

Insights from Fannie Mae Single-Family Loan Dataset



Capstone Project I: Data Visualization

Springboard Data Science Career Track | Nikki Seegars



DATASET OVERVIEW

Fannie Mae has made available on their website a subset of the single-family loans that were acquired by the agency since 2000.

The dataset is available in .txt format by quarter and each quarter has the following two types of datasets:

- Acquisition Data (identifying data i.e. loan type, borrower credit score, original interest rate)
- Performance Data (monthly data i.e. current loan balance, delinquency status, loan age)



DATASET WRANGLING

The dataset is pretty well organized in the text files. However there were a few steps that were required to clean the data including:

- Adding header names
- Removing rows with missing data for borrower credit score
- Filling in number of borrowers with either 1 or 2 depending on credit score data
- Combining the two dataset files into DataFrame and saving to a text file for later use



DATA EXPLORATION

Once the data was cleaned and saved, the summary data was explored in search of discovering trends and insights about the dataset.

Reading through the supporting documents on the Fannie Mae website resulted in concentrating on exploring the following fields:

- Borrower Credit Score, Co Borrower Credit Score
- Zero Balance Code
- Original Interest Rate
- Original Unpaid Balance
- Original Loan to Value
- Original Debt-to-Income Ratio

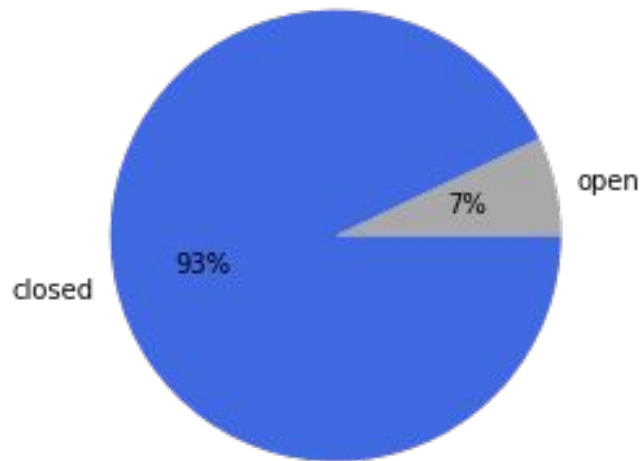


DATA VISUALIZATION

The 4th quarter of 2008 data set has nearly 10 years worth of records. Approximately 93% of those accounts are currently closed.

Total # of Accounts ~ 315,000

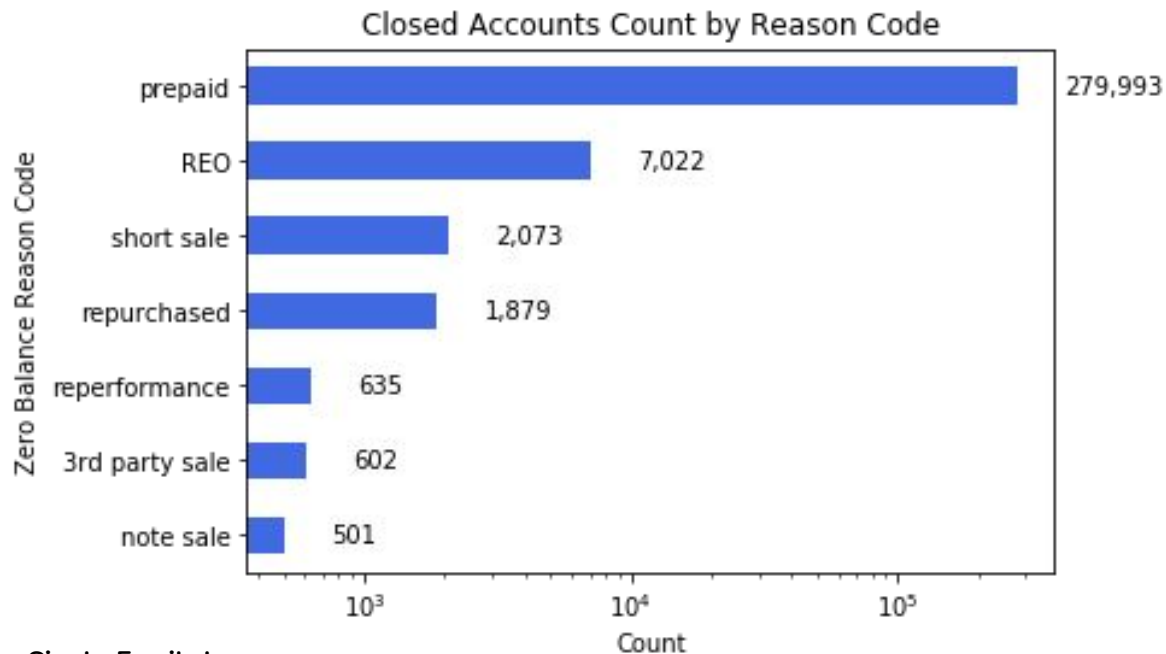
Status of Loan Accounts





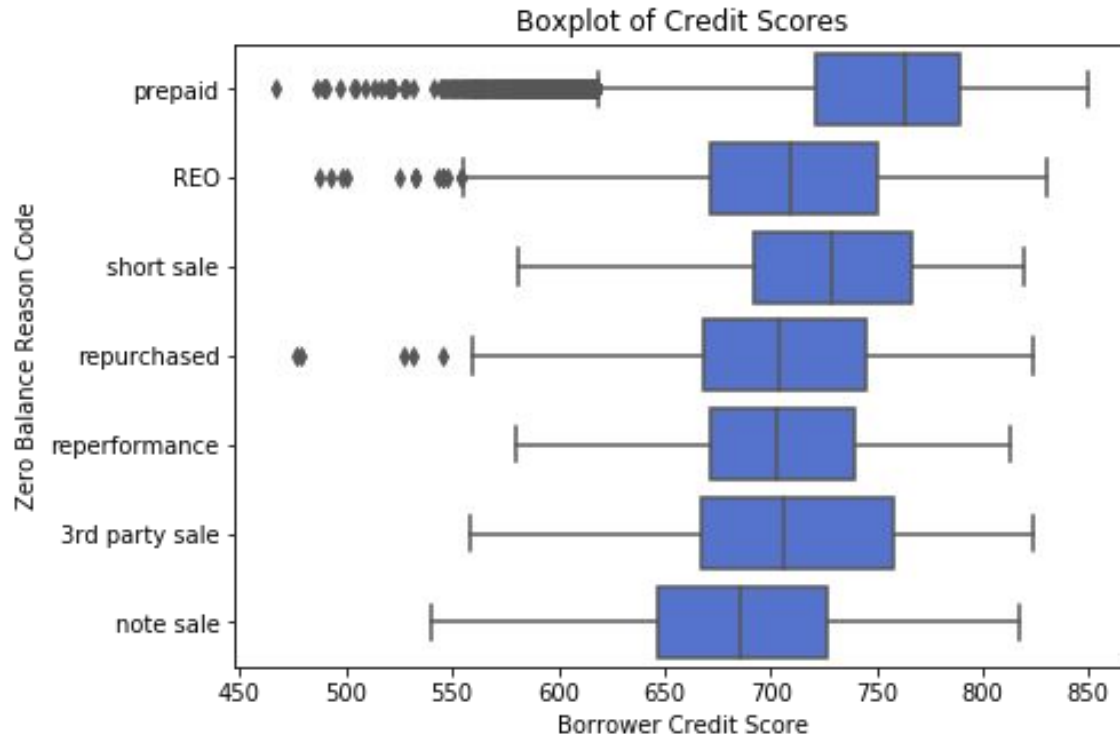
DATA VISUALIZATION

Of the 93% of accounts that are closed, most were prepaid (or matured).



DATA VISUALIZATION

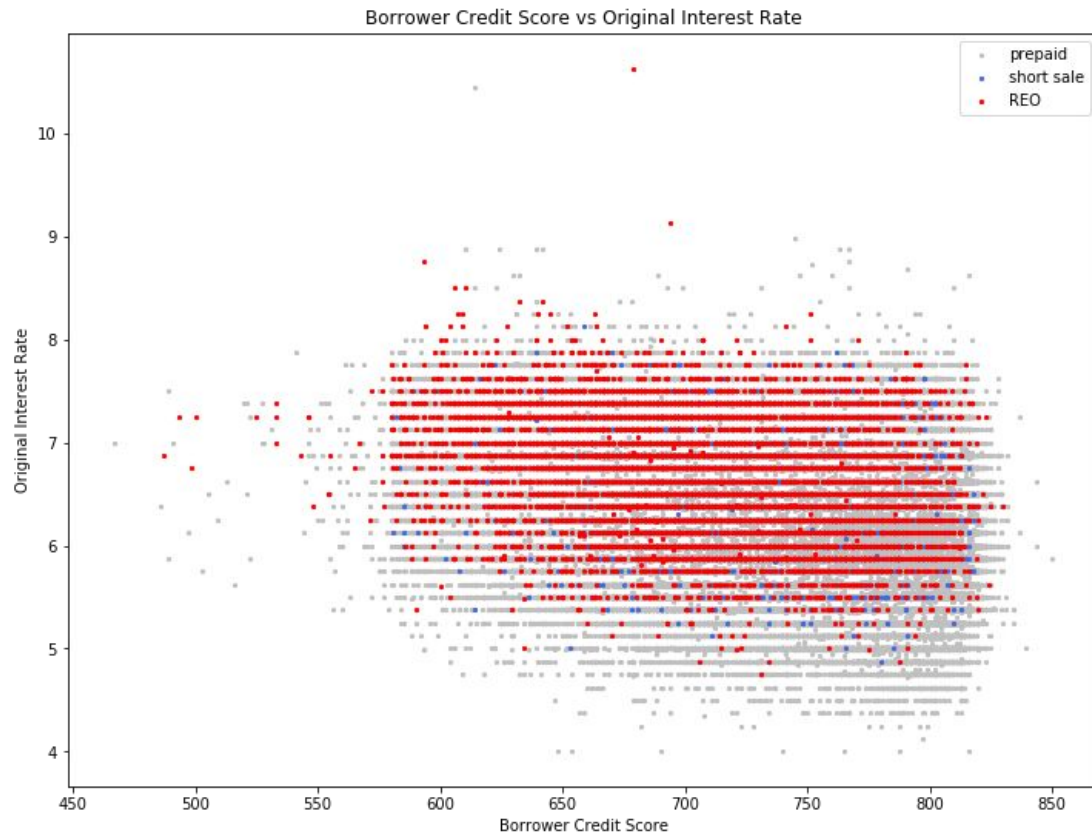
- ❖ With a larger portion of the data, the prepaid accounts have a credit scores with a larger variance and higher medians.
- ❖ Accounts closed coded as REO had significantly lower mean.



DATA VISUALIZATION

Narrowing down the data focus to those accounts that were either in default by borrower or sold short, comparison can be made to the larger subset of prepaid loans.

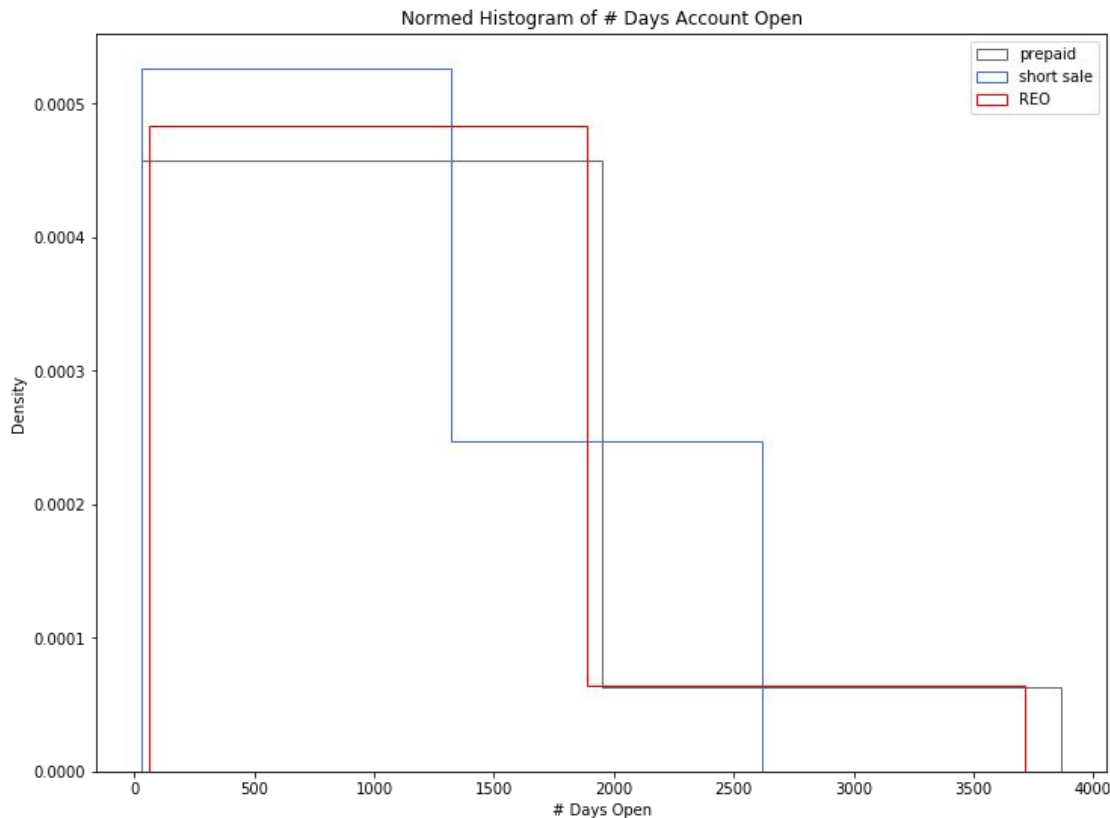
Accounts with lower credit scores tend to have only a slightly higher interest rate if any at all.





DATA VISUALIZATION

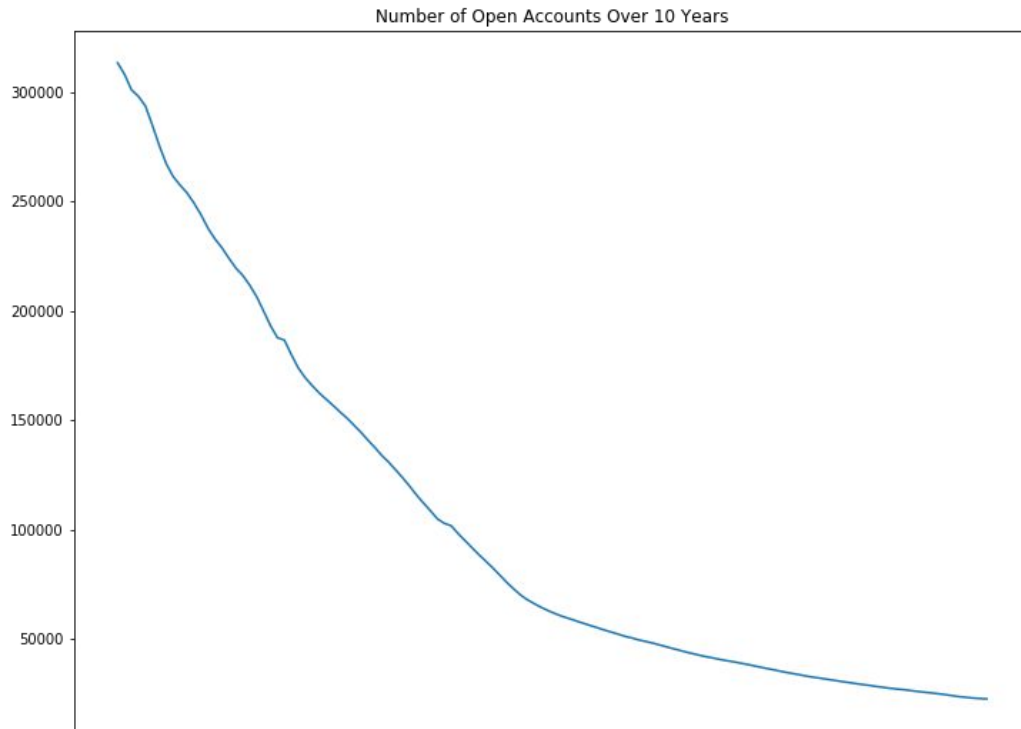
The vast majority of all the accounts are closed within 2,000 days ~ 7.5 years of the loan origination date.





DATA VISUALIZATION

Over 80% of loans are closed within five years of the origination date.





CONCLUSION

Preliminary data exploration suggests that Fannie Mae mortgage loan accounts

- Close out early, usually within five years of the origination date.
- Borrower credit score is somewhat in line with original loan interest rate, however there is not a strong correlation.
- Most of the loans close with borrower prepaying however there is a small portion of loans that default. These loans that default do tend to have lower median borrower credit scores.