

Exploratory Data Analysis



Lending Club Case Study





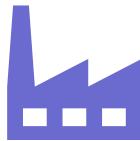
Contents

- Business Objective
- Problem Solving Approach
- Top View of Profit & Loss
- Conclusion Summary & Recommendation
- Data Analysis Slides
 - Univariate Analysis
 - Segmented Univariate Analysis
 - Bivariate Analysis



Business Objective

Company



The largest online loan marketplace

Business Challenge



Lending loans to 'risky' applicants is the largest source of financial loss (called credit loss)

Not approving loans to the clients who are likely to repay, results in loss of business

Historical Data



The complete loan data of the time period 2007 to 2011

Client's Ask

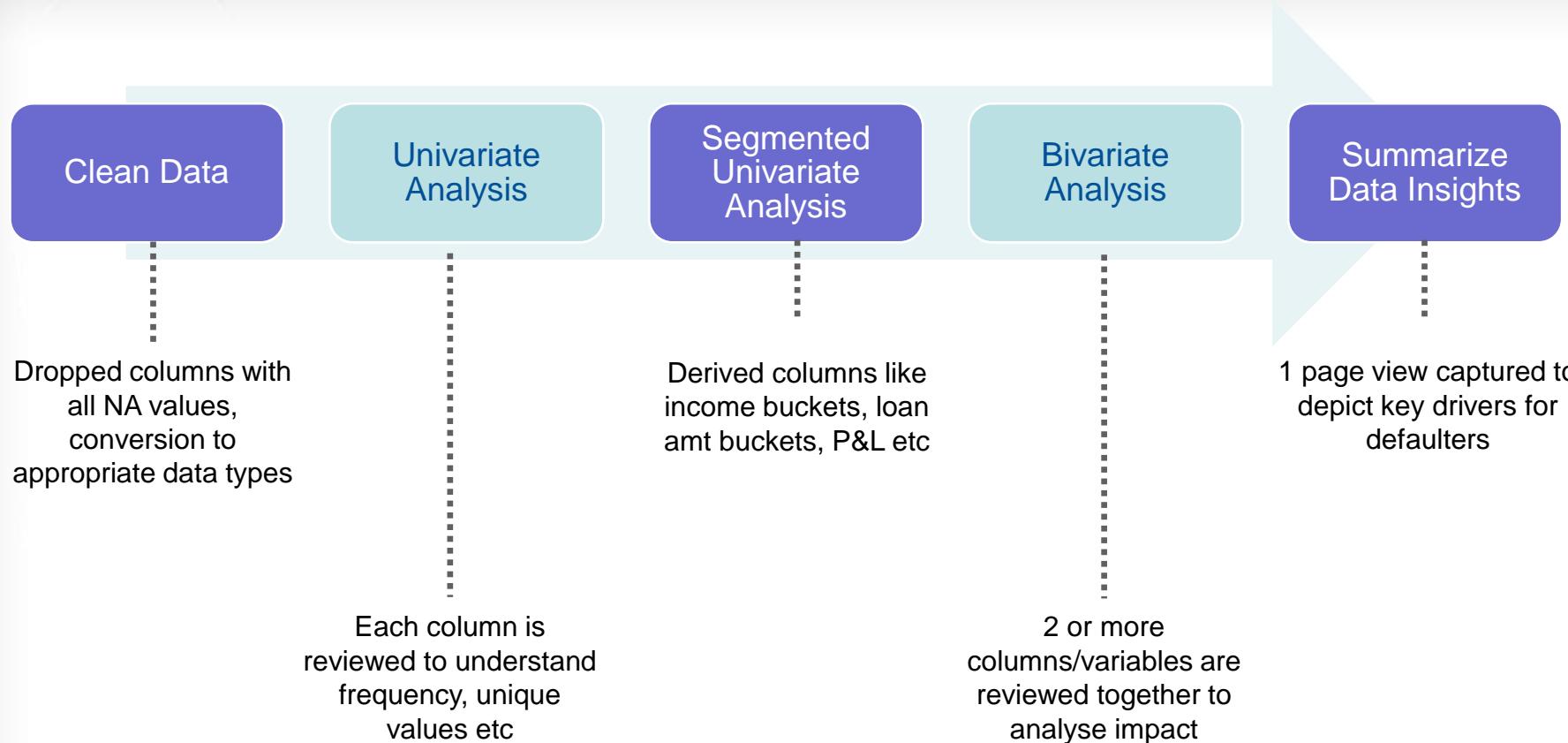


Identify driving factors behind loan default

It can be used for portfolio and risk assessment for reducing defaulters (credit loss)



Problem Solving Approach





Top View on Profit & Loss

Loan Data	Planned Recovery	Actual Recovery	Planned Profit	Actual Profit
Overall	528730306	457792116	24%	7%
Fully Paid	439321840	419314516		17%
Charged Off	89408466	38477600		-44%

- 44% Loss in defaulter cases which drastically reducing overall profit to the company and urgent need to take corrective/preventive actions



Defaulters - Driving factors & Recommendation



Term: Offer 36 months term more as with positive trend compare to 60 months



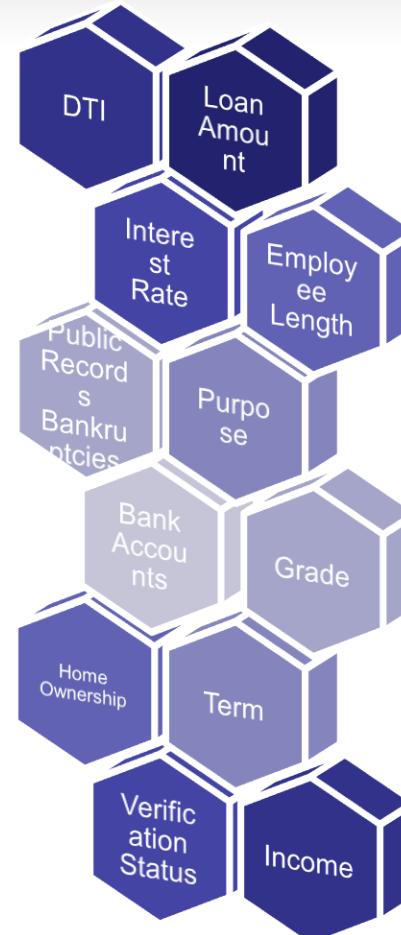
Home Ownership: Reduce loan request for Rent category or reduce loan amount



Interest Rate: Instead of keeping flat interest rate, apply more for boarder approval cases when higher **DTI/Public Records/ Bank Accounts**



Verification Status: Mandate source verification as no major loss when income source was verified



Loan Amount: Reject loan amount > 15% of **annual income** for income < 200K



Employee length: < 2 yrs or 10+ yrs with highest defaulter risks. Apply further stringent steps



Purpose: reduce loan sanction for 'Small Business' & 'Renewable Energy' or scrutinize deep



Grade: Lower Grades/Sub Grades with higher defaulter risks

Note: Use slide navigator to see detailed analysis for each driving factor mentioned above



Univariate Analysis

- Individual column was reviewed for values, unique categories, with frequency like below

Loan status:

- Fully Paid (32950)
- Charged Off (5627)
- Current (1140) this column

Pub Rec Bankruptcies:

- 0 (37339)
- 1 (1674)
- 2 (7)

Interest Rate:

14.65% (39707 loan cases)
7 other different rates (10 loan cases)

Term:

36 & 60 months

Grade:

A(10085), B(12020), C(8098), D(5307), E(2842),
F(1049), G(316)

Verification Status:

- Not Verified ()
- Verified
- Source Verified

Home Ownership:

- Rent (18899)
- Mortgage (17659)
- Own (3058)
- Other (98)
- Not Specified (3)

DTI:

count(39717), mean(13.31),
std(6.67),
min(0), 25%(8.17),
50%(13.40), 75%(18.60),
max(29.99)



Segmented Univariate Analysis

- Below columns are derived and analysed individually
 1. Annual income bucket (50 K – 400K)
 2. Loan Amount Bucket (5 K – 40K)
 3. P&L: (Total Payment – Loan Amount)/ Loan Amount
 4. Expected Recoveries: Installment * Term

Income Buckets

50	15389
100	17707
200	4911
300	399
400	81
600	66
500	24

P_L (Profit & Loss)

```
count    38577  
mean     7.80  
std      27.30  
min     -100  
25%      6.96  
50%      13.90  
75%      21.04  
max      68.84
```

Expected Recoveries

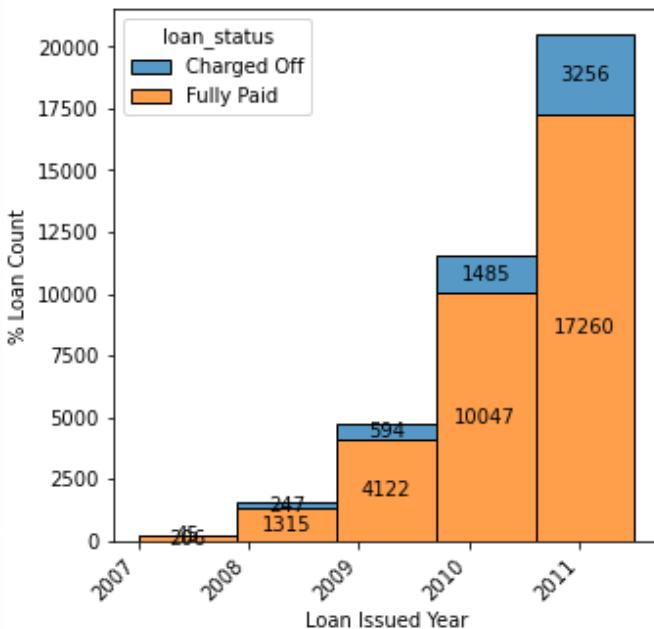
```
count    38577  
mean    13705  
std     9904  
min     564  
25%     6422  
50%     11199  
75%     18225  
max     60303
```

Loan Amount Buckets

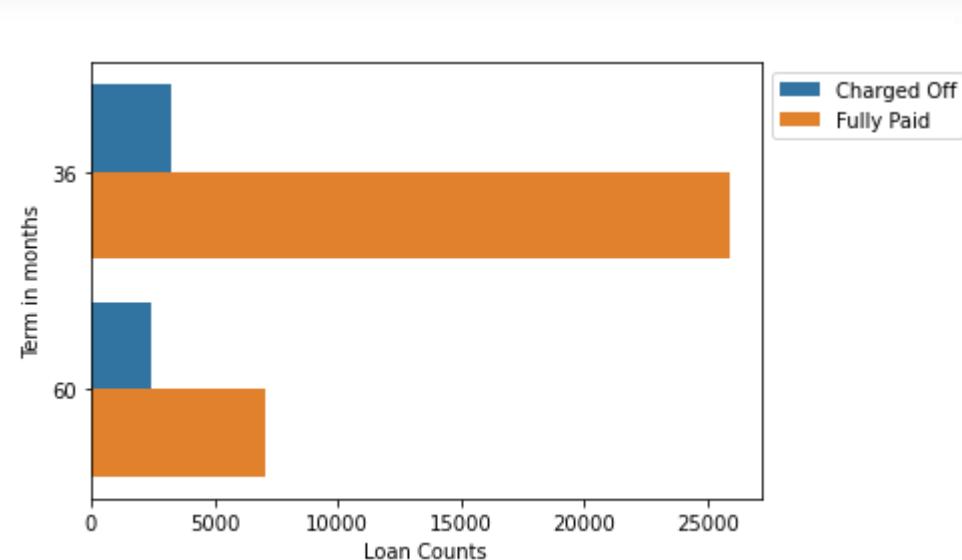
5	9472
10	12802
20	12032
30	3536
40	735



Bivariate Analysis – Year & Term



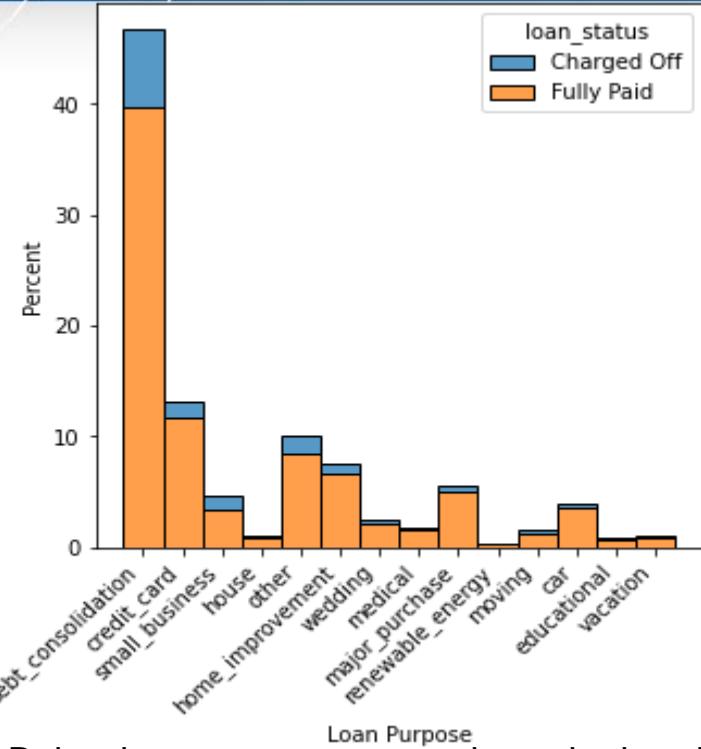
- Loan count is trending up but defaulter rate is consistent between 15% - 18% for last 3 years



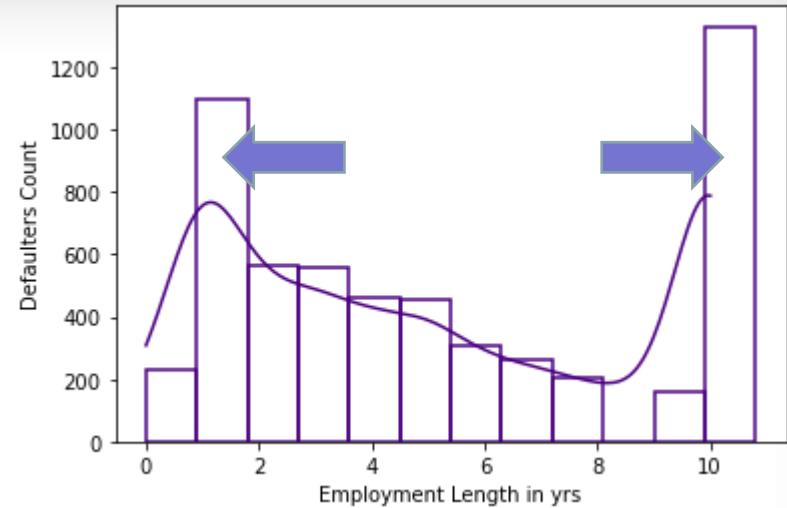
- 36 months loan term with better results than 60 months
- Better to offer more 36 months Term loans

[Go Back](#)

Bivariate Analysis – Employee Length & Purpose



- Below loan purpose categories to be handled cautiously as % defaulters are higher side
 - Dept Consolidation
 - Credit Card
 - Small Business
- Highest % defaulters are in Small Business Category

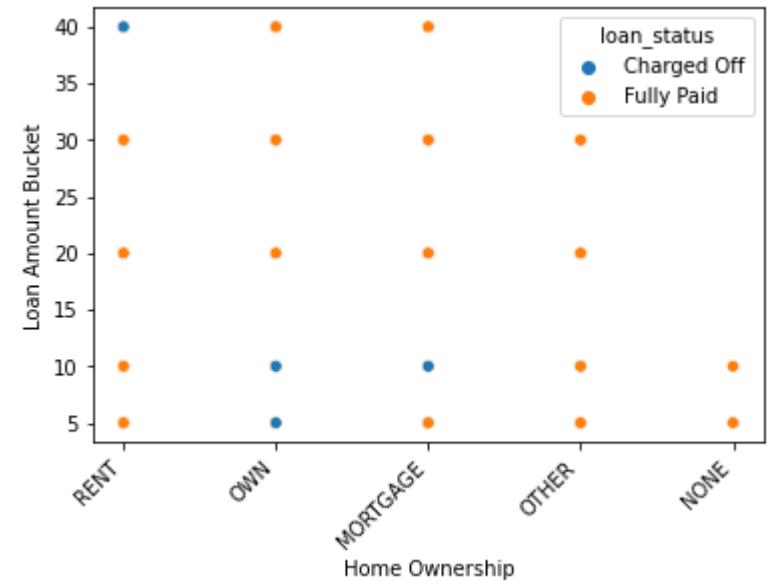
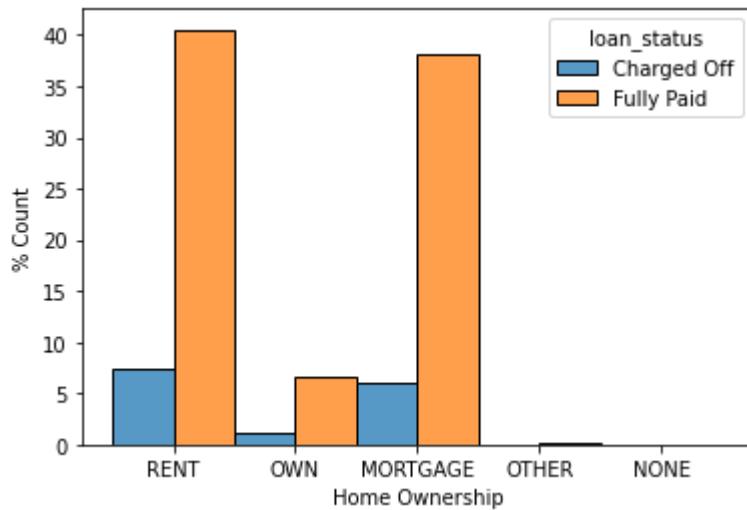


- Employment length < 2 yrs & 10+ years to be watched as deflater peaks

[Go Back](#)



Bivariate Analysis – Home Ownership

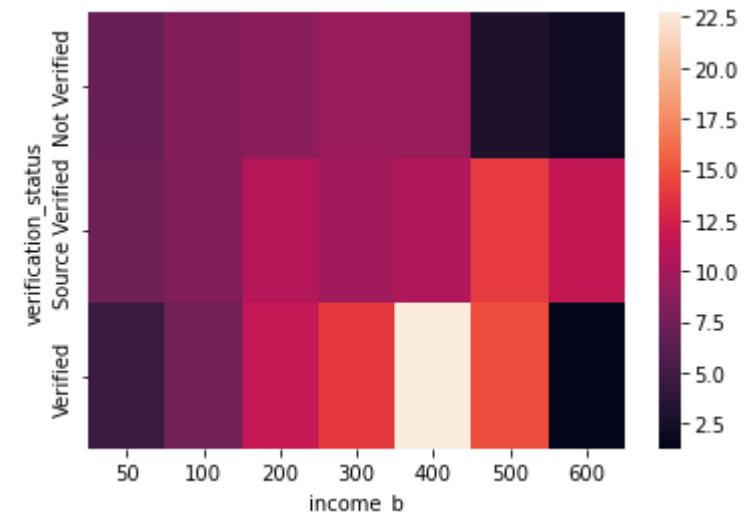
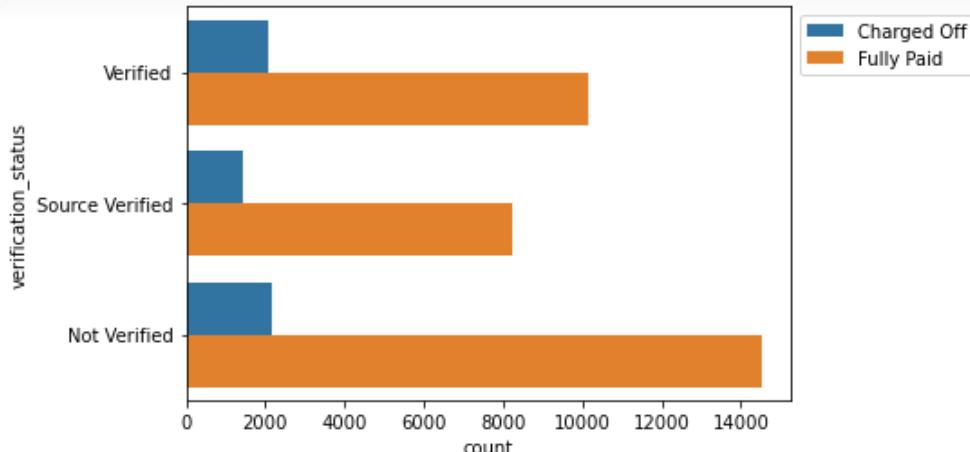


- % Defaulters high in case of Rent & Mortgage
- In case of Rent, defaulters with top loan amount bucket
- Reduce providing loans as well as reduce loan amount for Rent category

[Go Back](#)



Bivariate Analysis – Verification Status



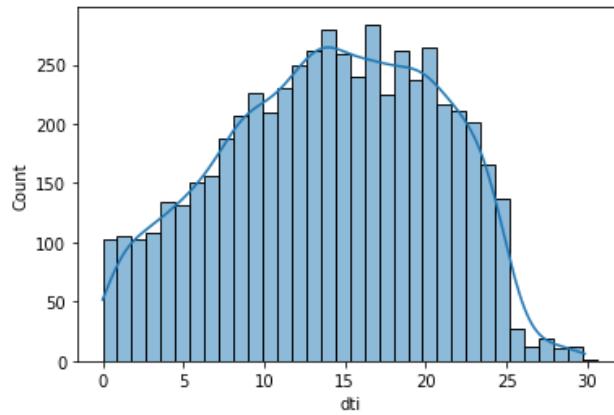
- Surprisingly defaulter rate is lower in 'Not Verified' state compare to 'Verified' & 'Source Verified'
- But, no major losses when source was verified. Better to mandate verification & source verification process

[Go Back](#)



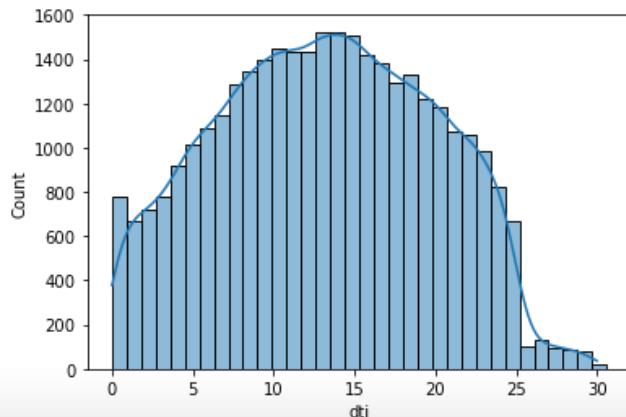
Bivariate Analysis – DTI Vs Loan Status

Defaulters



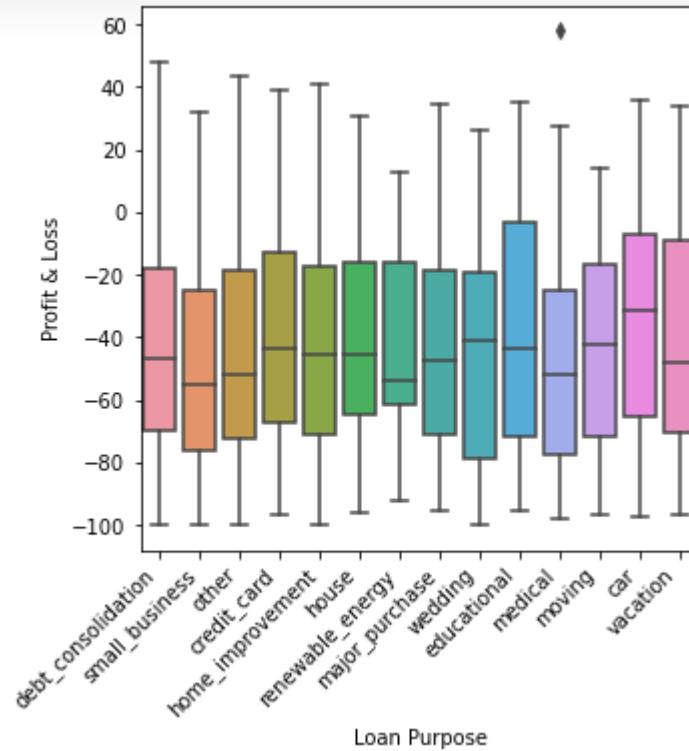
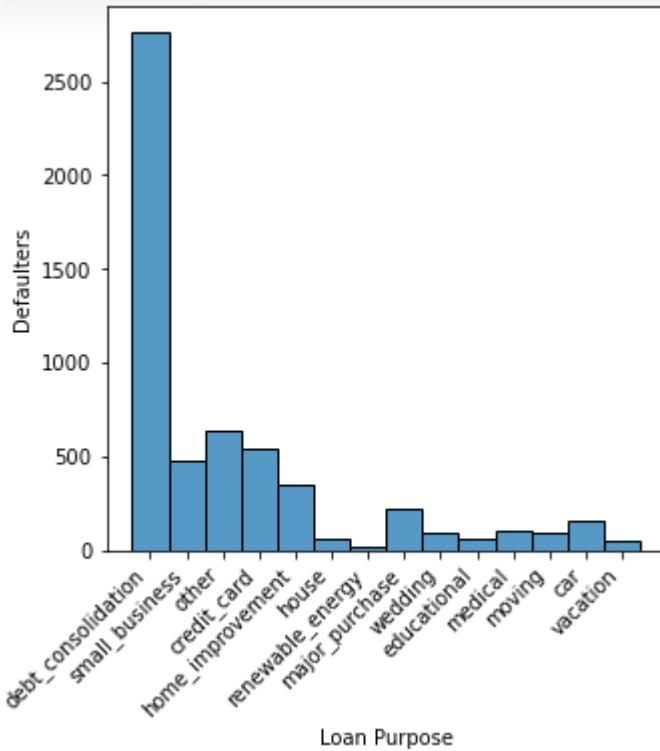
- Defaulter (loan status = “Charged Off”) count is trending up with DTI value
- Peaks observed DTI > 14; securitize deep for these loan requests

Non defaulters



[Go Back](#)

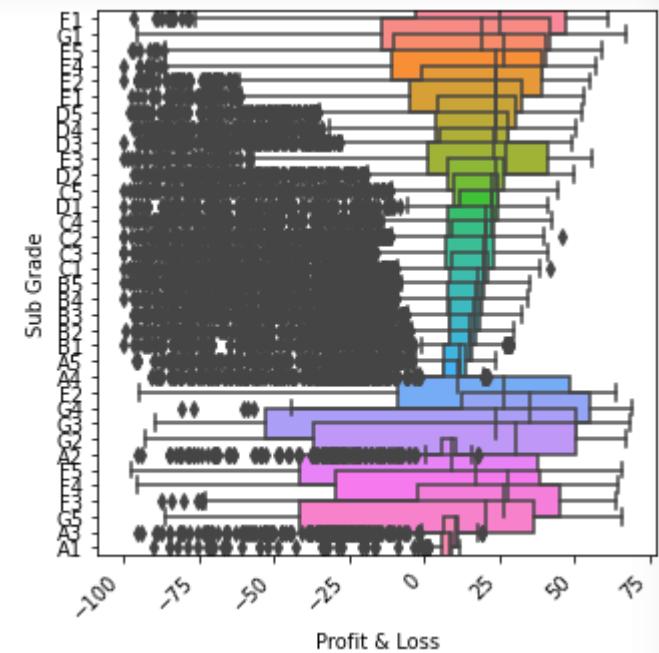
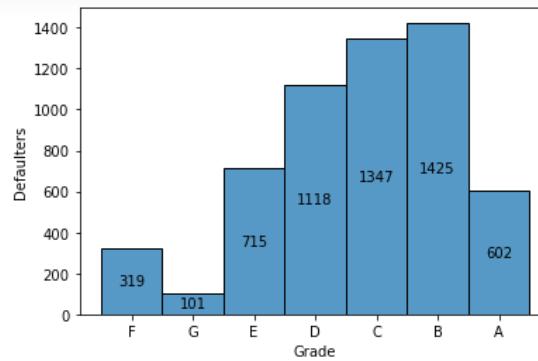
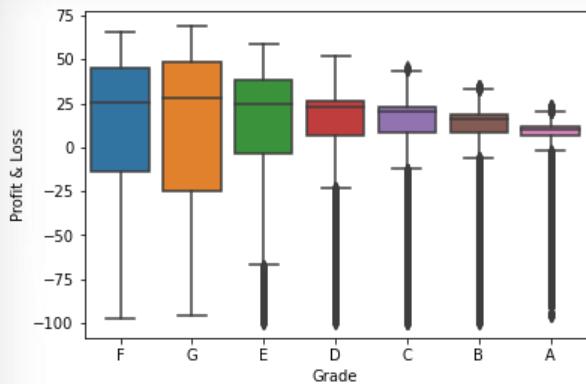
Bivariate Analysis – Loan Purpose & P&L



- 2 loan purpose categories – ‘Debt Consolidation’ & ‘Small Business’ with highest defaulters
- 2 industries – ‘Small Business’ & “Renewable Energy” with maximum average loss

[Go Back](#)

Bivariate Analysis – Profit & Loss across Grades

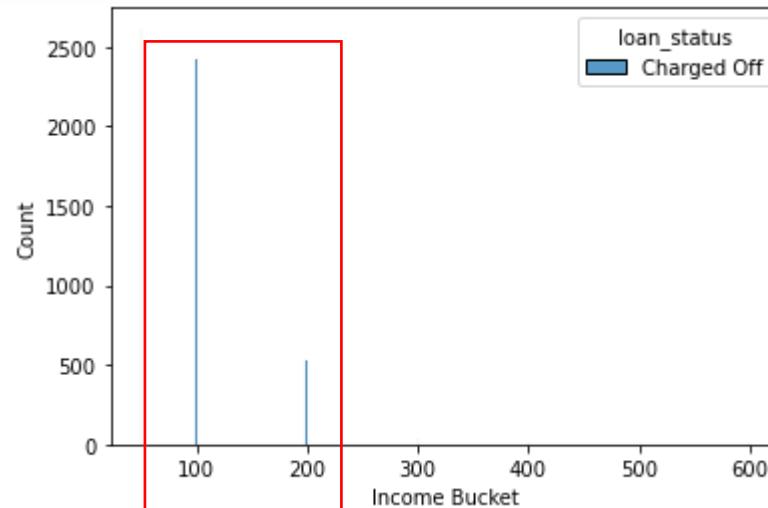
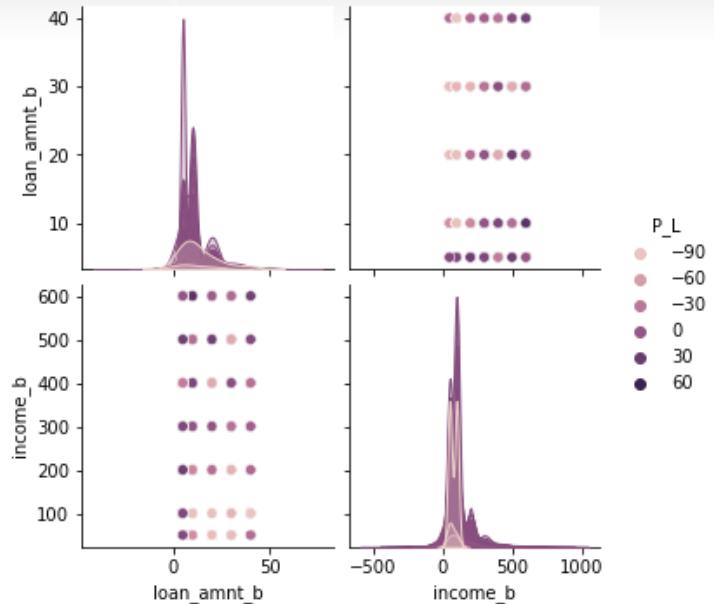


- Grade: profit improves as grade upscale, but high risk with lower grades
- Sub Grade: heavy loss making cases for middle sub grades though major cases are profit making
- Initial sub grades are with wide spread across loss and profit scale and to be reviewed closely

[Go Back](#)



Bivariate Analysis – Profit & Loss, Income, Loan Amount

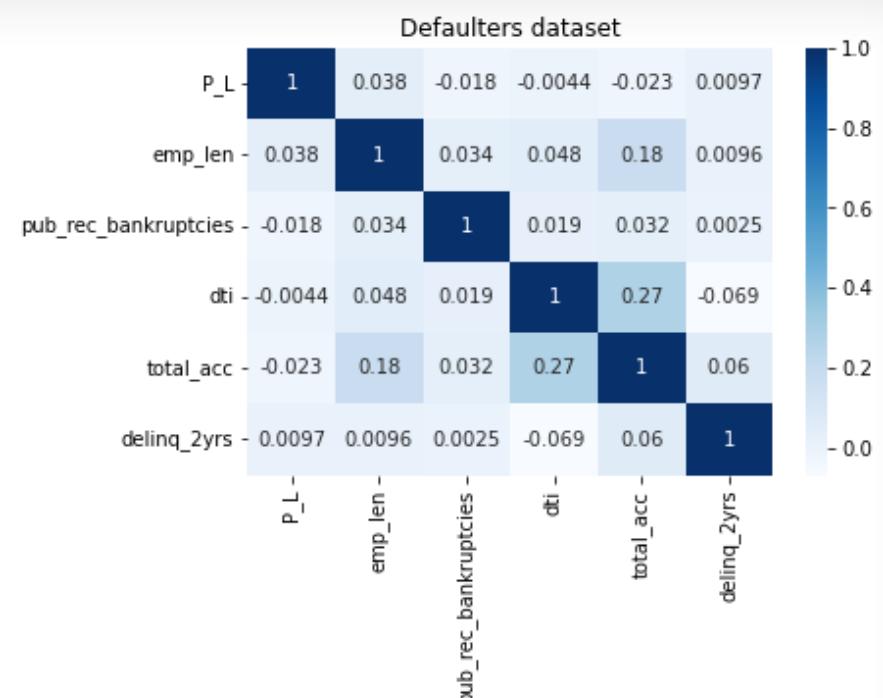
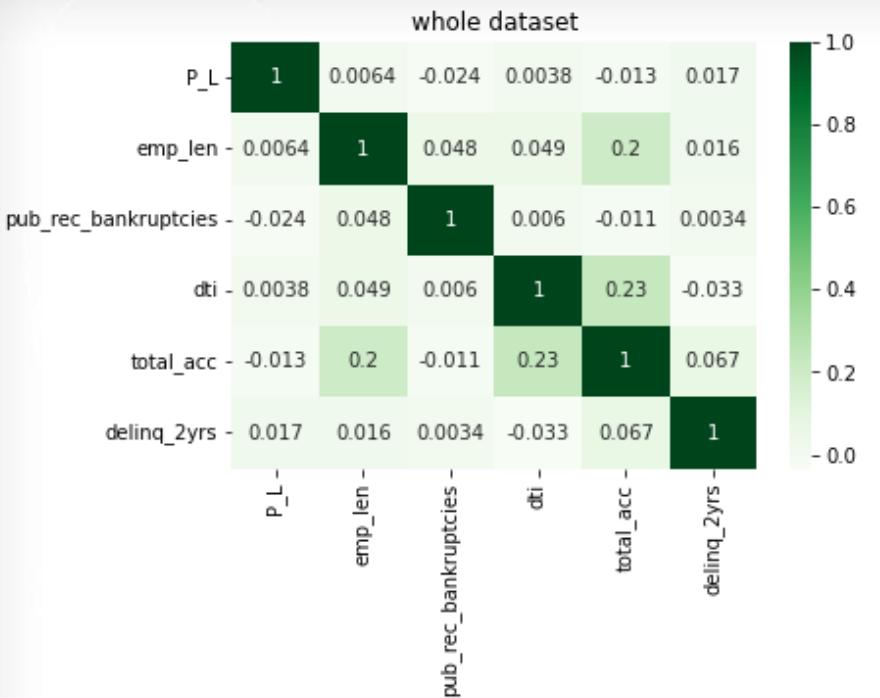


- Income < 200K & loan amount between 15K to 40K to be scrutinized to mitigate loss risks or increase interest rate to have better recovery
- Income < 200K, sanction loan up to 15% of annual income

[Go Back](#)



Bivariate Analysis – Watch Profit & Loss across key variables



- Overall not very strong corelation but certainly employee length is positively impacting whereas below 3 drivers for loss making
 1. Public Records (Bankruptcies)
 2. DTI
 3. Number of Bank Accounts

[Go Back](#)



The Goal was to turn data into information
and information into insights

