Detecting Fraud in BTC Blockchain

Sprint 2

Yael Brown

Overview

Detect possible fraudulent transactions on the blockchain (BTC)

Identify transactions with known and flagged addresses and amounts that will determine if the transaction could possibly be fraudulent

Fraudulent transactions used is Ransomware Transactions for PoC.

Software would be able to assist by notifying law enforcement of a possible transaction.

This is similar to detecting fraud transactions with credit cards

Overview

Download Bitcoin Transaction history from Harvard 2009 to 2020. (290 files, ~2 GB per file, 390 Gb total)

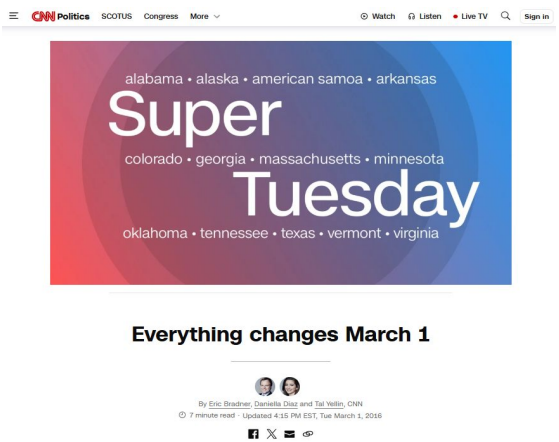
Bitcoin ransomware heist dataset from Kaggle (Year range is from 2016 - 2018)

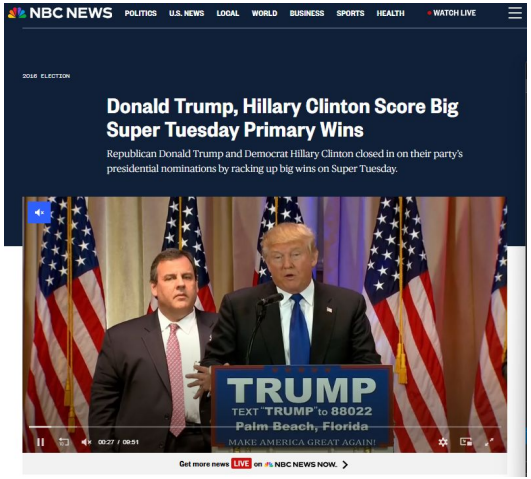
Sample transaction history to years 2016-2018 (43 Files) and sample ~200K transactions that range from March 1st to March 13st.

Insights

For this sample, most of the transactions were on the March 1st 2016.

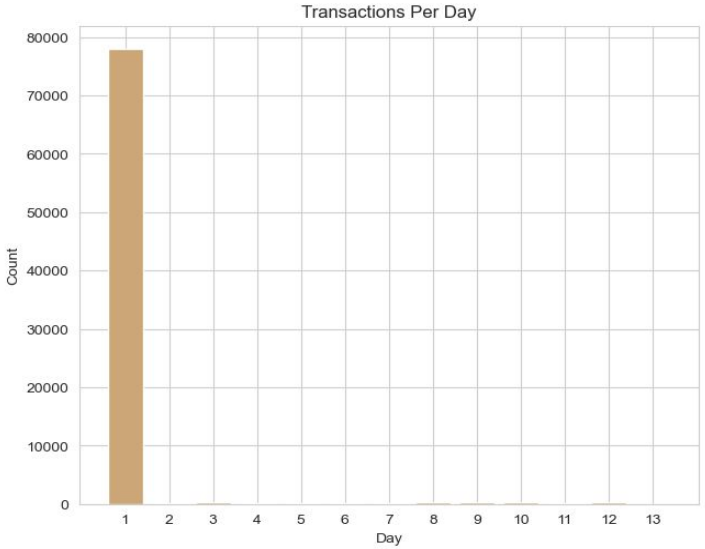
This is because this was the Primary election in the United States

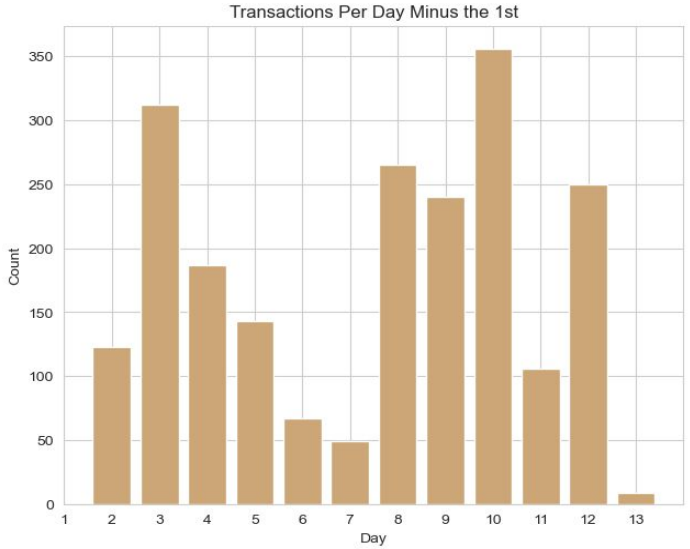




Insights

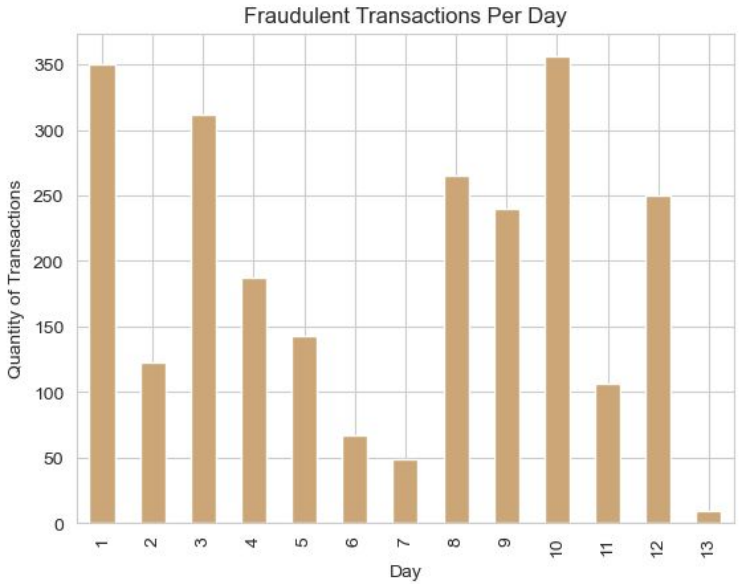
Political Events in the United States tend to heavily influence investor sentiment in crypto.





Insights

There was not an obvious amount of fraudulent transactions on the 1st despite the quantity of transactions on the first.



Insights

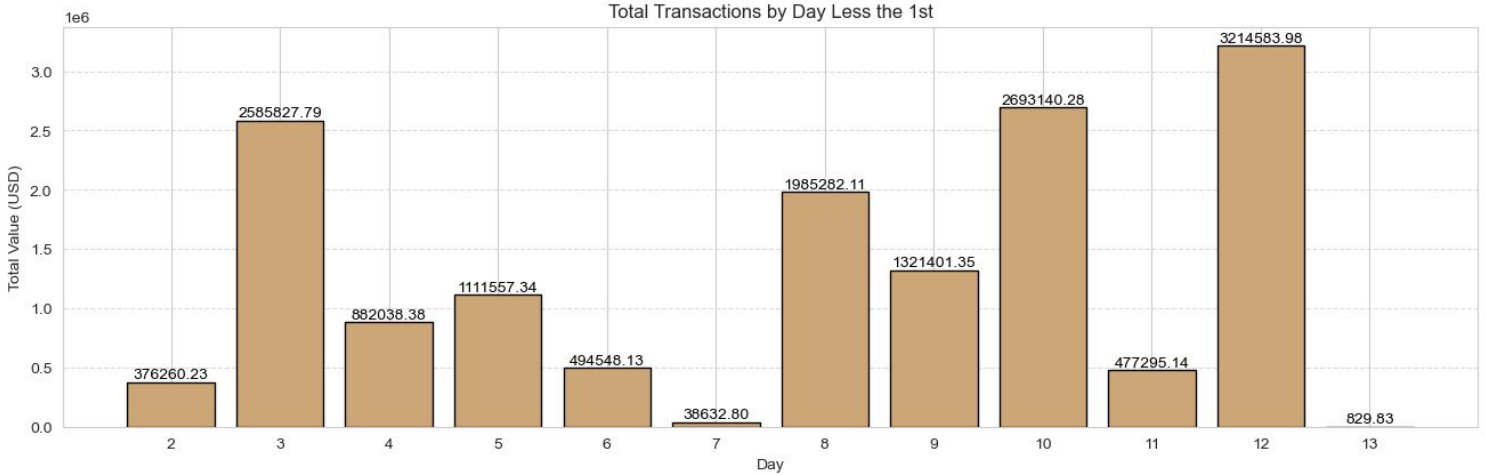
The value of fraudulent transactions for the sample data set is **$19,936,914.61** total.

On March 1st, the volume of transactions was **$445,208,718.30**

The average daily volume is **$1,265,116.44**

Insights


Total Transactions by Day
1e8
Total Value (USD)
Day
445208718.30
376260.23
2585827.79
882038.38
1111557.34
494548.13
38632.80
1985282.11
1321401.35
2693140.28
477295.14
3214583.98
829.83


Total Transactions by Day Less the 1st
1e6
Total Value (USD)
Day
376260.23
2585827.79
882038.38
1111557.34
494548.13
38632.80
1985282.11
1321401.35
2693140.28
477295.14
3214583.98
829.83

Model

Initial Model (Sprint 2) was very accurate.

Initial Model had **0.9996217851739788%** accuracy

With the new dataset and more data, **0.9996130803794309%**

This is a decrease of **.0000870479454 79%**

Model

Initial Model (Sprint 2) Precision, Recall, F1 Score, ROC-AUC scores:

Precision: 1.0000

Recall: 0.9353

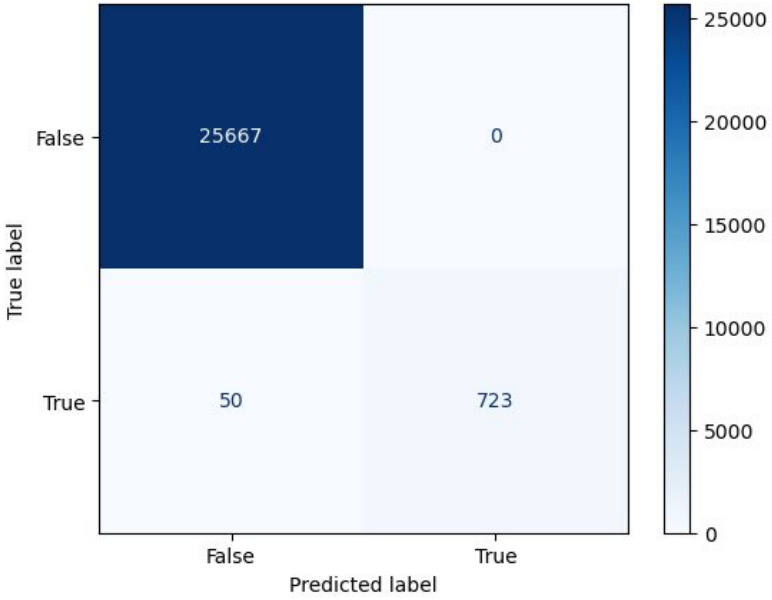
F1 Score: 0.9666

ROC-AUC Score: 0.9977

Model

Initial Model with Sprint 3 dataset.

Precision: 1.0000
Recall: 0.9353
F1 Score: 0.9666
ROC-AUC Score: 0.9977



Model

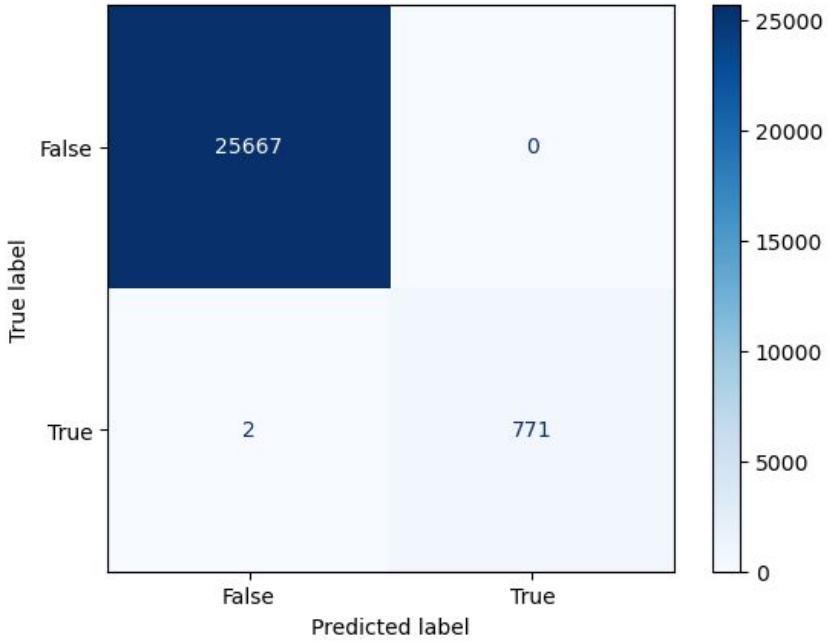
Final Model after Hyperparameter tuning

Precision: 1.0000

Recall: 0.9353

F1 Score: 0.9666

ROC-AUC Score: 0.9977



End