An Analysis of Mortgage Finances for 2007-2017

Jason “Scott” Person

Abstract

This project used the Home Mortgage Disclosure Act (HMDA) data from the United States Consumer Financial Protection Bureau. We aggregated the loan amounts to determine if the loan to income ratio had dropped since the financial crisis. The aggregate mean was calculated by year and graphed. In 2017 loan to income ratios exceeded those for 2007.

Motivation

I wanted to evaluate the trend of mortgage loan to borrower income over time. Debt to income is a common evaluation metric for lenders. It follows that the ratio would be indicative both of how affordable housing is as well as provide insight into the overall health of the mortgage market. Specifically, comparing recent loan to income ratio values with those in 2007-2008 (the height of the mortgage crisis) would be quite interesting.

Those considering investing in a home or mortgage related assets would be interested in this analysis. Policy makers at the Federal Reserve undoubtedly perform similar analysis.

Dataset(s)

This project uses the Home Mortgage Disclosure Act (HMDA) data provided by the Consumer Financial Protection Bureau (CFPB). The particular dataset used consists of all originated home mortgages in the United States for the years 2007 to 2017. It is a 52GB dataset with approximately 72 million records of which 62 million had data appropriate for analysis.

Data Preparation and Cleaning

Data preparation comprised the majority of the time available for this project.

The data was downloaded as zip files (one per year) to a Linux instance using the curl command. The zip files were extracted one at a time and moved to S3 using the AWS Command Line Interface (CLI).

As previously stated, the dataset is somewhat large at 52GB. To keep costs down, AWS Athena was used to query the data from S3. This allowed Athena to do much of the work and kept memory requirements for the notebook to an affordable level.

Research Question(s)

What is the historical trend of mortgage loan to income ratio? How does the recent value (year 2017) of this ratio compare with that of 2007-2008?

Methods

The data was aggregated using the mean function for each year for loan and income values. The mean was also calculated for the loan to income ratio after calculating the value for each detail record.

Graphs were then produced to allow visual comparison of the values.

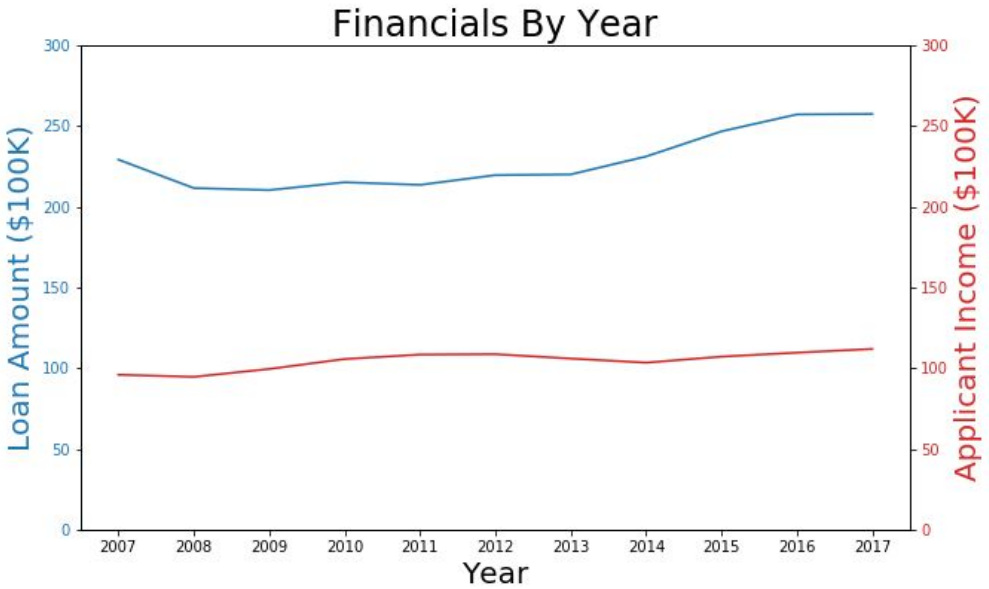
Jupyter Notebook

The Jupyter Notebook will be uploaded as a separate PDF.

Findings - Financials

The mean loan amount dipped considerably between 2007 and 2008, leveled off, then rose considerably between 2013 and 2016.

Interestingly, mean borrower income has been largely flat from 2007 to 2017. The lack of a significant dip is probably due to the nature of the records. Workers who lost jobs or otherwise had a large decrease in income are unlikely to apply for and receive new mortgages.

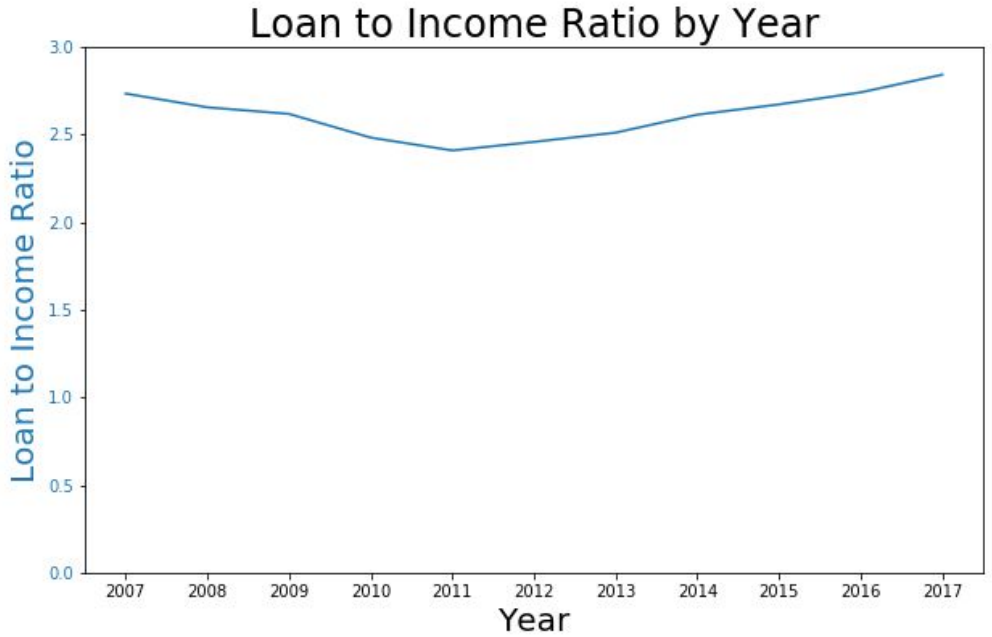


Findings - Loan to Income

Although apparent from the previous chart, the calculated loan to income ratio for the most recent data now exceeds that in the run up to the housing crisis.

This indicates that housing is now less affordable than in 2010-2013.

Note: Full sized charts and aggregated details are available in the supplemental slides.



Limitations

The data used for this project is quite detailed. There is a record for every home mortgage made in the United States between 2007 and 2017. The limitations of this analysis was primarily time. The size of the dataset required much more preparation and a plethora of tools to make it available. There are many features (race, location, co-borrowers etc.) that I would like to explore and will in further study.

Conclusions

The recent mortgage loan to borrower income for mortgages originated in the United States exceeds that from the height of the home mortgage crisis. Further analysis to determine actual affordability - overall population income to housing costs is certainly warranted.

Acknowledgements

Data Source: “Download Historic HMDA Data.” Consumer Financial Protection Bureau, www.consumerfinance.gov/data-research/hmda/historic-data/.

I have done extensive personal research in the past on personal finance and related topics.

No one reviewed this presentation.

References

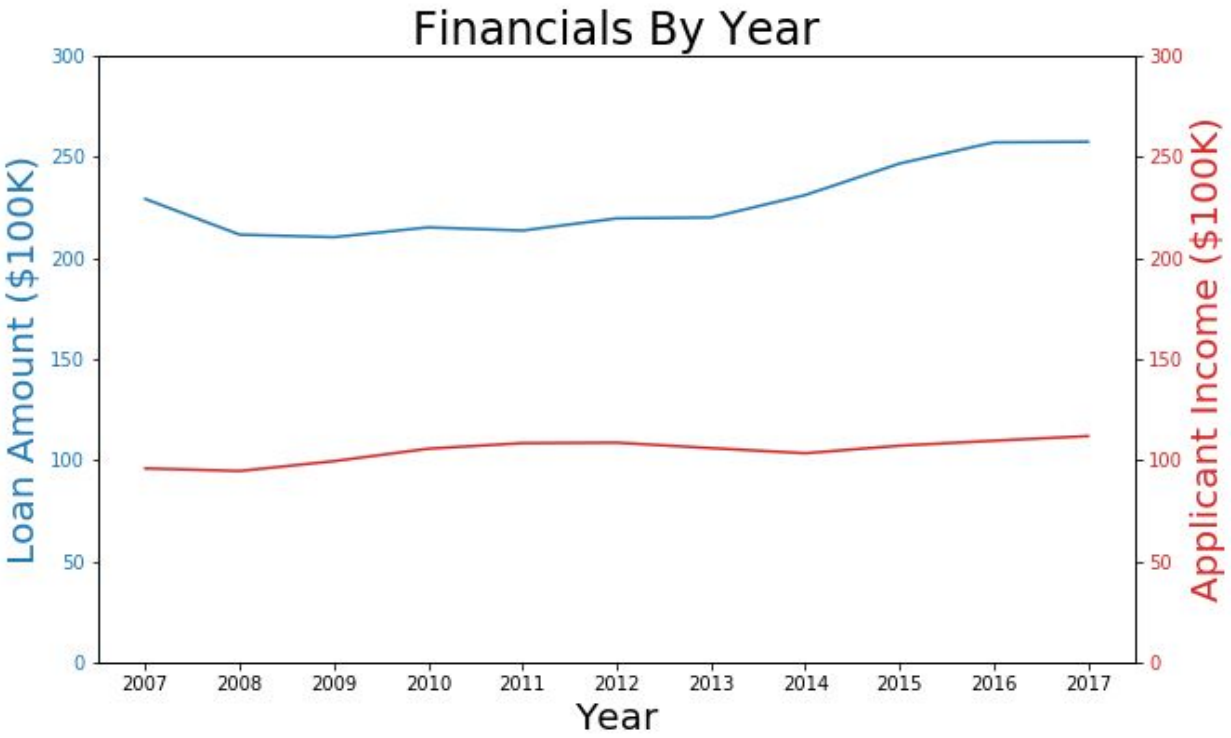
[1] Kandalam, Anjana. “Run SQL Queries from Your SageMaker Notebooks Using Amazon Athena: Amazon Web Services.” *Amazon*, Amazon, 13 Sept. 2018, aws.amazon.com/blogs/machine-learning/run-sql-queries-from-your-sagemaker-notebooks-using-amazon-athena/.

Supplemental Slides

Mean Data for Each Measure by Year

as_of_year	loan_amount_000s	applicant_income_000s	rate_spread	loan_to_income_ratio
2007	229.069417	95.931398	4.675704	2.732374
2008	211.4527	94.634002	4.214035	2.654111
2009	210.212739	99.505074	4.038547	2.616449
2010	215.051597	105.664634	2.499408	2.481347
2011	213.466322	108.402764	2.479931	2.408148
2012	219.481115	108.655808	2.478034	2.457055
2013	219.920301	105.955652	2.14503	2.509665
2014	231.035081	103.415194	1.992665	2.612156
2015	246.576477	107.138255	2.038234	2.670622
2016	257.072973	109.603304	2.013937	2.73911
2017	257.334154	111.901239	2.02135	2.84

Full Sized Graphs - Financials



Full Sized Graphs - Loan to Income