

# Detecting Fraud in BTC Blockchain Sprint 2



Yael Brown

# Overview

- I explained plan for acquiring data, merge data, and did preliminary analysis on all the data sets I was planning on working on.
- Settled with using 2 datasets; heist (ransomware) and transaction data (from Harvard)
- Collected data, filtered the data, then analyzing the sample subset.

# Data Collection

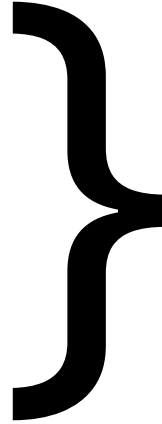


290 files  
~2 Gb per file  
390 Gb total

Date range: 2009-2020

Heist dataset is from  
2016-2018.

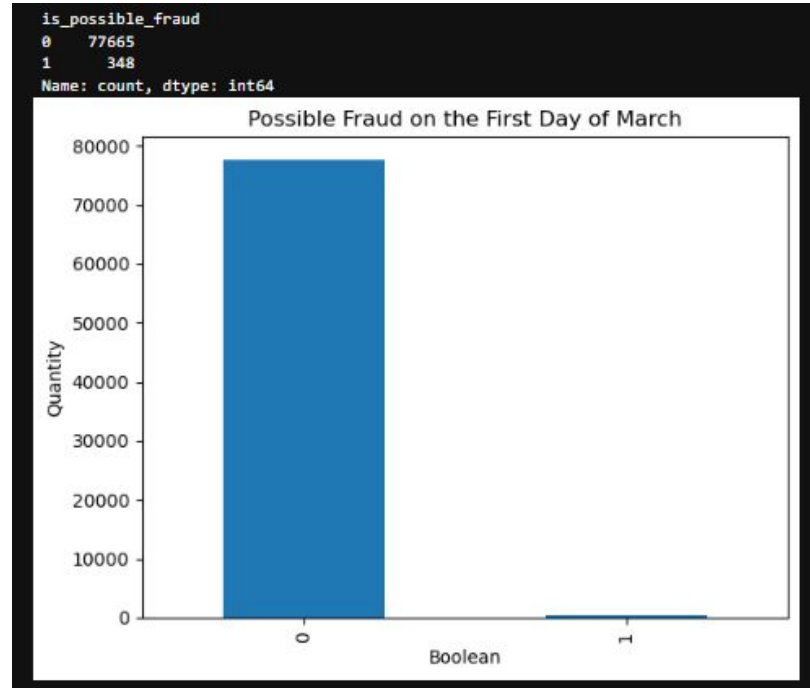
# Filter Data



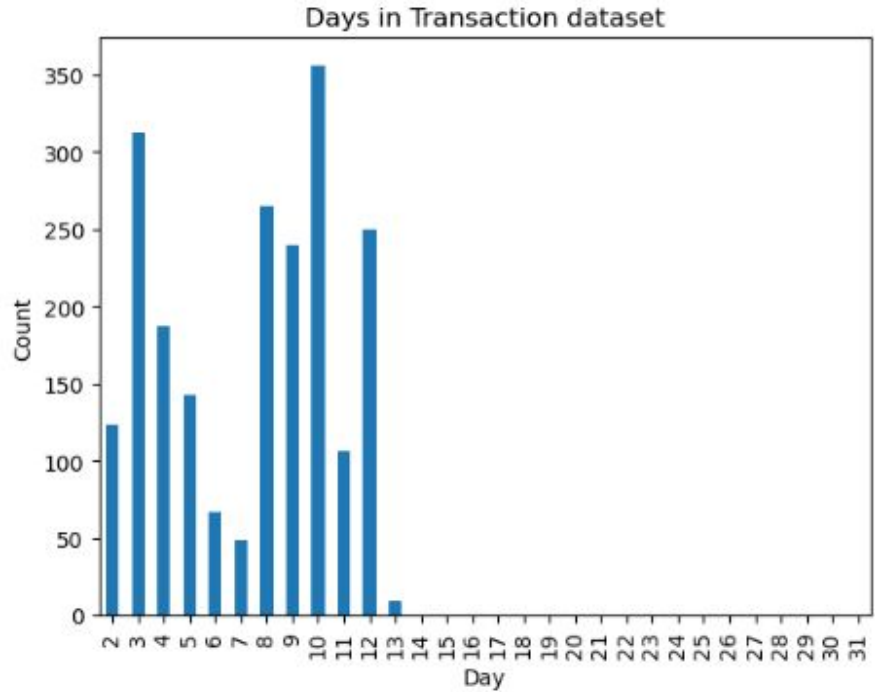
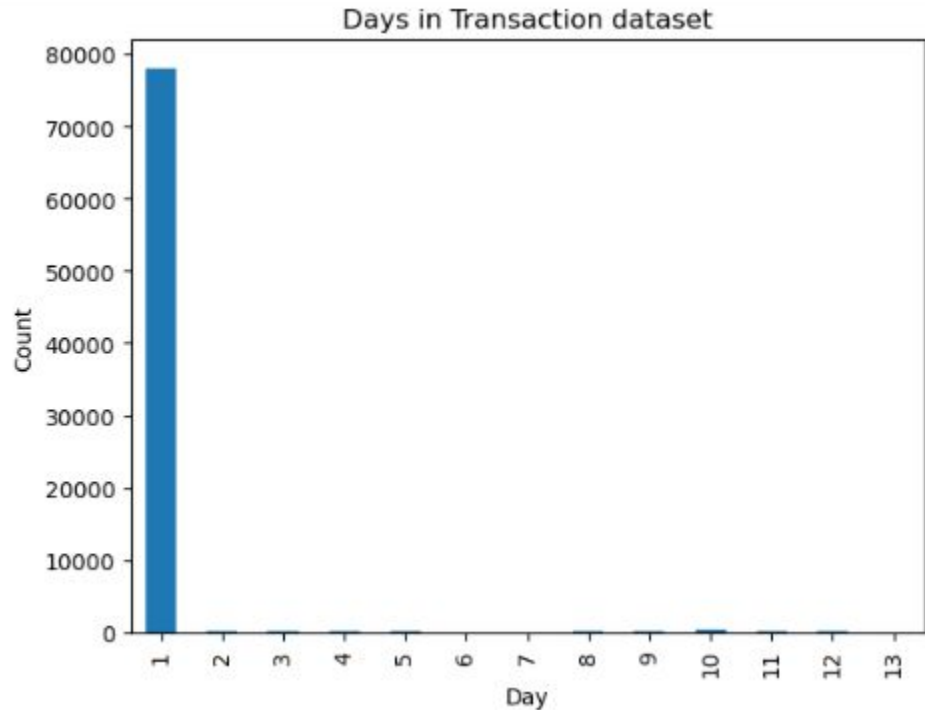
```
Btc_filtered_data = {  
    btc_tx_data[date] >= March 1, 2016  $\wedge$   
    Btc_tx_data[date] <= March 31, 2016  
}
```

43 Files from 290

# Check for poss. Fraud Transactions.



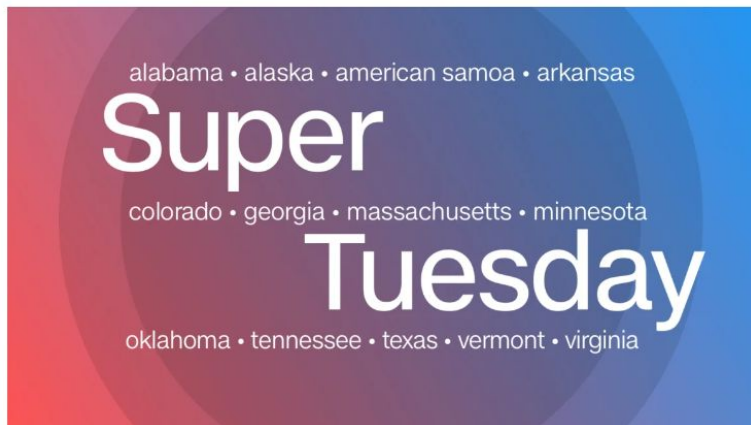
# Checking Dates of Poss. Fraud.



# What happened March 1, 2016 (cont)

☰ CNN Politics SCOTUS Congress More ▾

Ⓜ Watch 🎧 Listen 🔴 Live TV 🔍 Sign in



## Everything changes March 1



By Eric Bradner, Daniella Diaz and Tai Yellin, CNN  
🕒 7 minute read · Updated 4:15 PM EST, Tue March 1, 2016



**NBC NEWS** POLITICS U.S. NEWS LOCAL WORLD BUSINESS SPORTS HEALTH WATCH LIVE ☰

2016 ELECTION

### Donald Trump, Hillary Clinton Score Big Super Tuesday Primary Wins

Republican Donald Trump and Democrat Hillary Clinton closed in on their party's presidential nominations by racking up big wins on Super Tuesday.

A video player showing Donald Trump speaking at a podium. The podium has a sign that says "TRUMP TEXT 'TRUMP' to 88022 Palm Beach, Florida MAKE AMERICA GREAT AGAIN!". The video player includes standard controls like play, pause, and volume, and a progress bar showing 00:27 / 09:51.

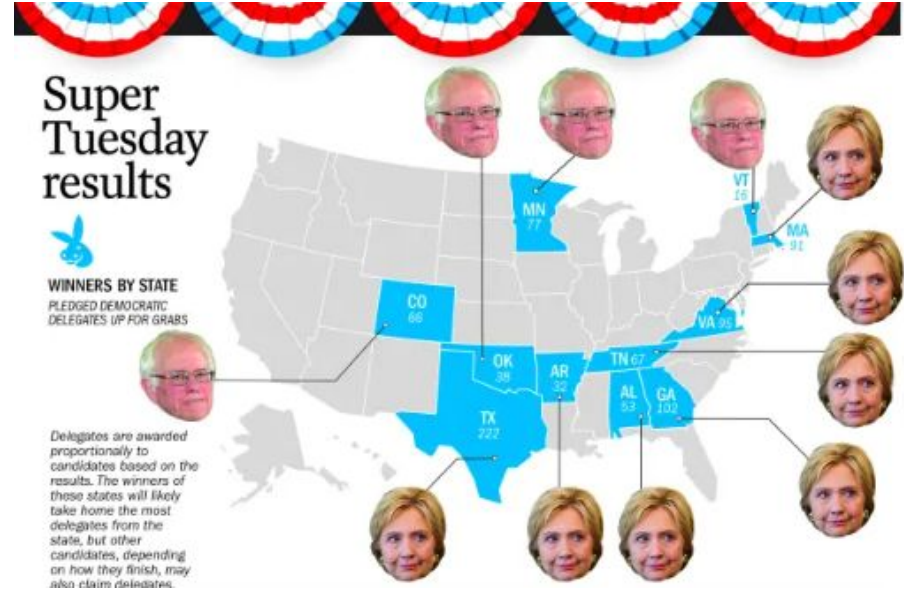
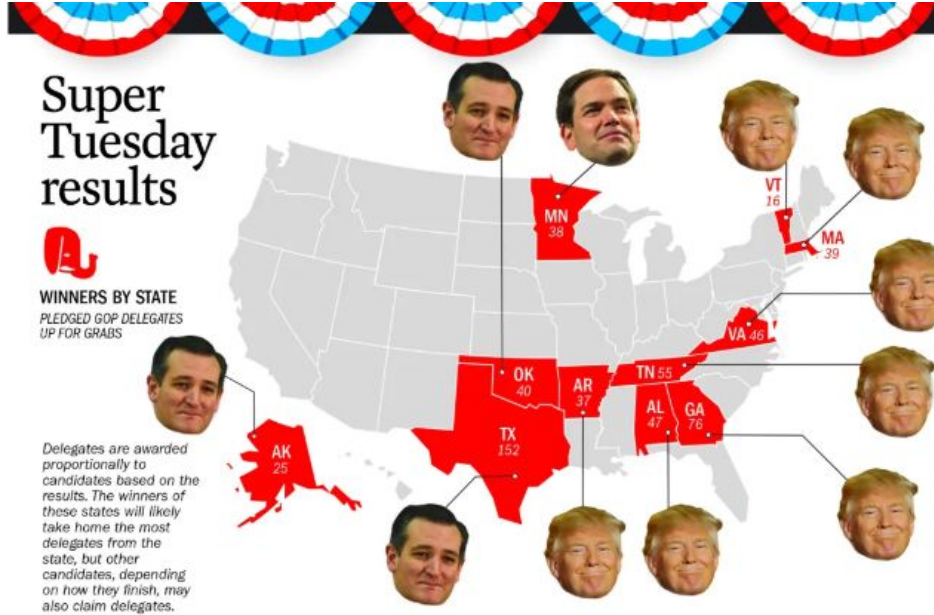
Get more news **LIVE** on **NBC NEWS NOW.** ➔

# What happened March 1, 2016

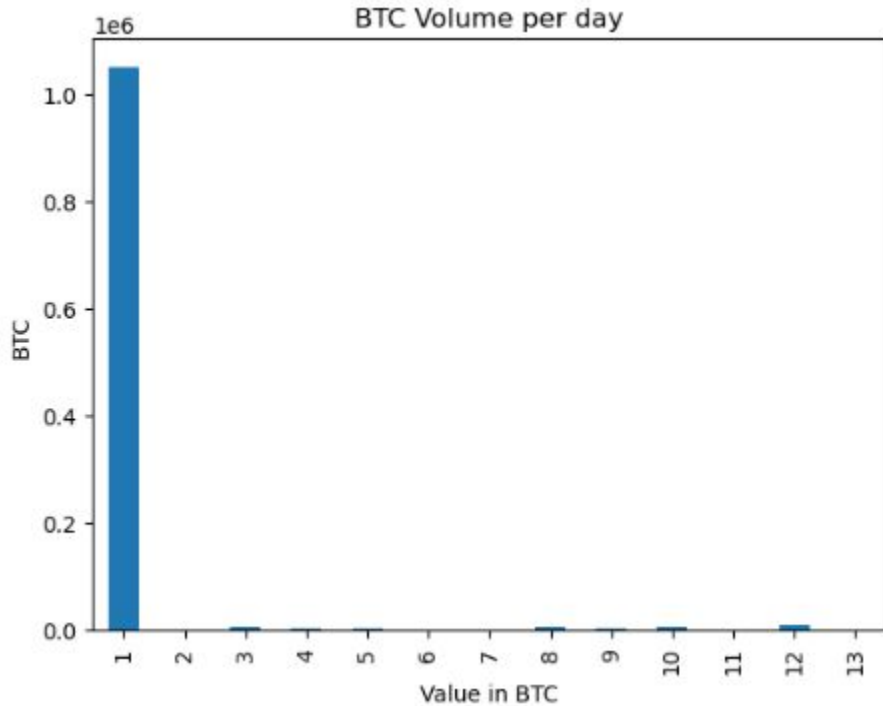




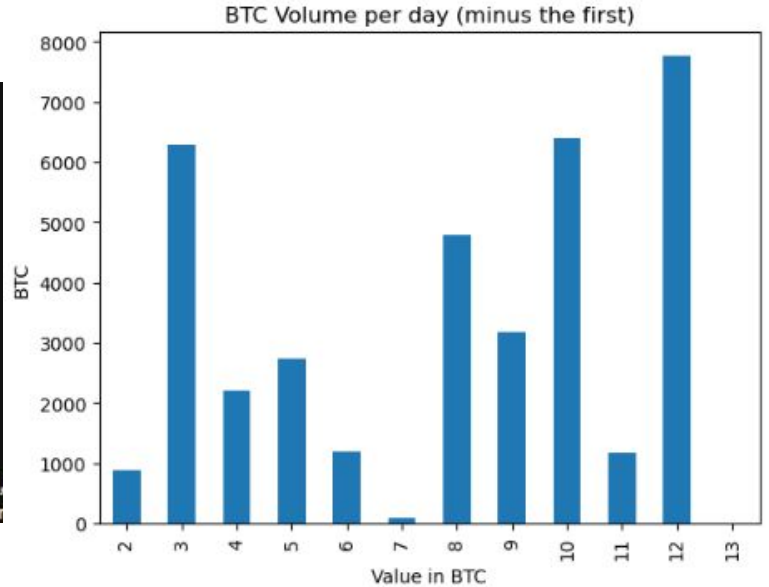
# What happened March 1, 2016 (cont)



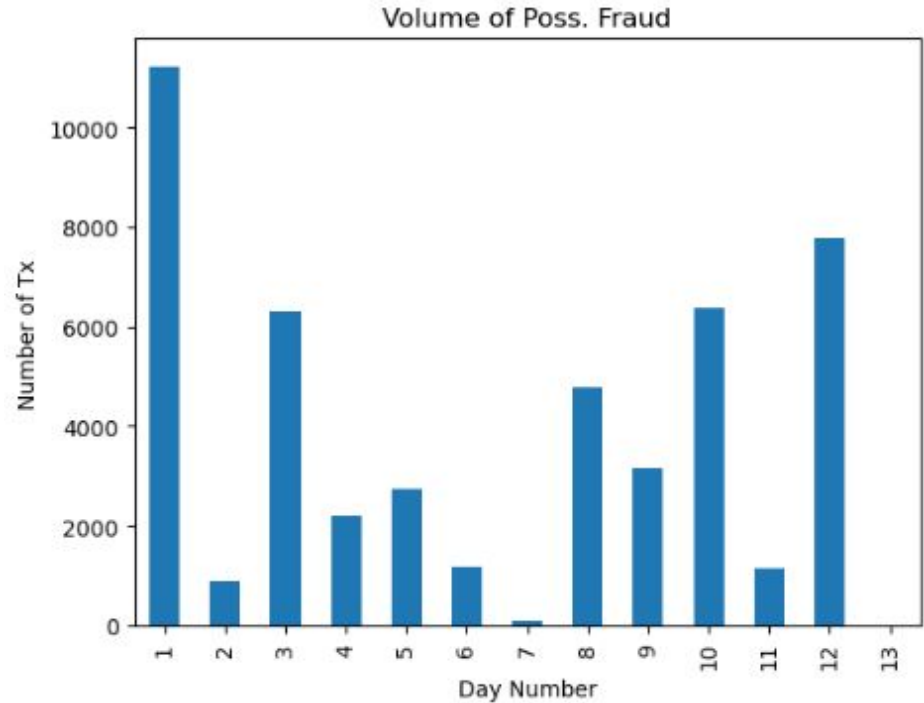
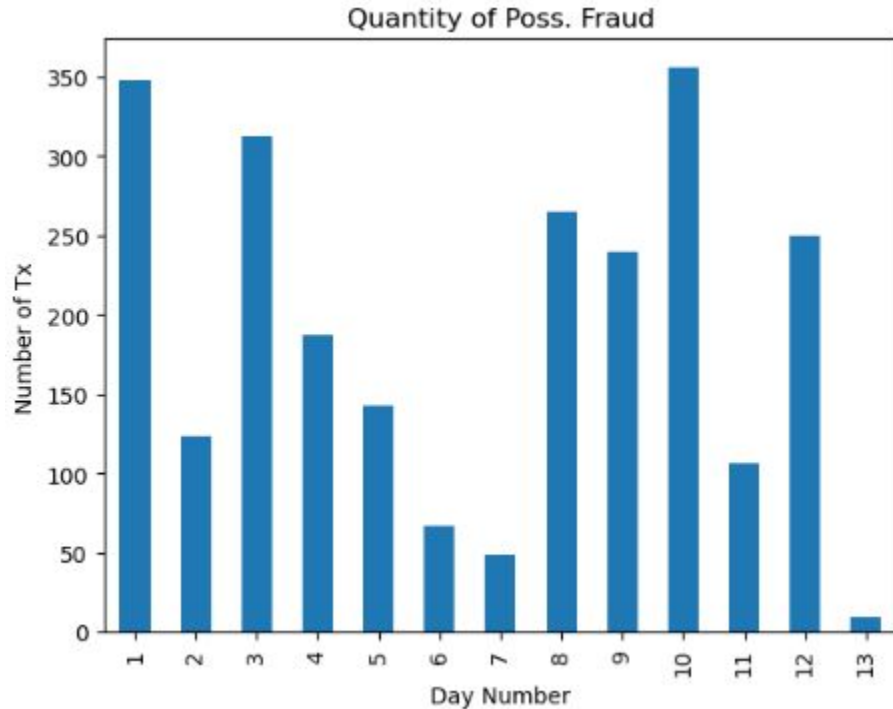
# Explore Volume in BTC



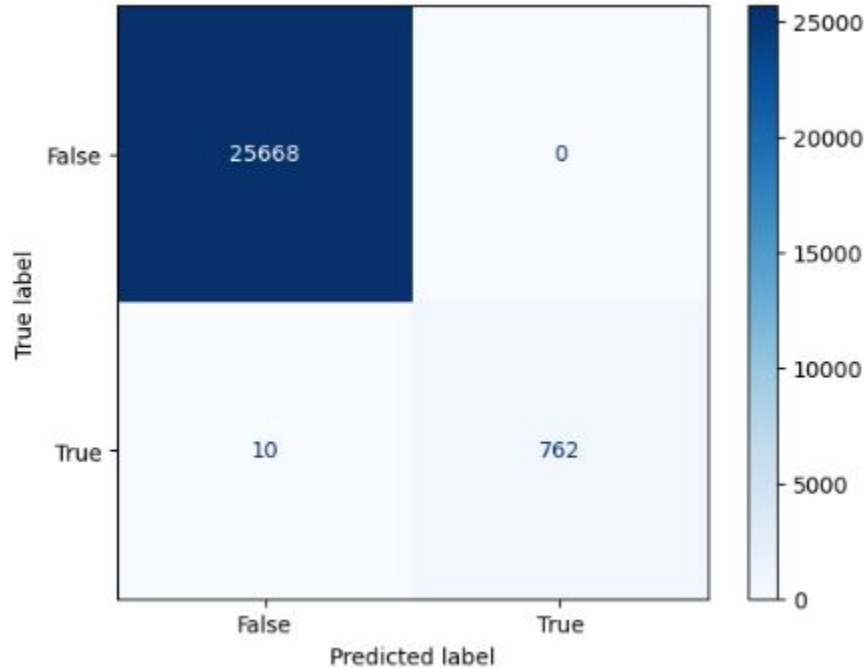
```
day
1 1050048
2 892
3 6292
4 2202
5 2726
6 1194
7 93
8 4785
9 3168
10 6387
11 1160
12 7763
13 2
Name: satoshi,
Total btc's in
```



# Quantity and Volume of Fraud



# Preliminary Modeling



Basic Logistic Regression model

Y target variable is "is\_possible\_fraud"

Accuracy: 0.9996217851739788

May need to do multi-label classification.

Goal is to classify by both sender and receiver addresses, and value.

# End

