

Mortgage Default Visualization and Risk Scoring

Bo Chen, Cody Nguyen, Daniel Mower, Harris Ashraf,
Ryan Wong, Travis Jefferies, Sharmila Baskaran

CREATING THE NEXT®

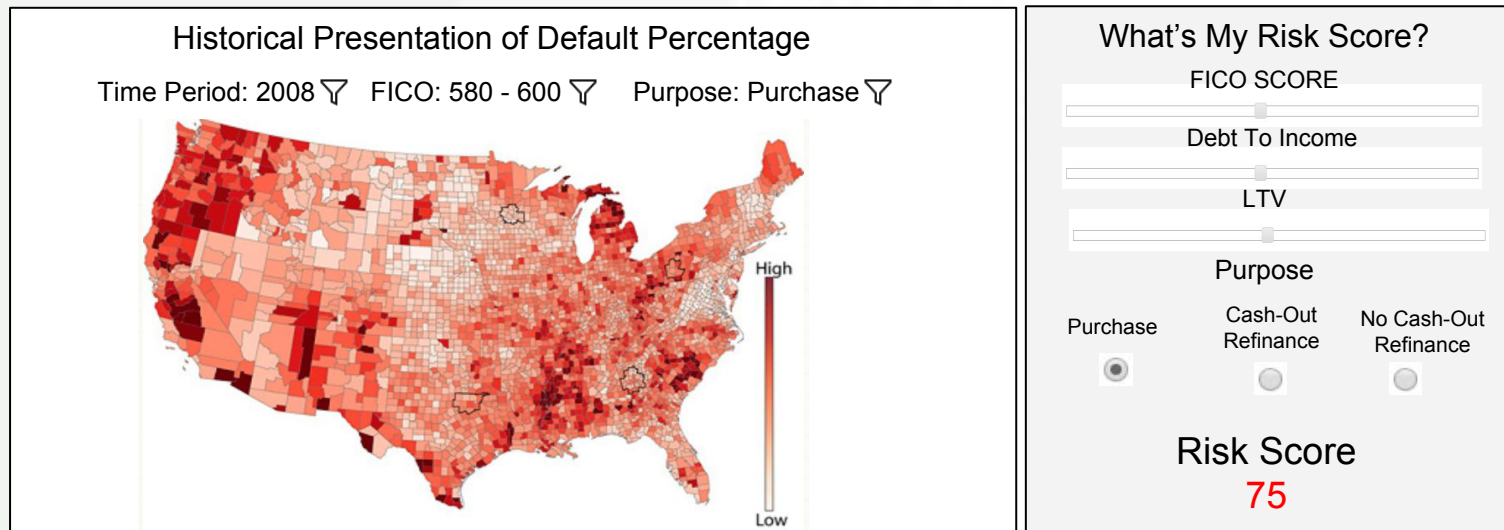
A series of parallel orange diagonal lines extending from the right edge of the slide.

Motivation

- Loan defaults can result in huge losses for banks.
- They were the cause of the 2007-2008 global financial crisis.
- Traditional approaches to predicting risk are subject to shortfalls.

Our Goal

- Create interactive visualization tools for users to view historical mortgage default characteristics and gauge their level of riskiness.



Literature Survey

- Analyzed the pros and cons of various prediction models such as K-Nearest Neighbors, Random Forest, and Deep Learning.
- Gathered information on different ways to visualize our findings.
- Researched efficient methods used to process large datasets.

How Is It Done Today? Limits?

- Banks employ large teams of analysts and economists.
- Use portfolio-level models to forecast 1-2 years out.
- Upcoming changes to *FASB rules will require modeling on an individual loan basis.

Our Approach

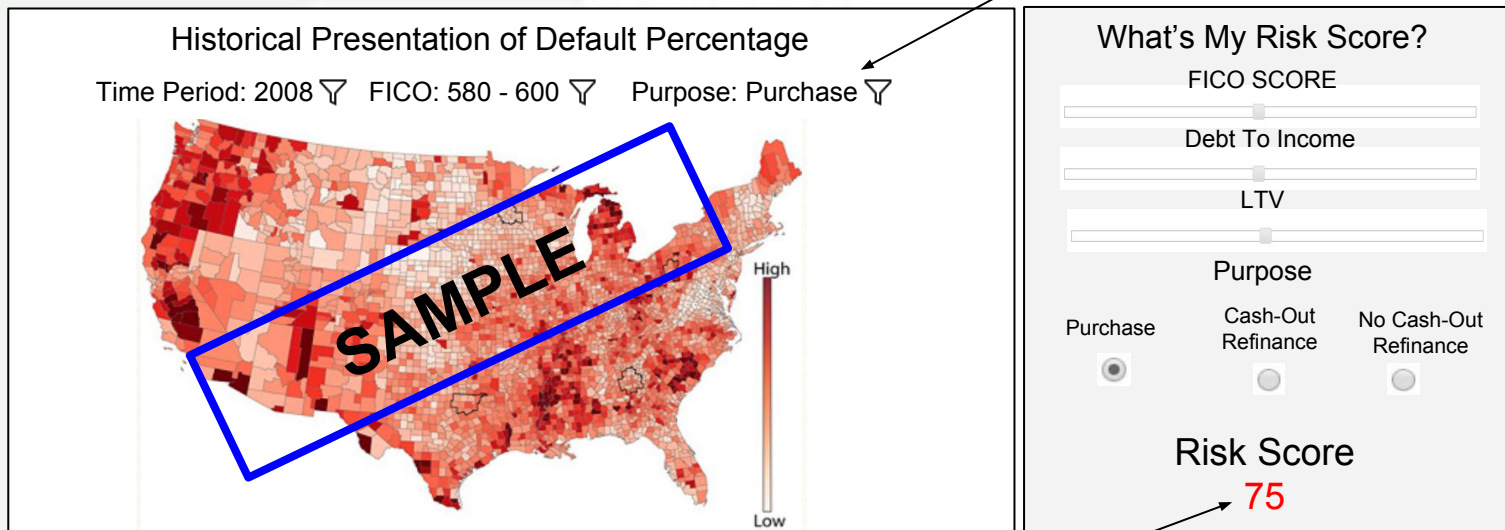
- 25 million loans from Freddie Mac and historical macroeconomic data from *FRED.
- Build a loan-level model for mortgage default probability.
- Present historical default statistics and default probability using a web based *UI.

Note:

- FASB: Financial Accounting Standards Board
- FRED: Federal Reserve Economic Data
- UI: User Interface

Our Approach Cont'd

Users can change filters to explore different segments.



Users can examine their risk rating.

Our Approach Cont'd

- Who Cares - Financial institutions, researchers, and home buyers.
- Impact - Provide banks with a risk rating tool and support in adhering to *CECL.
- Measure of Success: An informative and interactive visualization tool based on a stable and accurate model.
- Risks - Accuracy and time.
- Payoffs - Allow our users to vet their loan opportunities with less cost and time.
- Cost - None (open source and free credits if using cloud computing).
- Time - 45 days.

Note

- CECL: Current expected credit loss standard

Milestones

- Milestone 1
 - Data Cleaning, Feature Extraction
 - Preliminary Data Analysis and Visualization For One State
- Milestone 2
 - Fine Tuning Data Analysis
 - Final Visualizations For entire United States

Tools

- Hadoop - clustering
- Python (pandas, numpy, sklearn) - modeling
- SQLite - database querying
- Javascript (D3, React), Tableau - interface
- Selenium - web scraping

Group member contributions

- Work distribution should be equal (~14% per person), using multiple teams
- Web scraping: Sharmila
- Data modeling: Daniel, Travis, Sharmila
- Visualization: Ryan, Bo, Cody
- Clustering/Cloud: Harris, Bo, Cody