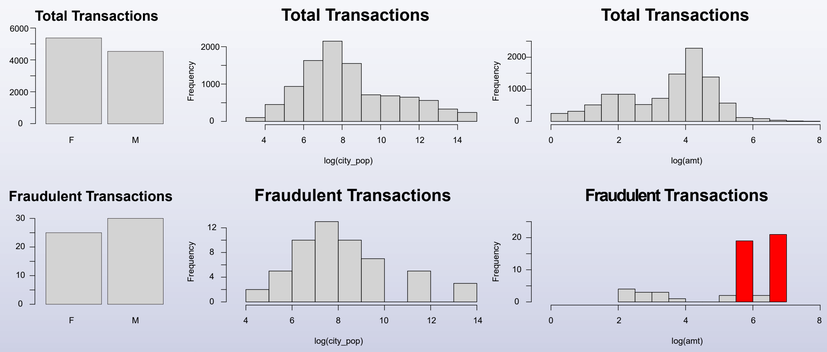
# Can We Predict Credit Card Fraud?


## Report by Melissa Mosier
## for IST 719 at Syracuse University



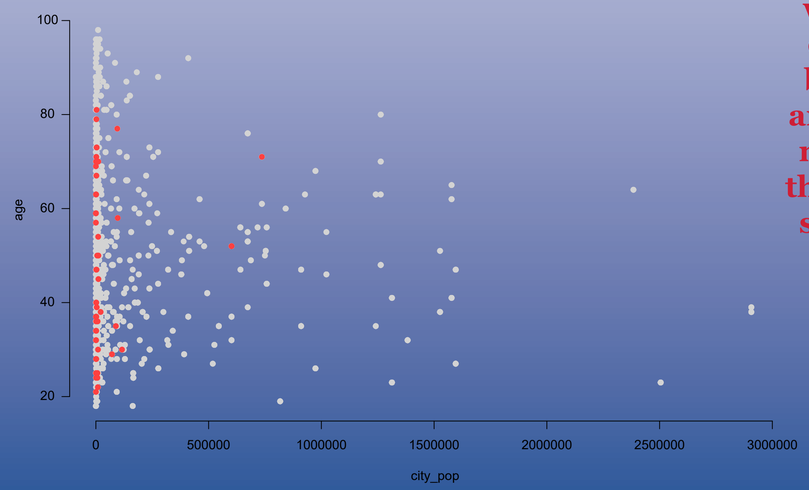
This “Credit Card Fraud Detection” dataset acquired from Kaggle contains 10,000 credit card transactions, including the location, amount, date of the transaction, and information about the person who owns the card. We have some transactional and demographic variables that might be related to fraud. Are fraudulent transactions more likely for certain groups of people - the elderly, city folk, or a specific gender? Can we use those trends to predict when someone might become a victim of credit card fraud?

**Preliminary analysis did not show any trends across gender, age or size of the city where the transaction took place. Transaction amount for fraudulent purchases, however, do seem to differ from your usual transaction.**



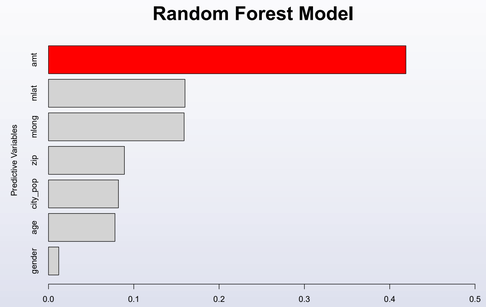

**If a charge had 6 or more zeroes - the amount exceeded hundreds of thousands of dollars - it is likely a fraudulent charge.**

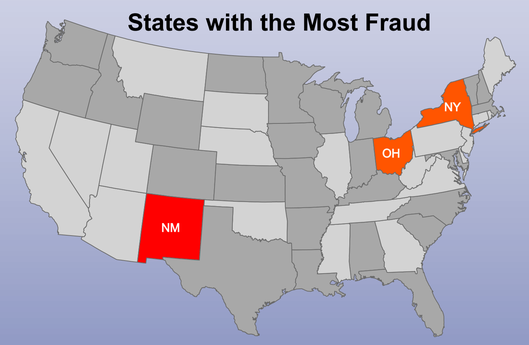
**We see no significant trend across age (y-axis) or city population (x-axis).**

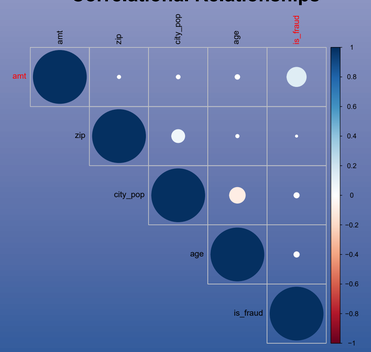



**There were states with more incidents of fraud, but correlation and decision tree models show us that those are not significant relationships.**

**Creating a predicitve model gave us the variables most likely to indicate fraud. The best predictive variable is amount.**

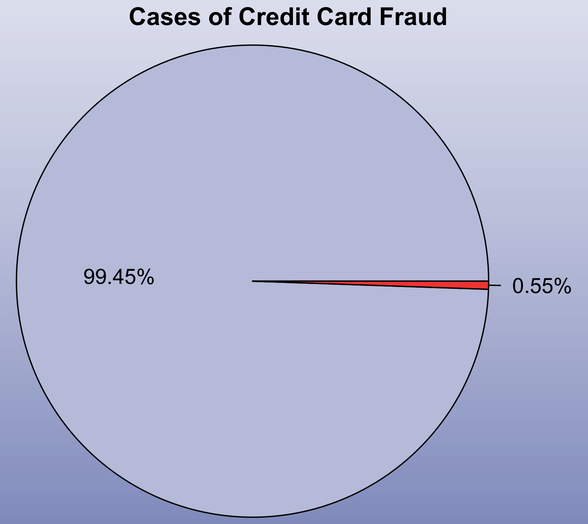



**So to answer the question:**
**Not really.**
**Credit card fraud is extremely rare.**

**Fraud only makes up 0.55% of all transactions. Of those transactions, our best indicator that it is fraudulent is the amount of the purchase.**




This dataset required minimal cleaning and had 23 variables after cleaning. Transformations included converting “date of birth” variable to age and creating dummy variables for gender.