



Project 3: Mortgage Default Prediction

Team members: Lisa Cannon, Brian Labelle, Oswald Vinueza

Data: Fannie Mae Single-Family Acquisition and Performance data. Files are released quarterly but contain monthly information. Each of the quarterly files (Acquisition and Performance) contains information on loans that originated in that quarter and all history to the most recent quarter.

Fannie Mae is the nickname of FNMA, the Federal National Mortgage Association. Fannie Mae was established in 1938 by congress as part of the New Deal to stimulate the housing market by making mortgage more attainable for low- and middle-income families. Fannie Mae does not originate loans, but it does back or guarantee them.

Goal: The goal of this project is to use Fannie Mae's mortgage performance data to make predictions about a given customer. We want to forecast 1. the probability a borrower will go into default in the next quarter, 2. the probability a borrower will not pay their next mortgage payment, 3. how long until a borrower goes into default, and 4. how these trends vary geographically.

Proposed Methods:

1. The probability a borrower will go into default in the next quarter.
 - a. Logistic regression on default indicator
 - b. Other ML techniques
 - c. automate selecting best model—display optimal model, graphics relevant to chosen model
 - d. Chose state to build state level model
2. The probability a borrower will not pay their next mortgage payment.
 - a. Logistic regression on missed payment indicator
 - b. Other ML techniques
 - c. automate selecting best model—display optimal model, graphics relevant to chosen model
 - d. State level
3. How long until a borrower goes into default.
 - a. Hazard model
 - b. State level
4. Forecast page—enter your info and get probability of default and time to default
5. How these trends vary geographically—pretty map picture