



# Statistical Analysis

---

Jonathan Hilgart

# Lending Club Data

Lending Club is a peer-to-peer marketplace for buying debt. Investors can set criteria for loans (annual inc. , FICO score, length of employment...over 74 features in total). The data is composed of over 800k loans from 2007-2015.

## 1) Loans were grouped into

- a) Default/Charged off (more than 120 days without a payment)
- b) Paid off

## 2) Main focus areas:

- a) Predicting the chance of a person paying off their loan
- b) Investigating default rates across states
- c) Modeling which factors have the largest impact on interest rates for loans
- d) Understanding the number of loans paid off over time

# Will I be able to pay off my Galvanize loan?

- Started with 59 features (that would be available at the onset of a loan)
- Ended with 13 features that predicted the chance a person will pay off loan
- Used GLM with logistic link function to predict binary 'Paid Off' columns.
  - 1 = Loan was paid off
  - 0 = Loan was defaulted/charged off
- Surprising results
  - Loan grade is not significant for determining percent chance of paying loan
  - Employment length is not significant for determining percent chance of paying loan



# Will I be able to pay off my Galvanize loan?

```
probability()
```

What is the size of loan you want to take out? 50000

What interest rate is your loan at? (Write as 13.8, not .138) 7

What is your annual\_inc? (current or future) 100000

How much are your monthly debt payments? (credit/ debt, do not include mortgage) 2700

How many 30+ days notices of delinquencies have you had in the past two years? 0

How many credit inquiries have you had in the last six months (not including auto and mortgage)? 1

How many months since your last credit delinquency? 100

How many open credit lines do you have? 3

What percent of your total credit available are you using? (Write as 30.1, not .301) 30

How many credit lines do you have in your credit history? 5

How many collections (for credit) have you completed in the past 12 months? 0

Is your FICO score above 660 (1=yes 2=no) 1

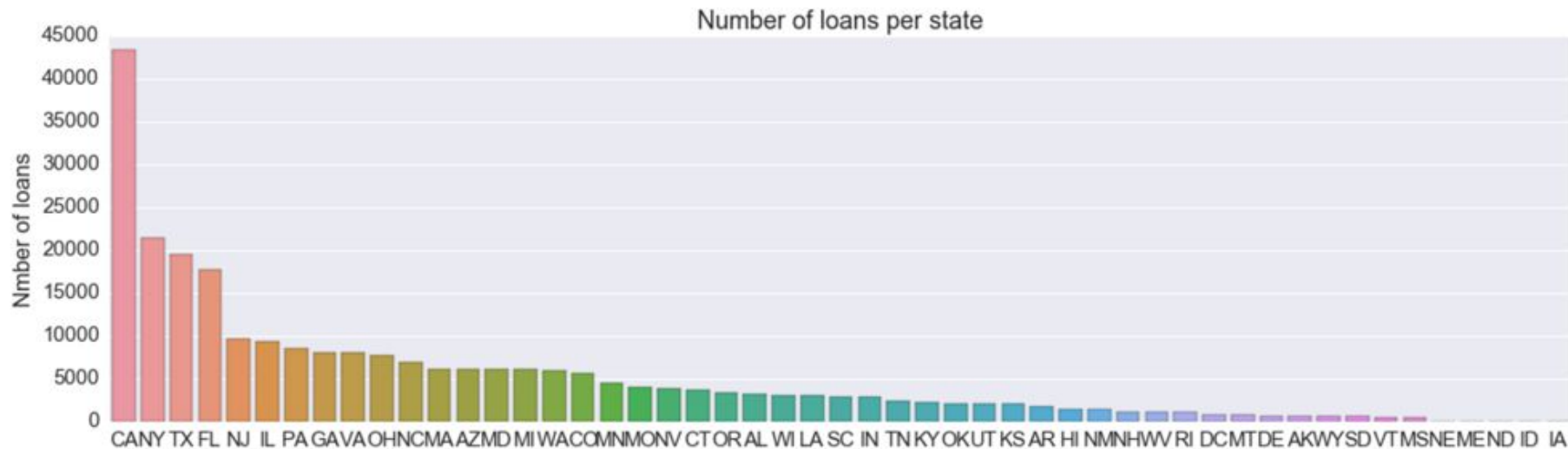
What is the current balance of all your accounts (savings, investments ..etc)? 20000

You have a [ 0.90907899 ] chance of repaying your loan.



# Default Rates Across States (Default & Paid loans)

- Do the ratios of paid off loans vary across states?
- $H_0$  = all states have the same ratio for percent of total loans that have been paid off.
- $H_a$  = states do not have the same ratio for the percent of total loans that have been paid off (Note: This does not say anything about individual states).
- Method: Chi square test



# Default Rates Across States (Default and Paid)

## Steps

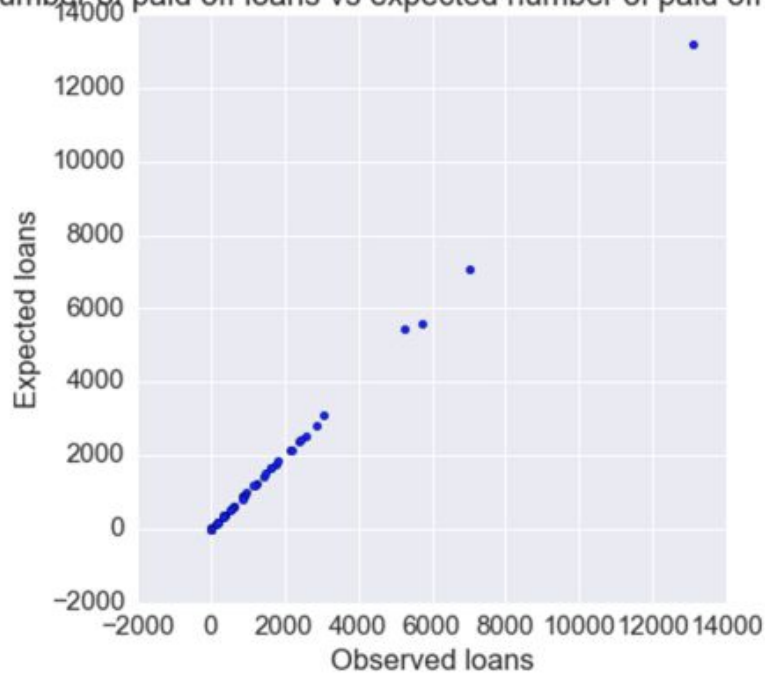
- Remove loans that originated in the years 2013 or later (keep 2007-2012).
- Determine national percentage of loans that were paid off for these years
  - $\text{sum of paid off loans} / \text{count of total loans} = \mathbf{84.3\%}$
- The total number of paid off loans per state = Observed
- The total number of loans per state times national paid off percent = Expected

```
observed = (state_count_of_paid.count_of_paid_off_loans)
expected = (state_cout_of_total_loans.total_number_of_loans) * national_default_rate
stats.chisquare(observed,expected)
```

```
Power_divergenceResult(statistic=31.129246937648936, pvalue=0.97821205630224073)
```

# Default Rates Across States (Default and Paid)

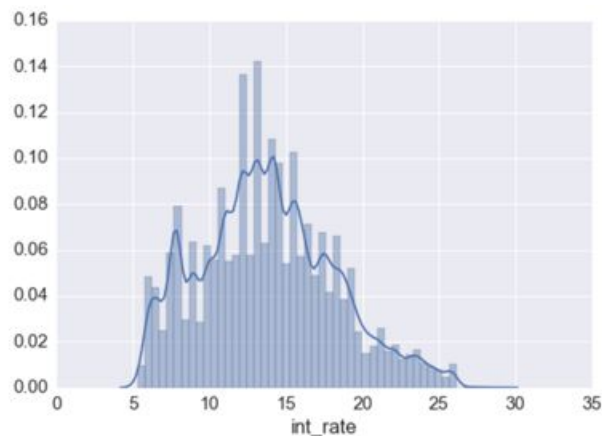
Observed number of paid off loans vs expected number of paid off loans (2007-2012)



- *Strong linear relationship. Minimal difference between observed and expected.*

# Predicting Interest Rates (Default and Paid loans)

- Can we predict what a potential borrower's interest rate will be?
- Steps
  - Remove highly correlated features
  - Remove insignificant coefficients
  - Remove features that we would not have at the onset of a loan (i.e. total payments, loan status)






# Predicting Interest Rates - Results

- Results
  - 45%  $R^2$  and adjusted  $R^2$
- What would my interest rate be?

```
In [728]: model14.predict([40000,100000,15,0,1,100,100,3,0,0,30,5,0,100,1,0,1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,1])
Out[728]: array([ 16.22741939])
```

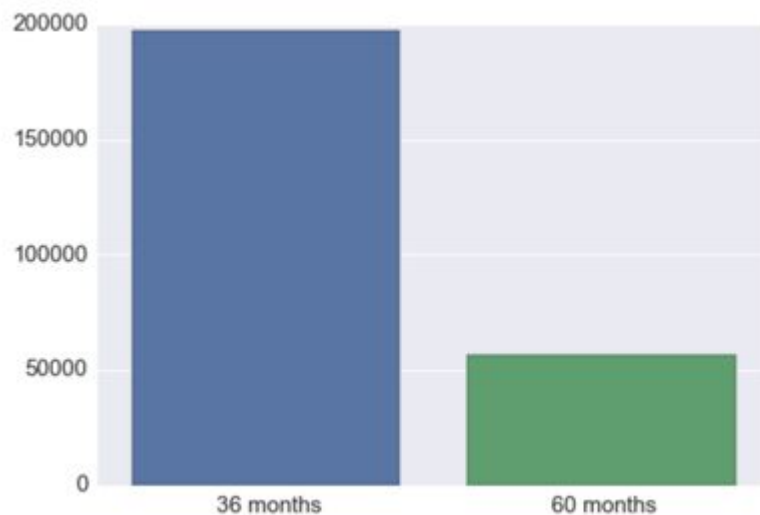
- Compare to Lending Club

<b>\$29,875</b> Loan Amount	<b>\$1,079.91 (36 payments)</b> Monthly Payment	<b>17.99%</b> Interest Rate	<b>22.51%</b> APR 
<a href="#">Get Loan</a>			

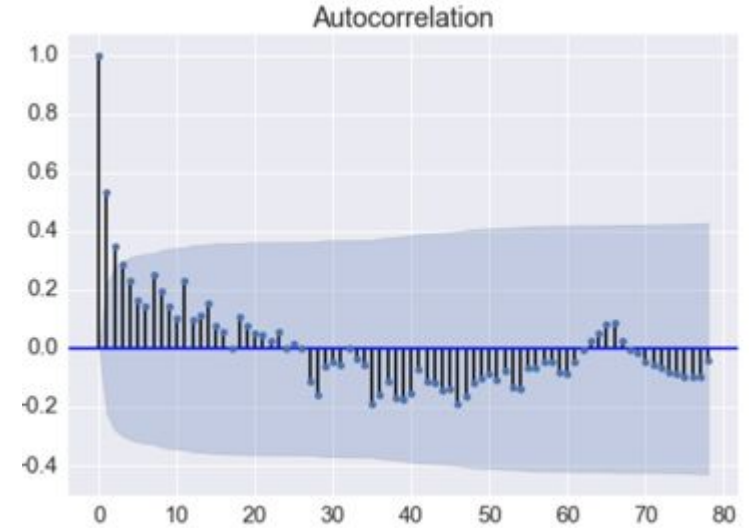
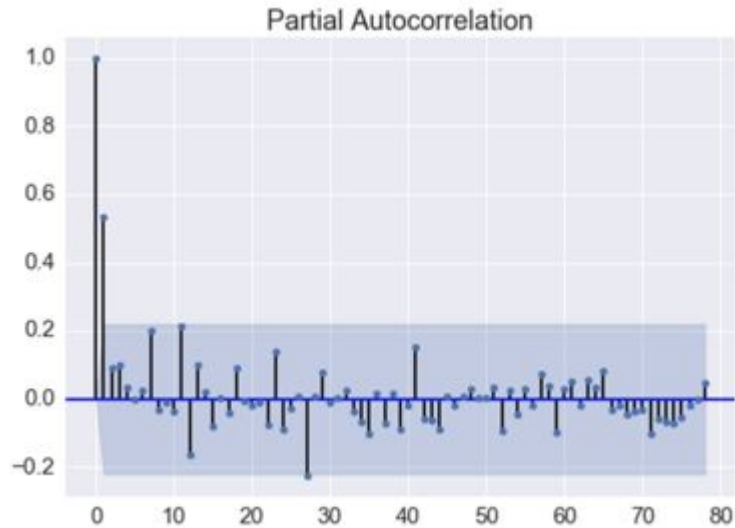
	coef	std err	t	P> t	[95.0% Conf. Int.]
loan_amnt	6.033e-05	9.94e-07	60.682	0.000	5.84e-05 6.23e-05
annual_inc	-4.088e-06	1.34e-07	-30.460	0.000	-4.35e-06 -3.83e-06
dti	0.0406	0.001	42.509	0.000	0.039 0.042
delinq_2yrs	0.5994	0.009	66.983	0.000	0.582 0.617
inq_last_6mths	0.9379	0.006	151.612	0.000	0.926 0.950
mths_since_last_delinq	0.0117	0.000	35.584	0.000	0.011 0.012
mths_since_last_record	0.0053	0.000	13.928	0.000	0.005 0.006
open_acc	0.1083	0.002	57.549	0.000	0.105 0.112
pub_rec	0.7006	0.024	29.692	0.000	0.654 0.747
revol_bal	-1.65e-05	4.02e-07	-41.035	0.000	-1.73e-05 -1.57e-05
revol_util	0.0637	0.000	220.186	0.000	0.063 0.064
total_acc	-0.0550	0.001	-69.352	0.000	-0.057 -0.053
collections_12_mths_ex_med	1.1593	0.073	15.952	0.000	1.017 1.302
mths_since_last_major_derog	0.0150	0.000	38.708	0.000	0.014 0.016
policy_code	6.5291	0.041	159.013	0.000	6.449 6.610
acc_now_delinq	2.0277	0.108	18.846	0.000	1.817 2.239
term_36 months	-4.1742	0.017	-243.291	0.000	-4.208 -4.141
emp_length_1 year	-0.5040	0.041	-12.152	0.000	-0.585 -0.423
emp_length_10+ years	-0.4680	0.035	-13.333	0.000	-0.537 -0.399
emp_length_2 years	-0.4748	0.039	-12.095	0.000	-0.552 -0.398
emp_length_3 years	-0.4976	0.040	-12.406	0.000	-0.576 -0.419
emp_length_4 years	-0.5122	0.042	-12.267	0.000	-0.594 -0.430
emp_length_5 years	-0.4684	0.041	-11.443	0.000	-0.549 -0.388
emp_length_6 years	-0.4667	0.042	-10.983	0.000	-0.550 -0.383
emp_length_7 years	-0.3644	0.043	-8.496	0.000	-0.448 -0.280
emp_length_8 years	-0.4149	0.045	-9.322	0.000	-0.502 -0.328
emp_length_9 years	-0.4049	0.047	-8.639	0.000	-0.497 -0.313
emp_length_< 1 year	-0.5741	0.040	-14.348	0.000	-0.653 -0.496
home_ownership_MORTGAGE	-0.9471	0.014	-67.968	0.000	-0.974 -0.920
purpose_car	-1.4032	0.088	-15.967	0.000	-1.575 -1.231
purpose_credit_card	-1.7591	0.071	-24.766	0.000	-1.898 -1.620
purpose_debt_consolidation	-0.7460	0.070	-10.652	0.000	-0.883 -0.609
purpose_educational	-0.8931	0.194	-4.607	0.000	-1.273 -0.513
purpose_home_improvement	-0.5766	0.075	-7.735	0.000	-0.723 -0.431
purpose_house	0.6762	0.106	6.377	0.000	0.468 0.884
purpose_major_purchase	-0.8498	0.081	-10.542	0.000	-1.008 -0.692
purpose_medical	1.0086	0.092	10.926	0.000	0.828 1.190
purpose_moving	1.3906	0.100	13.885	0.000	1.194 1.587

# Loans Completed Over Time (Paid vs Default loans)

- Use an ARMA model
- Removed loans that originated after 2013 (to allow ~three years for borrowers to repay). Lower percentage of loans with 60-month terms.

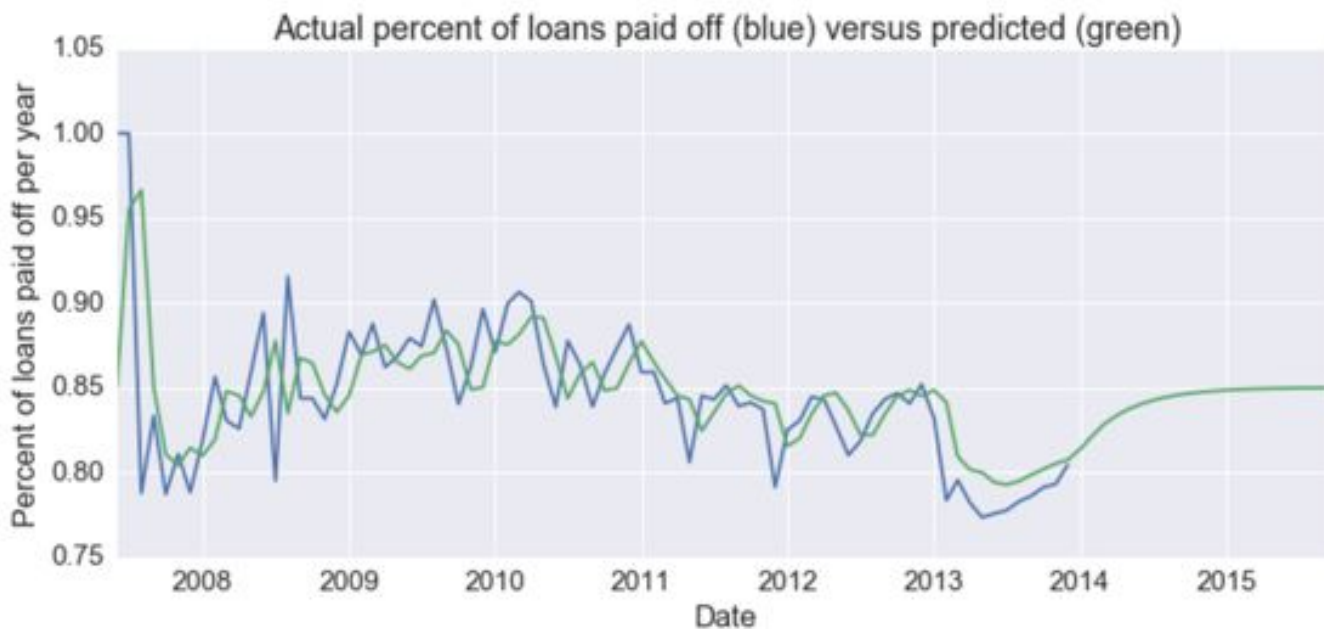


# Loans Completed Over Time



- $PACF(1) = AR(1)$
- $ACF(2) = MA(2)$

# Loans Completed Over Time



- ARMA quickly converges to long term mean without any seasonality terms.

Questions?

---

