Lending Club Assignment

Submission

Name: Jennis M. Vicente-Feliciano

LendingClub Data Analysis

LendingClub is a company that recommends clients for lending partners. They use data analysis in order to match the lender that will most likely fund a loan. Data for loans issued between 2007 and 2011 was analysed in order to provide the findings in this assignement.

Data properties

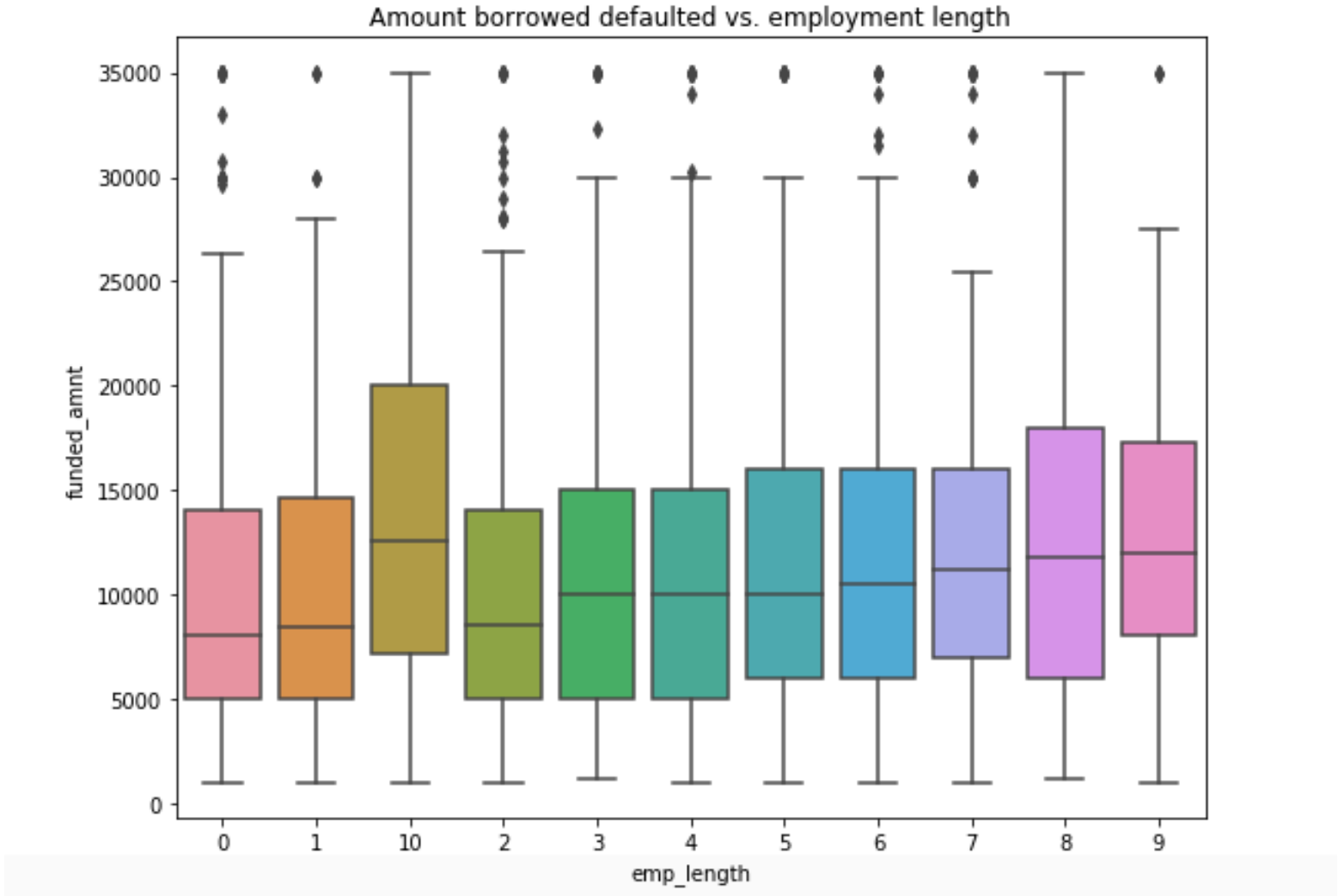
- Rows: 39717 entries
- Columns: 111 entries
- dtypes: float64(74), int64(13), object(24)
- memory usage: 33.6+ MB

Methodology

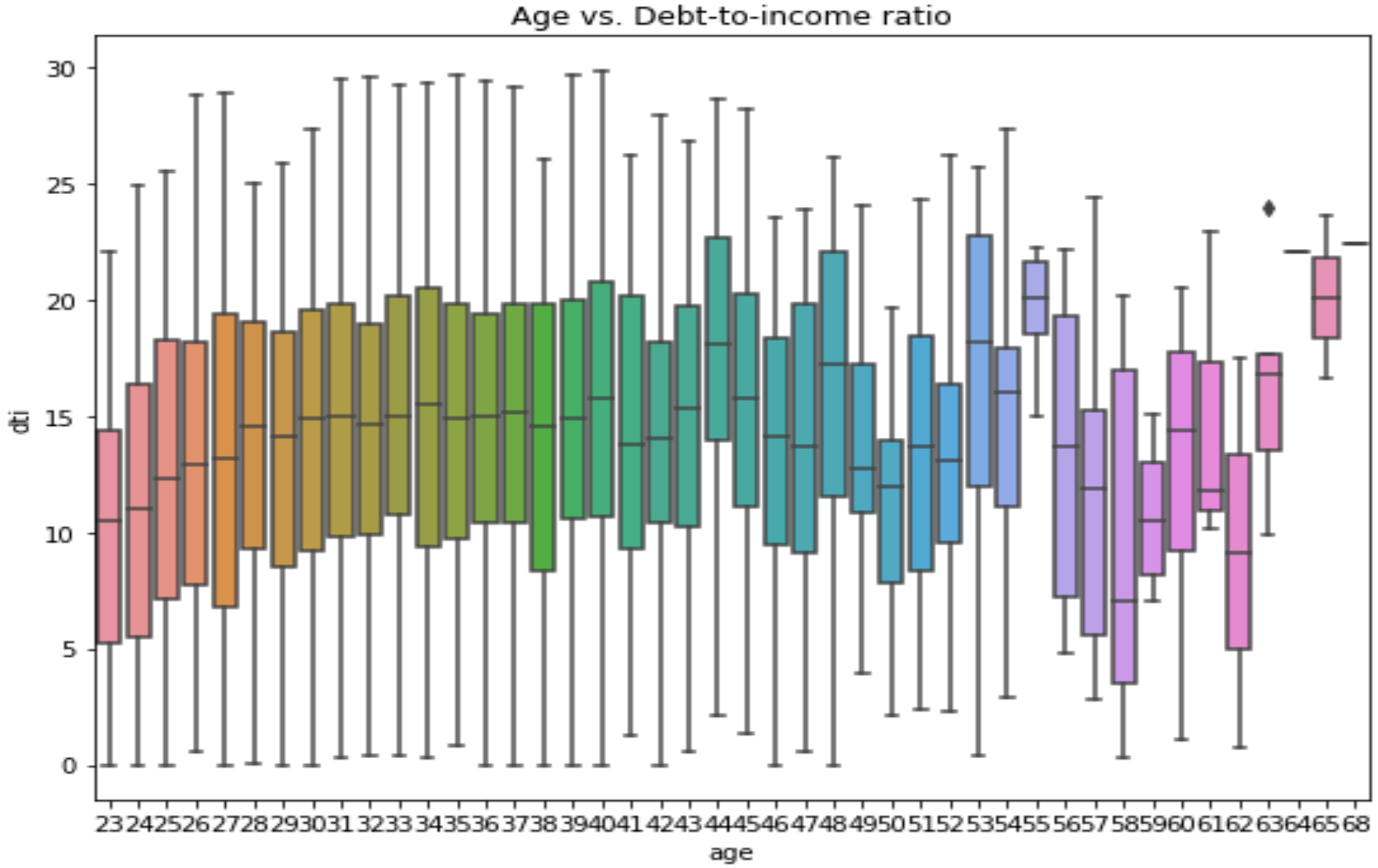
- Inspect and drop missing values: left with 37823 rows a 53 columns.
- Replace objects and change to int or float.
- Split data in columns to calculate oldest credit line’s age and borrower’s age.
- Plot and analyze data

Loan Characteristics- defaulted

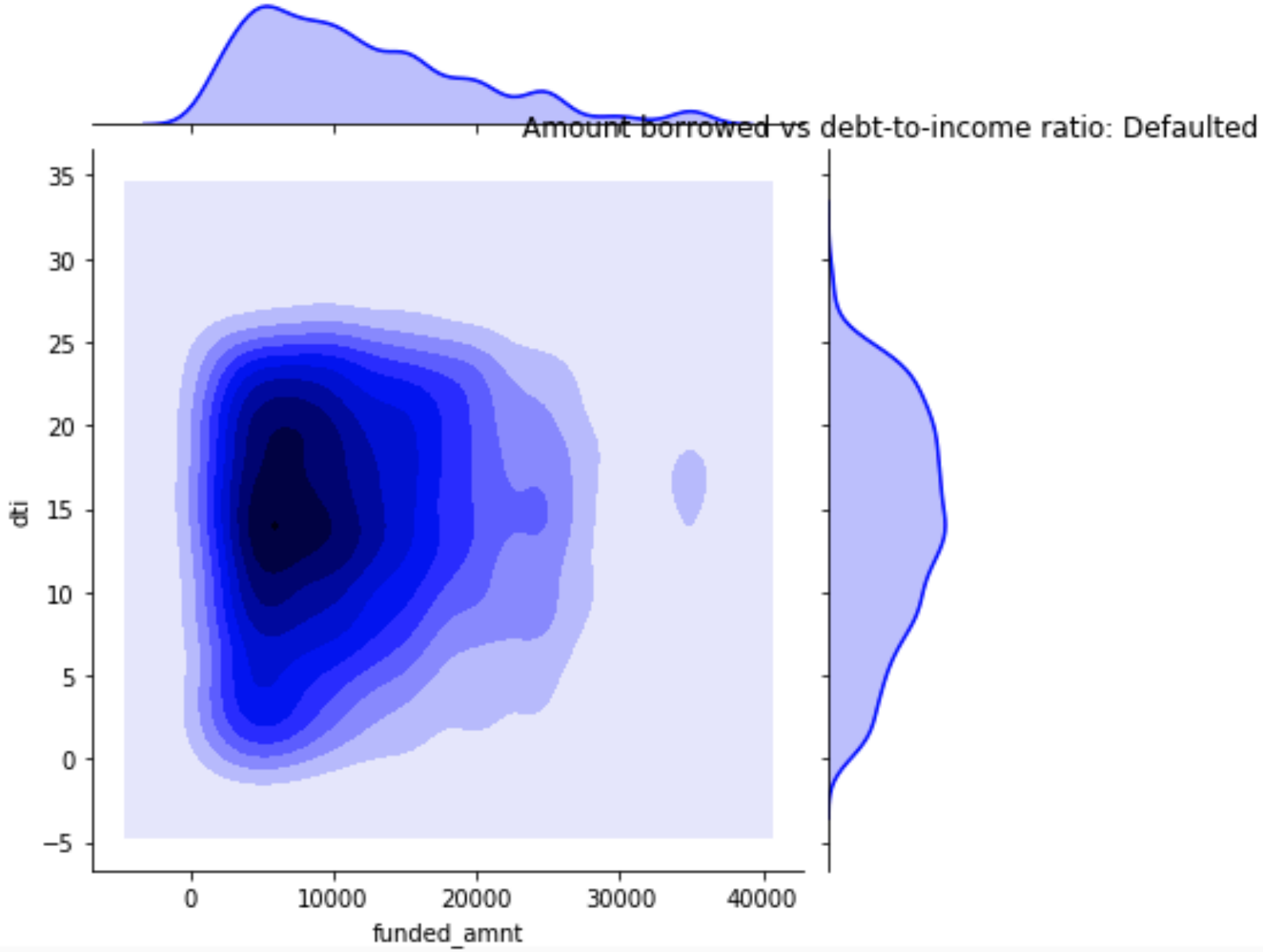
	member_id	issue_year	loan_amnt	funded_amnt	term	int_rate	installment	out_prncp	recoveries
count	5.201000e+03	5201.000000	5201.000000	5201.000000	5201.000000	5201.000000	5201.000000	5201.0	5201.000000
mean	8.804708e+05	2010.412228	12293.433955	11926.624688	46.645645	13.925574	339.516268	0.0	687.771202
std	2.583625e+05	0.813844	8096.734468	7759.359919	11.924473	3.657039	215.858288	0.0	1749.200323
min	1.328890e+05	2007.000000	1000.000000	1000.000000	36.000000	5.420000	22.790000	0.0	0.000000
25%	6.951510e+05	2010.000000	6000.000000	6000.000000	36.000000	11.490000	172.380000	0.0	9.770000
50%	8.965240e+05	2011.000000	10000.000000	10000.000000	36.000000	13.790000	297.530000	0.0	182.760000
75%	1.070018e+06	2011.000000	17000.000000	16000.000000	60.000000	16.450000	460.100000	0.0	613.350000
max	1.314167e+06	2011.000000	35000.000000	35000.000000	60.000000	24.400000	1305.190000	0.0	29623.350000



The trend is that the longer you are employed, the larger the loan amount.

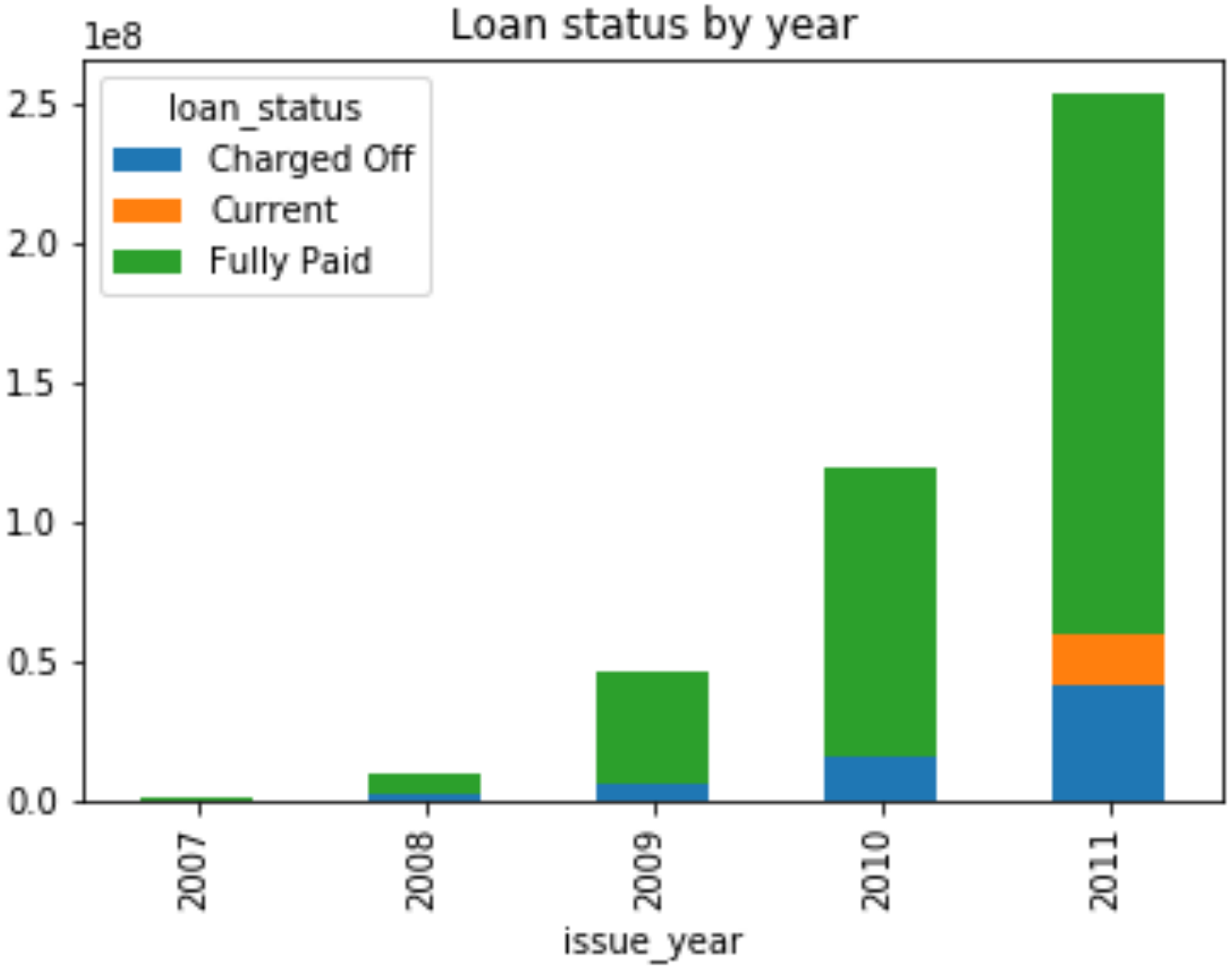


Assuming the 1st credit line was opened at 20 yrs old, the age is estimated at the time of funding the loan. There is not a clear trend in whether the age determines your debt ratio, however generally is higher.



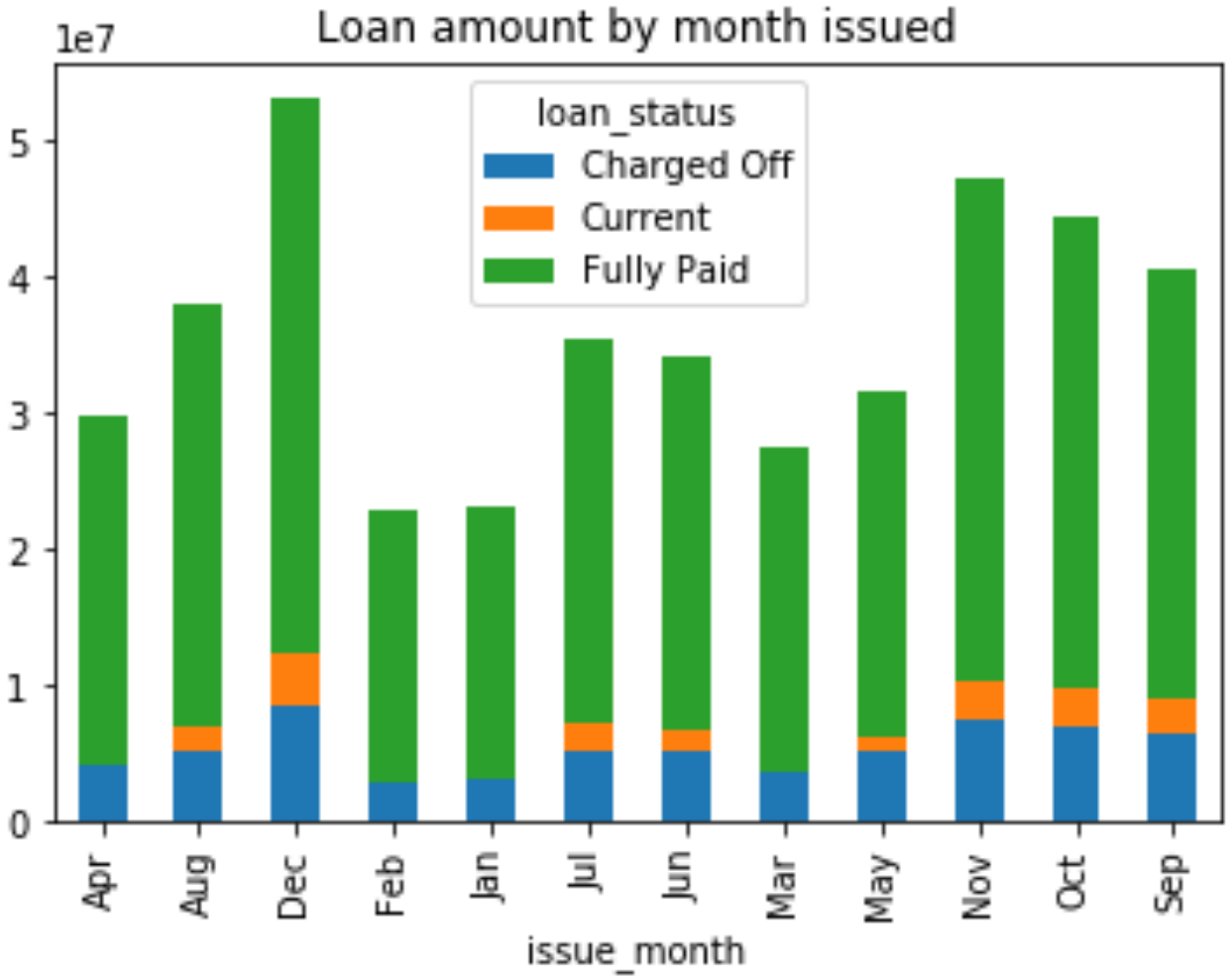
According to the graph, the most density is in borrowers with a 10-20 debt to income ratio and amounts from $5k-$10k.

Understanding all borrowers



More loans where charged off in 2011 than other years. Possibly due to the recession on the previous years and the concessions given by lenders (payment plants, deferment, etc).

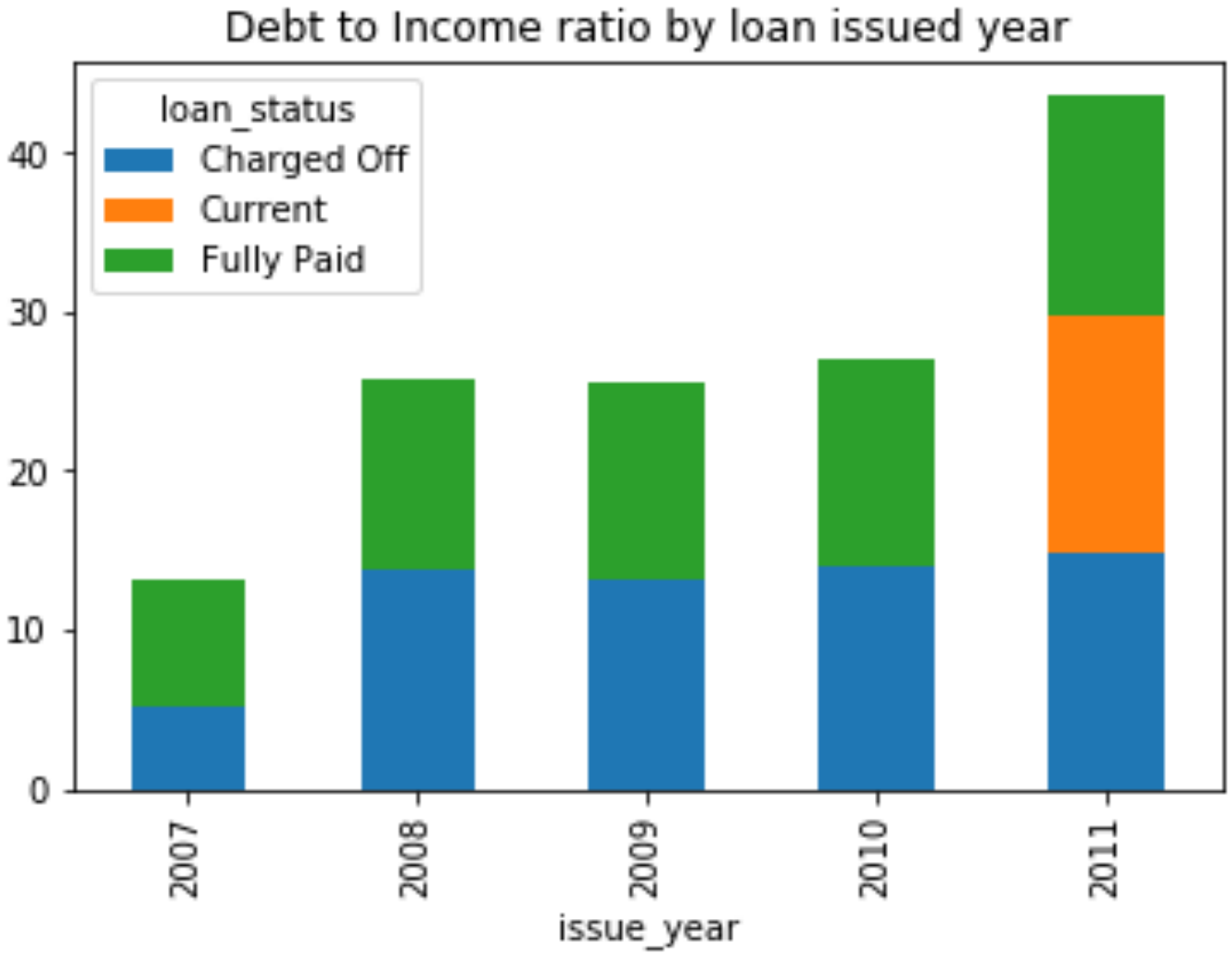
Understanding all borrowers



Significant increase of loans leading the holidays.

More loans were issued and charged off in December.

Understanding all borrowers

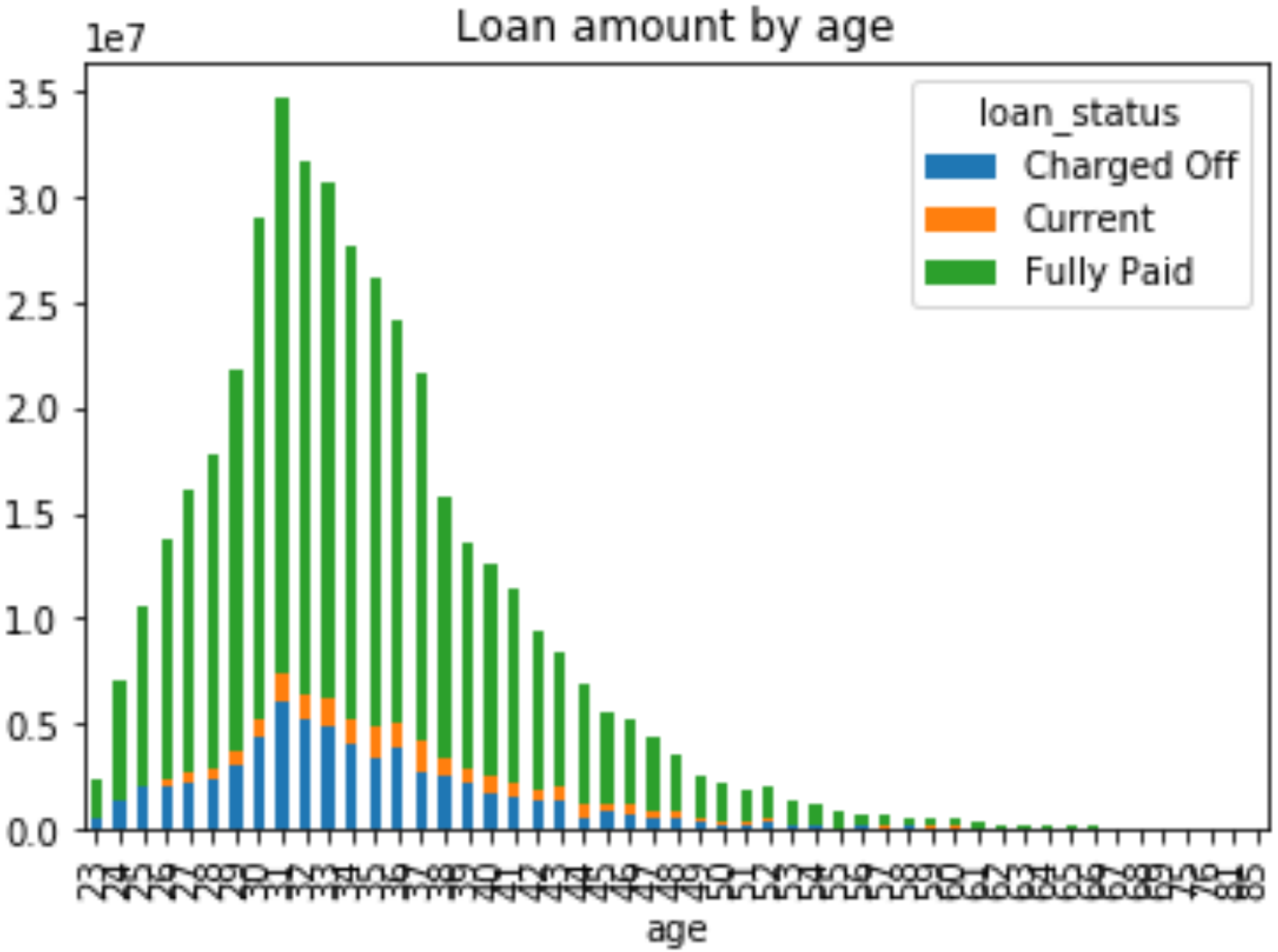


Debt income ratio extremely low in 2007 consistent to the cause of recession.

From 2008 to 2010 seems like half of the funded loans will default.

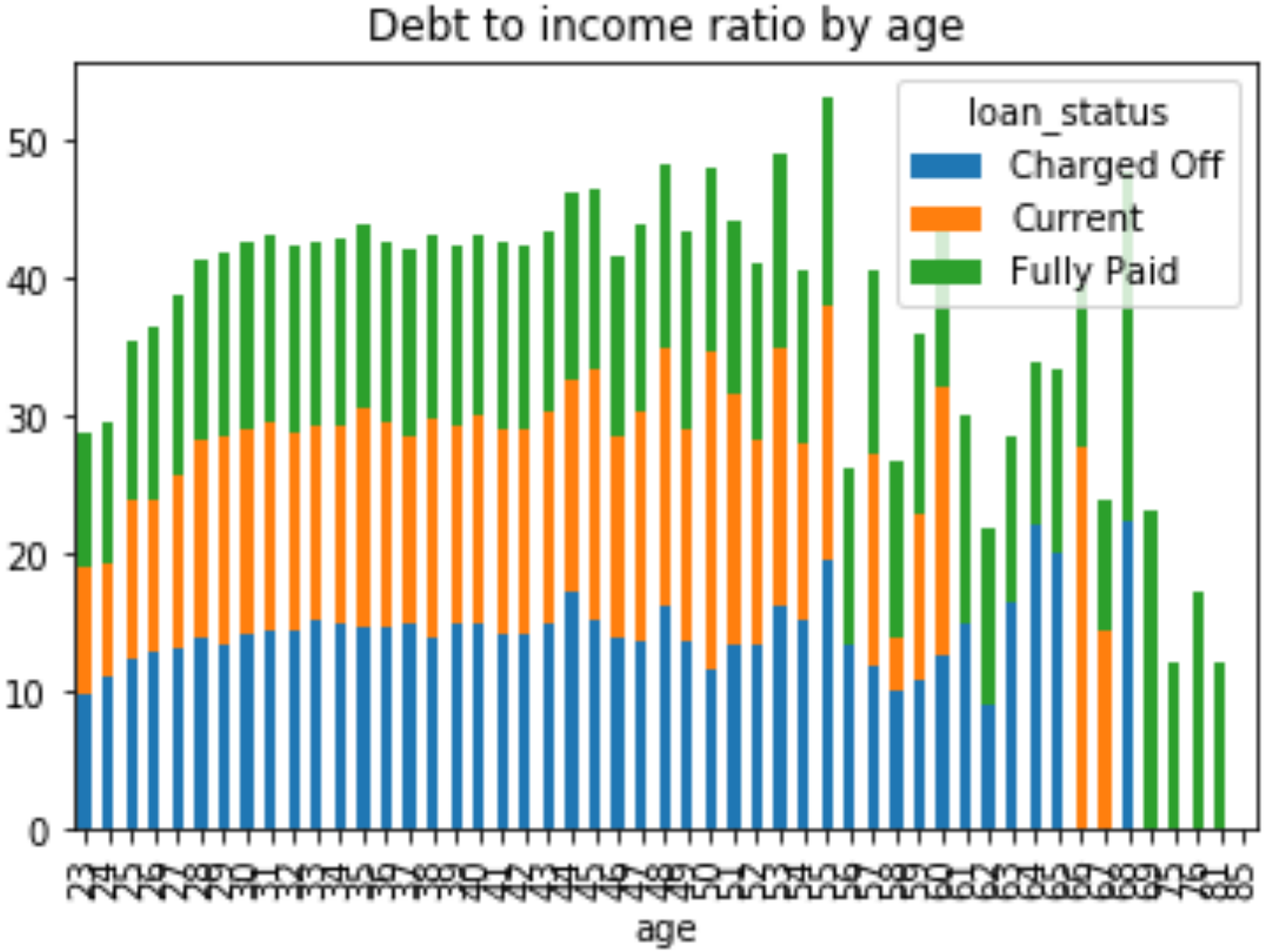
In 2011, the debt to income ratio is considerably higher than previous years.

Understanding all borrowers



Look like people on their 30s (based on assumption) take more loans. Majority were fully paid but also higher charged off loans are in this age group.

Understanding all borrowers

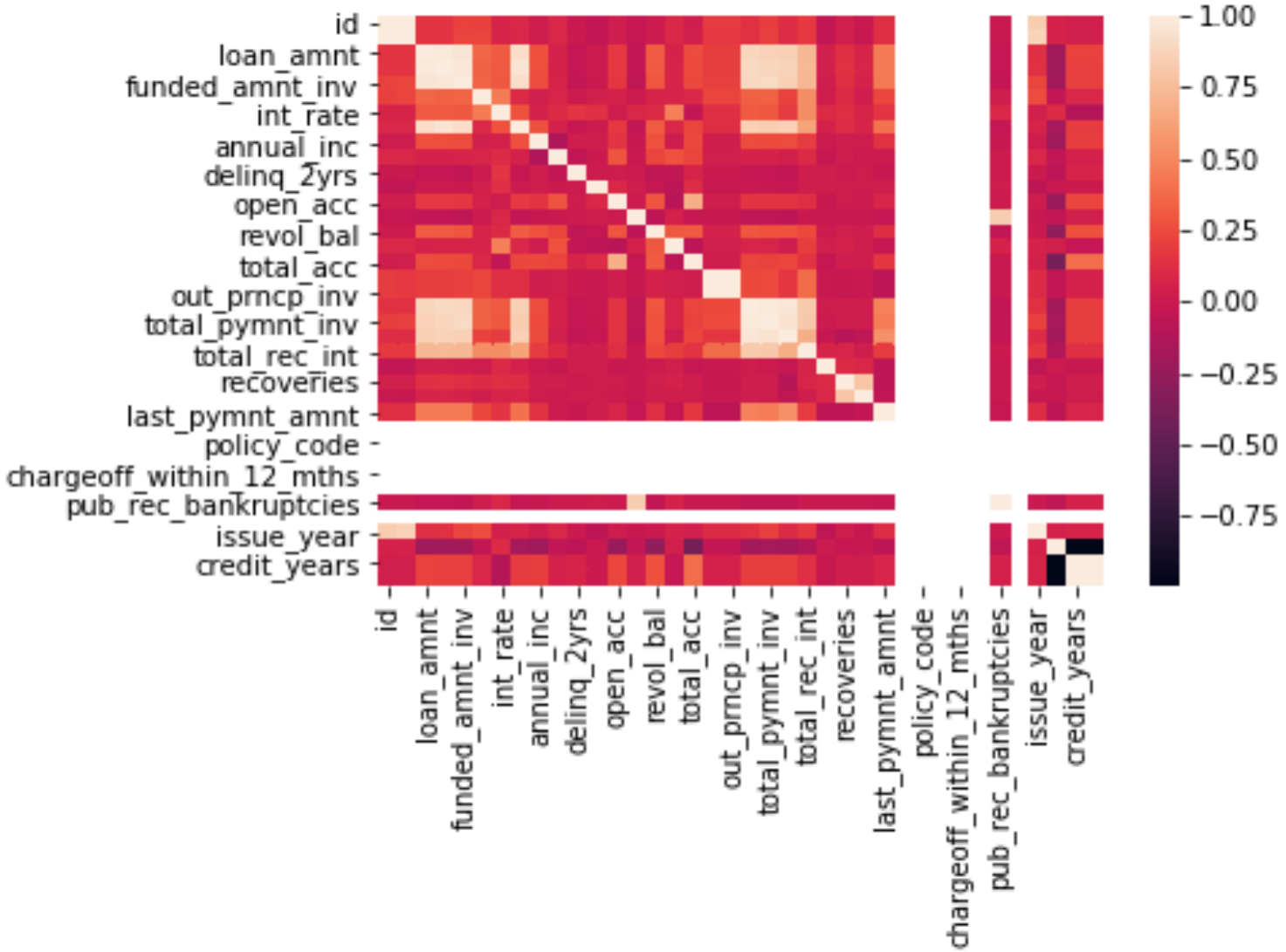


The higher the age, the most likely to be paid in full.

Understanding all borrowers

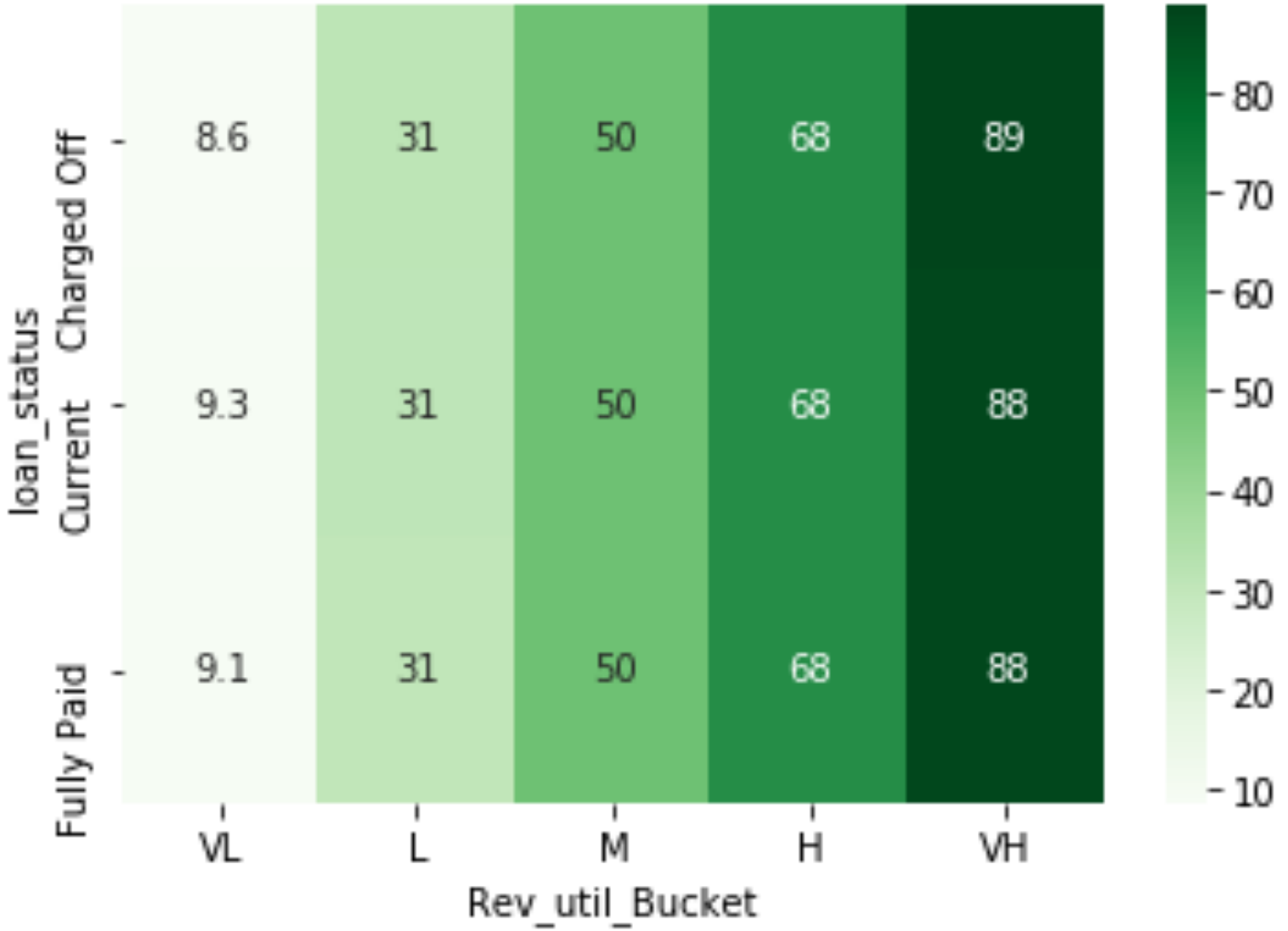
In states with less debt to income ratio, higher amounts of loans where fully paid.

Conclusions-Correlation Matrix



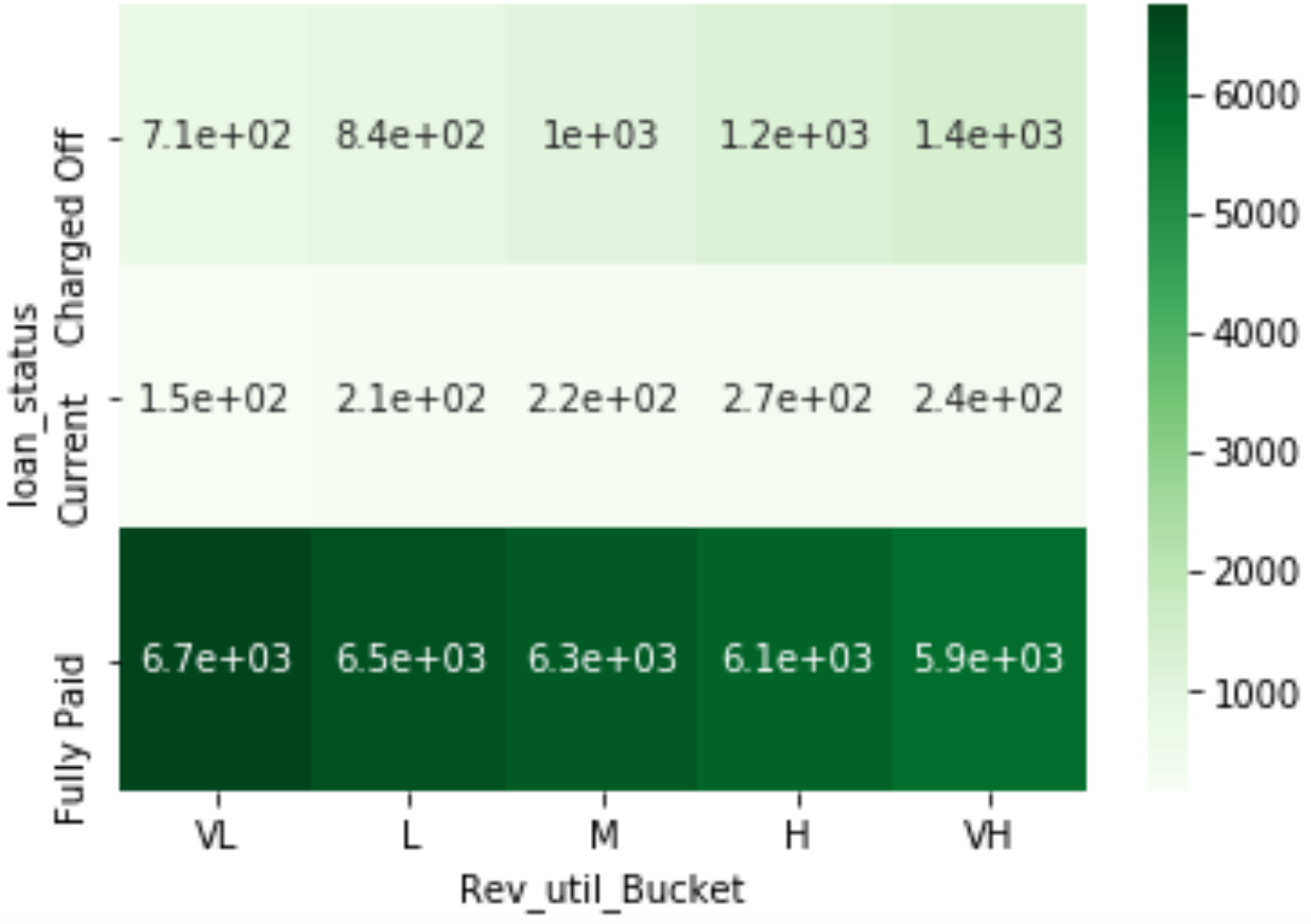
Using the correlation matrix, the only clear correlations are the loan amounts, Funded amounts, interest rate and installments. There is no clear correlation between variables that are not immediately related.

Conclusions-Revolving Utilization



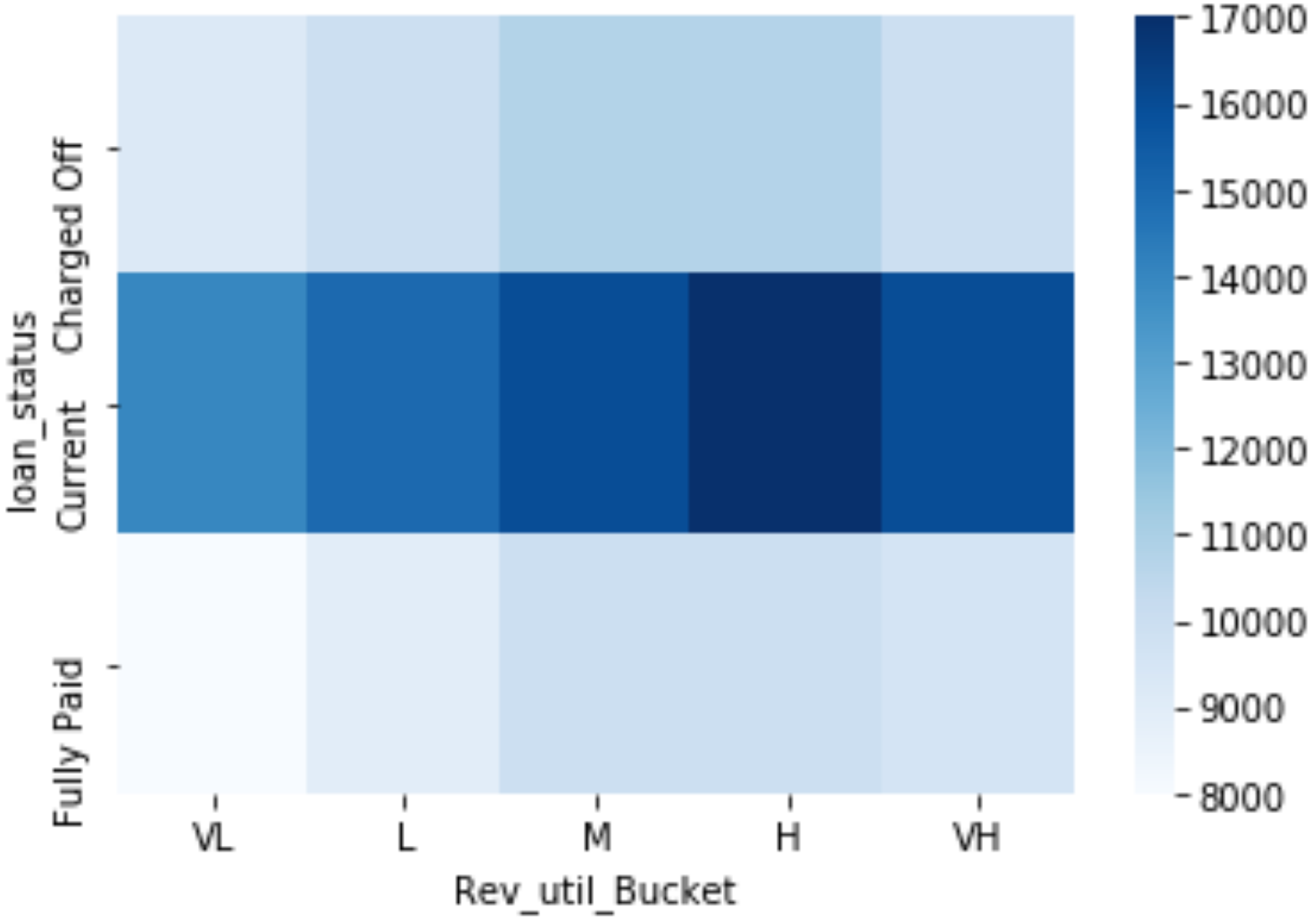
Credit card debt usage is not necessarily a predictor for defaulting on itself.

Conclusions-Revolving Utilization



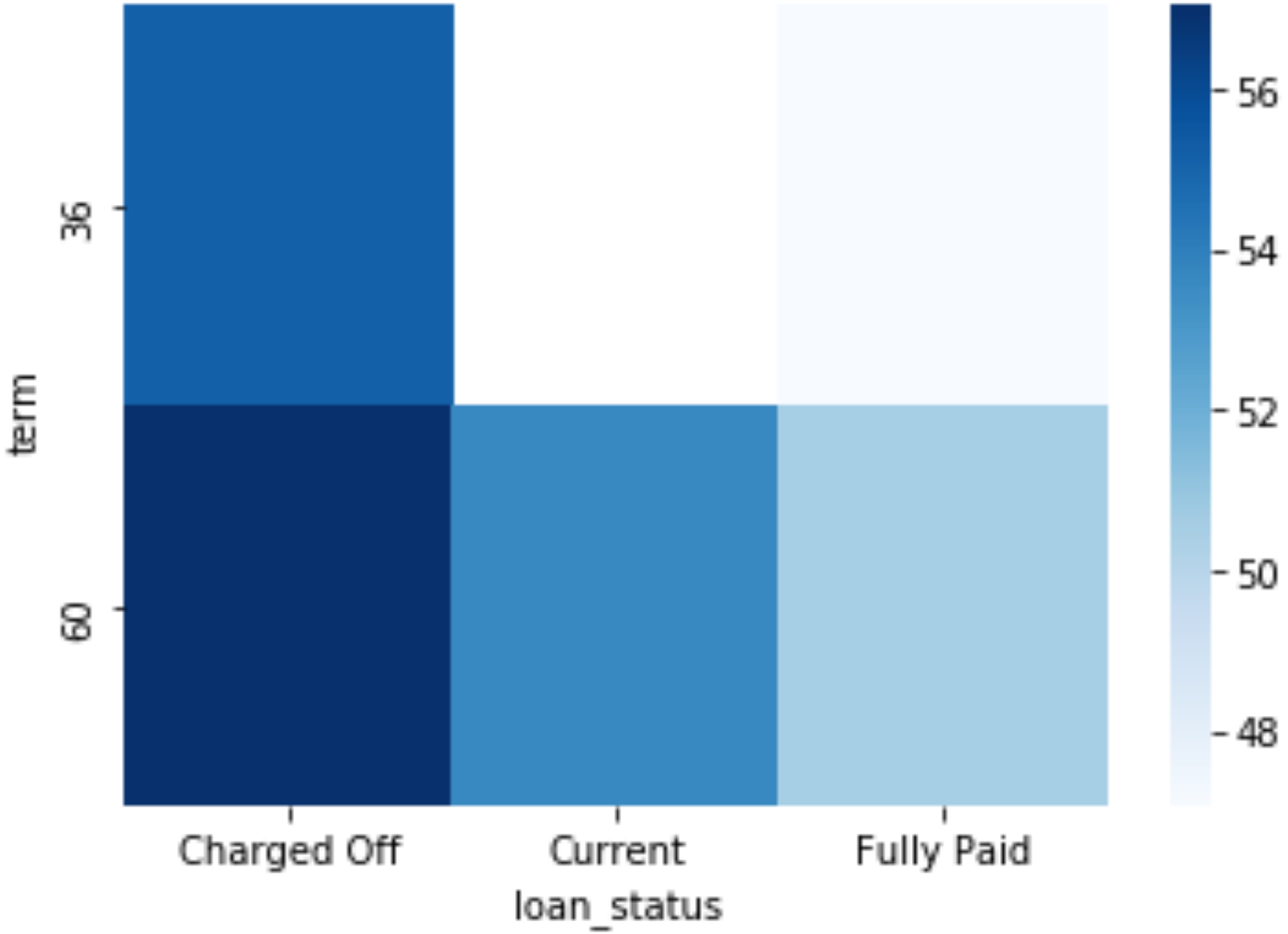
If we count the amount of loans charged of and how high the credit card usage is, we can see that the higher the usage, the likelihood of the loan being defaulted increases.

Conclusions-Revolving Utilization vs Loan Amount



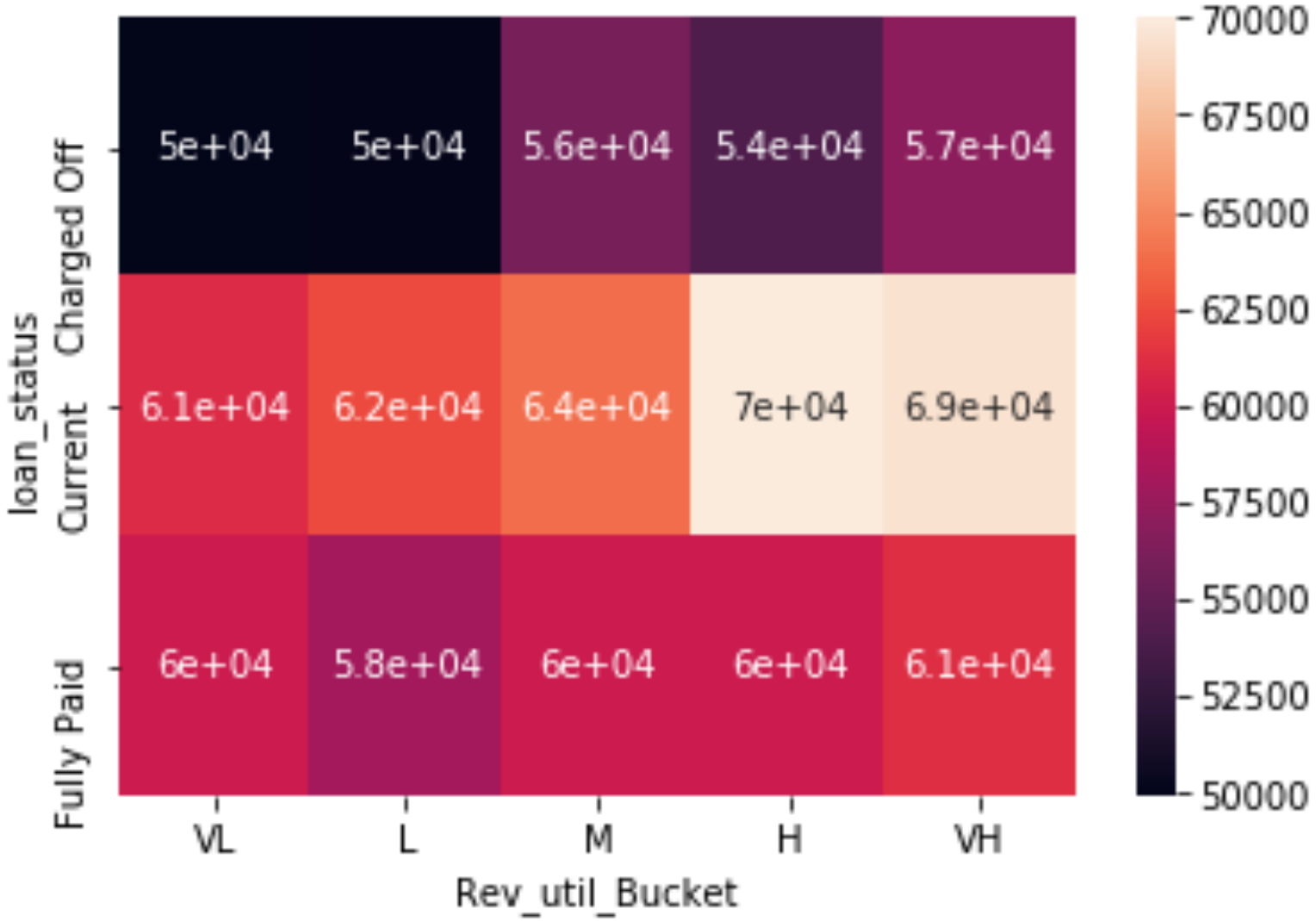
We can assume borrower's are most likely to default in smaller loans if their credit card utilization is very high.

Conclusions-Revolving Utilization vs Term



We can assume borrower's are most likely to default in 60 month loans if their credit card utilization is very high.

Conclusions-Revolving Utilization vs Annual Income



We can assume borrower's are most likely to default if their income is less than $60k regardless on how high their credit card utilization is.