## Basic stats

# 300 paidoff, 100 collection, 100 paid off after collection

-- 60% pay off, 40% overdue(20% collection

# insight: 272 30-day terms, 207 15-day terms, 21 7-day terms

* most loan are made in 15(41.4%) or 30 days(54.4%) term

# what are the age, edu difference in each term?

* The group of term 7 consist of 76 % Male, 24 % Female
* The group of term 15 consist of 86 % Male, 14 % Female
* The group of term 30 consist of 85 % Male, 15 % Female
  + The Gender ratio stays close across group except of 7-day term
* The group of term 7 consist of 33% High School or Below, 19% Bechalor, 38% college, 10% Master or Above,
* The group of term 15 consist of 38% High School or Below, 15% Bechalor, 46% college, 1% Master or Above,
* The group of term 30 consist of 46% High School or Below, 11% Bechalor, 43% college, 0% Master or Above
  + With the longer term, the lower it is the ratio of people with higher education level

# what are the ratio of overdue in each type of loan term?

* The group of term 7 consist of 0 % COLLECTION, 5 % COLLECTION\_PAIDOFF, 95 % PAIDOFF
* The group of term 15 consist of 20 % COLLECTION, 19 % COLLECTION\_PAIDOFF, 61 % PAIDOFF
* The group of term 30 consist of 22 % COLLECTION, 22 % COLLECTION\_PAIDOFF, 56 % PAIDOFF
* The group of term & has 0% chance of collection and 95% of paying off before due.

# insight: the ratio of paidoff decreases as the term increases

* hypothesis: longer terms, higher probability of overdue

# insight: College, high School or below are extremely large, with bachelor lower, and master or above extremely small

* This resembles the base rate in the general population, but is it identical to it?

# insight: the effective date is centered around 11 Sep 2016

# insight: age is centered on 28-30

# Null Hypothesis: No significant difference in age distribution between different loan\_status

# insight: the distribution of PAIDOFF AND PaidOff after collection are similar,

# with college education being the dominant, and high school or below and bechalor comes after

# whilst master or above being extremely rare in all group, which resembles the base rate in general population

# but in COLLECTION group, high school or below education level becomes the dominant

* Hypothesis: The lower education level, the more likely to overdue

# insight: distribution of age is somewhat similar but there is no spike at around 35 in COLLECTION group

# insight: past due days are on average 36.01 days

# 29.3 standard deviations, quite large

# consider the fact that terms are mostly 15, 30 days, the past due days are quite long

# insight : if people pay defaulted loan, some pay it quickly (within 10 days of due date), others do not pay it until compulsory collection

# note that there is a small increase at 20 + for loan of 30-day term

# insight: number of male loan borrowers is way larger than number of female borrowers

# insight: males take up more portion of Collection and Paid After Collection

--> Hypothesis: Females tend to overdue less on loans

# does the date when the borrower borrowed the loan affects?

* In loans borrowed on 1 consist of 52 % COLLECTION, 18 % COLLECTION\_PAIDOFF, 31 % PAIDOFF
* In loans borrowed on 2 consist of 66 % COLLECTION, 32 % COLLECTION\_PAIDOFF, 3 % PAIDOFF
* In loans borrowed on 3 consist of 100 % COLLECTION, 0 % COLLECTION\_PAIDOFF, 0 % PAIDOFF
* In loans borrowed on 4 consist of 100 % COLLECTION, 0 % COLLECTION\_PAIDOFF, 0 % PAIDOFF
* In loans borrowed on 5 consist of 100 % COLLECTION, 0 % COLLECTION\_PAIDOFF, 0 % PAIDOFF
* In loans borrowed on 6 consist of 33 % COLLECTION, 33 % COLLECTION\_PAIDOFF, 33 % PAIDOFF
* In loans borrowed on 7 consist of 41 % COLLECTION, 15 % COLLECTION\_PAIDOFF, 43 % PAIDOFF
* insight: people borrow less on Tue,Wed,Thu,Fri, most on Sun and Mon
* but poeple who borrowed on Tue,Wed,Thu will almost never overdue their loan

# Preliminary conclusions

1. The longer the term of loan, the higher the chance of the borrower not repaying the loan before due date.
2. People with higher education level tend to pay back loan more on time.
3. Females tend to pay back loan more on time.
4. Age does not seem to matter much on how likely a borrower is to pay back loan later then due date.
5. People who borrowed on Tue, Wed, Thu have a extremely low probability of not repaying on time.

## Suggested Actions

* Take into consideration factors of loan term, gender, education level, the day in week when the borrowing happens when considering the interest rate and approval of a loan.

## Problems with the analysis and further actions

* Look into behaviour of borrowers not repaying debt if they do not pay it within 10 days of being put into collection.
* Perform hypotheses testing on causal relationships of gender, education level, loan term, day of borrowing to whether the borrower will pay back on time.
* College, high School or below are extremely large, with bachelor lower, and master or above extremely small. Compare this fact with the population base rate, and see if anything interesting comes up.