Model Review Notes

NOTE: this file is being used to capture my notes while reading through the model writeups. I'll turn this into a cohesive summary once my analysis is complete.

Purpose

To compile a list of analyses already performed on the open source data and to learn the interesting relationships discovered in those analyses.

Lending club - predicting loan outcomes

ORourke (2016)

Data Period: 2007-2011

Author Notes:

- LinkedIn profile shows NO background in p2p lending or credit markets.

- Education at the University of Virginia not yet completed.

Analysis:

- Divided characteristics into two categories: those about the loan and about the borrower. Did not formally define the groups.
- Filtered out late loans leaving just fully paid, defaulted, and charged off. Defaulted and charge-off loans combined into a single group.

- Filtered characteristics by reading the descriptions in the data dictionary. Criteria seems to be intuition and the level of documentation. There was no analysis for correlation with outcome to support decisions.
- Chargeoffs focused in higher interest rates, fully paid more evenly distributed

- distribution of loans by grade

- 36 vs 60 month loan chargeoff rates

- Current.Account.Ratio number of open credit lines divided by total number f accounts

Modeling:

- Decision tree- The AUC from this model was 0.68 but only 60 charge-off loans were correctly identified. After resampling to increase the proportion of charge-offs, the confusion matrix increased to identifying 700 of the 1100 charge-offs.
- Logistic Regression- AUC: 0.697. Confusion matrix ID'd 730 of the 1100 charge-offs.
- Ensembled- Using bagging

Analysis review:

- The exploratory analysis was not particularly insightful. The author did not document the basis for several key decisions such as which variables to include/exclude or why he chose the specific algorithms. The charts were not styled beyond the ggplot defaults.

Changes in lending club underwriting

Wu (2015)

Data: June 2015 - May 2016

Notes:

- Response to LC's quarterly earnings release, where LC announced changes in underwriting
- Used commentary from LC's 10-Q.

Analysis:

- Scatter plot of average DTI for D grade loans by issue date. The plot attempts to show that less high DTI loans have been issued since LC changed its underwriting in late April 2016. The article hypothesizes that the loans are removed from the direct pay program. Direct pay requires the loan proceeds be paid directly to the borrower's existing debts.
- Scatter plot of number of inquiries by issue date for F and G grade loans. The plot shows that in the past month there have been less loans with a high number of previous inquiries.

Analysis of Lending Club's data

Darre (2015)

Data: through June 2015.

Author Notes:

- Strong educational background for data science: MSc in Statistics from Stanford University and MSc in Applied Math
- LinkedIn profile shows work experience as a quant for UBS in structuring CDOs.

Data cleaning:

- Removed loans where the :
  - policy code is not 1 meaning they are new products not yet publicly available.
  - high FICO score was 0 meaning the data was not valid.
  - Revolving credit utilization data was missing
  - Low FICO score was less than 600. Per the new underwriting policy, borrowers with a score less than 600 aren't allowed to receive a loan.
- New features engineered:
  - Year from the issue date
  - Buckets for the:
    * FICO scores
    * years, quarters, and months of issue dates
    * number of inquires in the past 6 months
    * DTI
    * Revolving balance
- Filtered to remove loans that have not matured.
- Grouped statuses into 2 categories: debt and purchases. Credit card and debt consolidation were grouped into the debt category and all others to purchases.
- Simplified the number of delinquencies to binary: 2+ delinquencies in 2 years.
- Simplified the number of public records to binary: threshold +1.

Analysis:

- FICO score by loan grade showing that most subgrades have average FICO scores within a 10 point range. The initial conclusion is that FICO score doesn't have a linear relationship with the loan grade. Expanded to show density around the average score by year which shows the densities started with a

tight distribution around the mean and then widened with time. The ending conclusion is that LC initially relied heavily on the credit score in the loan grading model and then loosened their reliance on it.

- Look at several variables for a relationship with defaults, loan grades, and fico scores:

Variable Name	Conclusion
Home ownership	'mortgage' defaulted at 11.5% and 2% less than other values. No analysis by loan grades
Purpose	'education' and 'small business' have higher charge off rates but education is a very small population. At the total population level, the FICO score does not reflect the higher risk
Revolving balance & employment length	no or minimal link to default. A bump in the charge-off ratio when the employment length is NA which is in contrast to the higher average credit score.
Number of delinquencies	Majority of borrowers do not have delinquencies leaving a small population which did. Delinquencies did have a lower average credit grade.
Number of open accounts	No clear trend
Debt to Income	Author notes a relationship with dti and higher credit score but that does not hold true for scores below 700.
Public Records	Small population but a 2% increase in charge-offs if already had a public record.
Age of credit history	Default rates decrease with credit history age. There was a 5% drop in defaults from borrowers with <6 years to >22 years. A secondary analysis could be performed to show the impact of borrower age on credit history age.
Revolving utilization	Higher usage leads to higher default rates
Annual income	High income defaults less, but does not eliminate the risk. Likely a significant variable.
Inquiries in 6 months	An indication of desperation to get credit. 7+ inquires are nearly 4x more likely to default than 0 inquiries.
Geography	Nevada and Florida are top 2 states for defaults.

Model Review:

- A very comprehensive analysis with several approaches that can be updated and reused. Use of violin charts to show the distribution of credit scores was very effective and can be used for dti, income, and credit score analysis. Plotly pie charts were less effective in transferring information because the default rates by category were embedded in the pop-ups instead of being directly visualized. Having just read Tufte's work, I think the bubble charts were too large to convey so little information and may have been better suited in a table. The analysis was a little disjointed the way he moved between the complete dataset and the subset of matured loans. It may have been useful to do them sequentially or enable a left-right comparison.

Lending Club Loan Analysis: Making Money with Logistic Regression

Davis (2012)

Data: through 2011

Author Notes:

- LinkedIn shows at the time he was the director of Search and Data at Etsy
- PhD in machine learning U of TX- Austin

Notes:

- Interesting discussion in the comments

Analysis:

- Default rate by description character length. He did not define how he measured defaults i.e. did he remove loans that haven't matured yet or were defaults included with the charge-offs.
- Logistic regression using first 50% of loans and evaluated on 2nd 50%. The model output is the probability of default which is then used to weight the interest rate. 12 variables were used in the model with no commentary discussing how they were chosen.
- Expanded to determine sensitivity of each of the variables by re running the model after holding out one of the variables. Noted that the amount requested had the biggest impact on investor return.

Review: - An interesting take on the model by flowing it through to a pricing model. So far in my review the authors have stopped at a classifier model. This is an important shift towards application.

Analyzing Historical Default Rates of Lending Club Notes

Toth (2015)

Data: 2012-2013

Author Notes:

- Data scientist at Orchard, a 'provider of data, tech and software to the online lending industry'.
- Undergrad at Wharton in stats/finance and a minor in math

Data Cleaning: - Classified statuses into performing and non-performing. Current and fully paid went into the performing group and all others to non-performing. - Removed variables that would not have been known at the time of issuance. -

Analysis:

- Share of loans performing/non-performing by grade and a further break down by subgrade
- Variable analysis:

Variable Name	Conclusion
Home ownership	prop.test for significance of mortgage vs owners and another for owners vs renters. Both tests were significant.
Debt to Income	Bucketed by 10% increments and intuitively states defaults increase as dti increased.
Revolving Utilization	Bucketed by 10% increments and intuitively states defaults increase as utilization increased.
Purpose	Calculated stats for each loan but summarized as 3 categories: 1) debt, 2) major purchase such as home improvement and cars, 3) luxury such as vacations and weddings.
Inquiries	Curve peaking at 3 inq and decreasing again with 4+ inquiries.
Total accounts	He cites a curve peaking at 20 accounts and diminishing effects above that. There's no analysis documenting the the bin size choice.
Annual Income	Bin by quintiles- top 20% of annual income are 6% more likely to be performing loans. The top cutoff for the top and bottom quintiles was 42K and 95K.
Loan Amount	Binned at 0-15,15-30, and 30-35K. Noting a decrease in performance for 30-35K loans. *This contrasts with numbers showing in my Loan Amount exploratory analysis?*
Employment Length	Binned as NA, <10, >10 years. No commentary but data suggests 1.25% improvement for >10 years over <10 years and another 1.25% over NAs.

Variable Name	Conclusion
Delinquencies past 2 years	Binned to 0, 1, 2, 3+ based on tail of distribution. No difference in 0-2 delinquencies but a fall-off of loan performance at 3+.
Number Open Accounts	No strong indication
Verifed Income Statements	Unexpected worse performance as verification status increased. Author suggests confounding variable as verification increases as credit score degraded.
Public Records	0, 1, 2+. Better performance at 1+ than 0
Non-Significant Variables	Months since last delinquencies, months since last major derogatory note, collections in past 12 months

Review: Extensive analysis of many variables. This variable-by-variable format has been used in other blogs but can be very boring to read without a central theme. The variables don't build on each other or follow a line of reasoning. Is there a better way? Also, there is no analysis of interaction between variables.

Author proposes a hypotheses for differences in results by variables but there's no testing for confirmation. (and how would you confirm?) Also, minimal documentation around decisions like bin cutoffs.

Gradient Boosting: Analysis of LendingClub's Data

Davenport (2013)

Data: Not stated, but likely through Q1 2013.

Author Notes:

- Data scientist at Amazon

- MS Computer Science from Georgia Tech

Data Cleaning:

- Replace missing values with NA
- Anova analysis listed Amount Requested, Debt/Income Ratio, Rent or Own Home, Inquiries in Last 6 Months, Length of Loan, and Purpose of Loan as being significantly correlated with MeanFICO and Interest Rate.
- Looked to the credit score info to understand link to other variables

Modeling:

- Objective: predict interest rate using RMSE as accuracy measure

- Straight to gbm: used to prevent overfitting

Analysis:

The author presents a fairly simple analysis which has more to do with interpreting the results of the gbm function than actually predicting the interest rates. This model has minimal utility in predicting default probability or expected cash flow from a loan. There are two possible insights from the blog post:

1) Loans have already been priced by the time they are on the market but it may be useful to understanding how loans were priced. You could then make a local optimization argument that there are arbitration gains where the interest rate doesn't reflect risk. The author did not attempt to explore, much less prove, that line of thinking.
2) Composition of the FICO score is a combination of amounts owed, payment history, credit mix, length of credit history, and new credit. This is a good starting point to identify confounding or high correlation variables.

Lending Club Data Analysis Revisited with Python

Davenport (2015)

Data:

Author:

- see above

Cleaning:

- Question relationships to identify risks, but no proposed solutions or conclusions without confirmation or analysis.
- Exploratory analysis was generally just distributions

- ID'd these risks:

Variable Name	Risk
Employment Title	Too many unique values to include as variable. Need to group somehow.
Employment Length	May confuse buying power, Job tenure doesn't indicate purchasing power.
Earliest credit line	Longer period could indicate age and earning potential.

Modeling:

- Objective was to explain the differences in interest rate
- Removed correlated variables >0.55
- Used gradient boosting regression, tuned with grid search
- Accuracy gets better with more iterations. Is there an over fitting concern?

Analysis: This is the first model that mentioned the risks of being wrong instead of being assertive towards assumed explanations. Most other models had been cavalier towards their hypotheses around deviances in the variables.

Putting the Python code in the analysis was a bit distracting. While it's good for reproducibility, it took away from the goal of understanding the author's message and what's happening with the loans. I think its preferable to include a link to Github or hold everything to the end.

Lending Club Deep Learning

Summers (2016)

Data: 2015 & 1H 2016

Modeling:

- Defined objective as "Use machine learning model to create a portfolio of 100 loans with a higher return than a Lending Club baseline". Note this is stated as measuring against profitability- shouldn't be just a default score. Used an average 10% return for the total population- no hold out.
- Links to external analyses- less duplicated work.
- Three strategies:

Strategy	Notes
1	Model assigns the probability of default. Scans through the range of probabilities calculating the average return, interest rate, and default rate for all loans above the cutoff. Find point of maximum return. Risk rate is the actual default rate. Model reduces risk of default but doesn't compensate for lower interest rate.

Strategy	Notes
2	Same model as above but applied to subsets of loans by grade. The model shows minimal improvement over a random sample from A & B grades but the results improve as the grades increase.
3	Change optimization from focusing on default rates to total return. Weight the payback probability by the interest rate. Minimal discussion, but the return percentage is higher than approach #1.

- The margin over average LC returns was highest in approach 3 but 2 applied to higher interest rate loans would be the best overall return. Approach number 3 maximizes the risk adjusted return.

- Used 10 fold cross validation

Analysis: Approach was written to maximize profit but the objective is written as a machine learning task. He starts as a demo type post but then does a better job of framing the writeup like a business driven model.

Author lets the code do much of the talking which makes interpreting the model difficult since my Python is not as strong. Recommend using a more verbose commentary to let non-Python readers understand what is happening in the code.

Mining Lending Club's Goldmine of Loan Data Part I of II − Visualizations by State

Cashorali (2011)

Data: Before Q3 2011

Author:

- BS in computer science and biology from Northeastern

- Worked in various analytics & data science roles starting in 2012 (1 year after the post was written)

Cleaning:

- Visualize the number of loans by state- No loans issued in North Dakota
- Animation of interest rates by year, state. Showed that rates were heterogeneous in 2007 than in 2011.
- Looked at fully paid vs late/defaulted loans by state. Florida had 40% default rate. TX, PA and NJ had 20% performing loan rate.

Analysis: Fairly superficial analysis but the first, so far, to consider geography. Interesting to note that FL and CA had high default rates but TX, PA, and NJ were the lower rates.

Investing at Lending Club with Watson Analytics

Polena (2016)
Data: Jan 2009- Dec 2012

Author: - Was an analyst at IBM at the time of the post. - Master's in quantitative economics - Master's in Financial Markets and Banking

Cleaning: - Nothing described in the article

Modeling:

- Passed the data to Watson, asked questions of it.
- Three insights:

No	Insight
1	Wyoming had the lowest default rate. The average rate by state was 14% but WY was only 9%.
2	Lowest default rates by purpose are Weddings and Cars. Small business has the worst rates. The proposals that weddings and cars or better performing are at odds with other analyses reviewed.
3	Employment length less than 5 years was more likely to repay than >5 years. No code or visualizations to support the claim

Analysis: Claims were not supported by code to reproduce the analysis or understand the exact assumptions and methods used. Claims were also not supported by a train of evidence and assumptions leading up the the final conclusion– the autor just cuts straight to the point. That may be useful as a demo of Watson's capabilities but not from an analytic reference.

Application of survival analysis to P2P Lending Club loans data

Mistry (2016)

Data: Not explicitly stated but assumed through Q2 2016

Author:LinkedIn profile shows

- Clinical researcher and modeling
- University of Manchester:
    - PhD in applied Math

    - Master's in Applied Computing

    - Bachelor in Math

Cleaning:

- Do not have to exclude current loans if using survival analysis.
- Set objective RR as the ratio of amount paid to amount lent. A value at the end of the loan greater than 1 is profitable and less than 1 is unprofitable.
- Referred to yhat blog for useful covariates.
- Looked for relationship of RR to the FICO score. Higher profitability score for FICO over 700.

Modeling:

- Stratified by FICO
- Used concordance rather than p-value to determine ranking of variables. Final model uses interest rate and term of the loan.
- Simple survival analysis resulting in profit probability curves. Give the probability of achieving a range of profitability ratios.

Analysis: Model gives a probability of reaching a given profitability ratio. This ratio isn't the same as the expected Annualized return but could probably be reworked to calculate for as such. The post was written simple enough to be understandable without knowing the math behind the survival analysis- no proofs or equations. Contrarily, there was no code to supporting so it's less reproducible. Seeing the shiny app as an output indicates it was done in R.

Concordance is a new concept and will need to be researched. Interesting to note that only interest rate and loan term were significant. These 2 seemed reasonable but I would have expected others to be significant.

Survival prediction (P2P loan profitability) competitions

Mistry (2017)

Suggested healthcare for survival analysis research. No model built in this post. Survival is used often to model disease states and how long a patient can live given a treatment variable. This is well developed science in healthcare and could be an opportunity to port over to finance specifically P2P lending.

Using genetic algorithms to maximize Lending Club performance

Patierno (2011)

Author:

- Software engineer at Google
- BS in Computer Science

Data: Written in Feb 2011 and claiming to analyze 4 years of data. Assuming through Q4 2010.

Modeling:

- Used genetic algorithm to select attributes. No exploratory analysis. No test or hold out dataset used to validate and prevent overfitting.
- Model source code on GitHub
- Claimed to result in a 12.5% return

Analysis: Author created a tool rather than a model. There are no conclusions for how to maximize returns, reduce defaults, or any other objective.

Peer Lending Risk Predictor

Tsai, Ramiah, and Singh (2014)

Authors: Students at Stanford, advised by Andrew Ng

Data: 2007- 2013

Cleaning:

- Removed loans that were fully paid or defaulted. Similar to filtering for matured loans but would include early pay-offs and early defaults. (Does this change the result?)
- Rebalanced the training set to be 50/50 defaulted vs paid loans. No discussion on how.
  - Feature selection using information gain in Weka and Matlab. Multiple methods consistently identified interest_rate, loan_amount, annual_income, and loan_purpose.
- Definition: Recall is number of loans paid and identified as paid
- Definition: Precision is number of loans correctly identified as not performing
- Used term precision to describe the amount of loans that were fully paid vs total loans in the category.

Modeling:

- Linear regression- Penalty function added to logistic regression to avoid classifying defaulted as paid off. Preference for precision over recall or overall accuracy. Penalty factor increased precision from 88.9% to 95.9%.
- SVM- Minimal discussion, probably truncated for space- 5 page limit.
- Naive Bayes- No discussion
- Random Forest: training accuracy was high but was over fitting. Used out-of-bag error to measure success and minimized error at 37.25%.

- TF-IDF- Ranked words by TF-IDF for both paid and defaulted loans. Comparison for words with high rank in one but low in the other. *steady* and *God* ranked higher in loans that defaulted but *university* ranked higher in paid loans. Created binary variables for each an improved accuracy by 3%. [Are descriptions even required anymore?]
- PCA for visualization. Plotting loans by 1st and 2nd components. Running a logistic regression to create a decision boundary. Eliminating defaulted loans also eliminates most good loans- hence high precision but low recall again.

Conclusions:

- Acknowledged that models recommended only 0.6% of loans and wasn't practical
- Compare to LC: Used Modified LogReg to create portfolio with same risk profile as A1 subgrade but 2% higher return. Similarly compared model classifications to LC grades. Concludes LC is classifying high quality loans into lower credit grades/high interest. Opportunity knocks.
- Understand model sensitivity to population skew, rebalance test set as needed. Generally, know your algorithms assumptions and weeknesses.
- Achieved greatest success by using the penalty factor for incorrectly idenifying bad loans.

Analysis: The first 3/4 of the paper goes through several models and measuring performance. Ng's influence is apparent in the theoretical commentary and mathematical notation. The last page of the paper compares the best model against LC. Use their model to create similar risk tranches as LC, but with higher interest rates. This paper is the first I've read which quanifies the profitability impact of reducing defaulted loans. It is also the first to include a penalty function in the model- other papers have accepted models out of the box with no/minimal tuning.

I appreciated the idea of optimizing for precision instead of accuracy. Investment dollars are limited and at this point in my portfolio I am more likely to run out of funds before running out of investment opportunities. It is acceptable to significantly reduce the number of loans. I assume this is the case until a single investor is purchasing a material share of the market- say $> 10\%$.

This seems to be a recurring project in the course as there are papers for the 2011, 2014, 2015 and 2016 sessions.

Lending Club Default Analysis

Moy (2015)

Author:

- Attended NYC Data Academy in 2015 at the time the post was written
- MS in Operations Research from SUNY
- Worked as Industrial Engineer prior to writing. No prior experience with p2p loans

Data: 2007-2014

Cleaning:

- Removed loans not in current loan listing

- New feature: character count for employee title
- Removed 'Current loans', no discussion on how to handle late loans

Modeling:

- Objective: predict default and reduce 22% loss on investments
- Feature selection based on random forest, selecting variables above 20 'importance'. Importance was not defined.
- Did not remove interest rate

Remaining Models

http://andirog.blogspot.com/2012/10/lending-club-loan-purpose-default-rate.html
http://blog.yhat.com/posts/machine-learning-for-predicting-bad-loans.html

- Gupta (2012)
- Anon. (2015)

- Anon.

Trends and organization

- 3 tranches: ML demo, exploratory analysis, and practical application
  - ML demo: casual analysis and then quickly move to applying ML algorithm. Some feature engineering, but they don't tend to be helpful variables. This category can be bifurcated into citizen data scientists doing demos of applied algos and academics using Greek letters to explain algos.
  - Exploratory: product of a course or bootcamp. Showing relationship of multiple variables. No resulting model or classifier.
  - Practical application: Blog posts from investment service firms. Show deep knowledge of the industry but are not a complete and comprehensive review of all variables. Tend to be a deep dive on a very narrow topic. Likely useful for understanding the nuance of the market but not for getting started.
- Approach for those with DS experience is to go variable by variable and look for the effect. Practitioners tend to focus o current events and the impact on the current credit scoring model.

Major take-aways:

- Add a penalty function to adjust for loans falsely classified as performing. Optimize for precision
- Model for profitability, not accuracy of the model
- Reading a step-by-step analysis by variable is very boring but without the detail it's hard to accept the conclusions. Should make sure to have a strong executive summary or abstract.
- Follow intution- the expected variables are likely to have significance.

- Interest is a measure of default risk and to reward the investor for taking that risk. Including int rate is going to be highly correlated with default experience so needs further thought. Looking for arbitrage from interest rate, not purely avoiding defaults.

References

Anon., 2015, Lending club - predicting loan outcomes,. Blog *Peer Lending Server* (Analyzing Lending Club Payment History).

Anon., Give me some credit,. Blog *Kaggle* (https://www.kaggle.com/c/GiveMeSomeCredit).

Cashorali, Tanya, 2011, Mining lending club's goldmine of loan data part i of ii visualizations by state,. Blog *R-bloggers* (https://www.r-bloggers.com/mining-lending-clubs-goldmine-of-loan-data-part-i-of-ii-visualizations-by-state/).

Darre, Jean-Francois, 2015, Analysis of lending club's data,. Blog *Data Science Central* (http://www.datasciencecentral.com/profiles/blogs/analysis-of-lending-club-s-data).

Davenport, Kevin, 2013, Gradient boosting: Analysis of lendingclub's data,. Blog (http://kldavenport.com/

gradient-boosting-analysis-of-lendingclubs-data//).

Davenport, Kevin, 2015, Lending club data analysis revisited with python,. Blog *Machine Learning and Statistics Blog* (http://kldavenport.com/lending-club-data-analysis-revisted-with-python/).

Davis, Jason, 2012, Lending club loan analysis: Making money with logistic regression,. Blog *Data Startups* (http://drjasondavis.com/blog/2012/04/08/lending-club-loan-analysis-making-money-with-logistic-regression).

Gupta, anil, 2012, Lending club loan purpose: Default rate and bad-loan experience index,. Blog *Random Thoughts* (http://andirog.blogspot.com/2012/10/lending-club-loan-purpose-default-rate.html).

Mistry, Hitesh, 2016, Application of survival analysis to p2p lending club loans data,. Blog *Systems Forecasting* (http://systemsforecasting.com/2016/07/application-of-survival-analysis-to-p2p-lending-club-loans-data/).

Mistry, Hitesh, 2017, Survival prediction (p2p loan profitability) competitions,. Blog *Systems Forecasting* (http://systemsforecasting.com/2017/02/survival-prediction-p2p-loan-profitability-competitions/).

Moy, Philip, 2015, Lending club default analysis, (http://www.creditreportservice.info/article/95227030/lending-club-default-analysis/).

ORourke, Ted, 2016, Lending club - predicting loan outcomes,. Blog *Rpubs* (https://rpubs.com/torourke97/190551).

Patierno, David, 2011, Using genetic algorithms to maximize lending club performance,. Blog (http://blog.dmpatierno.com/post/3161338411/lending-club-genetic-algorithm).

Polena, Michal, 2016, Investing at lending club with watson analytics,. Blog *IBM* (https://www.ibm.com/blogs/business-analytics/lending-club-and-watson-analytics-1/).

Summers, Cameron, 2016, Lending club deep learning, (https://scaubrey.github.io/).

Toth, Michael, 2015, Analyzing historical default rates of lending club notes,. Blog *R-bloggers* (http://michaeltoth.me/analyzing-historical-default-rates-of-lending-club-notes.html).

Tsai, Kevin, Sivagami Ramiah, and Sudhanshu Singh, 2014, Peer lending risk predictor, (http://cs229.stanford.edu/proj2014/Kevin%20Tsai,Sivagami%20Ramiah,Sudhanshu%20Singh,Peer%20Lending%20Risk%20Predictor.pdf).

Wu, James, 2015, Changes in lending club underwriting,. Blog *MonJa* (https://www.monjaco.com/blog/changes-in-lending-club-underwriting-looking-beneath-the-headlines//).