As students we are lucky enough to be entitled to loan access even without a lengthy credit history or financial means. In the real world, Loan approval is a hassle. Most people are not even approved. In the sample data set,480/5000 had a personal loan approved, which is less than 10% of the respondents.

My initial hypothesis was that income would be the determining factor. I went to Kaggle [1](https://www.kaggle.com/datasets/vikramamin/bank-loan-approval-lr-dt-rf-and-auc) and downloaded a data set for personal loans. It was pretty evident from the start that indeed Income was the biggest factor in loan approval. Although data of the zip codes are important to lenders, I felt that leaving them out of a regression model was Important as it is hard to prepare the data properly.

The dataset, although helpful, is missing some factors that I would have been interested in exploring. Was gender or race a factor in the loan approval process? Did any of these borrowers have another personal loan besides the one they just got approved for? Income was a huge factor, but I would love to have known what kind of data was being used to determine income. Was it previous year income or based off current W2 pay stubs?

Another question that needs to be further explored is what other factors contribute to loan approval. Only 50% of loans for Incomes over 150K were approved. In a further analysis we should investigate what the contributing factors are.

Before I analyzed the data, I assumed that as your income improves, so do your chances of getting approved for a loan. It does however seem that once you hit over 100K the chances of approval are nearly identical.

[[1]](#endnote-1)

1. Resources

   1.https://www.kaggle.com/datasets/vikramamin/bank-loan-approval-lr-dt-rf-and-auc [↑](#endnote-ref-1)