Analyzing Patterns of Deception: A Comprehensive Analysis of Credit Card Fraud Detection Through Data Science

Prepared by

Ankit Malhotra, Hong Che, Nathaniel Zhu

STAT 232 Final Report

University of California - Riverside

**Introduction**

Background of Data

In this modern time, the phenomenon of credit card fraud has evolved into a sophisticated challenge that plagues financial institutions worldwide. Nowadays, especially after the pandemic, more people rely on online transactions, such as using credit cards on websites, shopping on Amazon or any other platform, and transferring money using online banking. The opportunities for fraudulent transactions have increased, leading to significant financial losses and eroding consumer trust. It has come to our attention that we have seen a lot of news online about people stealing other people’s credit card credentials by secretly recording their information and using them on their own.

This research delves into the complexities of credit card fraud by leveraging a rich dataset that includes multiple important variables. This dataset enables us to further test and train our theories while doing our research. These variables testify to fraudulent transactions' intricate patterns and behaviors, providing a solid foundation for analysis, understanding, and research.

Variable of Interest

The dataset at the heart of this study is characterized by a wide array of variables, ranging from transaction details such as amount, time, date, and location (using long and lat), to more detailed indicators that may signal fraudulent behavior. Among these, one variable is used to determine if the transaction is fraudulent: "is\_fraud,” with numeric 1 indicating positive fraud, while “0” indicates negative fraud. In the further analysis, we will also use feature engineering to further expand our numeric variables to train and test our hypothesis. The exploration of these variables not only aids in identifying the potential of fraudulent transactions but also enriches our understanding of the operational tactics employed by fraudsters.

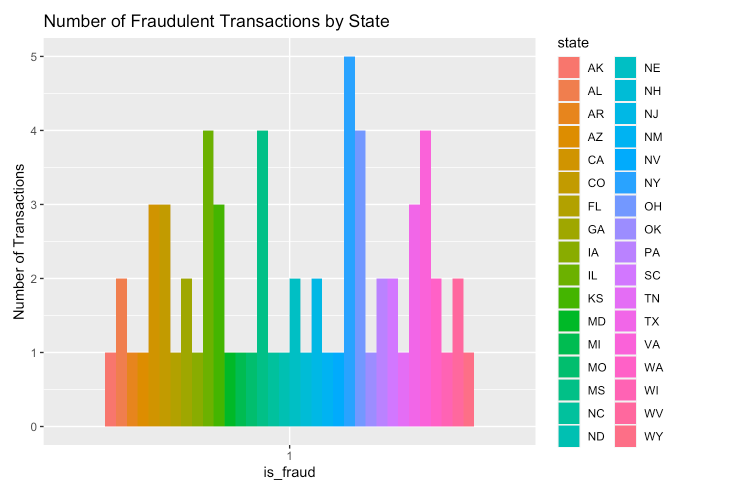
Business Operations:

This research is driven by a series of compelling business questions that seek to answer and illuminate the dark corners of credit card fraud. At this core, the study aims to answer the following research questions:

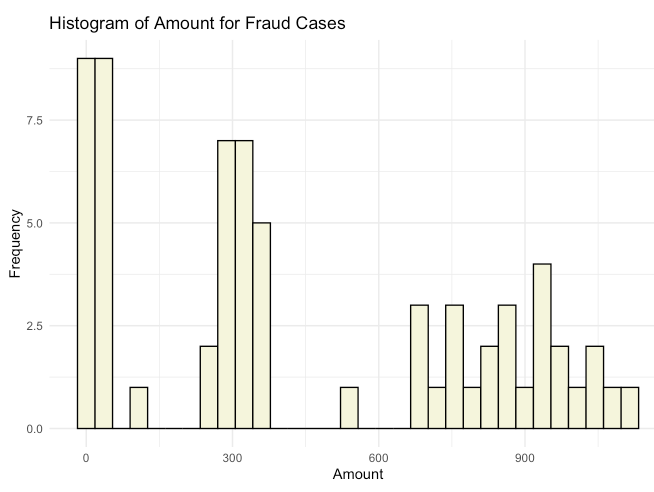
1. What are the distinguishing characteristics of fraudulent transactions as opposed to legitimate ones?
2. Can transactional patterns and behaviors predictive of fraud be accurately identified and quantified?
3. How can financial institutions leverage predictive modeling to identify, reduce, or mitigate fraud transactions?
4. What are the implications of fraud detection mechanisms on consumer trust and financial security?

Answering these questions not only contributes to the academic discussion about financial fraud but also provides actionable insights for businesses dealing with the threat of credit card fraud. This research strives to provide a beacon of hope in the fight against credit card fraud through a multi-staged and complex exploration of the dataset and the application of data science methods.

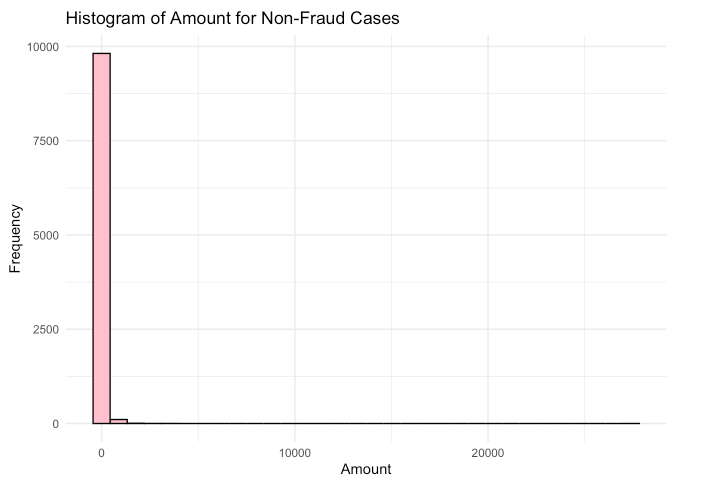
**Exploratory Data Analysis**



The bar chart shows a comparative view of fraudulent transactions across states. Each state is represented by a different color, making it clear at a glance which states have higher rates of fraud. The "is\_fraud" indicator is binary, with "1" representing a fraudulent transaction. It's worth noting that certain states have a higher frequency of fraud, which may indicate targeted areas or areas with weaker security measures. This visualization will serve as the basis for analyzing state patterns and developing region-specific anti-fraud strategies.



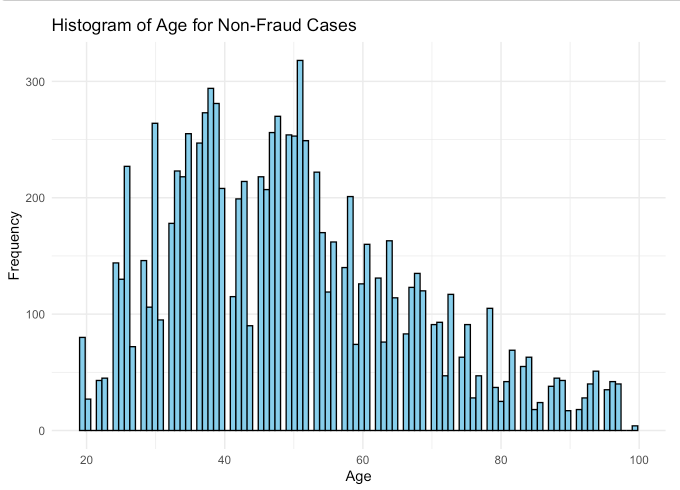
This histogram illustrates the distribution of transaction amounts for fraud cases. The x-axis represents the transaction amount, while the y-axis represents the frequency of transactions within specific amount ranges. Most fraudulent transactions cluster at lower amounts, suggesting that fraudsters may prefer smaller, less noticeable amounts to go around detection. This insight is crucial for adjusting fraud detection algorithms to be sensitive to transactions of varying amounts.



The histograms of non-fraudulent transactions show different patterns compared to fraudulent cases. Most legitimate transactions are concentrated in significantly lower dollar amounts. The distribution is heavily skewed to the left, indicating that the monetary value of regular customer transactions is generally smaller. The apparent difference in the distribution of amounts for fraud and non-fraud cases may indicate the behavior distinguishing legitimate transactions from fraudulent ones.



The age distribution of non-fraudulent transactions shows wide dispersion across age groups, with certain age ranges exhibiting higher transaction frequencies. This histogram provides insight into demographic patterns of credit card usage in the population and can serve as a baseline for understanding typical user profiles in the context of fraud detection.



The age histogram associated with fraudulent transactions presents a more diverse distribution than non-fraudulent cases, although the overall frequency is lower due to the relative rarity of fraud. Certain age groups appear to be more likely to be associated with fraudulent activity, or conversely, they may be the age group that fraudsters choose to impersonate. This data helps identify high-risk demographics, and fraud prevention measures may be developed accordingly.