Detecting Fraud in BTC Blockchain

Yael Brown

Overview

- Identify wallets and transactions to or from those wallets that were used in fraudulent transactions

- Fraudulent transactions are classified as transactions from victims in ransomware attacks, black marketplace payments for illegal goods.

- Having a model trained and actively reviewing transactions on the blockchain, this could aid law enforcement and tip them off to possible fraudulent transactions.

- This would beneficial especially if the blockchain scales or technology evolves to handle more transactions.

Bitcoin Overview

- 1 BTC = 1,000,000 Satoshi’s. (Satoshi’s are the smallest unit of Bitcoin)
- “Bc1qzyda53xqwkqruex3mzwvpja04x23r572mygpgfc90qckdw2cwwaqr2h70u” is an example wallet address. 62 letters and numbers.
- Bitcoin is purchased on exchanges and can traded with exchanges and wallets.
- All the transactions are transparent on the blockchain

Overview

- Collect transaction data, known fraudulent transaction data, and daily price (close) data.

- Combined to one data set.

- Train a Logistic Regression model on the data.

- Test it with possible transactions

Estimate of Impact of Solution

- To assist law enforcement in detecting fraudulent transactions

- Help trace bad actors to exchanges to help identify them when they try to cash out the fraudulent funds.

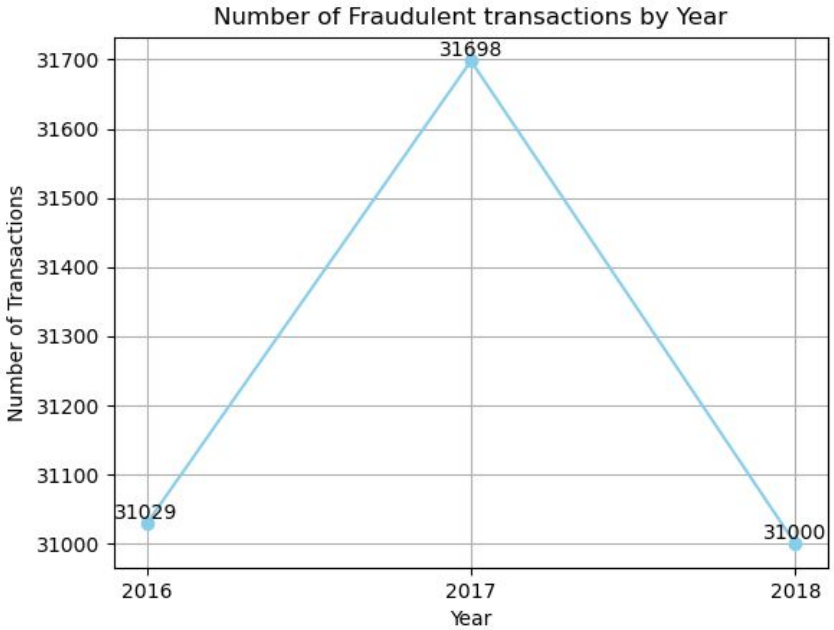
- Solution that is able to scale with the blockchain.

Intro to dataset

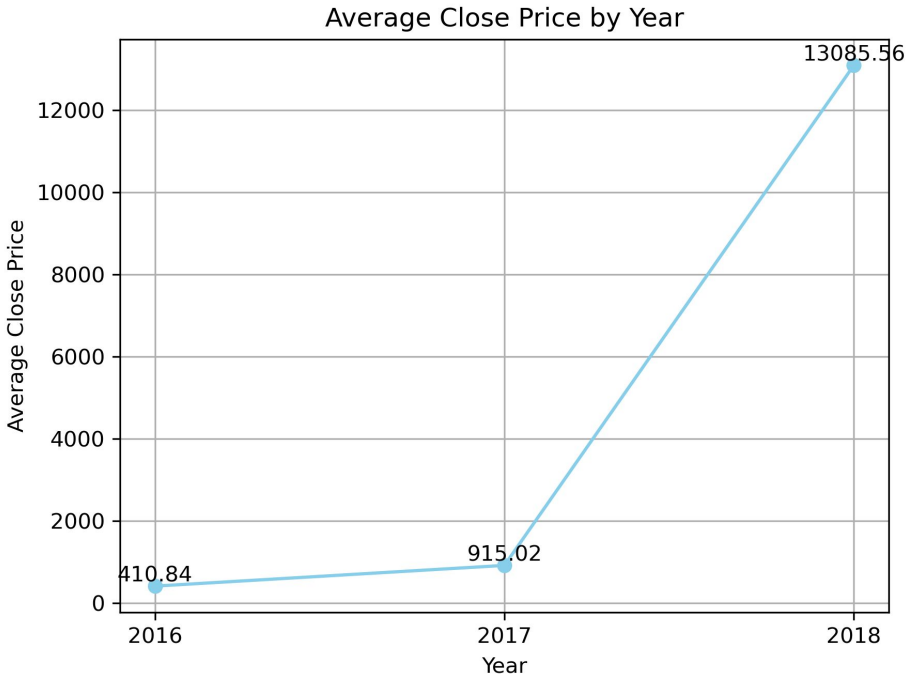
- Using transaction data, known fraud data, and coin price
- Combine the datasets and drop unnecessary columns from them.
- Normalize the date format. (a column for month, day, year)

- Transaction dataset is a lot (39 GB)
- Only 93k rows for fraud dataset

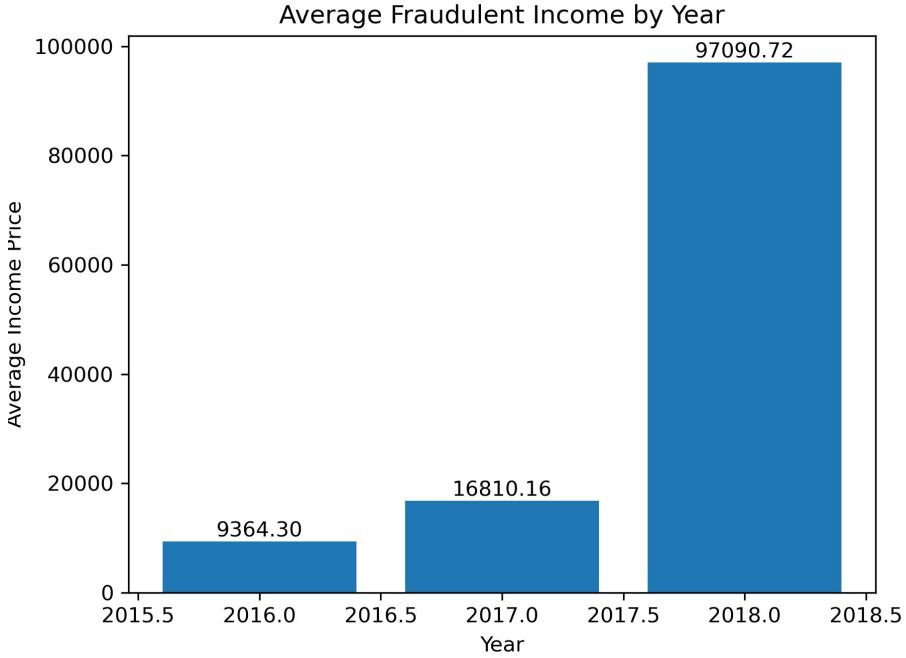
Number of Fraudulent Tx. in Dataset



Average Closing Price by Year



Average Fraudulent Income by Year



Next steps

- Finish downloading datasets
- Finish doing EDA and look for features

- Begin modeling phase

End