

# CS 209 A, Introduction to Data Science



## Milestone #2, group 52

### 1) Group ID

Team #52

#### Members :

Mariah Barber

Josh Feldman

Daniel Janini

Simon Sebbagh

### 2) Team Communication

The whole team met to discuss the project and how we will split the work. In order to facilitate the interactions within the team, we set up a Slack Channel for regular chat and will set up an Asana account for managing the task repartition

### 3) TF Communication

We have contacted our assigned TF, Filip Michalsky, as we had some questions about the scope and the outcome of this project.

Namely, we were wondering if we should either proceed to an analysis to the current investment analysis and highlight its strengths and weaknesses, or develop our own lending engine to drive our investment strategy.

### 4) Question Formulation

Here, we will build a model to advise an investor reviewing loan applications on the LendingClub network. This model will maximize the investor's ability to profit from the loans that they grant.

Importantly, the LendingClub network is an "Equal Housing Lender" barred from discriminating against applicants based on race, ability, sex, religion, familial status, or nationality. Currently, federal legislation does not adequately prevent this discrimination from taking place. For example, minority applicants are denied mortgage loans at higher rates than white applicants in many metropolitan areas in the US. While banks claim not to consider race when making loan decisions, there is a lack of transparency regarding the models they are using to make their decisions. A central aim of this project will be to propose a model that is superior to this discriminatory state of affairs. Thus, the ideal model for making loan decisions would be fair and interpretable, while also maximizing profit. Through our attempt to build such a model we will aim to address the following central question:

If we want to maximize the profitability of our lending model, do we need to sacrifice fairness and interpretability?

To answer this question, we will need to define the metrics by which to measure profitability, fairness, and interpretability. After defining these metrics, we'll build two models:

1. A model that maximizes profitability without concern for fairness or interpretability
2. A model that attempts to match the profitability of model 1, while simultaneously optimizing fairness and interpretability

Compare performance of models in interactive data visualization

5) Work split for Milestone 3

Thus far, we've downloaded our data and loaded it into a jupyter notebook.

Task	Member assigned	Deadline	Status
Exploratory data analysis with respect to profitability-iteration 1	Simon, Josh	Nov 1	In progress
Exploratory data analysis with respect to fairness-iteration 1	Dan, Mariah	Nov 1	In progress
Define our fairness metric-wrt fair housing act and in general-iteration 1	Dan, Mariah, Josh	Nov 1	In progress
Define our profitability metric-iteration 1	Simon	Nov 1	In progress
Define our interpretability metric-iteration 1	Josh	Nov 1	In progress
Exploratory data analysis with respect to profitability-iteration 2	Simon, Josh	Nov 8	In progress
Exploratory data analysis with respect to fairness-iteration 2	Dan, Mariah	Nov 8	In progress
Define our fairness metric-wrt fair housing act and in general-iteration 2	Dan, Mariah, Josh	Nov 8	In progress
Define our profitability metric-iteration 2	Simon	Nov 8	In progress
Define our interpretability metric-iteration 2	Josh	Nov 8	In progress

Build profit-only model - iteration 1	Full team	Nov 15	In progress
Build profit/fair/interpretable model - iteration 1	Full team	Nov 15	In progress
Build profit-only model - iteration 2	Full team	Nov 22	In progress
Build profit/fair/interpretable model - iteration 2	Full team	Nov 22	In progress
Write milestone 3 - draft 1	Full Team	Nov 22	In progress
Submit milestone 3	Full Team	Nov 28	In progress