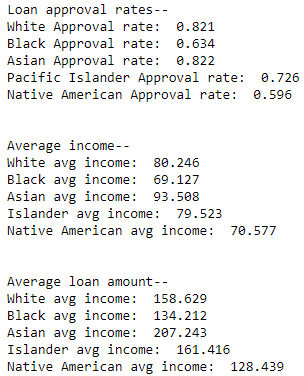
Dan Clayton

DSC-530 Final Project

This project began with the goal of determining if racism was taking place in the mortgage lending process. Specifically, the goal was to use Home Mortgage Disclosure Act data available [here](https://www.consumerfinance.gov/data-research/hmda/historic-data/) to see if different lending outcomes were experienced by loan applicants because of the color of their skin (all other things being equal). The analysis does confirm that different lending outcomes are taking place for applicants of different skin colors, however the data is insufficient to say whether this is because of the applicant’s skin colors, or because of other factors. Unfortunately, the data set can only answer part of the question, because it fails to address each applicant’s credit score, outstanding debts, monthly debt payments as a portion of income, and down payment amount all of which would be vital inputs in making a lending decision. The addition of these missing variables to the data set would have allowed for a true “apples to apples” comparison between similarly qualified applicants of different colors.

The hypothesis, that minority applicants are denied mortgages more than their white counterparts was partially confirmed, but only in the most rudimentary of statistical measures. As the summary statistics below indicate, the loan approval rate is much lower for applicants of color, particularly for Native Americans and blacks. It is noteworthy, however, that Asian applicants have a higher approval rate than white applicants, although perhaps less surprising given that Asian applicants have the highest average income of any group sampled.

**Summary Statistics:**



One challenge was cleaning and appropriately grouping the data. For example, the data provided 8 different loan statuses, ranging from “Approved but not accepted” to “File closed for incompleteness”. For the purposes of this analysis, only applications that were either approved or denied were considered relevant, so applications with loan status of “Loan originated” and “Application approved but not accepted” were both lumped together into a new “Approved” category. Only “Approved” and “Denied” applications were included in the final analysis.

If I were a federal regulator attempting to make policy decisions based on this data set, I would request that going forward additional data points including “FICO score”, “down payment”, “total outstanding debt”, and “monthly debt service amount”, along with any other relevant factors in making a lending decision, be included. This would allow a more rigorous analysis answering not only if there are different lending outcomes (which are essential to running a lending business), but if those outcomes were because of the applicant’s race (which would be totally illegal, and immoral).