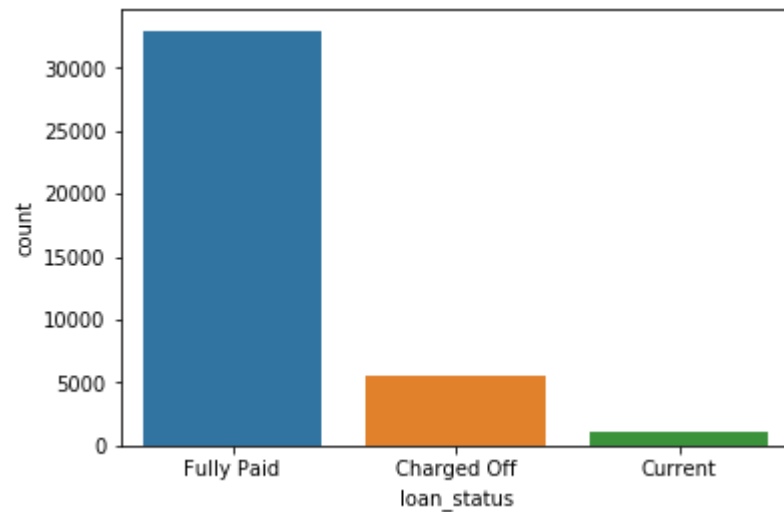


This shows "CA" has the highest number of loan takers

UNIVARIATE ANALYSIS

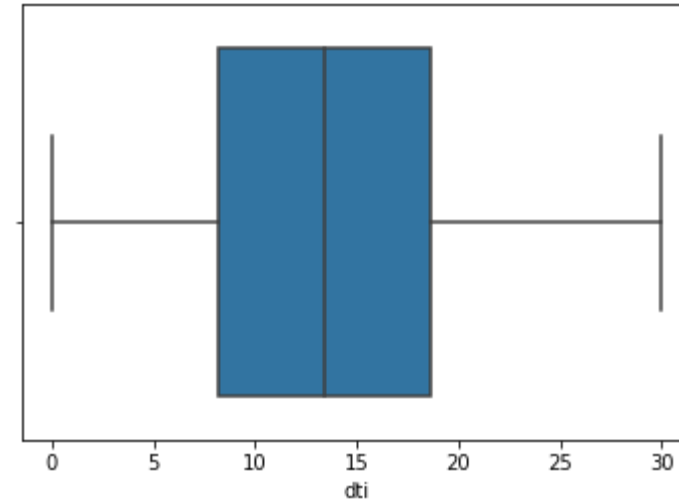
On Numerical Data:

- Count loan status wise
- in which Charged Off = 5000



Loan amount dti wise

Loan amount range: 8 to 18



On the basis of all univariate analysis, we were able to see the trends in data of loan taken for purpose wise, state wise or dti wise. But to conclude we need to do some more analysis.

SEGMENTED UNIVARIATE ANALYSIS

Mean loan Amount = loan_status

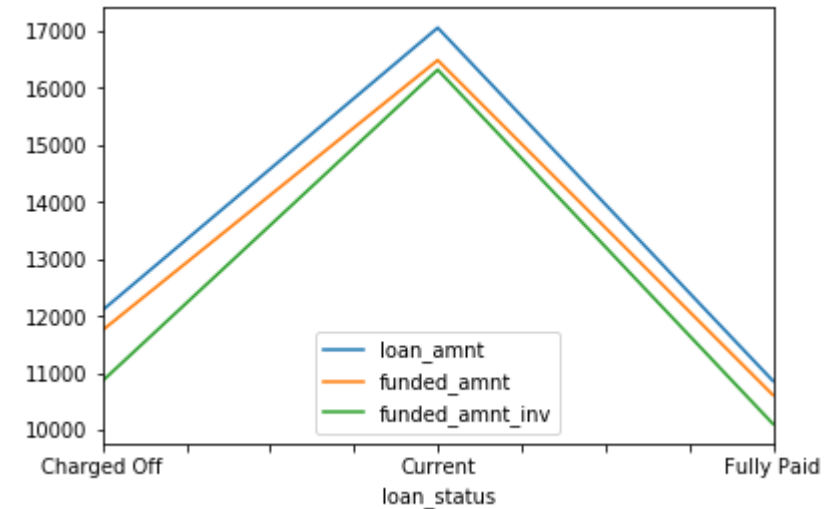
Charged Off	12104.385108
Current	17053.991228
Fully Paid	10866.455994

Mean loan Amount committed= loan_status

Charged Off	11753.398792
Current	16485.701754
Fully Paid	10618.520486

Mean loan Amount committed by investor= loan_status

Charged Off	10864.521324
Current	16318.265247
Fully Paid	10112.837414

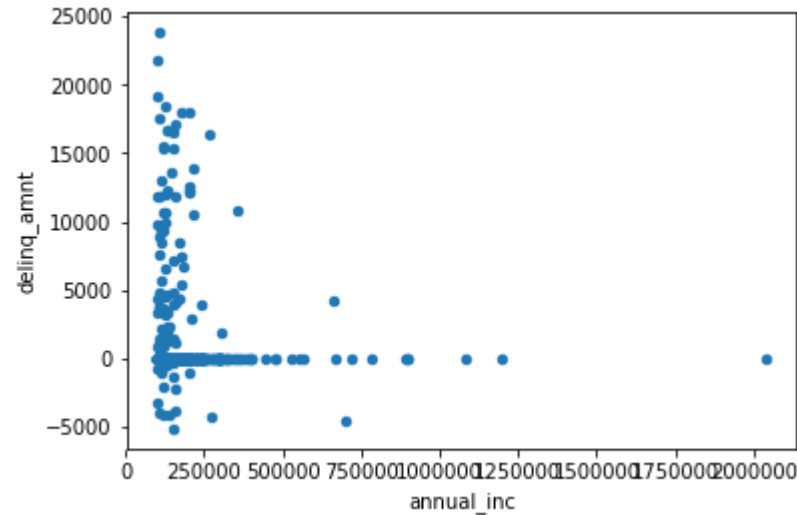


We can see from the data that mean loan amount is always less than mean loan amount committed for that loan or mean loan amount committed by the investor for that loan.

BIVARIATE ANALYSIS

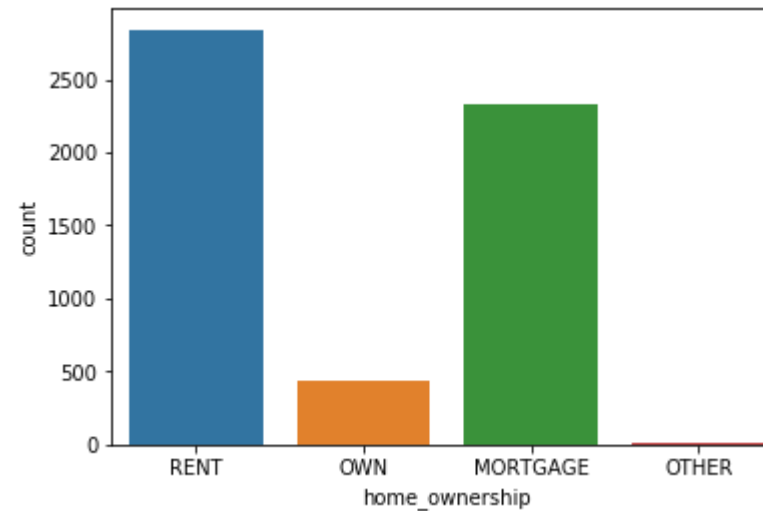
On Continous Data:

To find correlation between the income and delinquency



From the scatter graph,we can see that Loan amount upto \$100K are more likely to default.

Plotting "home_ownership" with count

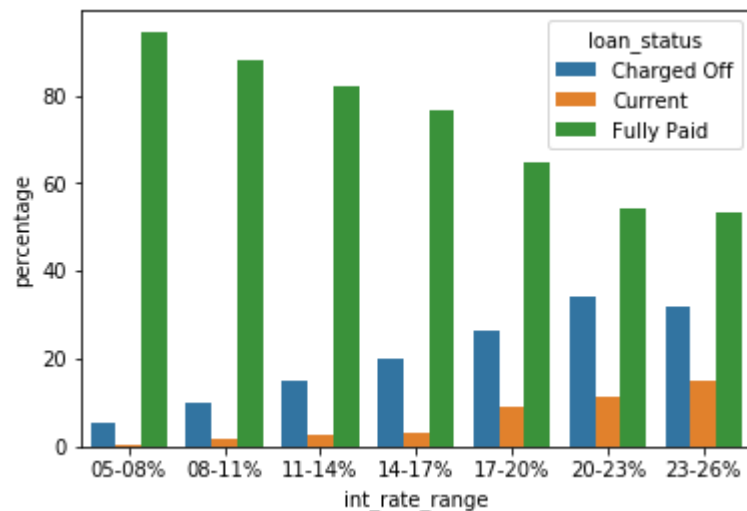


People with Own house are less likely to default in a loan

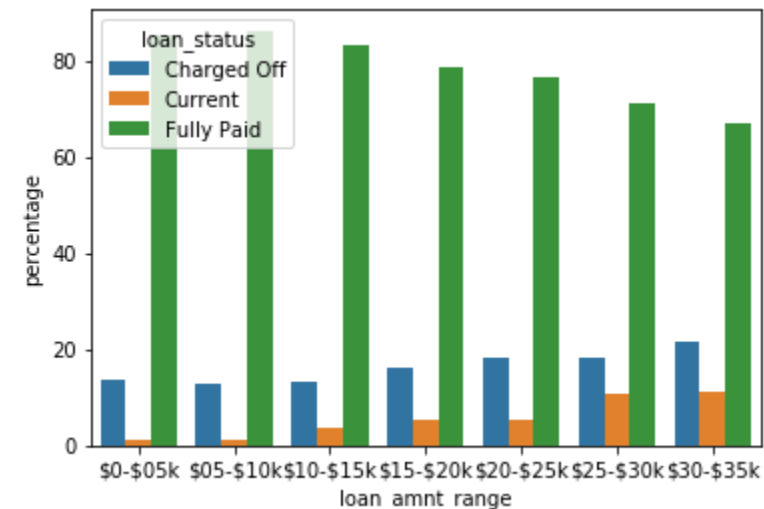
BIVARIATE ANALYSIS

On Categorical Data:

corelation between "loan_status" and various factors



This shows as "int_rate_range" increase, loan default is increased and fully paid is decreased

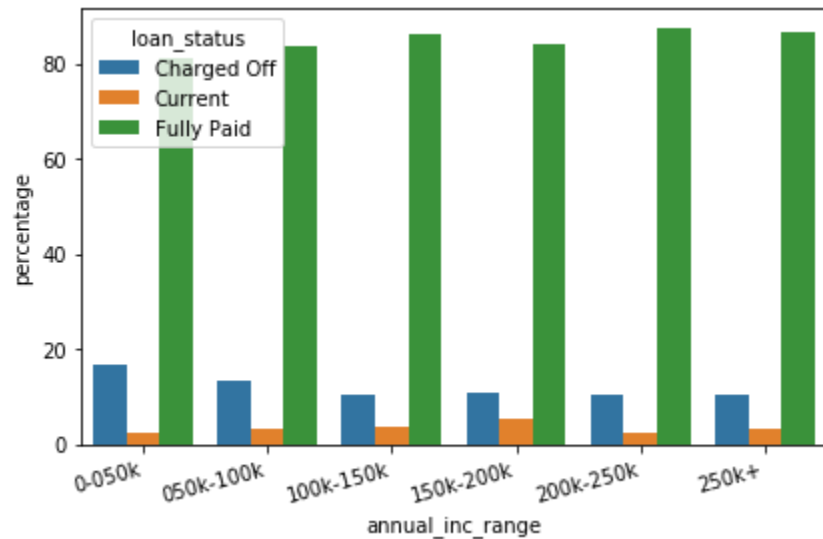


This shows as "loan_amnt_range" increase, loan default is increased and fully paid is decreased

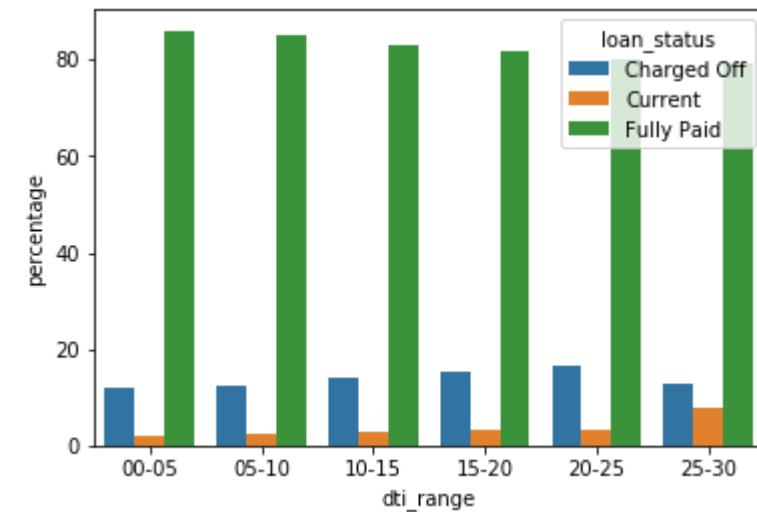
BIVARIATE ANALYSIS

On Categorical Data:

corelation between "loan_status" and various factors

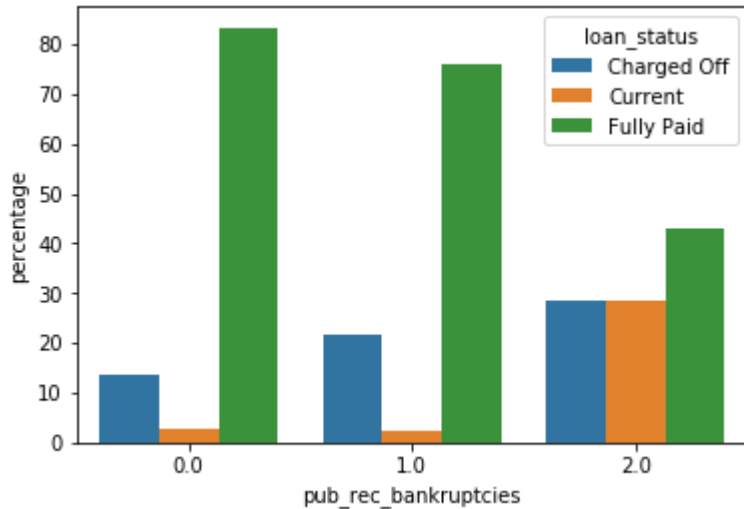


This shows as "annual_inc_range" increase, loan default is increased and fully paid is decreased



This shows as "dti_range" increase, loan default is increased and fully paid is decreased

BIVARIATE ANALYSIS



This shows as "pub_rec_bankruptcies" increase, loan default is increased and fully paid is decreased

	installment	annual_inc	dti	delinq_2yrs	inq_last_6mths	open_acc	int_rate	revol_bal	total_acc	pub_rec_bankruptcies
installment	1	0.270874	0.0541857	-0.0196565	0.00972221	0.172812	0.282703	0.312679	0.230824	-0.0337458
annual_inc	0.270874	1	-0.122732	0.0230833	0.0339079	0.1582	0.0531852	0.279961	0.235771	-0.0165136
dti	0.0541857	-0.122732	1	-0.0344516	0.00140517	0.288045	0.111162	0.228743	0.229881	0.00702108
delinq_2yrs	-0.0196565	0.0230833	-0.0344516	1	0.00809054	0.0116561	0.157916	-0.0551248	0.0678923	0.00295357
inq_last_6mths	0.00972221	0.0339079	0.00140517	0.00809054	1	0.0917134	0.133013	-0.0223814	0.111499	0.0156523
open_acc	0.172812	0.1582	0.288045	0.0116561	0.0917134	1	0.0103949	0.288964	0.686635	0.00576265
int_rate	0.282703	0.0531852	0.111162	0.157916	0.133013	0.0103949	1	0.0971	-0.0435702	0.0835786
revol_bal	0.312679	0.279961	0.228743	-0.0551248	-0.0223814	0.288964	0.0971	1	0.313602	-0.048876
total_acc	0.230824	0.235771	0.229881	0.0678923	0.111499	0.686635	-0.0435702	0.313602	1	-0.0107575
pub_rec_bankruptcies	-0.0337458	-0.0165136	0.00702108	0.00295357	0.0156523	0.00576265	0.0835786	-0.048876	-0.0107575	1

A heat map is plotted to find the correlation between various parameters

RESULTS

By all the analysis we observed that, Driving variables/factors behind loan default are as follows:

- Annual Income
- Interest rate
- Public Bankruptcy record
- Debt-income Ratio
- Loan Amount

CONCLUSION

The aim of this case study was to identify patterns and conditions which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

By using Exploratory Data Analytics we studied the data patterns and found the correlation between them with the help of univariate and bivariate analysis.

We also had to add some derived metrics to infer proper results. The analysis helped us to find the driving factors that lead the loan applicants to default.