LendingClub

Case Study

August 9, 2023

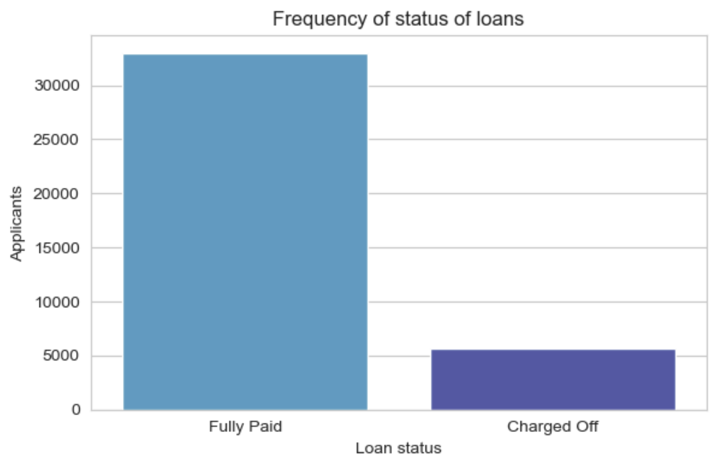
Team
- Moneca Mallidi
- Nagaraju Oruganti

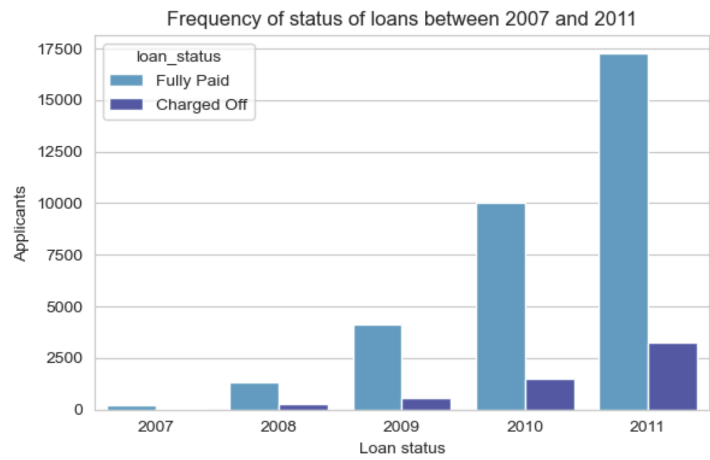
Problem Statement

LendingClub is a consumer finance company which specializes in lending various types of loans to urban customers.

The objective of the exploratory analysis to investigate variables driving the default rates.

Until 2010, the company reduced default rates to 12.9% but in the year 2011, the rate increased to 15.9%.





Data

The dataset is comprised of loan approved from 2007 to 2011. It included loan related attributes and the borrower attributes.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 39717 entries, 0 to 39716
Columns: 111 entries, id to total_il_high_credit_limit
dtypes: float64(74), int64(13), object(24)
memory usage: 33.6+ MB
```

The dataset contains 39,717 records with 111 features.

Pre-processing (Clean-up)

```
Features with NULL or NA values: 54 or (48.65%)
Features with constant value: 9 or (8.11%)
Features with missing more than 90% of values: 2 or (1.80%)
Shape of the dataset after removing null and constant value features (39717, 46)
```

```
Removed id and url columns because it is redundant to member_idUrl of all the record are same except they are
suffixed by id
```

```
behaviour_var =  [ "delinq_2yrs", "earliest_cr_line", "inq_last_6mths", "open_acc", "pub_rec", "revol_bal",
                  "revol_util", "total_acc", "out_prncp", "out_prncp_inv", "total_pymnt",
                  "total_pymnt_inv",
                  "total_rec_prncp", "total_rec_int", "total_rec_late_fee", "recoveries",
                  "collection_recovery_fee",
                  "last_pymnt_d", "last_pymnt_amnt", "last_credit_pull_d"]
```

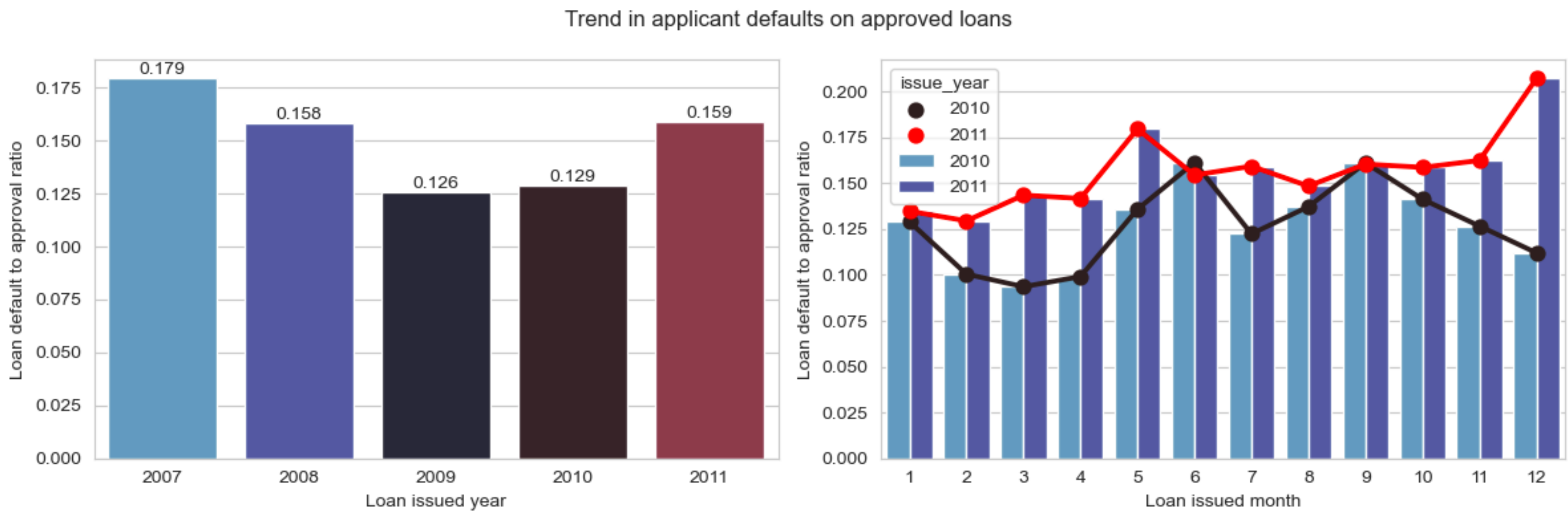
The features reduced from 111 to 24

Features by category

```
Indentifier:                   member_id
Target:                        loan_status
# of features:                 21
List of categorical features:   (10)['home_ownership', 'verification_status', 'purpose', 'title', 'addr_state',
'term', 'grade', 'sub_grade', 'emp_length', 'pub_rec_bankruptcies']
List of numerical features:     (8)['loan_amnt', 'funded_amnt', 'funded_amnt_inv', 'int_rate', 'installment',
'annual_inc', 'dti', 'mths_since_last_delinq']
List of text based features:    (2)['emp_title', 'desc']
```

Text based features are not considered in this investigative study

Loan Status (target)



Questions to explore

1. What factors influencing the defaults?
2. What factors drove the sharp increase in defaults in 2011?

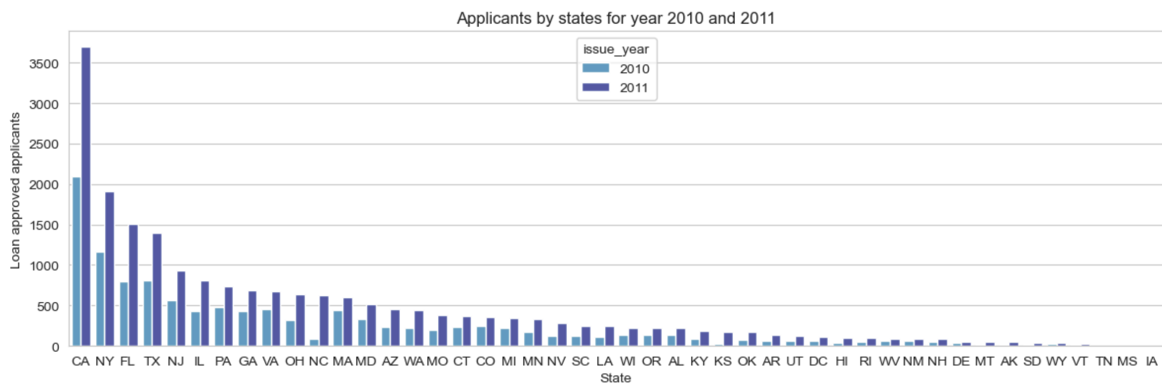
(What factors drove divergence in trend in the last quarter of the calendar year 2010 and 2011?) <- Not considered.

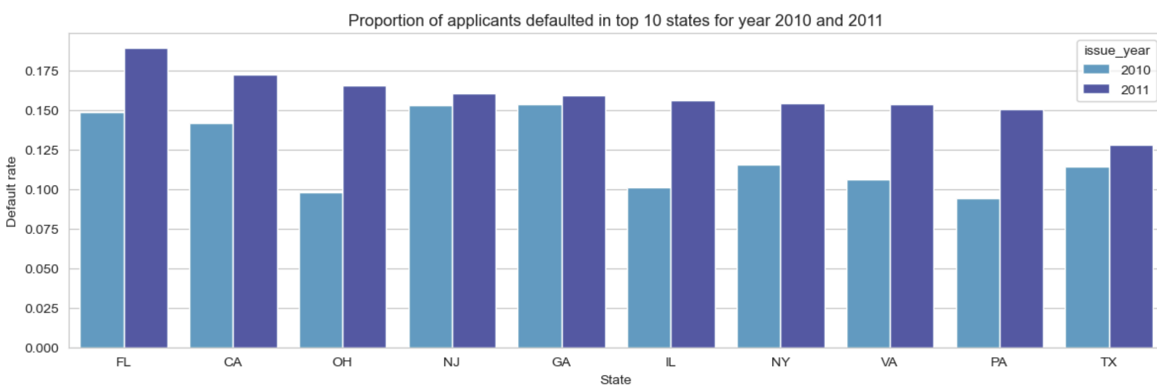
Categorical Features

```
(nominal) home_ownership      : (5)     ['MORTGAGE', 'NONE', 'OTHER', 'OWN', 'RENT']
(nominal) verification_status : (3)     ['Not Verified', 'Source Verified', 'Verified']
(nominal) purpose             : (14)
(nominal) title               : (19297)
(nominal) addr_state          : (50)
(ordinal) term                : (2)     [' 36 months', ' 60 months']
(ordinal) grade               : (7)     ['A', 'B', 'C', 'D', 'E', 'F', 'G']
(ordinal) sub_grade           : (35)
(ordinal) emp_length          : (12)
(ordinal) pub_rec_bankruptcies: (4)     [-1.0, 0.0, 1.0, 2.0]
(ordinal) issue_month         : (12)
(ordinal) issue_year          : (5)     [2007, 2008, 2009, 2010, 2011]
```

*Replace null with -1 for public record bankruptcies

Region

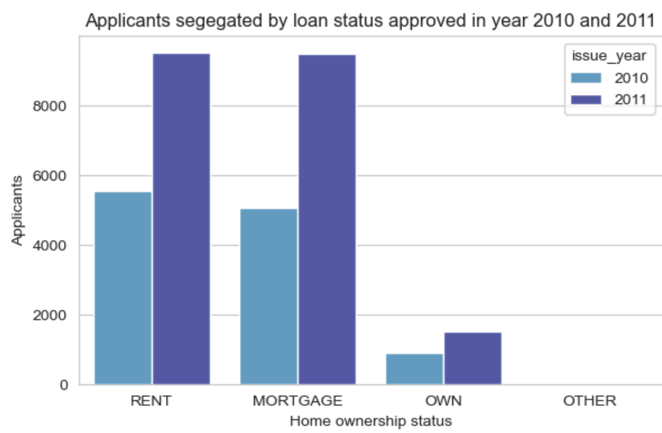


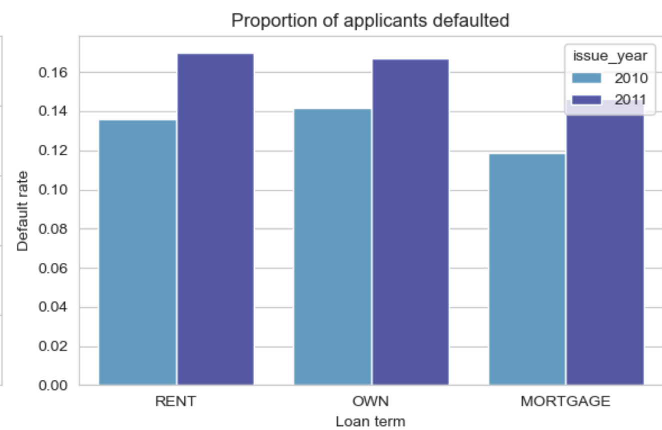


Majority of applicants from the wealthy states in the US including CA, NY, FL, TX, NJ, and IL. LeadingClub might have stated its operations in these area on gradually propagated over the country.

The marginal increase is defaults in the year 2011 is uniform across the top 10 states by applicants.

Home Ownership

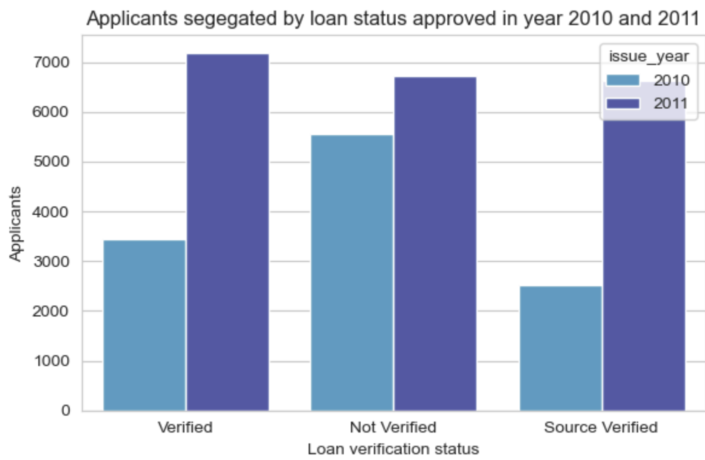


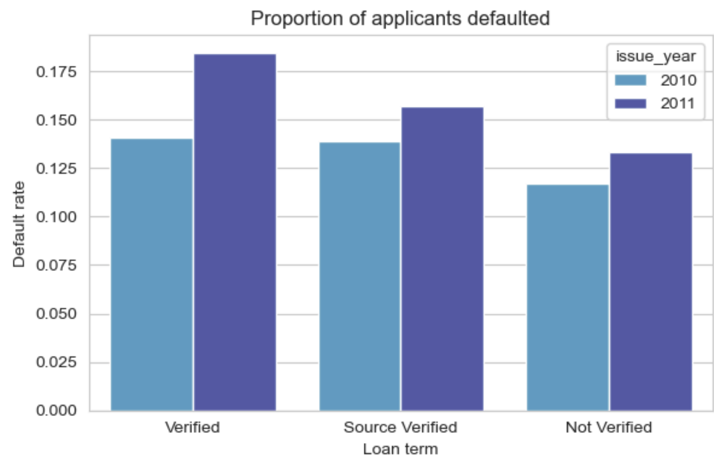


In the year 2010 and 2011, the OTHER category for home ownership is 2 and 1 respectively. It seems the field in an application form made mandatory to avoid OTHER category applicants.

The proportional increase in homeownership is fairly consistent between the years.

Verification Status



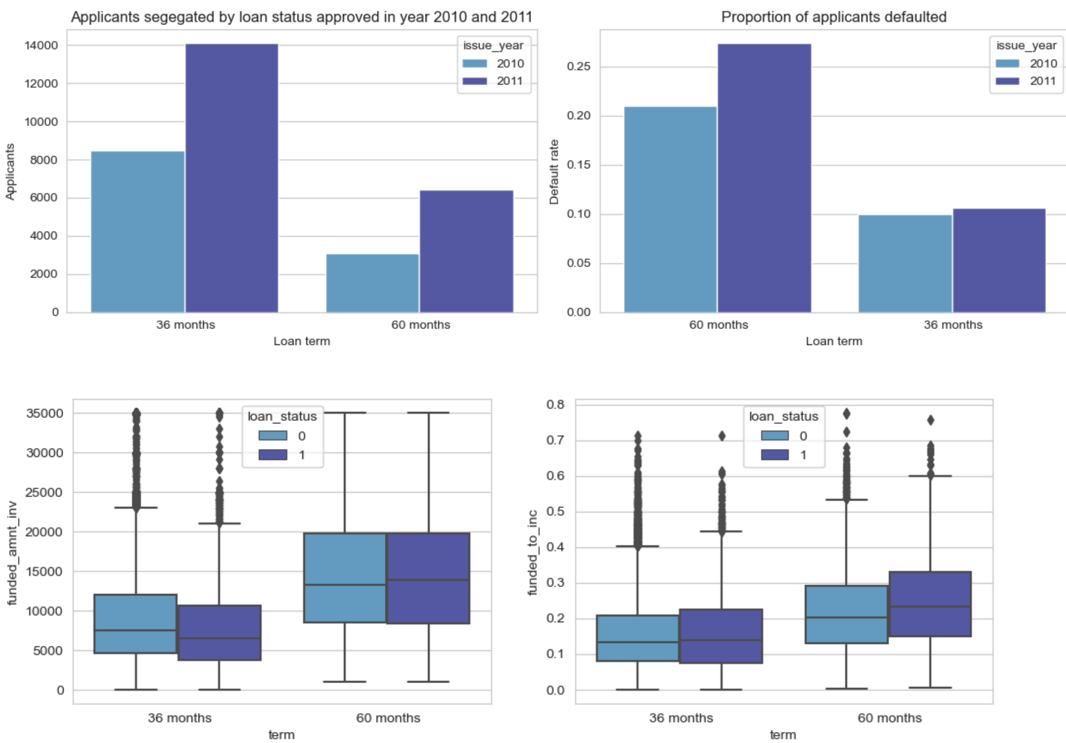


The default rate for those applicants' income verified by the business, and the income source verified is significantly higher than non-verified.

This observation is contrary to normal wisdom. Loans that are verified are expected to have low defaults than non-verified ones.

It is likely that the verification could be the aftermath of delinquency/default. In that case, the verification process should be integrated into the application process.

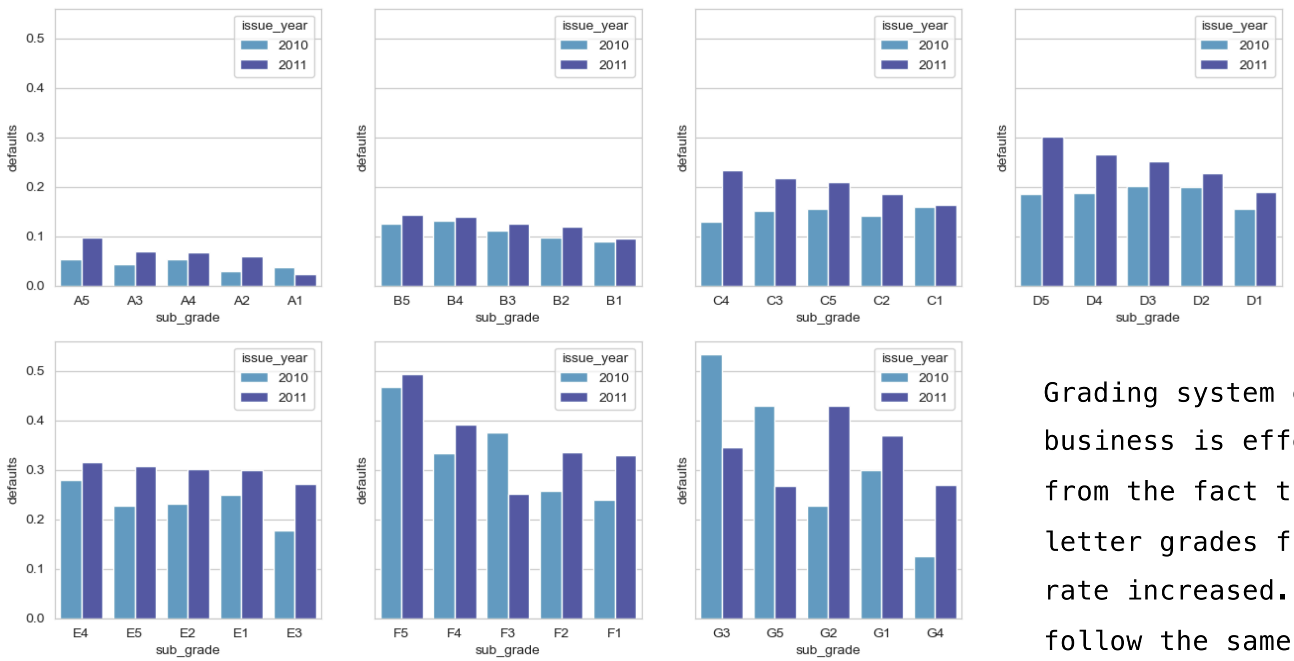
Loan Term (term)



Long-term loans of 60 months have higher defaults than 36 months.

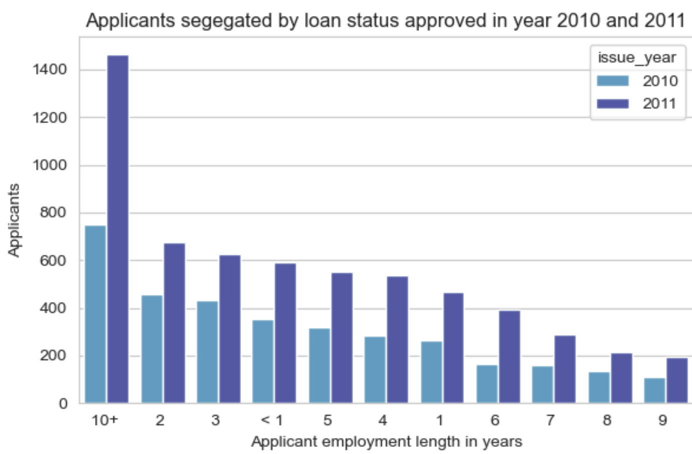
While those who took long-term loans defaulted when the funded amount is more than 0.2 of their income on average.

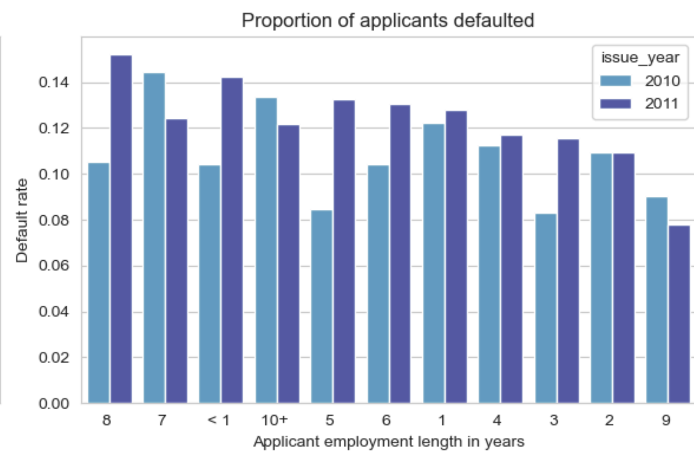
Grade / Sub-grade



Grading system employed by the business is effective. It is evident from the fact that with an increase in letter grades from A to G, the default rate increased. The sub-grades also follow the same trend.

Employment Length

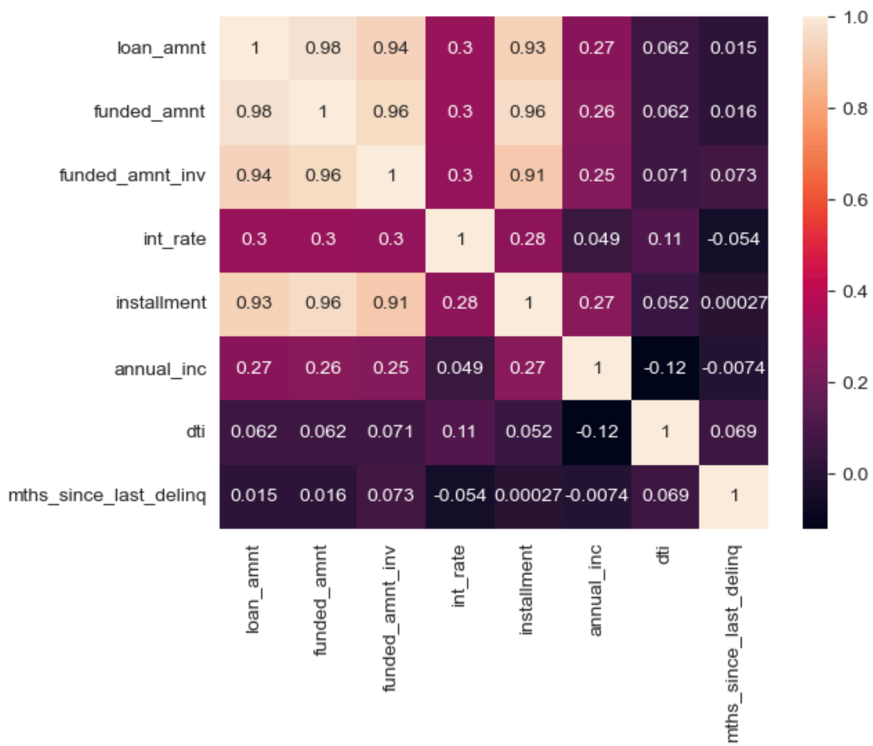





```
Large number of applicants with more than 10-years of experience got loan
approved and the increase in default in the group is slightly higher.

The large increase in applicants could be affect of great recession in
2008, that lasted until 2011.
```

Numerical Features



It is no surprise that loan amount, funded amount, funded amount inventory and installment are correlated strongly.

These feature are related.

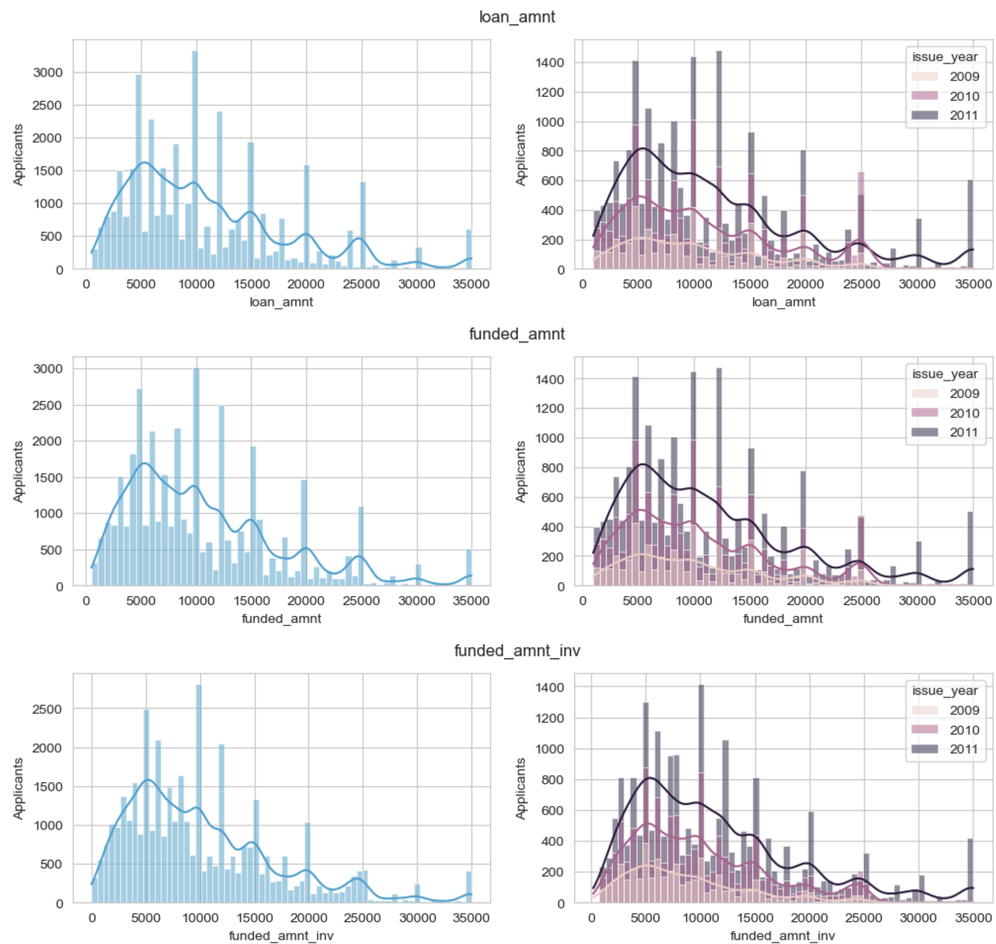
Amounts

- Loan mount: The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.

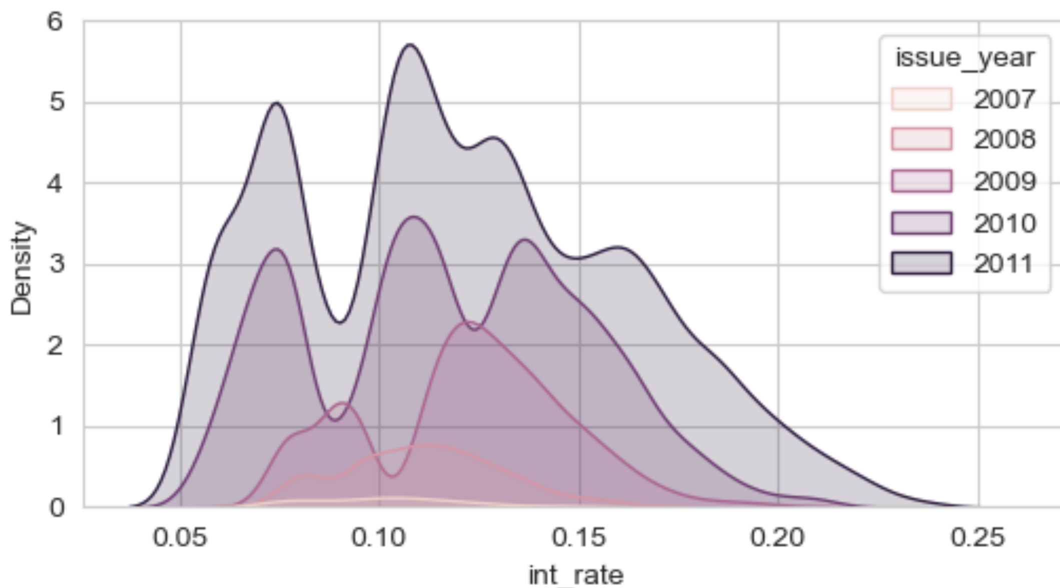
- Funded amount: The total amount committed to that loan at that point in time.

- Funded amount by investor: The total amount committed by investors for that loan at that point in time.

Large portion of loan amounts are rounded to nearest 1000 dollars.
There is clear spike on multiple of 5000 including 5k, 10k, 15k, 20k, 25k, 35k.
The distributions are right skewed.



Interest Rate

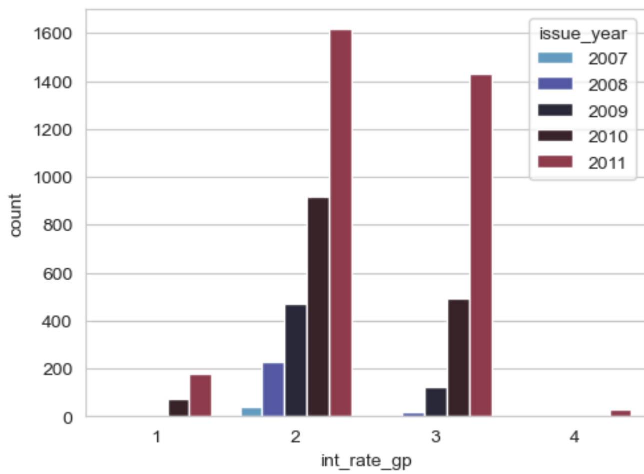


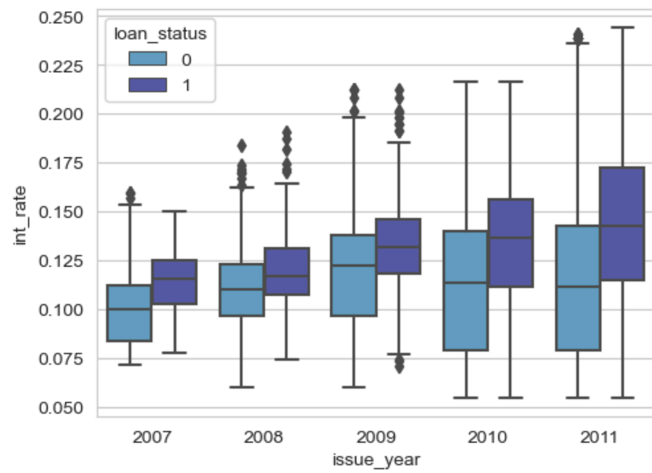
Interest rate charged to the borrower seems to follow bimodal distribution in 2009 and tri-model in 2010 and 2012.

The number of applicants who secured with higher interest rates (more than 20%) increased considerably in 2011.

Interest rate

Interest rate is binned with 0.075 interval

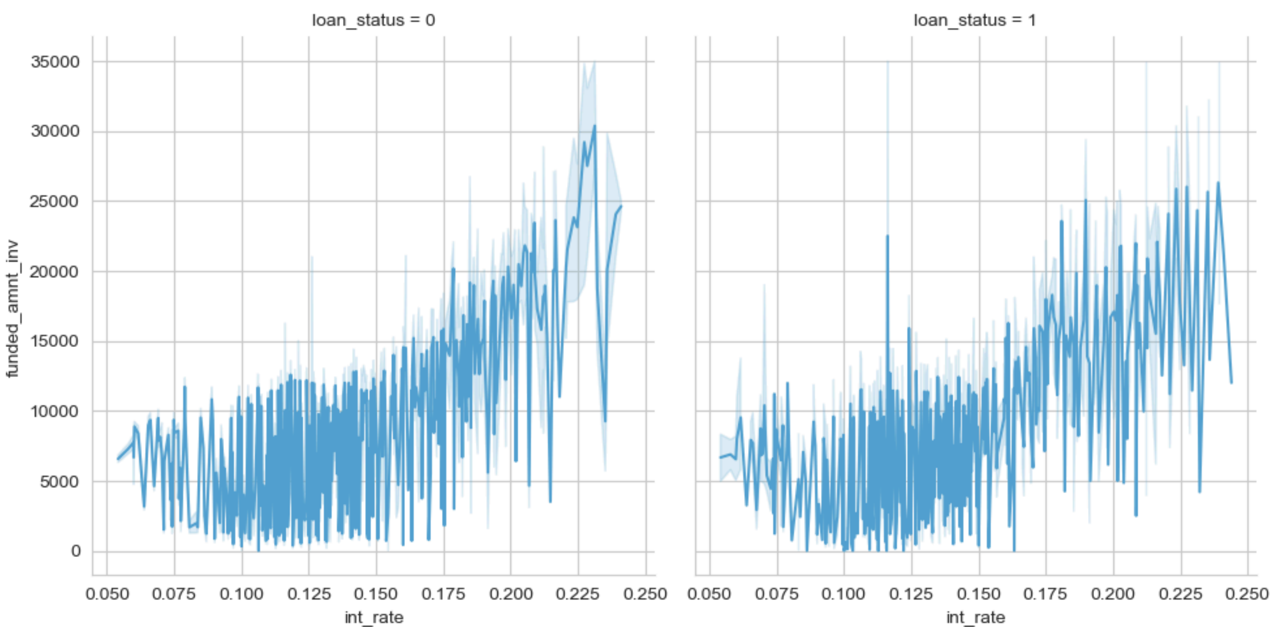




Interest rate charge to the defaulted borrowers is considerably higher.

It might be possible that the management aware of the risk by other factors such as credit score and charged higher interest rate. The median difference is as much as 5% in 2011.

Funded Amount vs. Interest rate

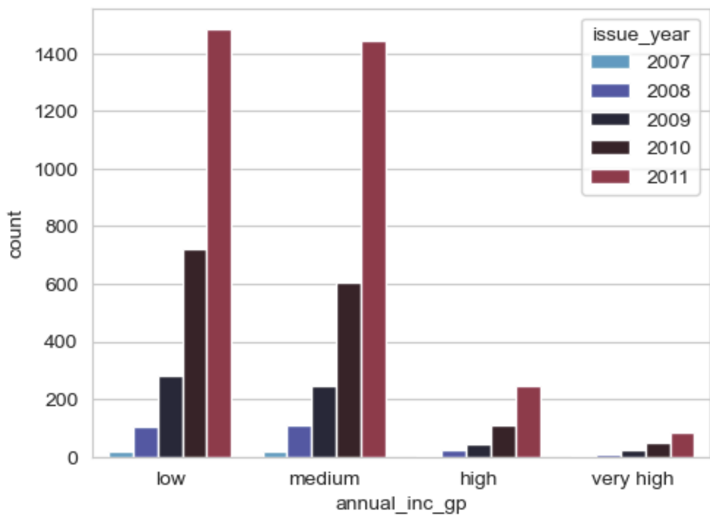


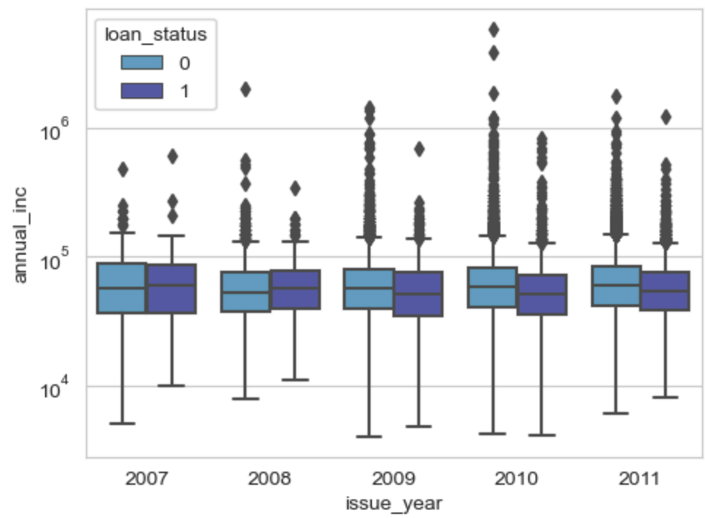
Non-linear relationship between interest rate and amount funded.

Irrespective of the borrower risk, the rate is higher when large amount of loan is funded.

Annual Income

Income is grouped by 50k interval.

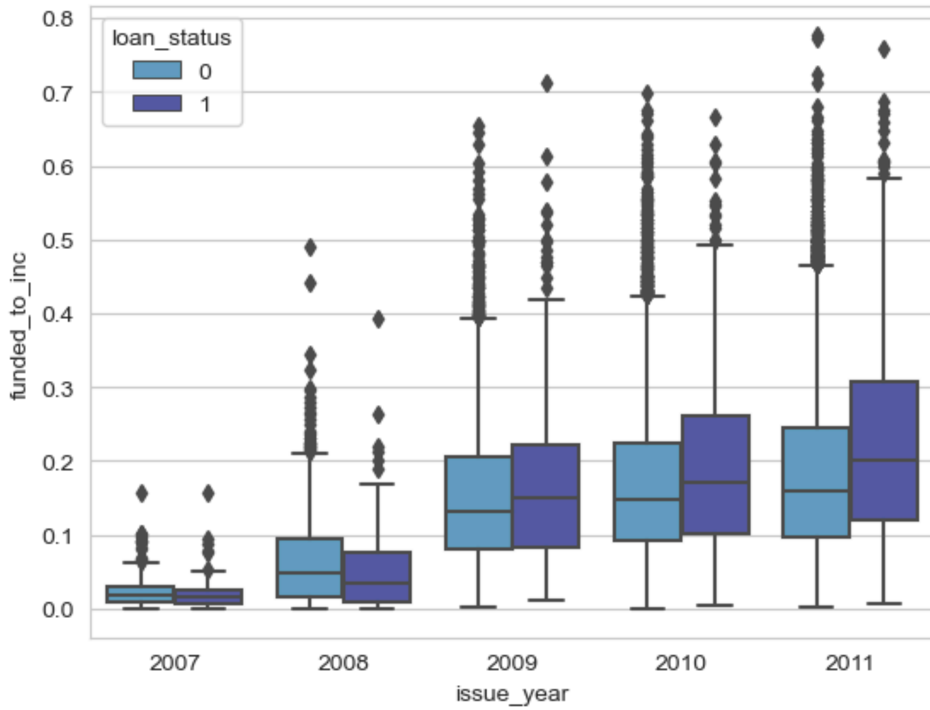




Borrowers who defaulted are with lower mean annual income. The difference is statistically estimated using t-test at 95% COI.

Funded Amount to Annual Income

```
df['funded_to_inc'] = df['funded_amnt_inv'] / df['annual_inc']
```



Funded amount to the borrowers income increased over the years.

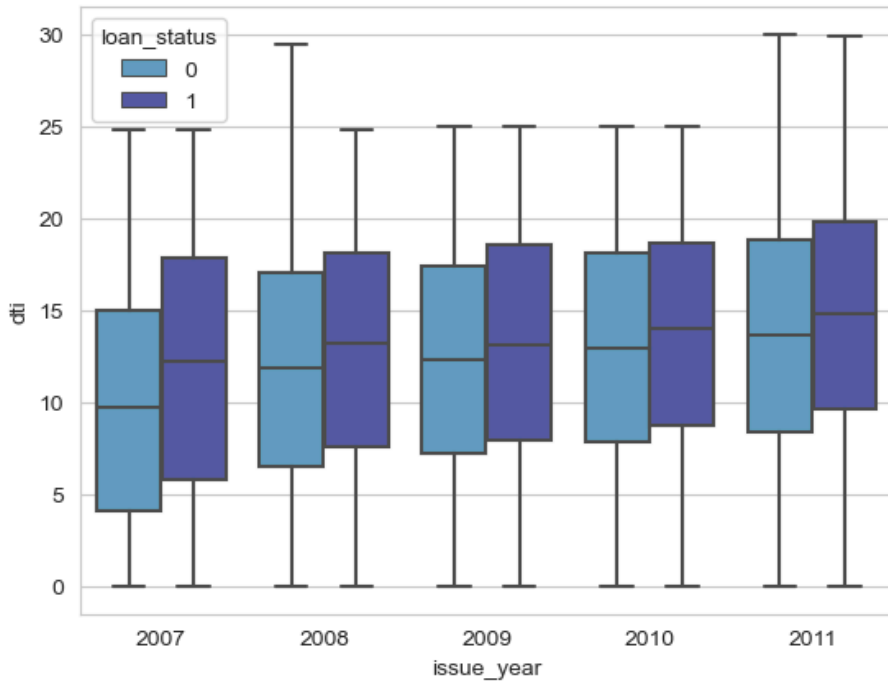
With reaching more customers between 2009 and 2011, the LeadingClub started giving loan their nominal 0.15 times for annual income to 0.2 times in 2011.

T-test also confirms that the mean between the applicants defaulted and the fully paid is statistically significant.

Some of the loans were as high as 80% of the annual income.

dti

```
dti : A ratio calculated using the borrower’s total
monthly debt payments on the total debt obligations,
excluding mortgage and the requested LC loan, divided
by the borrower’s self-reported monthly income.
```



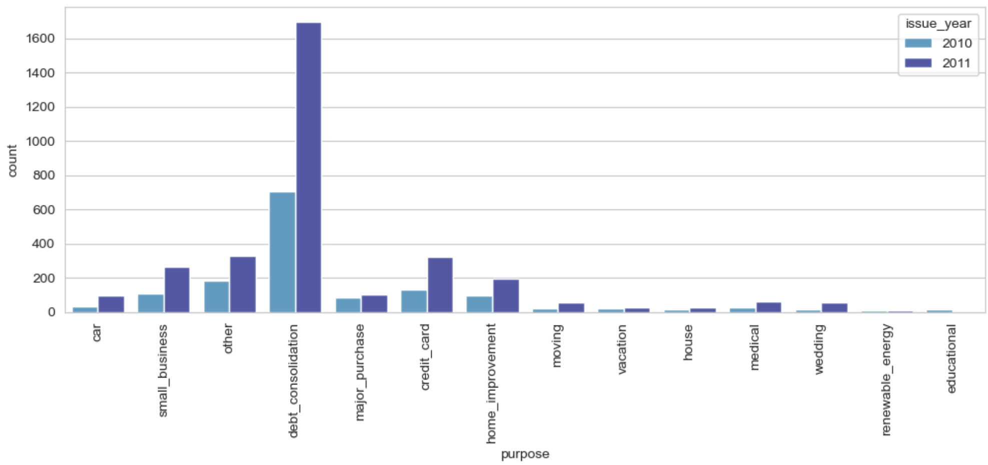
The observation is in sync with installments

The borrowers who default pays more debt payments to the obligation in comparison

The trend is moving up gradually, but not an alarming rate.

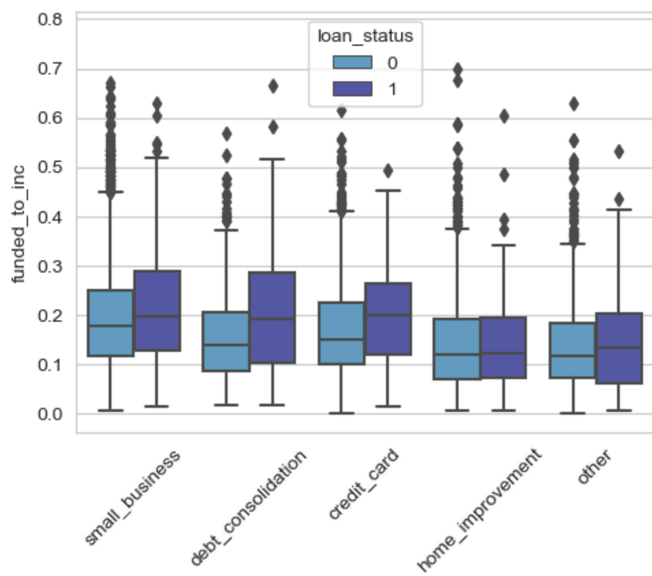
dti is more likely an companion or resulting variable of interest rate, installment, and risk in general.

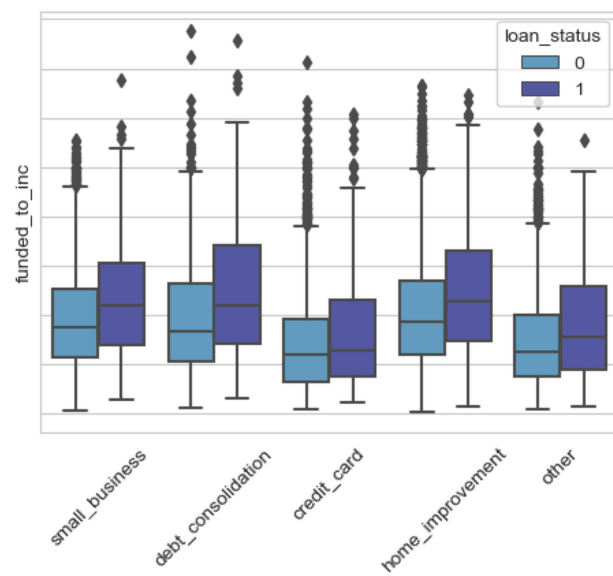
Purpose recorded by the borrowers



Applicants who took loans on the purpose for debt consolidation, small business, credit card, other, and home improvement, saw greater increase in defaults in 2011 compared to 2010. Medical and wedding purpose loans also saw increase in defaults.

Funded Amount to annual income by purpose





On average, the those defaulted applicants are funded
more for the annual income across the selected purpose

Summary (1 of 2)

The two main questions we investigated are

(1) Identify factors driving the defaults and

(2) Shift in the defaults from 2010 to 2011.

- The primary factor in the loan default is those applicants who secured loans on debt consolidation followed by small businesses, credit cards, home improvements, and others.

- Divergence between the mean funded to annual income ratio for the applicants' defaults is higher.

- Contrary to the above observation, it is observed that there exists a non-linear relationship between the interest charge and the funded amount irrespective of the riskiness of the default. Interest rates are as high as 25% funded in 2011, which is about 500bps higher than in 2010.

Summary (2 of 2)

- Over the years, with the business expansion, the risk appetite also increased. Until 2009, the funded amount to annual income ratio of the applicant was 0.15, and this increased to 0.2 in 2011.

- Grading system employed by the business is effective. It is evident from the fact that with an increase in letter grades from A to G, the default rate increased.

- The business implementation of the verification process can be improved. The default rate for those applicants' income verified by the business, and the income source verified is significantly higher than non-verified. This observation is contrary to normal wisdom. Loans that are verified are expected to have low defaults than non-verified ones. It is likely that the verification could be the aftermath of delinquency/default. In that case, the verification process should be integrated into the application process.