

Lending Club Case Study

EDA and Recommendations

CL-48 Debarshi Chakraborty

Objective

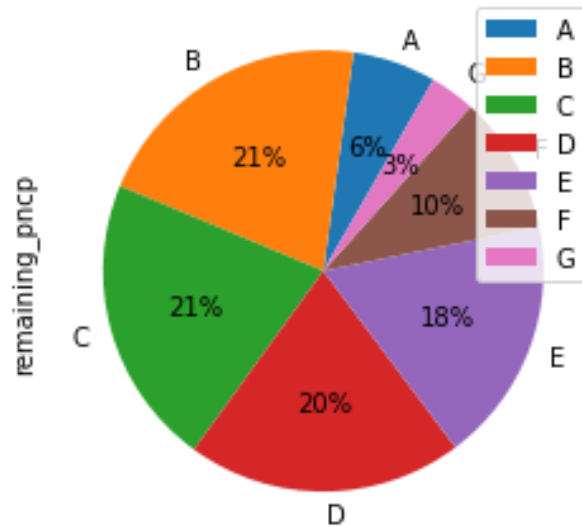
- The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Data Cleaning

- Customer behavioural columns were dropped as they will not be available during approval process.
- Some description columns were also dropped that either held same values or were just informative.
- Only “Charged-Off” rows were taken.
- New column was added for the “Remaining Principal” amount. This column reflected the amount that is still outstanding to be received.
- Based on this newly added column some rows were identified where the remaining amount was either zero or negative.

Data Analysis

LC issued Loan grade

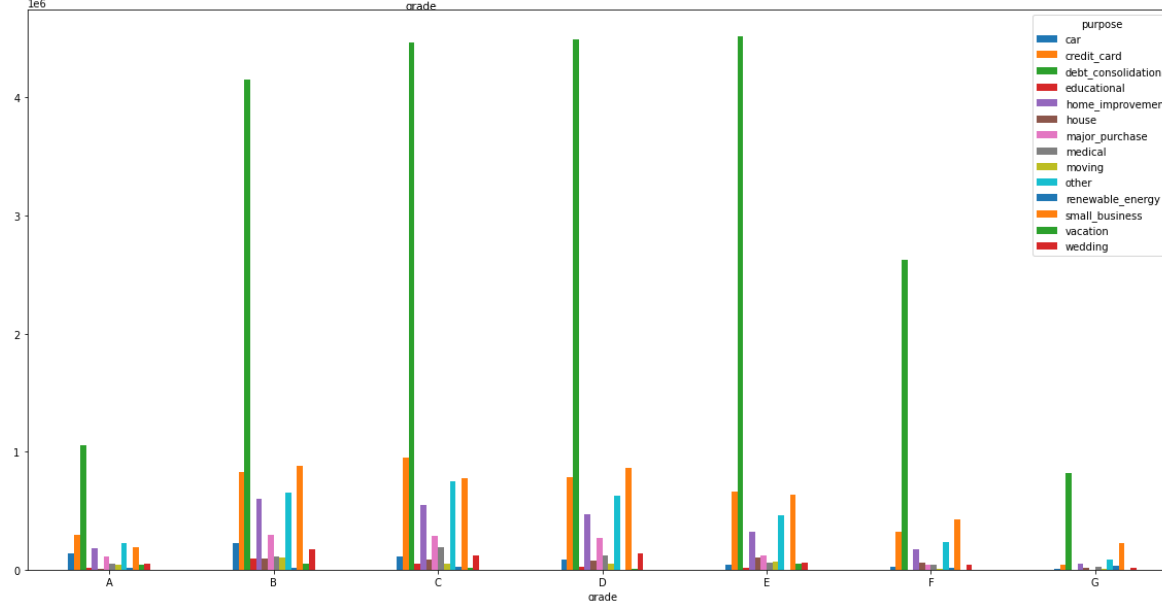
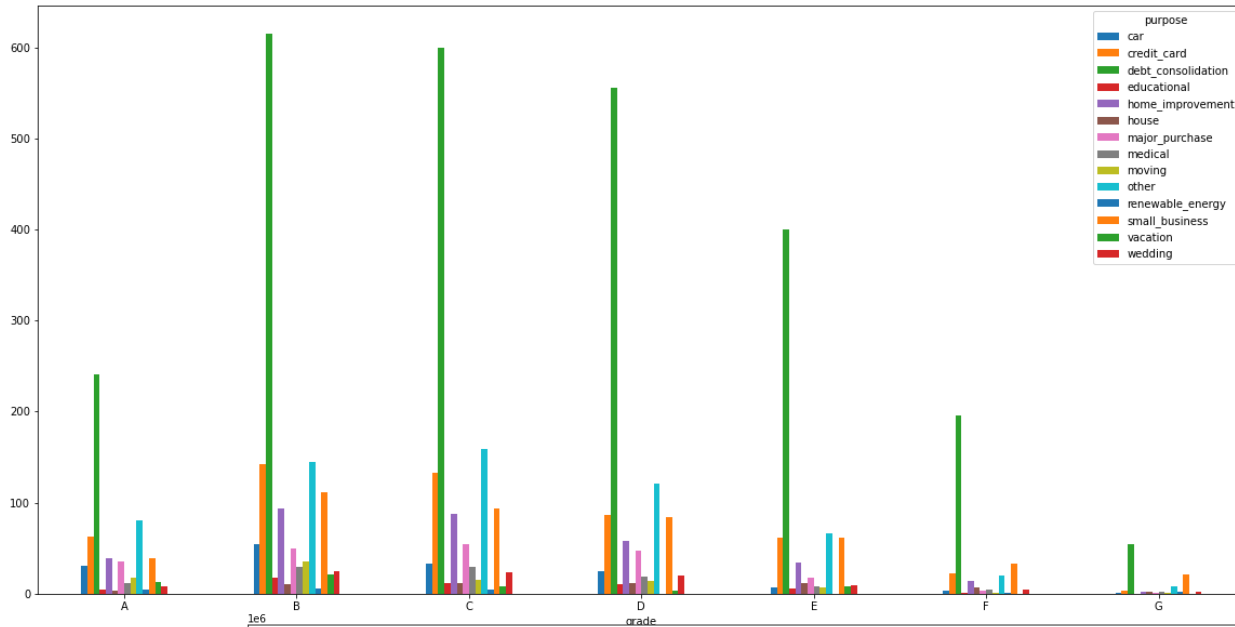


loan_status Charged Off

Row Labels	Count of member_id	
A	602	
B	1,425	25%
C	1,347	24%
D	1,118	20%
E	715	13%
F	319	
G	101	
Grand Total	5,627	

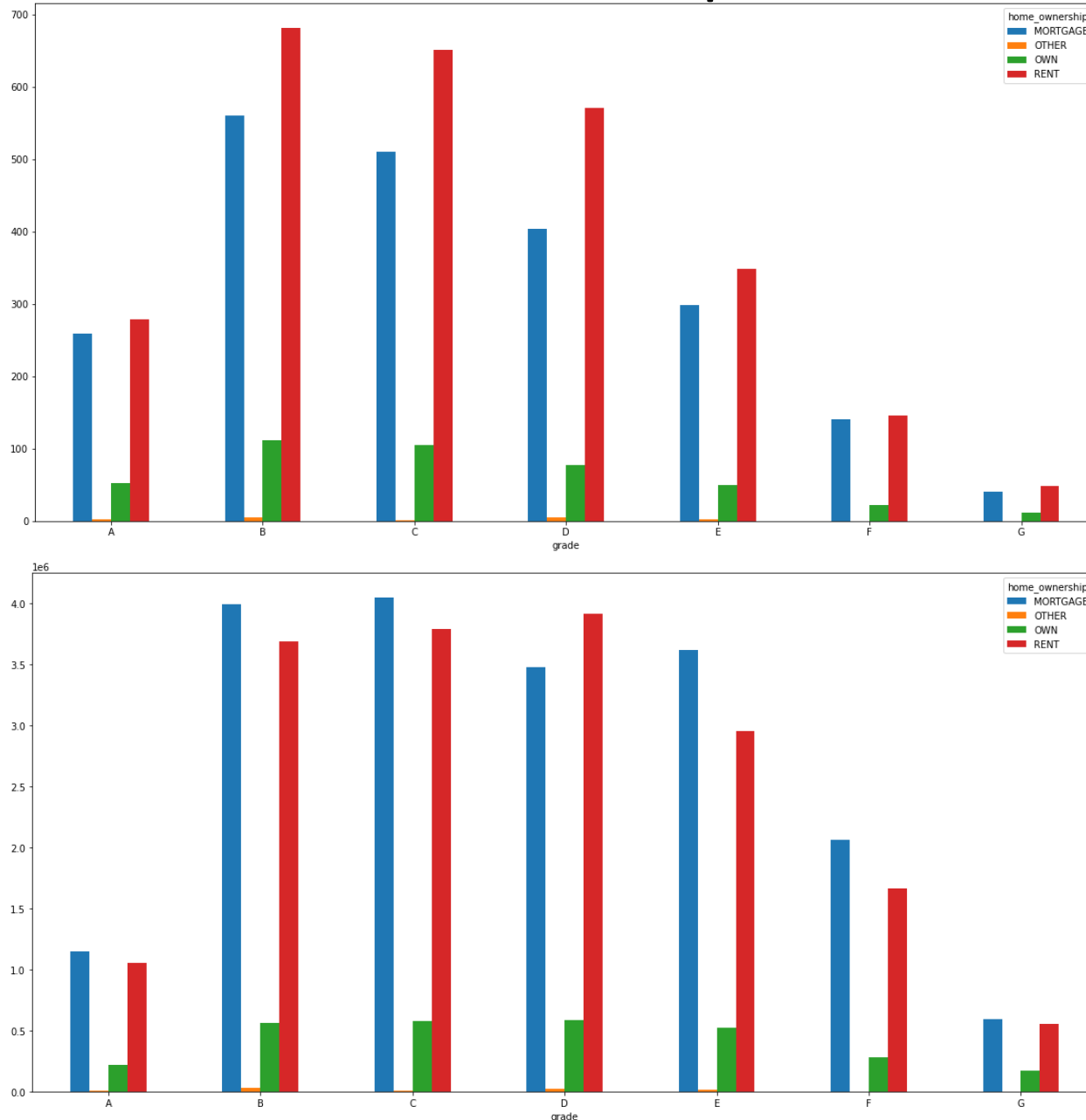
- The pie shows the percentage of remaining amount to be recovered across different loan grades.
- Loan grade B, C, D and E together contributed to **80%** of the amount to be recovered.
- Even in terms of numbers, these four categories contribute the numbers
- **5627** are the number of defaulters

Purpose of Loan



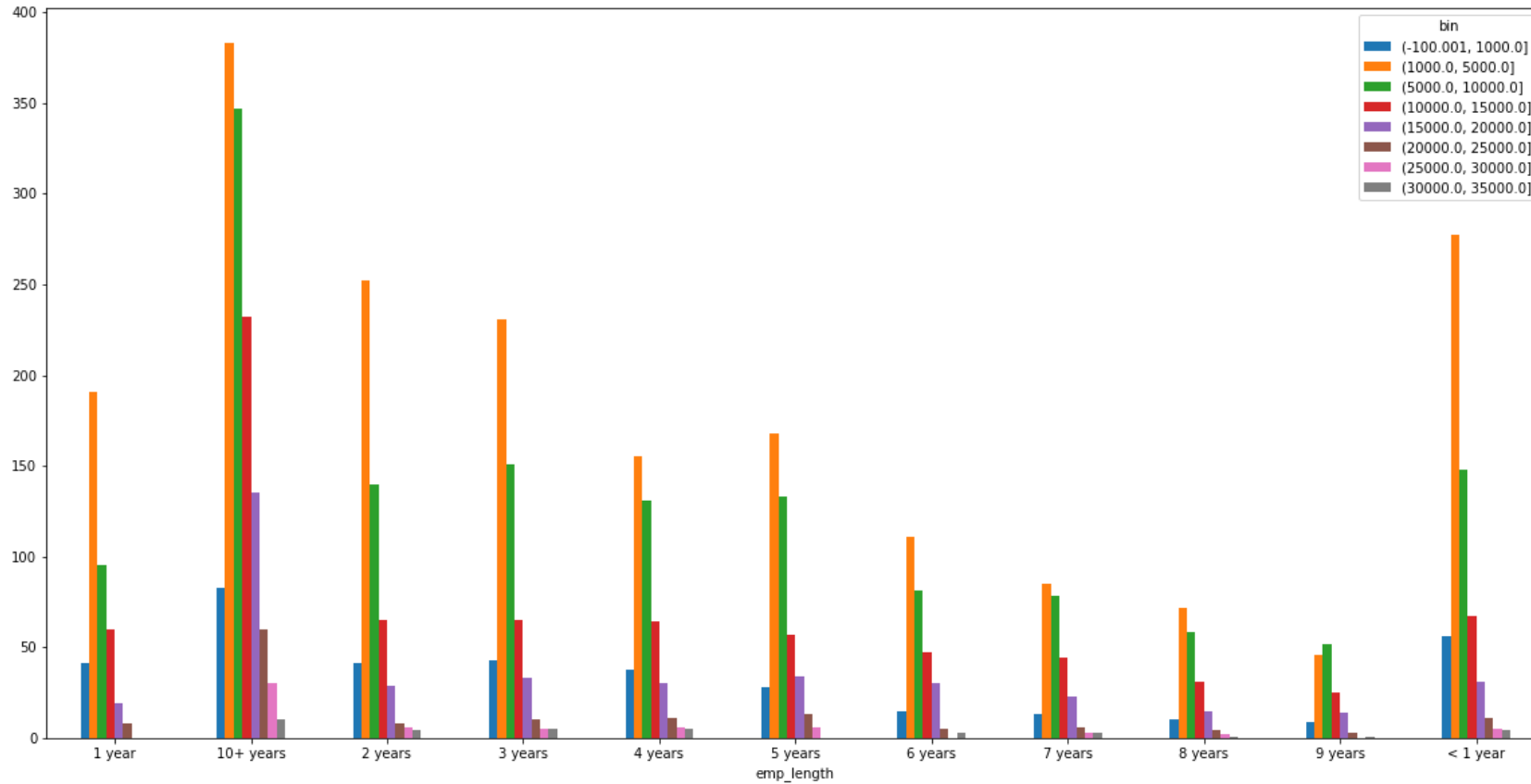
- The bar chart – *top* one shows the number and *bottom* one shows the default amount remaining – across different loan grades, when measured against the purpose of loan.
- The loan purpose of “**debt-consolidation**” is the bar standing out, both in terms of number of defaulters and the amount they defaulted, across all grades.

Home Ownership



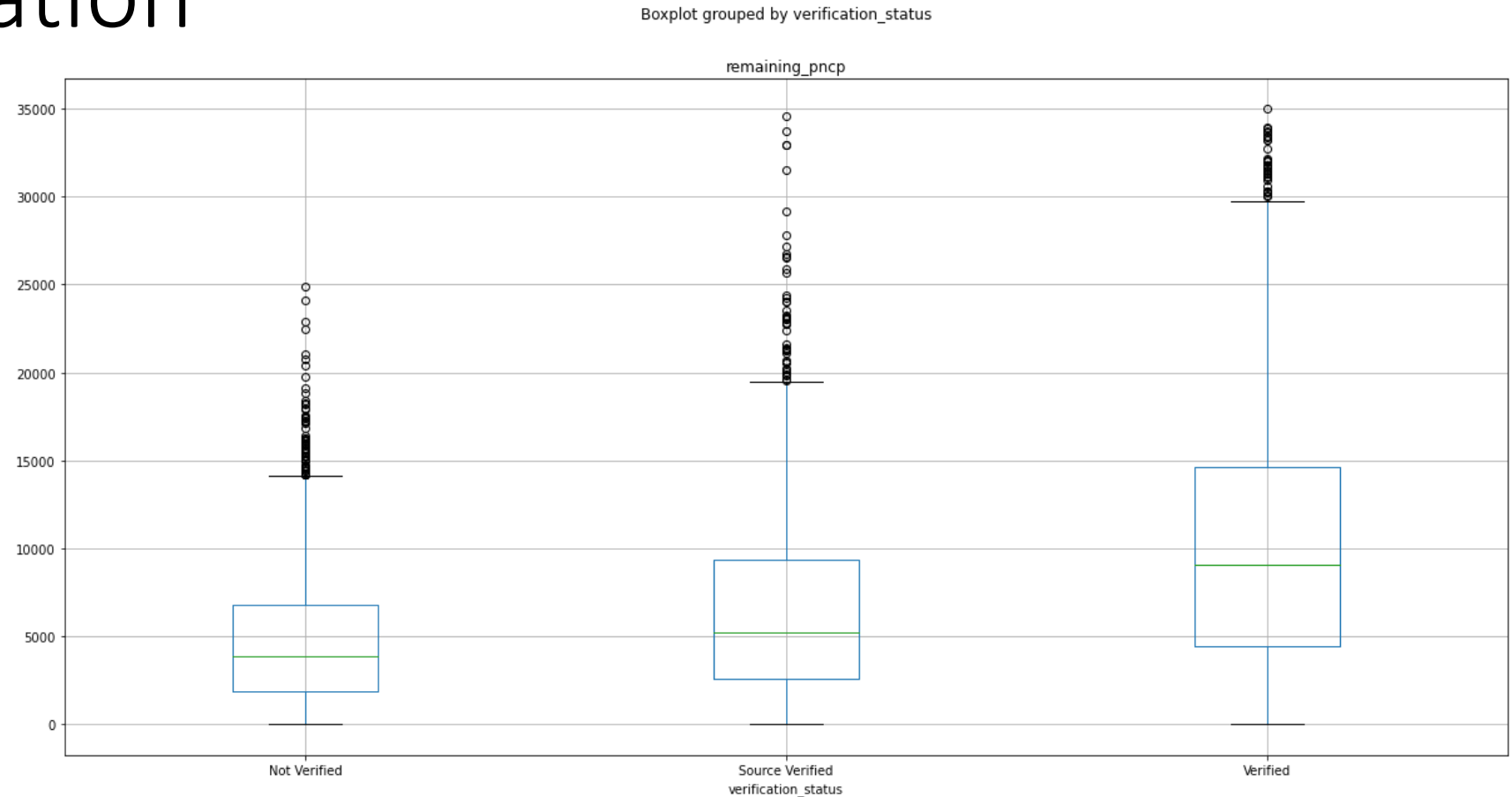
- The bar chart – *top* one shows the number and *bottom* one shows the default amount remaining – across different loan grades, when measured against the ownership of the home property.
- The properties that are on “**Rent**” or “**Mortgage**” is the bar standing out, both in terms of number of defaulters and the amount they defaulted, across all grades.

Length of Employment



- The bar chart –shows the number of defaulters – across different bins of amount remaining to recovered, when measured against the length of employment.
- The people having experience less than a year or more than 10 years, have defaulted more.
- Majority of the amount defaulted is in the range of 1000 to 15000

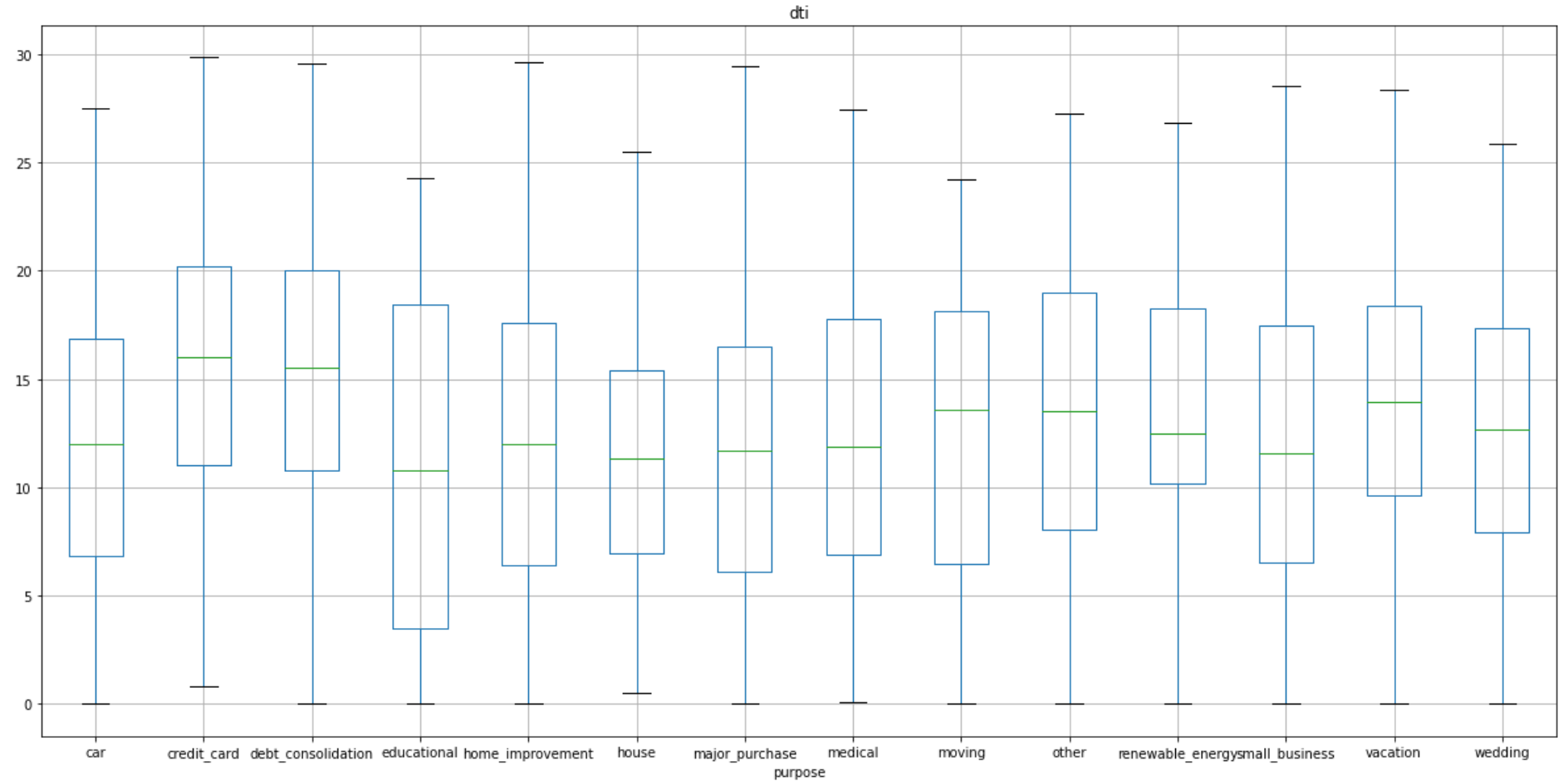
Income Verification



- The box plot shows the remaining amount to be collected back, across the verification status of Income.
- Surprisingly, the “**Verified**” sources had higher number of defaulters under different percentile. Even the outliers amount are more for Verified sources.

Dept to Income Ratio - DTI

Boxplot grouped by purpose



- The box plot shows the dti ratios across different purpose for taking the loan.
- Reaffirming the earlier conclusion the “**dept-consolidation**” has higher dti.
- Even the “**credit-card**” box has higher dti

Recommendation

- Loan category of B, C, D and E should have a deeper scrutiny before approving the loans.
- The purpose of 'Dept-consolidation' often ends up being defaulted irrespective of other parameters
- The people whose houses are on RENT or MORTGAGE has higher chances of being default
- Income sources should be verified