```
In [2]:   from sqlalchemy import create_engine
          import pandas as pd
          import numpy as np
          from datetime import datetime
```

```
In [3]:   pd.set_option('max_columns', 30)
```

Data munging

```
In [9]:   !head -n 2 AML\ homework\ dataset.csv
```

user_id,request_id,target_recipient_id,date_user_created,addr_country_code,add
r_city,recipient_country_code,flag_personal_business,payment_type,date_request
_submitted,date_request_received,date_request_transferred,date_request_cancell
ed,invoice_value,invoice_value_cancel,flag_transferred,payment_status,ccy_sen
d,ccy_target,transfer_to_self,sending_bank_name,sending_bank_country,payment_r
eference_classification,device,transfer_sequence,days_since_previous_req,first
_attempt_date,first_success_date
77656618388e134648db01eecf7e79ee,2e9f911150e3a79e8d71a35779706e4c,992e0a729d
6380d3b50aef5aa7c22572,27/01/2014 15:02,DEU,Berlin,GB,Personal,Direct Debi
t,26/08/2016 07:35,01/09/2016 08:31,01/09/2016 09:28,NULL,155.8123389,NULL,1,T
ransferred,EUR,GBP,Other Recipient,Other/unknown,Other/unknown,gift,Desktop We
b,55,0,27/01/2014 16:01,27/01/2014 16:17

```
In [7]:   df = pd.read_csv('AML homework dataset.csv')
          df.head()
```

Out[7]:

	user_id	request_id	target_recipient
0	77656618388e134648db01eecf7e79ee	2e9f911150e3a79e8d71a35779706e4c	992e0a729d6380d3b50aef5aa7c22!
1	a2497e0c763a7e5640fbf05e53fe0466	69cdf2f9ab2f59d10b636dc86bc9d7b7	02878ea857dbc90b2ed89b8f3488d!
2	759735d092819085c125a5cf81faf24b	5d7d30d709268f22f1a73ddbd6601690	927d3808cdc31d61226ae7c80bc8d(
3	df9627db375322e65f4648ca72f4c630	0df2fc0a4a31595678cd1de3fad57e15	0411b7eb4c220a14876e77da8125f.
4	5672b2f16063ed75fbb304fee57c024b	7dcbf32659ae5ef61e10e5174a314d7d	dc248a1266709a71e45c40f33056bl

Transform date fields

```
In [8]:   date_fields = ['date_user_created', 'date_request_submitted', 'date_request_rec
          eived',
                         'date_request_transferred', 'date_request_cancelled',
                         'first_attempt_date', 'first_success_date']
```

```
In [9]:   def prettifyDate(dt):
              try:
                  dt_datetime = datetime.strptime(dt, '%d/%m/%Y %H:%M')
                  dt_string = datetime.strftime(dt_datetime, '%Y-%m-%d %H:%M')
                  return dt_string
              except:
                  return np.nan
```

In [10]:
```python
for col in date_fields:
    df[col] = df[col].apply(lambda x: prettifyDate(x))

df.head()
```

Out[10]:

	user_id	request_id	target_recipient
0	77656618388e134648db01eecf7e79ee	2e9f911150e3a79e8d71a35779706e4c	992e0a729d6380d3b50aef5aa7c22!
1	a2497e0c763a7e5640fbf05e53fe0466	69cdf2f9ab2f59d10b636dc86bc9d7b7	02878ea857dbc90b2ed89b8f3488d!
2	759735d092819085c125a5cf81faf24b	5d7d30d709268f22f1a73ddbd6601690	927d3808cdc31d61226ae7c80bc8d
3	df9627db375322e65f4648ca72f4c630	0df2fc0a4a31595678cd1de3fad57e15	0411b7eb4c220a14876e77da8125f
4	5672b2f16063ed75fbb304fee57c024b	7dcbf32659ae5ef61e10e5174a314d7d	dc248a1266709a71e45c40f33056b

Put the data into a DB

In [4]:
```python
engine = create_engine("mysql+pymysql://root:rootpass@localhost/transferwise",
echo=False)
```

In [5]:
```python
conn = engine.connect()
```

In [6]:
```python
connection = engine.raw_connection()
cursor = connection.cursor()
```

In [ ]:
```python
cursor.execute('create database transferwise')
```

In [14]:
```python
df.to_csv('tw.csv', index=False)
```

In [12]:
```python
df.to_sql('tw', con=engine, index=False, if_exists='replace')
```

In [13]:
```python
pd.read_sql('show tables in transferwise', con = conn)
```

Out[13]:

	Tables_in_transferwise
0	tw

In [14]: `pd.read_sql('desc transferwise.tw', con = conn)`

Out[14]:

	Field	Type	Null	Key	Default	Extra
0	user_id	text	YES		None	
1	request_id	text	YES		None	
2	target_recipient_id	text	YES		None	
3	date_user_created	text	YES		None	
4	addr_country_code	text	YES		None	
5	addr_city	text	YES		None	
6	recipient_country_code	text	YES		None	
7	flag_personal_business	text	YES		None	
8	payment_type	text	YES		None	
9	date_request_submitted	text	YES		None	
10	date_request_received	text	YES		None	
11	date_request_transferred	text	YES		None	
12	date_request_cancelled	text	YES		None	
13	invoice_value	double	YES		None	
14	invoice_value_cancel	double	YES		None	
15	flag_transferred	bigint(20)	YES		None	
16	payment_status	text	YES		None	
17	ccy_send	text	YES		None	
18	ccy_target	text	YES		None	
19	transfer_to_self	text	YES		None	
20	sending_bank_name	text	YES		None	
21	sending_bank_country	text	YES		None	
22	payment_reference_classification	text	YES		None	
23	device	text	YES		None	
24	transfer_sequence	double	YES		None	
25	days_since_previous_req	double	YES		None	
26	first_attempt_date	text	YES		None	
27	first_success_date	text	YES		None	

```
In [15]: pd.read_sql('select * from transferwise.tw limit 10', con = conn)
```

Out[15]:

	user_id	request_id	target_recipie
0	77656618388e134648db01eecf7e79ee	2e9f911150e3a79e8d71a35779706e4c	992e0a729d6380d3b50aef5aa7c2
1	a2497e0c763a7e5640fbf05e53fe0466	69cdf2f9ab2f59d10b636dc86bc9d7b7	02878ea857dbc90b2ed89b8f3488
2	759735d092819085c125a5cf81faf24b	5d7d30d709268f22f1a73ddbd6601690	927d3808cdc31d61226ae7c80bc8
3	df9627db375322e65f4648ca72f4c630	0df2fc0a4a31595678cd1de3fad57e15	0411b7eb4c220a14876e77da812
4	5672b2f16063ed75fbb304fee57c024b	7dcbf32659ae5ef61e10e5174a314d7d	dc248a1266709a71e45c40f33056
5	8550806de0bde5a60150a052a93de13a	ca2842215ad3a7e63898ddb0c82c5685	4823a2e09b9befb7464f1ebf05aa
6	e21239d6e50808b5140978b71e91c220	b6ebb7c201d404b7c3dac9491ed7d136	514af1becc24b50960acd482255
7	197a367545e95dbfd97c3b31bda1dd6d	fbbe5e8023cc9e9eb575aa1b930d949e	e10ba5e923822ae656ee53272ee
8	5ff38ed71fd104808abcea1e6461e3a7	b0133ceabc9be686861db8b4ec0431e0	1cc81356a3b37aabbfd5c2014858
9	56321d4dd46b8b0e765fda26b6dd53db	78aa92f51d9078598a798c8f8ef35995	41500fd80ce546a5125f6c05b907

First time users - Blacklists

User on blacklist

- first transfer
- low amount
- never succeeded
- money already sent to TW
- took several days before cancelled -> either the documents were not legit or person was on blacklist

In [11]:
```python
pd.read_sql("""
    select *
    from transferwise.tw
    where 1=1
        and transfer_sequence = 1
        and date_request_received is not null
        and first_success_date is null
        and invoice_value_cancel < 100
    limit 10
""", con = conn)
```

Out[11]:

	user_id	request_id	target_recipie
0	7451d7a0471ba942f72583d5ff30ee4d	e37fca7e77e5c76f717167470e578682	11518b9e16c60c8b1567b584e539
1	3052cbb40b0bc0e32b2059f8ee44e92e	ca2b45af20e61631579916e8476f8cee	6b09c0ba825e08e1ca4698d8682d
2	bed435a6c7f23d7464da5dca7d7cf9a6	8bcbc59c6cff71ce79abcafcd383679b	a1d4751ba2800f88a8023658998d
3	ee7b06869aeb8a958cac83517d2eee81	c9400387a7c667e31b6d54f76ea907cc	69568f76da3ee62dd05c6506beaf
4	5a5365c5f2400a44c7d1dc43ed8f21f5	4dabb7b4a2a5572112f32c51b4abdc16	a2eec530323887648962c0854c88
5	2ec18ce71d89341598ed373a28f29b21	551bc41d180c7b97f4ac49ca87b1cb79	b211f9216bd691f29766be94808e
6	b79ec7846e41bf5bdfd726c2b71a8980	ff19ddfe7b52795d94b1a4a0d572c4db	809289ca1912d802a0af7169d536
7	2c77733c83a0433c334d24d755915726	ccbb76b92d3a3c36bf1c88677695dca9	4dbced1f4772721de0500c5b0a54
8	74c4b8bcab0906d7c3813da6e312f83f	98f0f8fd647471523670809f5207d4b3	70b00b9997701aba7a290ba25b50
9	9c582de7f3e4d626350f9f05e14c767b	fcbae32a54f0a62e2a181a6ff51eb776	03c88f4fdc76059ecb100936fb0

Problematic countries

- service not available

In [12]:
```python
pd.read_sql("""
    select recipient_country_code, count(*) as cnt
    from transferwise.tw
    where 1=1
        and recipient_country_code in ("IQ", "IR", "SY", "AF")
    group by recipient_country_code
""", con = conn)
```

Out[12]:

recipient_country_code	cnt

```
In [13]: pd.read_sql("""
    select addr_country_code, count(*) as cnt
    from transferwise.tw
    where 1=1
        and addr_country_code in ("IRQ", "IRN", "SYR", "AFG")
    group by addr_country_code
""", con = conn)
```

Out[13]:

addr_country_code	cnt

Recipient on blacklist

```
In [14]: pd.read_sql("""
    select target_recipient_id, count(distinct user_id) as cnt
    from transferwise.tw
    where 1=1
    group by target_recipient_id
    order by cnt desc
    limit 10
""", con = conn)
```

Out[14]:

	target_recipient_id	cnt
0	3b66475e924ac6f5cdb845930d772529	6
1	55602ba1d012407008f2c555b98fa9ef	6
2	f144f57b8e574a181463aed16a6092e3	5
3	0aa79d6a6ca2c2770ee188d62b162516	4
4	828cf566b7a7a6704f3d938c4611e28b	3
5	3d3cf197e19aa043632a819c6dc61fea	3
6	abb175b16259c2ae14a7931b3eac8179	3
7	3d58c46a20137588a4e3fff013e55975	2
8	4b4ea83c0225cb4d920ee7e00decfee2	2
9	db1315683c634ff7ba7797c97c7e6511	2

```
In [ ]: pd.read_sql("""
    select *
    from transferwise.tw
    where 1=1
        and target_recipient_id in (select target_recipient_id from transferwis
e.tw group by 1 having count(*) > 3)
    order by target_recipient_id, date_request_submitted
""", con = conn)
```

```
In [ ]: pd.read_sql("""select target_recipient_id from transferwise.tw group by 1 havin
g count(*) > 3""", con = conn)
```

First time users - Amount limits

- first transfer
- amount > €1000
- some never succeeded, others took a couple of days (first_success_date is not empty)

```
In [36]: pd.read_sql("""
    select *
    from transferwise.tw
    where 1=1
        and transfer_sequence = 1
        and payment_status = 'Cancelled'
        and ccy_send = 'EUR'
        and invoice_value_cancel > 1000
        and date_request_received is not null
    limit 10
""", con = conn)
```

Out[36]:

	user_id	request_id	target_recipien
0	1a8ebb56a246a22b8dc164049c703e92	e9dc93fc679a50772c78f65c4d09318e	a9f5e98fbd35961bb7bae7acf6a2
1	77fc14c6ba9a8046df80a6f1b85f87ab	93cf7410611dd8d353b1ebbb54d05bdf	36af8df61bab5a736a7e9d185785
2	6008779d978132785979f29361c18bbe	b147ae6e019e649b0bee4f009c027d6b	81bd284ab34d94f47f95f1edb915
3	b0b7297b8fad84ca157ce5dbaf7796a3	229b4cf6f33fe5a4c54084fd7cbd7af1	e3ee24955b28e488e14243f0d3c2
4	3b2220f4c7655b2e62885c6faea8fcd8	9389bcd80e8e55aa1afa3ca60c75cf9b	cd11a131a7d79eefd9ffdaab071b
5	ca1dc9a2be24a6b6082112456a67bad1	b5cf8e9e471833cf99d98f045c9345ac	b2f649ca894c870715371c3c245
6	7723cc80c01e5927be31e88e822b85e5	617d2934a631d9a625e0756510c346db	1b7cccfffc261c1b6961e2f2156b
7	dc8b557a57ba2004ab82508fd25a907d	256b659b54f05a732b5da6095e91807b	42bca77d4c60b8b857c8066f8df2
8	9af82e24953a18e4d607d3b263745134	4d589240dd29e10a0c41b6972984d6d2	0e78331dfaf4f154a90a6257163c
9	6fab464b57a0b0fbe2d09d345fa32d4f	957274f8e58d80130e35761698e96c09	2f64f25af802dd0ee8f1e243726c

Regular users - Too frequent attempts

- the first 5 definitely have problems -- 0% success rate, a lot of transfers

```
In [71]: pd.read_sql("""
             select
                 user_id
             , sum(flag_transferred) as successful
             , count(flag_transferred) as attempts
             , sum(flag_transferred) / count(flag_transferred) as prop_received
             from transferwise.tw
             where 1=1
             group by user_id
             having count(flag_transferred) > 5
             order by prop_received
             limit 10
             """, con = conn)
```

Out[71]:

	user_id	successful	attempts	prop_received
0	39ded89a9e1d1a7ee64742c701657716	0.0	6	0.0000
1	8917c69f0a728519439d93ba9b49005a	0.0	89	0.0000
2	b78c8ce3f52611c3df21da6b9effe911	0.0	72	0.0000
3	b83a45db00014f69a91f9777eaf93c0c	0.0	7	0.0000
4	d966e072fab4f783c66d30fa2ed4a723	0.0	47	0.0000
5	d6fb6f9d8ca89d20d9db824b845159d0	2.0	6	0.3333
6	09e77338448191698acbc7799d6b779f	4.0	8	0.5000
7	d82e95523bbef24c01732dbd26ccd23f	10.0	18	0.5556
8	db9646f784842973bc07ab39fb1331d0	5.0	8	0.6250
9	cfef40996e0de2c0f3fe1031421403fd	7.0	11	0.6364

some sort of a test-account:

```
In [75]: pd.read_sql("""
             select *
             from transferwise.tw
             where 1=1
                 and user_id = '8917c69f0a728519439d93ba9b49005a'
             limit 3
             """, con = conn)
```

Out[75]:

	user_id	request_id	target_recipient
0	8917c69f0a728519439d93ba9b49005a	cda57949c68fefb809477617c2059e38	81be15cefd8da370373fd4105fea0
1	8917c69f0a728519439d93ba9b49005a	e63b1f600572dfcffb983b751dbded4d	81be15cefd8da370373fd4105fea0
2	8917c69f0a728519439d93ba9b49005a	f207e7d9301f624591cb7c2a94a3d120	0d4c8a842f3b6c6d2dc0a0d01d03d7

This user has just registered, and started a lot of transfers, within one day, all of them cancelled

In [78]:
```
pd.read_sql("""
    select *
    from transferwise.tw
    where 1=1
        and user_id = 'b78c8ce3f52611c3df21da6b9effe911'
    limit 10
    """, con = conn)
```

Out[78]:

	user_id	request_id	target_recipient_
0	b78c8ce3f52611c3df21da6b9effe911	0522e4faedda820fad6529b31e070166	a1fe5ab59bc08890817cebe28eb3fbl
1	b78c8ce3f52611c3df21da6b9effe911	24c1ccfda9a700931789b66e53e75779	329119d03e9695c2da010830341191a
2	b78c8ce3f52611c3df21da6b9effe911	adea821d6976517d67b9a1f98b786792	58fe230a9fee72fa16b0672eebf4e9l
3	b78c8ce3f52611c3df21da6b9effe911	940cdbb64d738d5ff957ab446bda6a27	7e51b399b68c3f8af5a7059dd65f6al
4	b78c8ce3f52611c3df21da6b9effe911	93f0e09ffceb923930d297abce3fcd78	e9136f78458430ed99f6878025a59a!
5	b78c8ce3f52611c3df21da6b9effe911	cb7ac092e2d85c78f40b3fd9050c8f2f	3db5cc2f963e7f9caf46b59e5d6e11(
6	b78c8ce3f52611c3df21da6b9effe911	9d9b2786b93f281a23be1e61ce79a092	2f32003f57cb9578ad16335b9ab4ca
7	b78c8ce3f52611c3df21da6b9effe911	cc014fb700c2187c5ea312927262362a	cf8624c74f17557cad5b419183c3c4:
8	b78c8ce3f52611c3df21da6b9effe911	1dfec01c16109397d29cd456e954f34d	7062e2fbefdd0633d125fb95939e36!
9	b78c8ce3f52611c3df21da6b9effe911	c203eac78b04e58eeb062e51c2612c73	65a97954685381c7ae12f4c2de002a;

In [77]:
```
pd.read_sql("""
    select *
    from transferwise.tw
    where 1=1
        and user_id = 'b78c8ce3f52611c3df21da6b9effe911'
        and date(date_request_submitted) > '2016-11-21'
    limit 10
    """, con = conn)
```

Out[77]:

user_id	request_id	target_recipient_id	date_user_created	addr_country_code	addr_city	recipient_country

Regular users - Unusual transfers

Top 1% of the amounts transferred in each currency

just an example query to find the top 1% of each currency. these are the frequent ones, having 100+ transfers

```
In [117]: pd.read_sql("""
          select ccy_send, min(transfer_value) as threshold
          from (
              select
                    ccy_send
                  , invoice_value
                  , invoice_value_cancel
                  , transfer_value
                  , ntile(100) over(partition by ccy_send order by transfer_value) as cou
          ntry_transfer_percentile
              from (
                  select
                        ccy_send
                      , invoice_value
                      , invoice_value_cancel
                      , coalesce(invoice_value,0) + coalesce(invoice_value_cancel,0) as t
          ransfer_value
                  from transferwise.tw
                  ) t
              where 1=1
              -- limit 10
          ) tt
          where country_transfer_percentile = 100
          group by ccy_send
          -- limit 10
          """, con = conn)
```

Out[117]:

	ccy_send	threshold
0	AUD	16659.18000
1	BGN	4569.71840
2	BRL	2558.95848
3	CAD	13469.76848
4	CHF	28806.72000
5	CZK	16521.00000
6	DKK	14470.95000
7	EUR	17772.75870
8	GBP	17500.00000
9	HUF	19531.38622
10	NOK	13757.76000
11	NZD	24311.15400
12	PLN	13501.13650
13	RON	11781.05400
14	SEK	13477.05000
15	USD	34835.90158

Regular transfers of low amounts, then one is much larger

example query to find users with some unusual transfers compared to their habits

```python
In [130]: pd.read_sql("""
    select
        user_id
      , ccy_send
      , count(*) as cnt
      , min(coalesce(invoice_value,0) + coalesce(invoice_value_cancel,0)) as
min_amount
      , max(coalesce(invoice_value,0) + coalesce(invoice_value_cancel,0)) as
max_amount
      , avg(coalesce(invoice_value,0) + coalesce(invoice_value_cancel,0)) as
avg_amount
      , sum(case when coalesce(invoice_value,0) + coalesce(invoice_value_canc
el,0) > 1000 then 1 else 0 end) as transfers_1000_plus
      , sum(case when coalesce(invoice_value,0) + coalesce(invoice_value_canc
el,0) > 1000 then 1 else 0 end) / count(*) as large_transfer_prop
    from transferwise.tw
    where 1=1
    group by user_id, ccy_send
    having count(*) > 4
        and large_transfer_prop between 0.01 and 0.5
    order by cnt desc
    -- limit 100
    """, con = conn)
```

Out[130]:

	user_id	ccy_send	cnt	min_amount	max_amount	avg_amount	transfers
0	69fd02c4fbd5bfa6533f7a5eac3bd81c	EUR	80	8.819153	9416.220271	142.503891	
1	d966e072fab4f783c66d30fa2ed4a723	EUR	22	12.752028	16027.336000	1749.765870	
2	d966e072fab4f783c66d30fa2ed4a723	GBP	18	10.000000	10000.000000	1128.047222	
3	19466121d8747bd79d1ec4d109b63c52	GBP	18	1.000000	7936.510000	679.770000	
4	e91068fff3d7fa1594dfdf3b4308433a	EUR	14	65.500223	10944.255400	1017.213242	
5	e608d060011ee543263f345f9887c6c4	EUR	13	22.477719	3605.119722	488.638008	
6	cb57cdb7cc459dc6fbbc33f91485b5e2	GBP	12	32.370000	1942.980000	454.753333	
7	71c064631fa61b22b45793b865fc560e	GBP	11	60.270000	1144.790000	208.597273	
8	ddb3a4dd1e14f550e64c449bd8b333c3	GBP	11	250.000000	2700.000000	1228.969091	
9	d045c59a90d7587d8d671b5f5aec4e7c	GBP	9	105.720000	1539.730000	592.855556	
10	781877bda0783aac5f1cf765c128b437	GBP	8	126.630000	2537.040000	571.256250	
11	09e77338448191698acbc7799d6b779f	GBP	8	8.240000	1642.030000	656.571250	
12	6afae039e85211b150bf08df79284ee0	USD	7	66.869590	1907.845670	497.457908	
13	8364f022baab60878f044061cdf596db	EUR	7	55.918532	1284.186079	603.280179	
14	db9646f784842973bc07ab39fb1331d0	EUR	7	79.780541	1148.129613	379.735664	
15	79286674060c5b8f6644bfff4ae069a2	EUR	7	342.874851	85131.000000	13103.402920	
16	5da8180d1cfa90364c87ad48083781aa	GBP	7	102.740000	2289.180000	500.888571	
17	e7a83afdc07b0f085095181de1c68b5f	GBP	7	11.440000	2461.540000	541.127143	
18	8ad1a53c92b468ebae567c66323ccc64	HUF	7	4.403850	1402.390000	504.031264	
19	8fe1a3942c9b455e1e0a0acecdc51378	GBP	7	77.990000	2500.000000	912.198571	
20	2e56d54a72d31e51e9e11c5ec36e2118	GBP	7	244.000000	1036.220000	529.697143	
21	6614b55cadb2430d59f44362c03c5bbd	GBP	7	177.240000	1714.860000	608.951429	
22	d8fc0b18987b457c30ecbae310072341	GBP	7	155.200000	1113.080000	551.248571	
23	55af04fb9bcfed3eaa9bcd25215f8c74	EUR	6	62.765810	2213.078119	1031.844434	
24	0cf9e05e144599e342541462a833f535	GBP	6	500.000000	1891.000000	1185.166667	
25	9b2164704913a0f1e53ee68fcc73a979	USD	6	142.946440	2290.224033	1217.047599	
26	59fae23e31dc90b03927c608acc7337a	GBP	6	30.000000	3304.220000	932.146667	
27	940297ccc2a6e32c4a2523e3ec8ec8a3	GBP	6	79.430000	6677.150000	1535.476667	
28	d43dad0f0b86caa895b5b8d84b368c75	GBP	6	476.000000	2141.440000	1059.730000	
29	e489386f007381debc2d07c57e47d484	GBP	6	245.520000	2596.940000	941.526667	
30	c308dbfc240759680e659bfccebc1bb2	GBP	6	331.000000	3286.830000	1181.286667	
31	492114f6915a69aa3dd005aa4233ef51	GBP	6	19.050000	1631.960000	681.061667	
32	f18398f819fa3e4d5f1f28ba6fd5e748	GBP	5	100.000000	1240.000000	379.800000	
33	b9a75914c53e488df9c2a07e78b56076	GBP	5	91.370000	1252.950000	385.236000	
34	b087d6b44ec5f79e06e190ea28886d25	EUR	5	258.734190	1905.327000	895.981774	
35	101a0069dcf9563af4db55c3bc596945	EUR	5	181.340813	1818.633168	735.917992	
36	f28317bc2145641a9cb2e8fa7572c4d5	GBP	5	192.520000	1880.250000	994.342000	
37	ea4a2cfad5393713e72930d3225d6330	GBP	5	143.120000	1578.260000	873.058000	
38	83041564f8b355de68145482d05023a1	EUR	5	382.911500	5196.366000	1528.237634	
39	98a9e02cbbd5cfd0869534f8a0737ff5	GBP	5	358.270000	1449.200000	799.622000	
40	9be534baba147ba01d760affe5cf89b8	GBP	5	49.530000	1434.000000	756.706000	
41		GBP	5				

we have a winner, all his unusually high transfers were cancelled

```
In [133]: pd.read_sql("""
              select *
              from transferwise.tw
              where 1=1
                  and user_id = 'd966e072fab4f783c66d30fa2ed4a723'
                  and (invoice_value > 1000 or invoice_value_cancel > 1000)
              -- limit 10
              """, con = conn)
```

Out[133]:

	user_id	request_id	target_recipient_
0	d966e072fab4f783c66d30fa2ed4a723	0da050713780885f99857d211a425158	8d857b90f0211b901839fcf772d3df
1	d966e072fab4f783c66d30fa2ed4a723	de9ed5d1e3d0c47914420c974cca2c4c	8d857b90f0211b901839fcf772d3df
2	d966e072fab4f783c66d30fa2ed4a723	92707f824927cc9c404c1a98d861a65a	179c212ed0bce6fc92b19f0a8fddf34
3	d966e072fab4f783c66d30fa2ed4a723	17d1b47ad581f1d42d5d53a7519eb963	ef99669ef32a980ad54d27f566496d
4	d966e072fab4f783c66d30fa2ed4a723	1a7dcda526c1d95eba7d630c6e4dfbd7	ef99669ef32a980ad54d27f566496d
5	d966e072fab4f783c66d30fa2ed4a723	1f77a128aac0b5997b34c2ff0baf4924	ef99669ef32a980ad54d27f566496d
6	d966e072fab4f783c66d30fa2ed4a723	3fbf67e6795647b5b557045079c2917a	ef99669ef32a980ad54d27f566496d
7	d966e072fab4f783c66d30fa2ed4a723	1f05791811f3c58000e86f7e749f324d	179c212ed0bce6fc92b19f0a8fddf34
8	d966e072fab4f783c66d30fa2ed4a723	39897f6dc3ca959f155f51756b8996d1	179c212ed0bce6fc92b19f0a8fddf34
9	d966e072fab4f783c66d30fa2ed4a723	4b621fef4cf561e9a5fb773dd15bf4a5	ef99669ef32a980ad54d27f566496d
10	d966e072fab4f783c66d30fa2ed4a723	05b5157fc4456c51c73d8631c37c9adf	ef99669ef32a980ad54d27f566496d
11	d966e072fab4f783c66d30fa2ed4a723	d92a140b8b684a8018f31112b5cd514e	ef99669ef32a980ad54d27f566496d
12	d966e072fab4f783c66d30fa2ed4a723	2107b8f19c7ac1b8922c391c04dc097b	179c212ed0bce6fc92b19f0a8fddf34

Regular users - Chains, loops, circles

Chains of 2 transfers

```
In [ ]: pd.read_sql("""
    select
          count(*) as cnt  -- 1707
        , count(distinct sender.user_id) as sender_cnt   -- 1432
        , count(distinct recipient.user_id) as recipient_cnt   -- 1436
        , count(distinct sender.request_id) as first_request_cnt   -- 1497
        , count(distinct recipient.request_id) as second_request_cnt -- 1641
    from tw sender
    join tw recipient
        on sender.target_recipient_id = recipient.user_id
        and sender.date_request_submitted < recipient.date_request_submitted
""", con = conn)
```

Chains of 3 transfers -> 16 records

nothing really suspicious

```
In [ ]: pd.read_sql("""
    select count(*) as cnt
    from tw first_transfer
    join tw second_transfer
        on first_transfer.target_recipient_id = second_transfer.user_id
        and first_transfer.date_request_submitted < second_transfer.date_reques
t_submitted
        join tw third_transfer
        on second_transfer.target_recipient_id = third_transfer.user_id
        and second_transfer.date_request_submitted < third_transfer.date_reques
t_submitted
""", con = conn)
```

Circle of 3 transfer

none :(

```
In [ ]: pd.read_sql("""
    select count(*) as cnt
    from tw first_transfer
    join tw second_transfer
        on first_transfer.target_recipient_id = second_transfer.user_id
        and first_transfer.date_request_submitted < second_transfer.date_reques
t_submitted
        join tw third_transfer
        on second_transfer.target_recipient_id = third_transfer.user_id
        and third_transfer.target_recipient_id = first_transfer.user_id
        and second_transfer.date_request_submitted < third_transfer.date_reques
t_submitted
""", con = conn)
```

Chains of 4 transfers

none :(

```python
pd.read_sql("""
    select count(*) as cnt
    from tw first_transfer
    join tw second_transfer
        on first_transfer.target_recipient_id = second_transfer.user_id
        and first_transfer.date_request_submitted < second_transfer.date_reques
t_submitted
    join tw third_transfer
        on second_transfer.target_recipient_id = third_transfer.user_id
        and second_transfer.date_request_submitted < third_transfer.date_reques
t_submitted
    join tw fourth_transfer
        on third_transfer.target_recipient_id = fourth_transfer.user_id
        and third_transfer.date_request_submitted < fourth_transfer.date_reques
t_submitted
""", con = conn)
```

Regular users - Business accounts

In [42]:
```python
pd.read_sql("""
    select
         payment_reference_classification
    , count(*) as cnt
    , sum(flag_transferred) / count(flag_transferred) as prop_received
    from tw
    where 1=1
        and flag_personal_business = 'Business'
    group by 1
    order by cnt desc
    limit 10
""", con = conn)
```

Out[42]:

	payment_reference_classification	cnt	prop_received
0	Other/unknown	2144	0.8535
1	invoice	2016	0.9097
2	blank	1201	0.8360
3	monthly	575	0.9061
4	salary	78	0.9487
5	expense	40	0.9250
6	family	40	0.8250
7	rent	39	0.8718
8	freelance	39	0.8974
9	deposit	31	0.8387

In [ ]: