Inca_Digital_exercise

December 1, 2020

[407]:
```
import pandas as pd
import numpy as np
import os
import time
from datetime import datetime, date, time, timedelta
```

[214]:
```
os.chdir('C:\\Users\\Irina\\Documents\\R')

bitcoin_2015 = pd.read_csv('2015.csv')
bitcoin_2016 = pd.read_csv('2016.csv')
bitcoin_2017 = pd.read_csv('2017.csv')

df_bit = pd.concat([bitcoin_2015, bitcoin_2016, bitcoin_2017])
```

[215]:
```
transactions_map = pd.read_csv('download_transactions_map.csv')
```

[216]:
```
# convert all dates to datetime format
transactions_map['begin_date']= pd.to_datetime(transactions_map['begin_date'])
transactions_map['end_date']= pd.to_datetime(transactions_map['end_date'])
df_bit['time']= pd.to_datetime(df_bit['time'])
```

[217]:
```
# Filter for transactions for the dates that are available in the bitcoin␣
 ↪dataset
df_trans = transactions_map[(transactions_map['begin_date'] >= '2015-01-02') &␣
 ↪(transactions_map['begin_date'] <= '2017-04-30')]

# Filter for transactions that are within the max and min amount range in the␣
 ↪bitcoin dataset
df_trans = df_trans[(df_trans['amount_transactions'] >=␣
 ↪min(df_bit['Transaction_amount_USD'])) & (df_trans['amount_transactions'] <=␣
 ↪max(df_bit['Transaction_amount_USD']))]

# Sort datasets by time
df_trans = df_trans.sort_values(by='begin_date')
df_bit = df_bit.sort_values(by='time')
```

[361]:
```
# Extract frist 3 numbers of 'amount_transactions'
strings = [x[:3] for x in df_trans['amount_transactions'].astype(str)]
```

[ ]:
```
# For every suspisious transaction in the FinCEN(df_trans) dataset output the␣
 ↪bitcoin transaction that matches the first 3 numbers of transaction
# amount for a date interval close to the suspisous transaction

## Code chunk may run for a minute or two

result = []

for i in range(0,len(strings)):
    tmp = df_bit.loc[(df_bit['time'] > (df_trans['begin_date'].iloc[i] -␣
 ↪timedelta(days=2))) & (df_bit['time'] < df_trans['begin_date'].iloc[i])]
    tmp['id'] = df_trans['id'].iloc[i]
    tmp['amount_transactions'] = df_trans['amount_transactions'].iloc[i]
    tmp['amount_diff'] = tmp['Transaction_amount_USD']-tmp['amount_transactions']
    tmp['begin date'] = df_trans['begin_date'].iloc[i]
    tmp['end_date'] = df_trans['end_date'].iloc[i]
    tmp = tmp[tmp.Transaction_amount_USD.astype(str).str.startswith(strings[i])]
    if not tmp.empty:
        result.append((tmp))


## Alternativley we could look for dates that are in between the 'begin_date'␣
 ↪and 'end_date' which will give more results

# result = []

# for i in range(0,len(strings)):
#     tmp = df_bit.loc[(df_bit['time'] > (df_trans['begin_date'].iloc[i] -␣
 ↪timedelta(days=2))) & (df_bit['time'] < df_trans['end_date'].iloc[i])]
#     tmp['id'] = df_trans['id'].iloc[i]
#     tmp['amount_transactions'] = df_trans['amount_transactions'].iloc[i]
#     tmp['amount diff'] =␣
 ↪tmp['Transaction_amount_USD']-tmp['amount_transactions']
#     tmp['begin date'] = df_trans['begin_date'].iloc[i]
#     tmp['end_date'] = df_trans['end_date'].iloc[i]
#     tmp = tmp[tmp.Transaction_amount_USD.astype(str).str.
 ↪startswith(strings[i])]
#     if not tmp.empty:
#         result.append((tmp))
```

[504]:
```
tmpp = pd.concat(result)

# Filter out the transactions that have a difference in value in the millions␣
 ↪range
df = tmpp[tmpp['amount diff'].round(2).astype(str).str.len() <= 10]
```

[506]:
```
df.head(10)
```

[506]:

	time	Sender
894	2015-01-25	12U7EKBW6mPDkgiSmd7FiperFfEXfKV174
895	2015-01-25	1DhPv4PZZRMUrtpchi8yjZSvqgSLwkBvUW
225	2015-03-09	39ngTt6DWQRUbzjmx6ucxz7pPvQcm52CL3
233	2015-03-09	12pCPrWvudnefJCtXQUBcm9z2NogtC3Rix
117	2015-03-17	18RMgKSeJryawnTNDeYty4gZsqki8jQxwD
118	2015-03-17	1L38tfTob637YEWvHM154z6P2D59ukUH4X
119	2015-03-17	1VTnq6RNvJEP7So5x5cvS2MY9f7JT6XZe
7554	2015-10-01	3HNSiAq7wFDaPsYDcUxNSRMD78qVcYKicw\n38DtmJWHc2...
23502	2016-01-19	1FQucMkWYi2GTeWyn813FdDrwsQ4ALxi3w
23503	2016-01-19	15Q9Tm4Yv2GdQd3kjz62Utfu4d2zGZSx2B

	Receiver	Transaction_amount	Price
894	1A18c4Sot94wF6rpytWEtMUUyPcV8Crmdd	4018.697100	249.080000
895	12U7EKBW6mPDkgiSmd7FiperFfEXfKV174	4019.697200	249.080000
225	3Q9ovFSZ77r4UPfDNdCyJqaQznmzxWxDGU	5000.000000	279.993394
233	1J2MJj4HqRaxCexWjiWaZQP94kNEwSos7A	50000.353000	279.993394
117	1LhxgRX4yLewcdwnyGwgHCHdNb5kgWLDkP	5066.714518	289.525036
118	18RMgKSeJryawnTNDeYty4gZsqki8jQxwD	5069.214618	289.525036
119	1L38tfTob637YEWvHM154z6P2D59ukUH4X	5070.214718	289.525036
7554	3BCxsMfcY2hL6tCzrPXJUnPsGovVPApteF	11558.228012	237.223546
23502	12UxFhWTk268GFRGev29dJxuASrsLHeuLk	3238.990627	383.883439
23503	1FQucMkWYi2GTeWyn813FdDrwsQ4ALxi3w	3239.972408	383.883439

	Transaction_amount_USD	id	amount_transactions	amount diff
894	1.000977e+06	225914	1000000.0	9.771000e+02
895	1.001226e+06	225914	1000000.0	1.226200e+03
225	1.399967e+06	235716	1399934.0	3.296751e+01
233	1.399977e+07	235716	1399934.0	1.259984e+07
117	1.466941e+06	235718	1467500.0	-5.592984e+02
118	1.467665e+06	235718	1467500.0	1.645432e+02
119	1.467954e+06	235718	1467500.0	4.540972e+02
7554	2.741884e+06	234293	2741500.0	3.838356e+02
23502	1.243395e+06	235265	1243025.0	3.698593e+02
23503	1.243772e+06	235265	1243025.0	7.467488e+02

	begin date	end_date
894	2015-01-26	2015-02-24
895	2015-01-26	2015-02-24
225	2015-03-10	2015-03-30
233	2015-03-10	2015-03-30
117	2015-03-18	2015-03-19
118	2015-03-18	2015-03-19
119	2015-03-18	2015-03-19
7554	2015-10-02	2015-10-13
23502	2016-01-20	2016-01-26
23503	2016-01-20	2016-01-26

[505]:
```
# number of suspisious transactions that could be linked to bitcoin activity
len(df.id.unique())

# ids of suspisious transactions that could be linked to bitcoin activity
print(df.id.unique())
```

```
[225914 235716 235718 234293 235265 238033 229404 241471 227832 239766
 234445 243286 234912 234911 229510 229509]
```