

# Project Proposal

Ray 2019

# The Problem

We need to reduce the default rate of the loans we write.  
What information can help us determine whether or not we'd think an individual will default on their loan?

- Which features most explain whether a person will pay their loan payments on time?
- Can we predict whether or not a person will repay their loan?

# The Data

Data source: Kaggle


- LendingClub loans
- ~100 columns
- ~2m loans


## Columns

 id

 member\_id A unique LC assigned Id for the borrower member

 loan\_amnt amount of money requested by the borrower


 funded\_amnt The total amount committed to that loan at time.


 funded\_amnt\_inv The total amount committed by investors at that point in time.

 term

 int\_rate The interest rate on the loan

 installment

 grade

 sub\_grade

 emp\_title

 emp\_length

 home\_ownership

 annual\_inc

 verification\_status

 issue\_d

 loan\_status

# Hypothesis

**Hypothesis:** Annual income and home ownership will be the best determinants for whether a borrower defaults on their loan.

**Challenges:** There are a sh\*t load of columns to go through and we haven't had practice working with such a large data set. One potential data flaw is that people that seek out personal loans might be inherently under a lot of financial pressure.