MACROECONOMIC FACTORS AND LOAN QUALITY

An analysis of Lending Club's portfolio

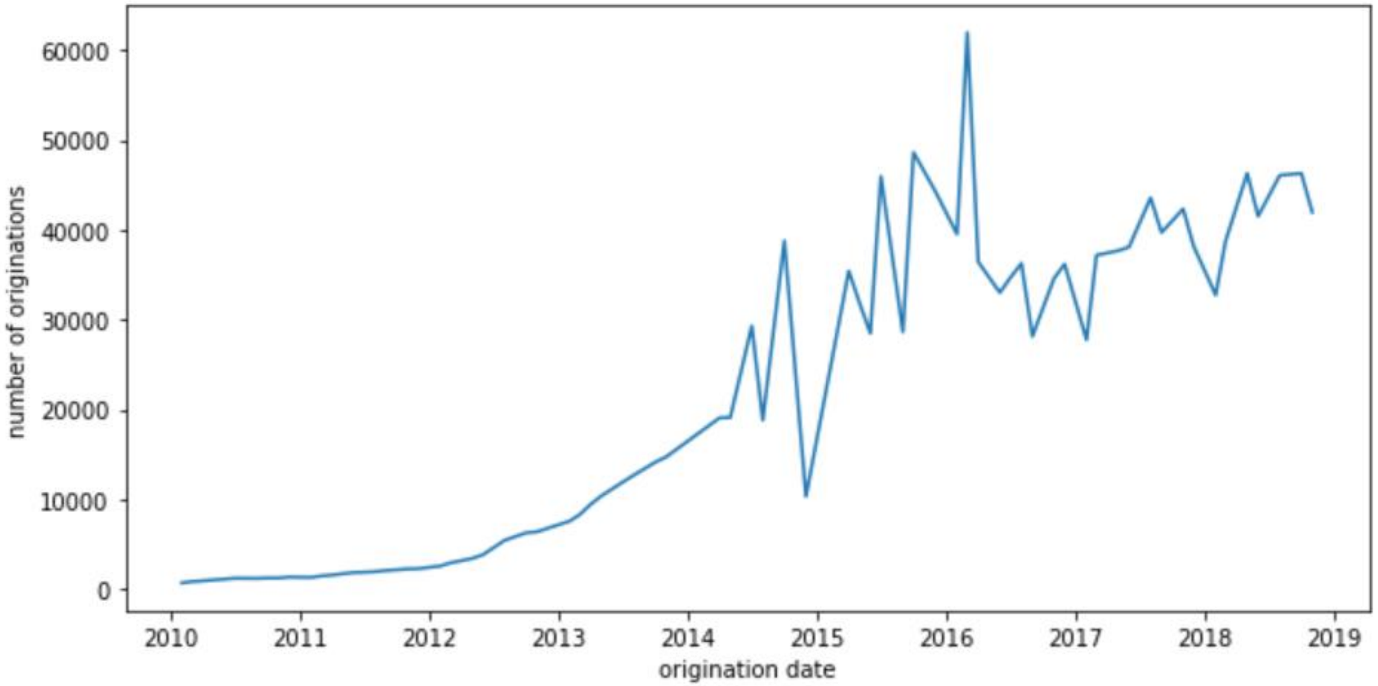
The Problem

- Unsecured personal loans are risky – companies need to price their loans to compensate for the default risks that they face, and decline customers who are unlikely to pay.
- Lending companies use sophisticated models to evaluate borrowers' loan default likelihood based on their credit profiles.
- Factors outside of an individual borrower's credit portfolio are likely to impact default likelihood – **can macroeconomic factors, like stock market fluctuations or the unemployment rate, be used to improve assessment of default risk?**

The Data

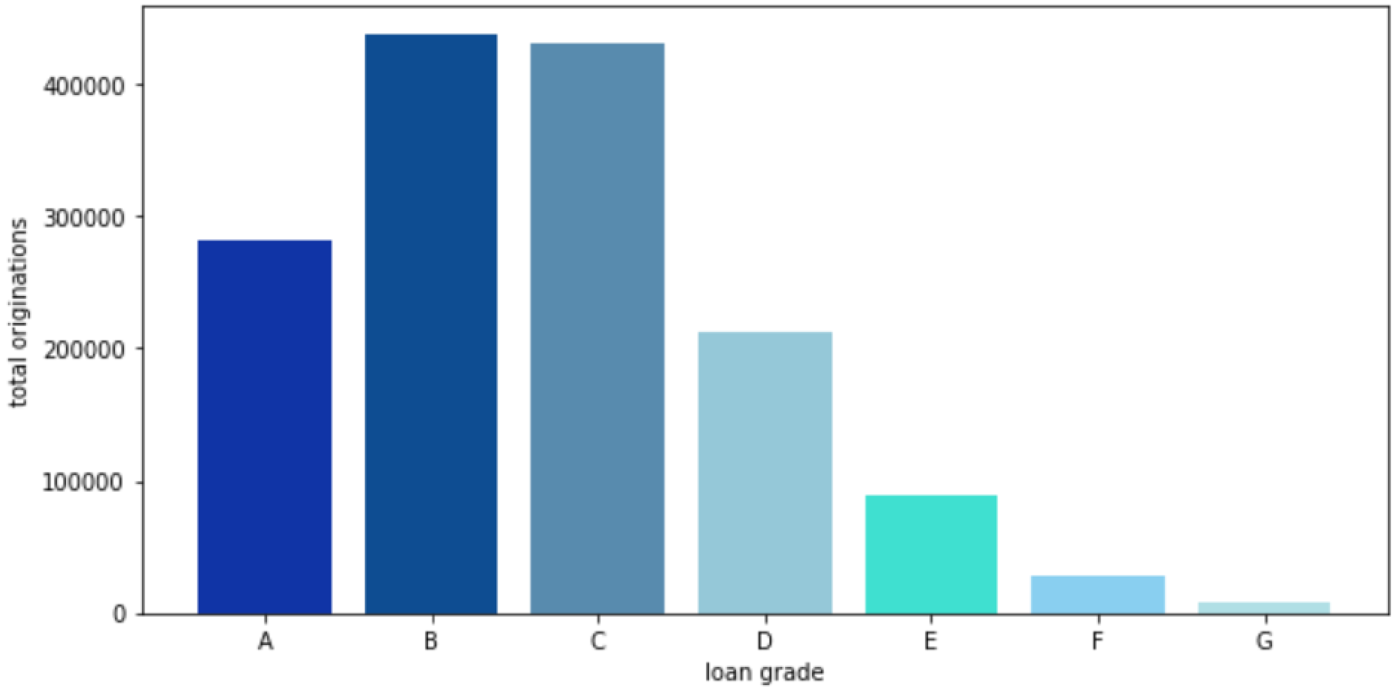
- Lending Club is a company that originates unsecured personal loans. Data on their loans is publically published and is the basis for this exploration.
- FRED – Federal Reserve Economic Data – is a US government-maintained database of economic data. A host of important macroeconomic factors are sourced from FRED.
- Epiq Global maintains statistics on the number of personal bankruptcies filed each month in the US, which may predict aggregate borrower quality.

Exploring the Data – Loan Originations



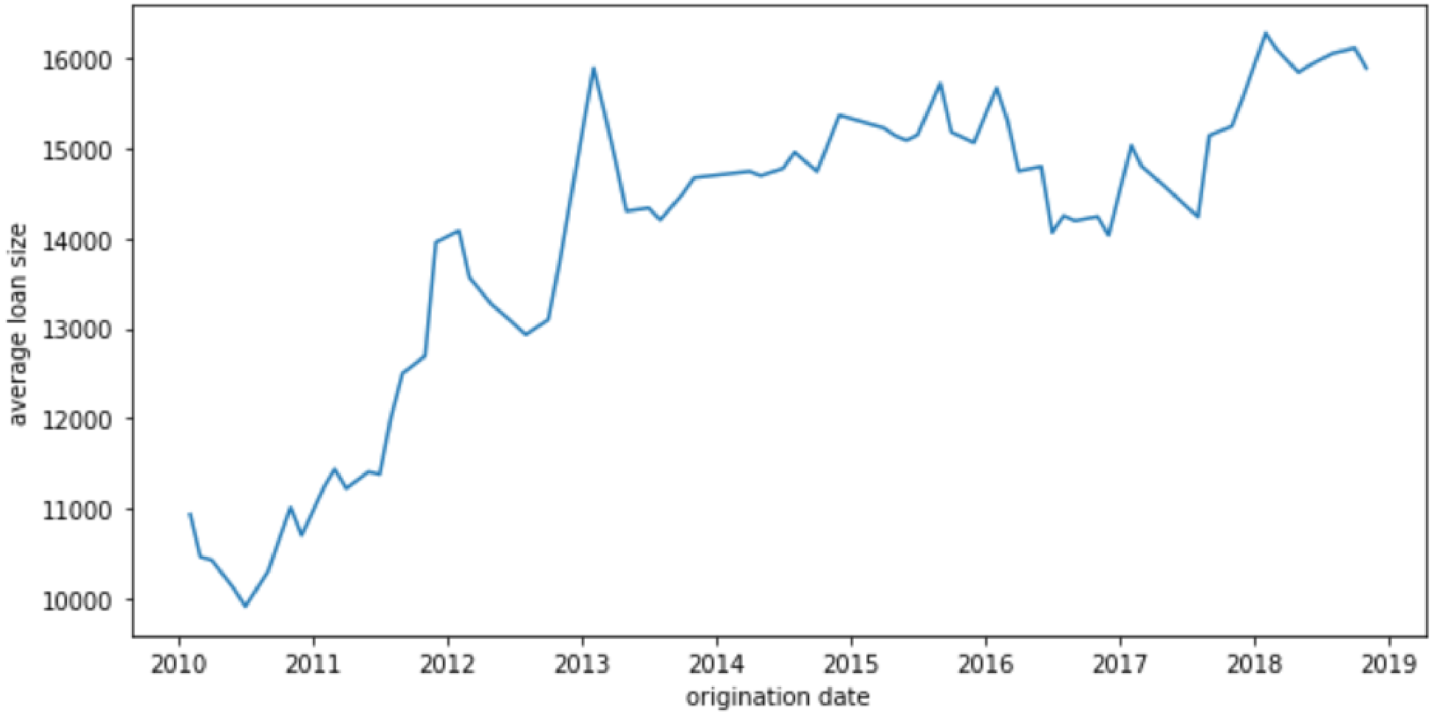
- Lending Club has originated tens of thousands of loans per month in recent years, with a trend of increased originations over time.

Exploring the Data – Grades



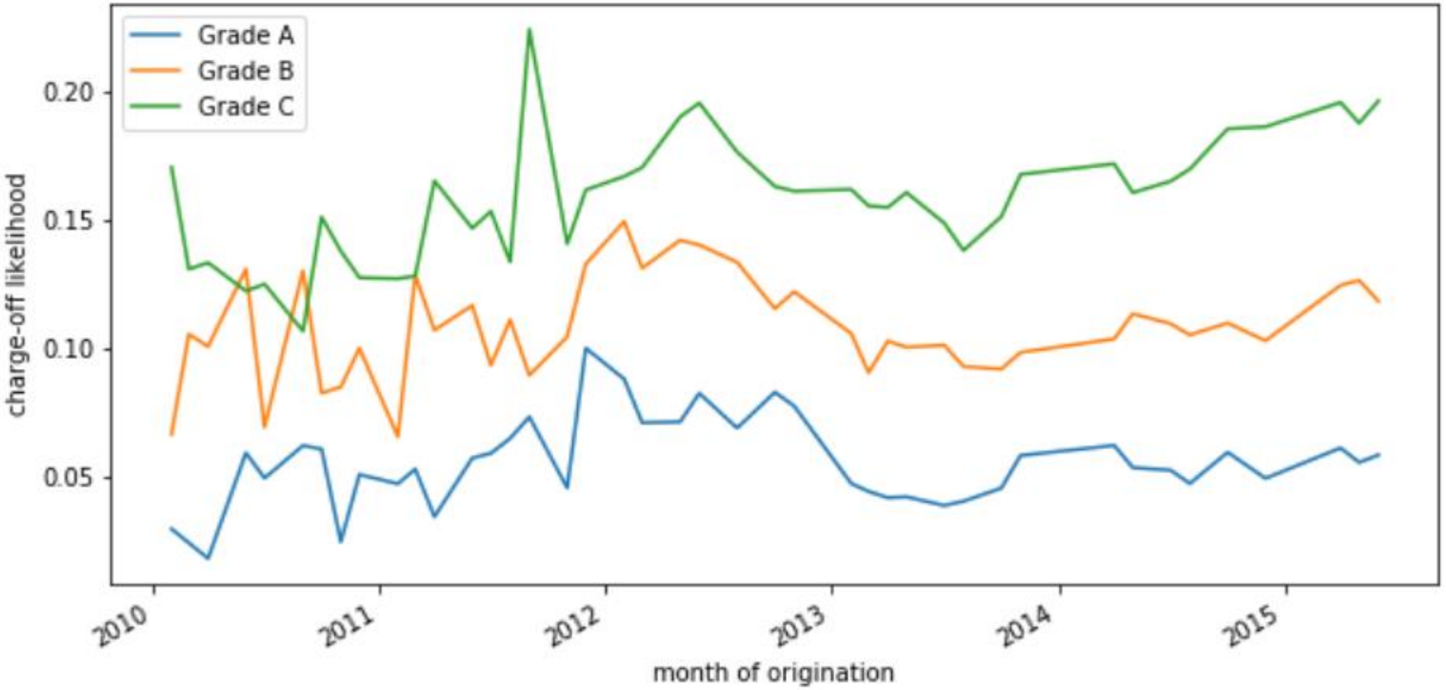
- Lending Club loans are graded A through G, with grade A loans being lower-risk and lower-interest rate, and grade G loans being the highest-risk, highest-rate loans. Total originations for each grade are shown in the bar graph above – the large majority are grades A, B, or C.

Exploring the Data – Loan Size



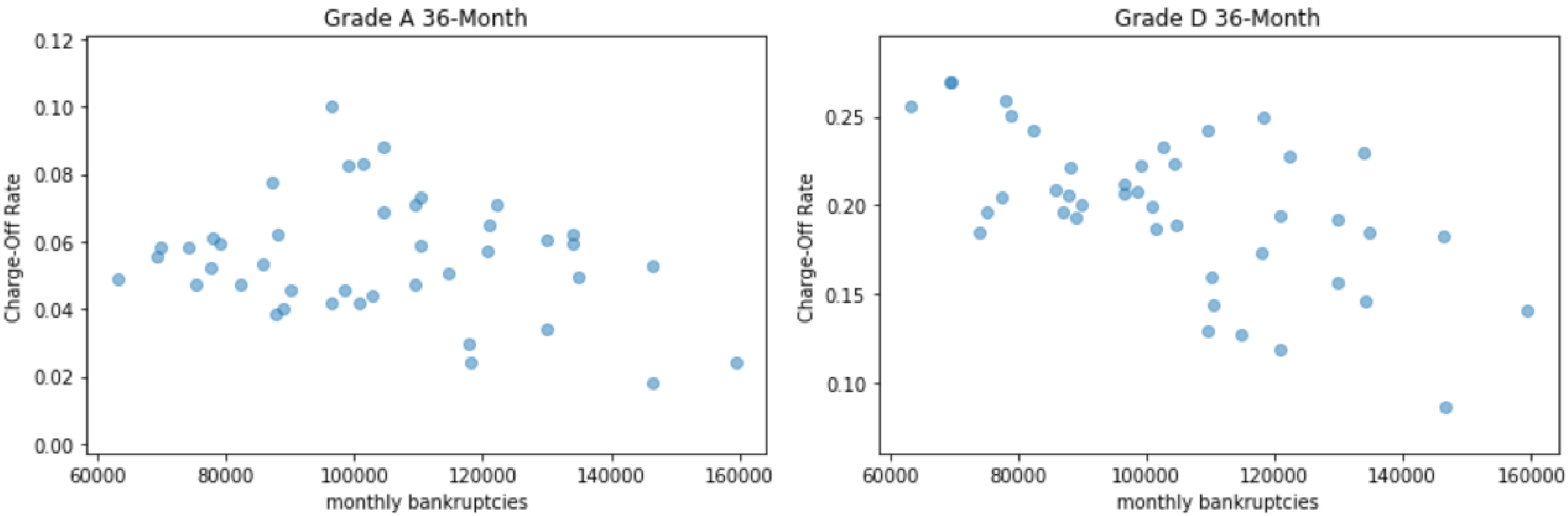
- Loans range in size from $5,000 to $40,000. Since 2014, they have averaged roughly $14,000 - $16,000 per loan.

Exploring the Data – Charge-Offs



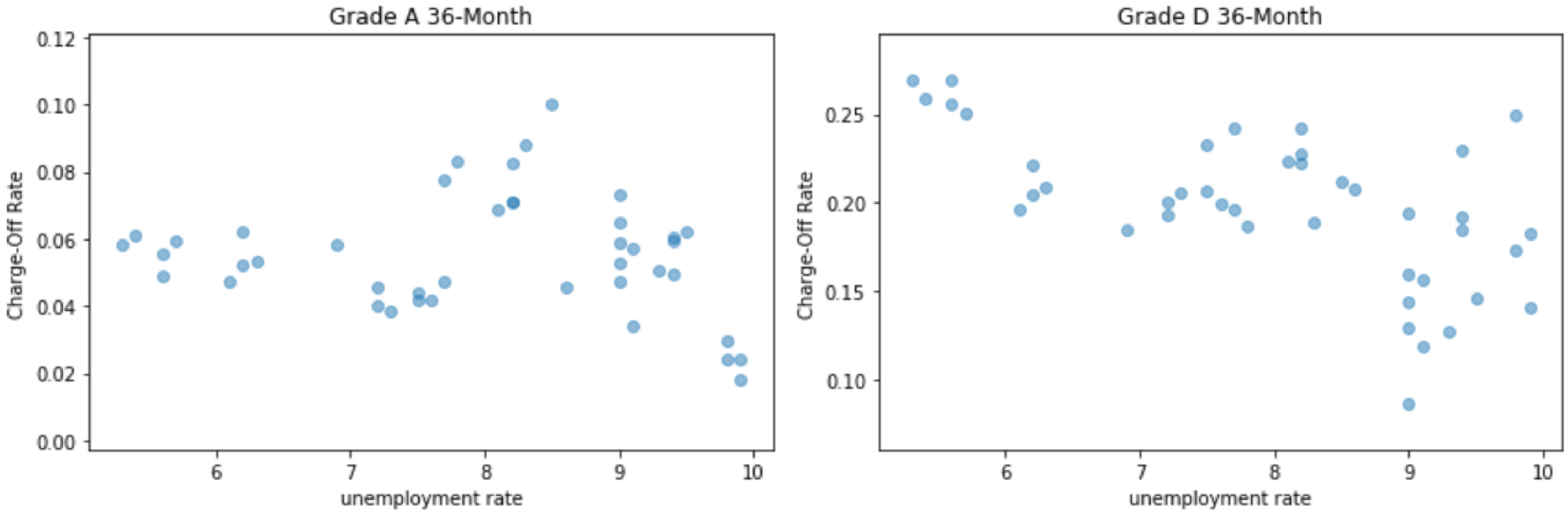
- Charge-offs are the variable of interest – a loan is "charged off" if the borrower did not make all payments required.
- The graph above shows that the loan "grading" by Lending Club is accurate in a relative sense – Grade C loans are significantly more likely to charge off than Grade A loans.

Macroeconomic Factors - Bankruptcies



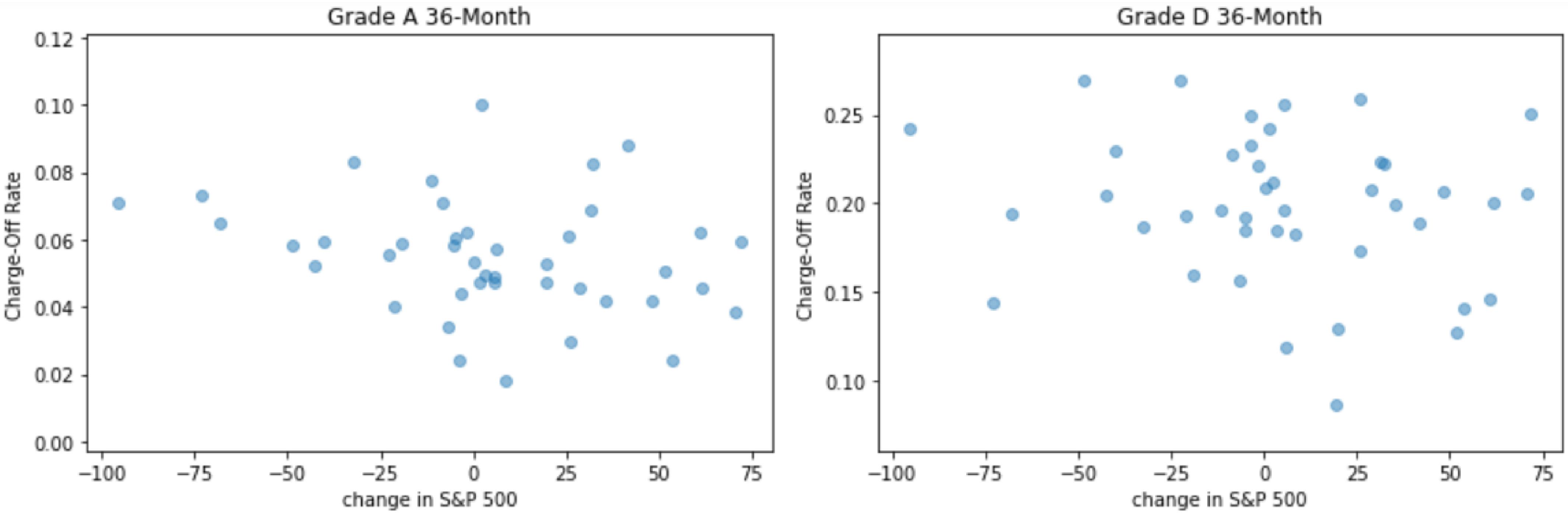
- It might be expected that higher national bankruptcy rates when loans are originated would predict higher charge-off rates throughout the life of the loan.
- This prediction does not appear to be borne out in the data, where the *opposite* trend is observed. The graphs above show that higher bankruptcies are correlated with lower charge-off rates.

Macroeconomic Factors - Unemployment



- One would expect a higher unemployment rate to predict a higher charge-off likelihood
- This does not appear to be the case in the data. Higher unemployment rates at origination appear to be associated with higher charge-off rates.

Macroeconomic Factors – S&P 500



- Decreases in the S&P 500 might represent economic deterioration that would predict higher default likelihood. The x-axis of the plots above shows the monthly change in the S&P 500 stock index.
- There is not a clear relationship between changes in the S&P 500 and default rates.

Signal from Noise

- Relationships between macroeconomic variables and loan default rates *appear to be* weak.
- But could machine learning models find signal in the noise that is not obvious in exploratory data analysis?
- Machine learning models incorporating macroeconomic data were tested to evaluate this question.

Machine Learning – The Baseline

- Applying machine learning models to lending data is not novel – a good baseline for testing whether macroeconomic data provides useful information is model performance *without* macroeconomic data
- Three classification models were used to predict whether a loan is "good" (repaid) or "bad" (charged off before being repaid): KNN, SVM, and Random Forest.
- Each model used 3-fold cross validation and a variety of different specifications. For each model, the most predictive set of parameters was used.

Machine Learning – Baseline Results

- Standard practice in lending is to evaluate loans by their "good-to-bad ratio" – the number of good loans that would be rejected based on a rule change for every bad loan that is rejected. The results for each of the three models is shown below. These are very favorable rates, as a loan that defaults typically involves a very large loss.

Model Name	False Positives	True Positives	Good:Bad Ratio
KNN	4121	4423	.9317
SVM	2225	4226	.5265
Random Forest	2236	4059	.5509

Machine Learning – Incorporating Macroeconomic Factors

- Several macroeconomic factors that are typically viewed as good predictors of macroeconomic health were then incorporated into the data.
- Macroeconomic factors used include measures of the unemployment rate, broad aggregates of economic health, a stock market aggregate, and others.
- The exact models performed on the baseline data were re-run on the new data.

Machine Learning – Baseline Results

- Good-to-bad ratios improved markedly by incorporating macroeconomic data.
- The best-performing model, Random Forest, saw large improvement: the number of false positives barely changed (from 2236 to 2203), but the number of true positives increased from 4059 to 4487.

Model Name	False Positives	True Positives	Good:Bad Ratio
KNN	3749	4079	.9191
SVM	2430	4698	.5172
Random Forest	2203	4487	.4910

Conclusions

- Even with a limited set of Machine Learning models utilizing a small amount of the available macroeconomic data, there was a significant improvement in identification of “bad” accounts.
- Overall, incorporation of macroeconomic variables is a potentially potent – and very profitable – tool for evaluating whether an applicant should be extended credit.