

--	--	--	--	--	--	--

# Loan Status Analysis part 1

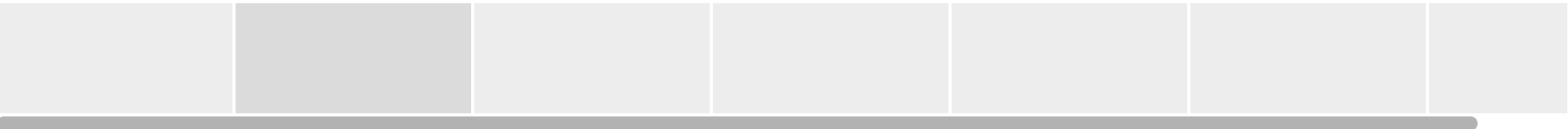
by

Tae Hyon Lee

For

Feng Tang  
Monjaco

5/31/2017



# Content

- 1. Introduction ..... 2
  - Background ..... 2
  - Purpose ..... 2
  - Assumptions ..... 2
  - Questions ..... 2
- 2. Analysis ..... 3
  - Probability ..... 3
  - Status A, B, C ..... 4
  - Additional Analysis ..... 5
- 3. Summary ..... 6
  - Findings ..... 6
  - Methodologies ..... 6

--	--	--	--	--	--	--

# Introduction

## Background

LoanStatus.csv file is given. It has the history of status of 55 loans from period 0 to period 11. LoanStatus.csv has the following variables:

- 1. ID:** Unique loan identifier
- 2. Period:** Time period represented by 0, 1, 2, 3..., n. Example: if 1 means May 2014, then 2 means June 2014.
- 3. Current Status:** A label indicating the status of a loan in the current period.
- 4. Next Status:** A label indicating the status of a loan in the next time period.

## Purposes

The purpose of this analysis is to analyze past loan data to create a prediction model on status of loans over time.

## Assumptions

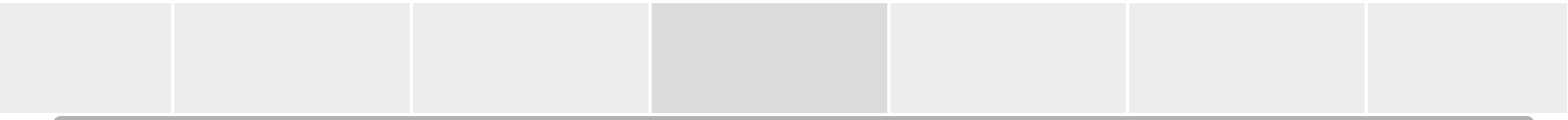
- The loan status analysis is based on 55 loans from period 0 to 12.
- We assume that the data is randomly selected from real life data.
- We also assume that the loan is made from a loaner who makes reasonable decisions, meaning the loaner will only lend the money to those who are financially fit based on background, credit score, etc.

## Limitations

Some of the analysis and the findings in the coming pages may **NOT** statistically significant and are only a visual interpretation. There is some A/B testing but t-test or chi-square test is not conducted. There is just too few data to do any kind of statistical test on it. So we may not use this analysis to make any confident decisions.

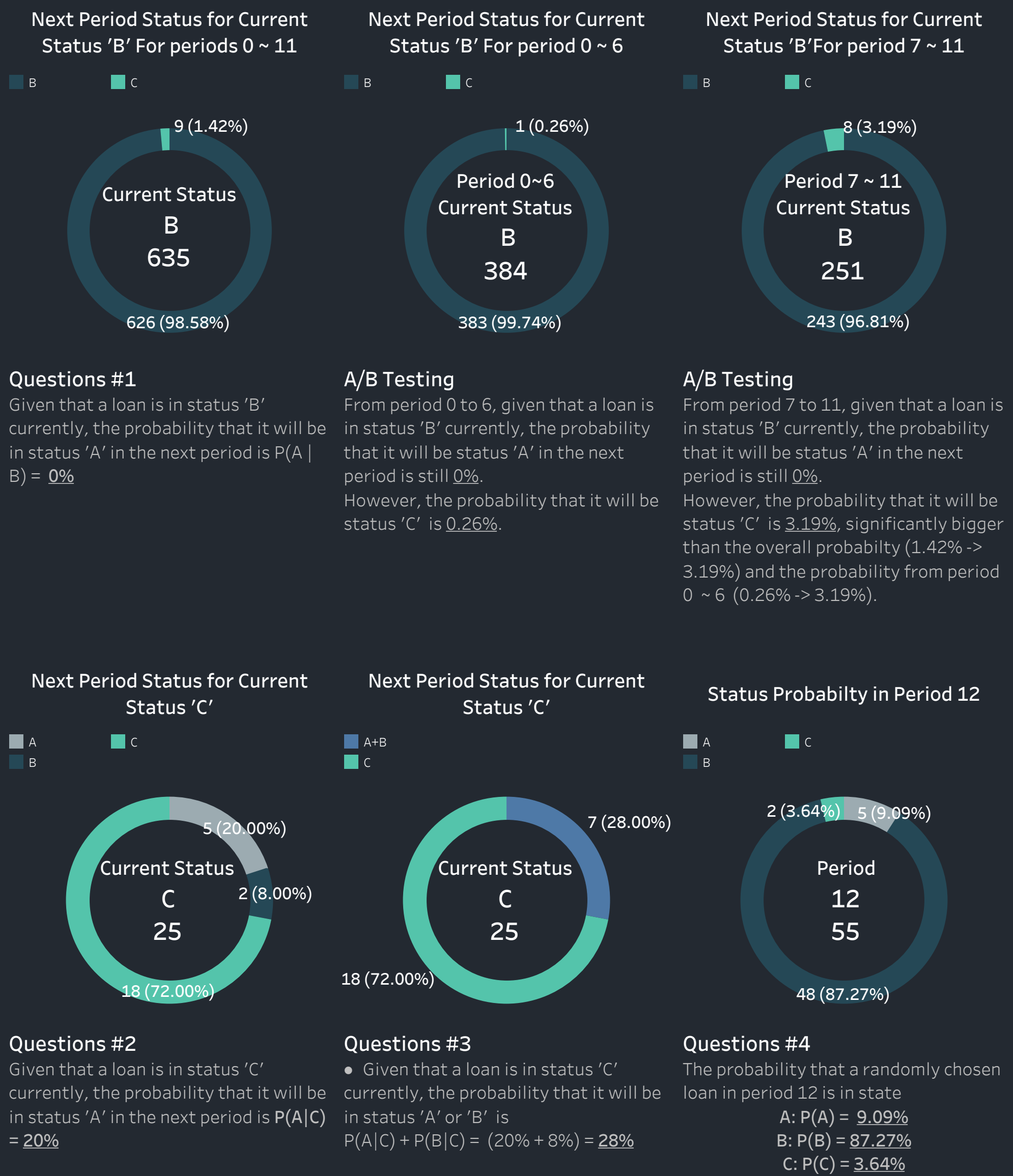
## Questions

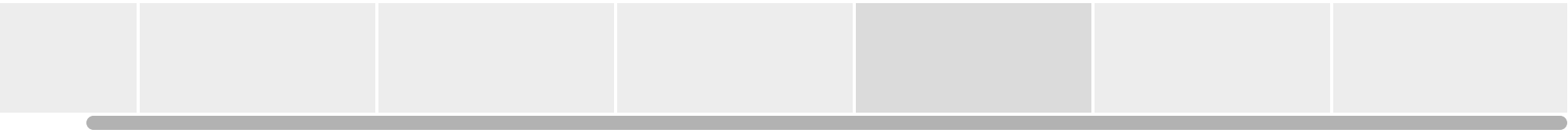
1. Given that a loan is in status 'B' currently, what's the probability that it will be in status 'A' in the next period?
2. Given that a loan is in status 'C' currently, what's the probability that it will be in status 'A' in the next period?
3. Given that a loan is in status 'C' currently, what's the probability that it will be in status 'B' or 'A'?
4. What's the probability that a randomly chosen loan in period 12 is in state A, B, and C respectively?
5. The labels A B C represent payment status of loans: 'paying on time', 'late', and 'default' (severely late) in real life. Can you guess which one-letter label belongs to what actual status?



# Analysis

## Probabilities





Status A, B, C

After analyzing the loan status data, I was able to find out what status belongs to each one-letter label:

Status 'B'	Status 'C'	Status 'A'
<p>B is 'paying on time' for the following reasons:</p> <ul style="list-style-type: none"><li>Assuming that the loaner makes reasonable decisions on who to loan the money, it is likely that the borrowers will pay back the loans on time at least for the first few periods .</li><li>It is highly likley that most of the loans throughout all twelve periods are paying on time, and B is the most common status.</li></ul>	<p>C is 'late' for the following reasons:</p> <ul style="list-style-type: none"><li>Status 'C' is the second most common status in the data. It is more likely that in a real life situation that there are more loans tgat are late than the loans that are severely late.</li><li>Status 'C' only happens towards the later periods (periods after 7), and the number of status C increases as period increases. In a real life situation, it is possible that borrowers become financially unstable over time and therefore make 'late' payments.</li><li>The loans that have current status of status 'C' can change to either status 'A' or status 'B'. It makes sense that 'C' is 'late' because it is between 'paying on time' and 'severely late'.</li></ul>	<p>A is 'severely late' for the following reasons:</p> <ul style="list-style-type: none"><li>It is very unlikely for a borrower who is 'paying on time' to make payments 'severely late' the next period. Status 'B' (Paying on time), is never turned into status 'A' (Severely late).</li><li>Status A only happens in period 12, the next period of current period 11. If a borrow becomes financially unstable and makes late payments, there is a chance that the financial situation becomes worse and therefore makes even more late payments in real life situation. Shown on the graphs from the previous page, 20% of the current status 'C' turns into status 'A' the next period.</li></ul>

Loan Analysis

--	--	--	--	--	--	--

Additional Analysis

Loan Status from Period 0 to 7

Next Status	Period / Current Status									
	0	1	2	3	4	5		6	7	
	B	B	B	B	B	B	C	B	B	
B	55 100.00%	55 100.00%	55 100.00%	55 100.00%	54 98.18%	54 100.00%	1 100.00%	55 100.00%	51 92.73%	
C	0	0	0	0	1 1.82%	0		0	4 7.27%	
Grand Total	55 100.00%	55 100.00%	55 100.00%	55 100.00%	55 100.00%	54 100.00%	1 100.00%	55 100.00%	55 100.00%	

- From period 0 to 6, all of 55 loans pay on time except for one loan on period 5 but it pays on time for the next two periods.

Loan Status from Period 7 to 11

Next Status	Period / Current Status							
	7	8		9		10		11
A			0		0		0	5 71.43%
B	51 92.73%	50 98.04%	0	47 94.00%	0	47 100.00%	1 12.50%	48 100.00%
C	4 7.27%	1 1.96%	4 100.00%	3 6.00%	5 100.00%	0	7 87.50%	0
Grand Total	55 100.00%	51 100.00%	4 100.00%	50 100.00%	5 100.00%	47 100.00%	8 100.00%	48 100.00%

- In period 8, 4 loans had late payments.
- In period 9, 5 loans, including the 4 from last period, had late payments.
- In period 10, 8 loans, including the 5 from last period, had late payments.
- In period 11, 7 loans, all of the loans that late payments except for one, had late payments. The one loan that changed paid on time.
- In period 12, out of the 7 loans that had late payments last period, 5 of them were severely late and 2 of them were late again.

Loans that had late payments next period

ID	Period											
	0	1	2	3	4	5	6	7	8	9	10	11
BX9199244	B	B	B	B	C	B	B	B	B	C	C	A
BXX4829XX	B	B	B	B	B	B	B	C	C	C	C	A
EX9219217	B	B	B	B	B	B	B	C	C	C	C	A
EXX981917	B	B	B	B	B	B	B	B	B	C	C	A
FX6240117	B	B	B	B	B	B	B	C	C	C	C	A
BX6220117	B	B	B	B	B	B	B	C	C	C	C	C
BX726X317	B	B	B	B	B	B	B	B	C	C	C	C
CX716X317	B	B	B	B	B	B	B	B	B	C	B	B

- 5 of the 8 loans that ever had late status from period 0 to 11 became severely late status in period 12 (period 11’s next status).
- 2 of them stayed late status in period 12.
- 1 of them paid on time until period 12 after one late status in period 11.

Loans that had severely late payments next period

ID	Period											
	0	1	2	3	4	5	6	7	8	9	10	11
BX9199244	B	B	B	B	C	B	B	B	B	C	C	A
BXX4829XX	B	B	B	B	B	B	B	C	C	C	C	A
EX9219217	B	B	B	B	B	B	B	C	C	C	C	A
EXX981917	B	B	B	B	B	B	B	B	B	C	C	A
FX6240117	B	B	B	B	B	B	B	C	C	C	C	A

- Before loan turns into status A, severely late, loan has to be late at least twice for the past two periods.
- The longer you’ve been consecutively late for the past couple of payments, the more likely uoi are going to be severely late.
- There has not been a case where a loan was on time and the next period, the loan was paid severely late.

--	--	--	--	--	--	--

## Summary

### Findings

#### Paying on Time

- From period 0 to 11, when a loan has a current status of 'B' paying on time, there is 1.42% chance that the next period loan will have a status of 'C' late payment.
- From period 0 to 6, there is 0.26% that loan that paid on time will pay late in the next period.
- From period 7 to 11, there is 3.19% that loan that paid on time will pay late in the next period.
- Starting from period 8, there is an increasing number of loans that used to pay on time but started to pay late over time.
- This difference between the two sets of periods is significant in the prediction model in that we can expect the payments to be made on time for the first 6 months, then we will experience more frequent late payments over time.

#### Late

- Loans that pay its payment late have a 72% chance that it will pay late again in the next period.
- There is only 8% chance that it will go back to paying on time and 20% that it will be severely late for the next payment.
- Loans that stayed late status for two consecutive periods never returned back to pay on time status.
- From this finding, it is important for a loaner to watch carefully of those loans that pay late; it is highly likley that the borrowers are experiencing financial difficulties and continue to pay late. Some even pay severely late which can be a sign that the borrower may not be able to pay back the money.

#### Severely Late

- It is very important to note that out of 8 loans that ever had late payments, 5 of them ended up paying severely late in period 12, suggesting that loan payments that are paid late are likely due to financial difficulties of the borrower and the borrowers will not likely recover from the difficulties.
- Also, those loans that had severely late status on period 12 had at least two late payments for two periods, so it is safe to assume that loans that never had a late status before will likely not have a severely late payment next period.
- The observation of severely late payment on period 12 can hint that the loans will not only have more frequent late payments, but also more severly late payments over time.

### Methodologies

- Checked the file format and found that the data provided, LoanStatus.csv, has 3 categorical variables (Period, Current Status, Next Status) and one label variable (ID).
- Imported the file to Tableau and checked for any corrupt data.
- Used Tableau to calculate the frequencies and found out the probabilities of each possible outcomes. Ex. Probability of next status 'B' given current status 'C'.
- The probabilities and frenquencies were displayed in table format or on donut chart for visual interpretation.
- The probabilities were also filtered by periods to find patterns and predict outcomes.
- Some loans that had different observations were analyzed to find the causes of certain patterns.