Peer to Peer Lending:
Optimal Decisions for
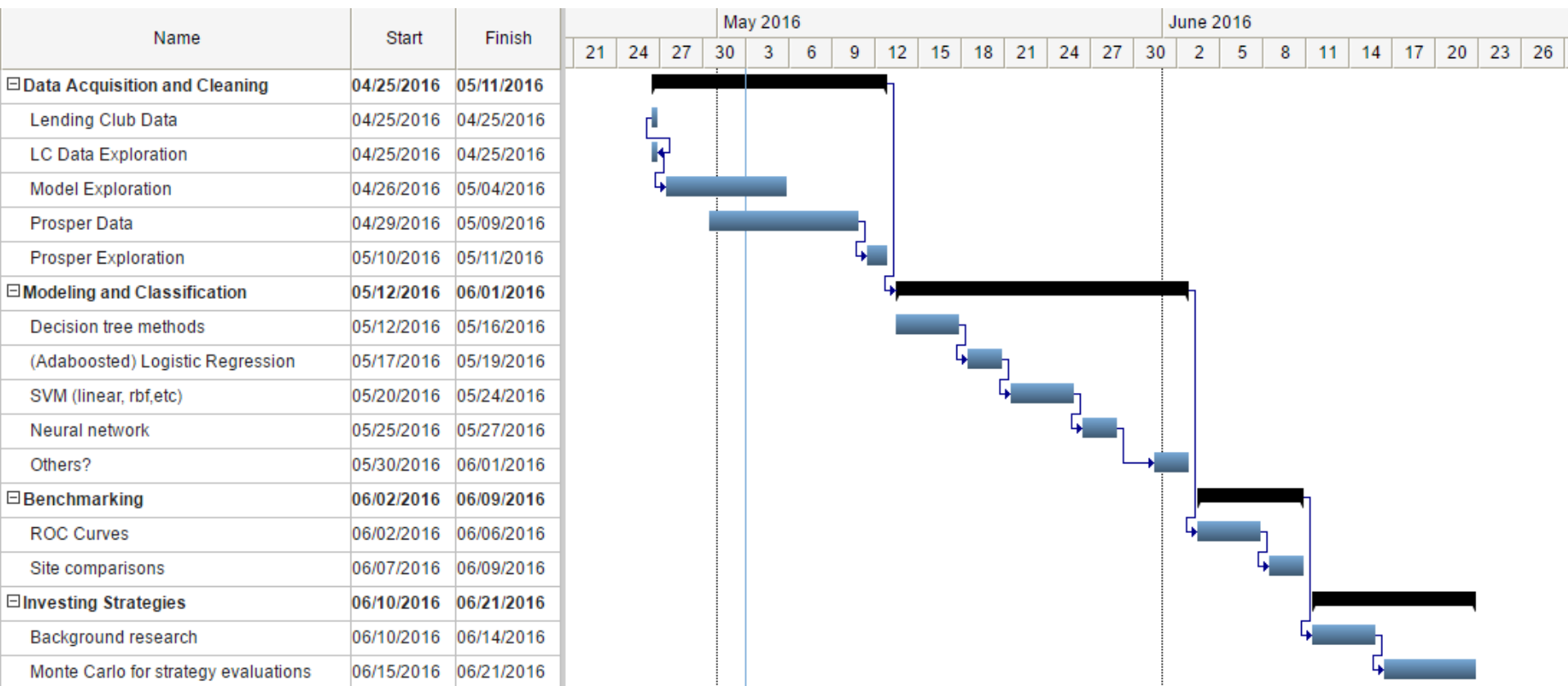# Investors and Lending Companies

Jacob Stinnett

May 3, 2016

Background

Various companies created to crowdfund personal loans (Lending Club, Prosper, etc).

- How to best identify good/bad loans?
- How do different sites compare?
- What information matters? What is the "best" site?
- Can we improve on company's internal score?
- How do investment strategies compare?

Research Plan

Name	Start	Finish
Data Acquisition and Cleaning	04/25/2016	05/11/2016
Lending Club Data	04/25/2016	04/25/2016
LC Data Exploration	04/25/2016	04/25/2016
Model Exploration	04/26/2016	05/04/2016
Prosper Data	04/29/2016	05/09/2016
Prosper Exploration	05/10/2016	05/11/2016
Modeling and Classification	05/12/2016	06/01/2016
Decision tree methods	05/12/2016	05/16/2016
(Adaboosted) Logistic Regression	05/17/2016	05/19/2016
SVM (linear, rbf,etc)	05/20/2016	05/24/2016
Neural network	05/25/2016	05/27/2016
Others?	05/30/2016	06/01/2016
Benchmarking	06/02/2016	06/09/2016
ROC Curves	06/02/2016	06/06/2016
Site comparisons	06/07/2016	06/09/2016
Investing Strategies	06/10/2016	06/21/2016
Background research	06/10/2016	06/14/2016
Monte Carlo for strategy evaluations	06/15/2016	06/21/2016

Note: plan does not include ~3 weeks of travel in May (conference, family stuff, etc), so actual completion time is mid- to late-July.
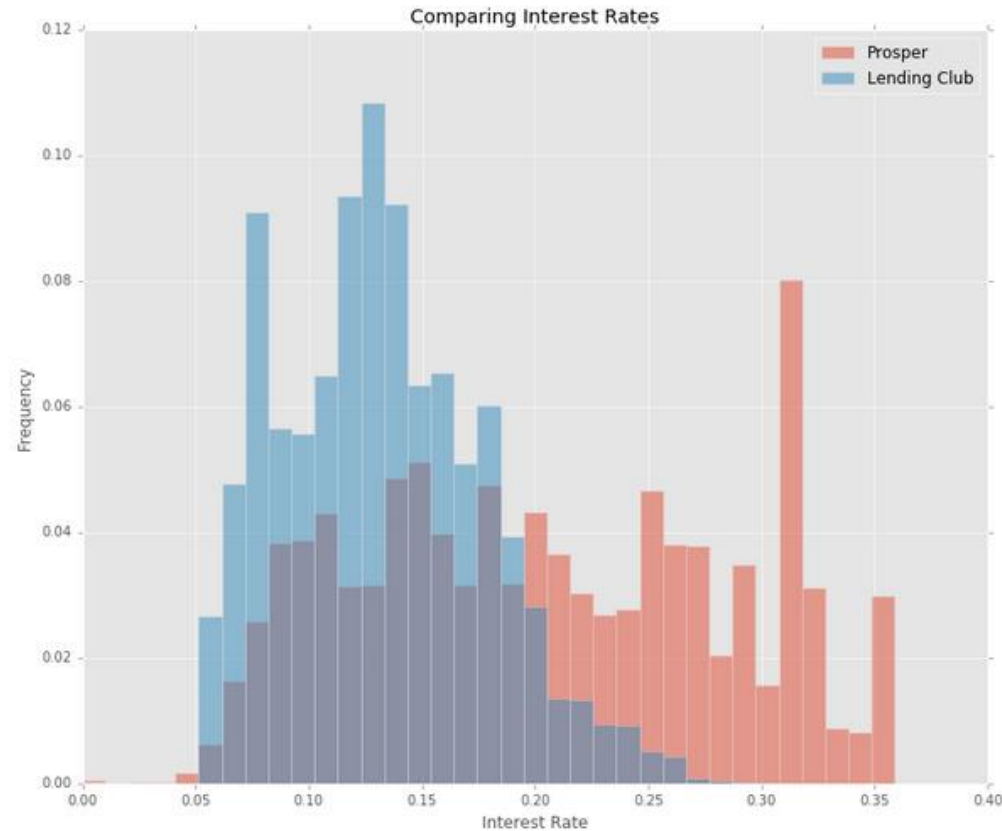
Previous Work

- Lending Club (LC) data acquisition/prep

- Exploratory analysis

- Early tests of various classifiers (limited dimensions)
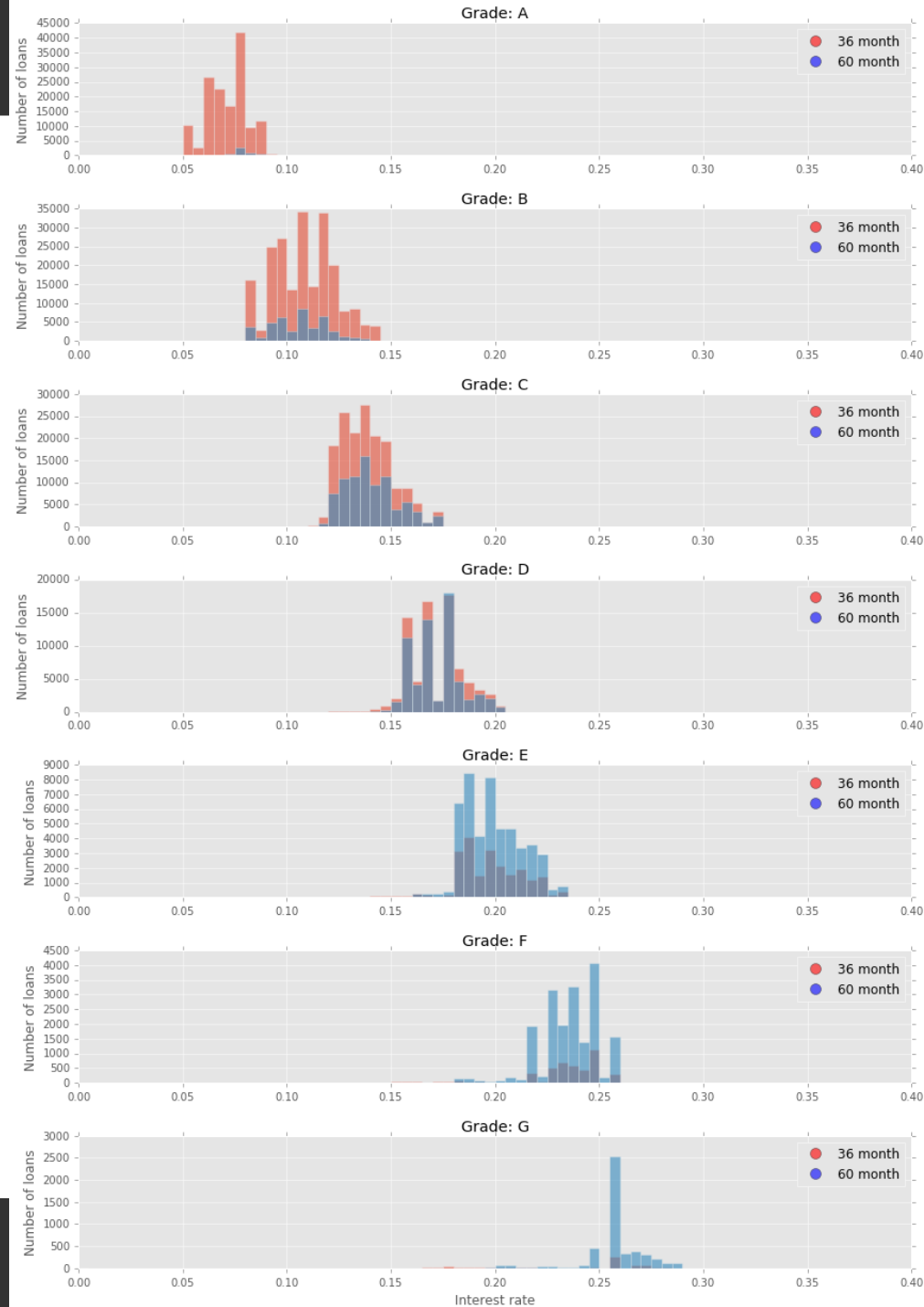
Progress

- Obtained Prosper data

  - Database merging problem: loan histories and loan applications don't have matching identifiers

  - Less records than LC (~100k vs 800k) but much more application data

	Lending Club	Prosper
Records	800k	100k
Useful columns	100	500
Max Interest	29%	36% *
Bad loan rate	6.10%	13.80%

Preliminary Classification Results

- Lending Club data only, with 32 columns

- However, ROC curve may not be the best metric for performance evaluation