Anomaly

February 26, 2021

[2]:
```
import pandas as pd
import numpy as np
import time
import datetime
from datetime import datetime, timedelta, timezone
NoneType = type(None)
```

[3]:
```
df_any = pd.read_json (r'test_frame_V1.json')
df_any.columns

# df_any['   '] = df_any['   '].astype(datetime)

dtn_ = datetime.now().isoformat()
dtn = datetime.strptime(dtn_+'Z', '%Y-%m-%dT%H:%M:%S.%fZ')

df_any['       ']=\
df_any['       ']\
.apply(lambda x: datetime.strptime(x, '%d-%m-%Y'))

df_any['   ']=\
df_any['   ']\
.apply(lambda x: datetime.strptime(x, '%Y-%m-%dT%H:%M:%S.%fZ'))

df_any['     ']=\
df_any['     ']\
.apply(lambda x: datetime.strptime(x, '%Y-%m-%dT%H:%M:%S.%fZ') if not␣
 ↪isinstance(x, NoneType) else 0)


df_any.loc[:, '     2'] = df_any['     ']
df_any.loc[df_any['     2'] == 0, '     2'] = dtn

df_any.loc[:, '    '] = (pd.to_datetime(df_any['     2']) - pd.
 ↪to_datetime(df_any['   ']))#.datetime.days

df_any.loc[:, '    '] =\
df_any['       ']\
.apply(lambda x: (dtn - x).days/365.25)
```

[4]:
```
df_any.head()
```

[4]:
```
                                              \
0   User_0    1989-03-02      Broker_account_0    Strategy_0
1   User_1    1981-07-15      Broker_account_1    Strategy_0
2  User_10    1988-08-14     Broker_account_10    Strategy_2
3  User_10    1988-08-14  Broker_account_14023  Strategy_238
4  User_10    1988-08-14  Broker_account_14024  Strategy_236

                                                                2  \
0 2020-06-17 12:28:00                    0  2021-02-26 02:23:07.263777
1 2020-11-27 16:39:12  2021-02-02 15:11:57         2021-02-02 15:11:57
2 2020-12-09 11:59:27                    0  2021-02-26 02:23:07.263777
3 2020-09-24 12:54:06  2021-02-04 16:49:55         2021-02-04 16:49:55
4 2020-11-10 15:52:13                    0  2021-02-26 02:23:07.263777



0 253 days 13:55:07.263777  31.989049
1         66 days 22:32:45  39.619439
2  78 days 14:23:40.263777  32.536619
3        133 days 03:55:49  32.536619
4 107 days 10:30:54.263777  32.536619
```

[5]:
```
# Native Check
df_any[['    ']].dropna().shape[0] == df_any.shape[0], df_any.shape[0],\
'-----MIN-----',np.min(df_any),\
'-----MAX-----',np.max(df_any),\
df_any[df_any['    ']>df_any['     2']] #          <
```

[5]:
```
(True,
 17271,
 '-----MIN-----',
                              User_0
              1981-03-01 00:00:00
                Broker_account_0
                          Strategy_0
                 2020-01-02 13:02:32
      2         2020-05-04 14:31:15
                     0 days 00:00:35
                            5.89185
 dtype: object,
 '-----MAX-----',
                                  User_9999
                     2015-04-07 00:00:00
                    Broker_account_9999
```

```
                               Strategy_99
                       2020-12-14 23:49:35
     2         2021-02-26 02:23:07.263777
                  420 days 13:20:35.263777
                                   39.9918
dtype: object,
Empty DataFrame
Columns: [   ,         ,         ,    ,    ,      ,
    2,    ,    ]
Index: [])
```

[5]:

```
# Criteria
#
#             ,

# 1 Se urity:
#
#       Date of birth
#           -           > 3 (   -          ) -          ,         .


# 2 Lawyer:
# Age < 18

# 3 Product:
#                 ( .       ),
#

# 4 Marketing:
#              ( .       -     ),
#           -



#                              . .
#
```

[6]:

```
df_any.head()
```

[6]:

```
                                            \
0   User_0    1989-03-02      Broker_account_0    Strategy_0
1   User_1    1981-07-15      Broker_account_1    Strategy_0
2  User_10    1988-08-14     Broker_account_10    Strategy_2
3  User_10    1988-08-14  Broker_account_14023  Strategy_238
4  User_10    1988-08-14  Broker_account_14024  Strategy_236

                                                         2   \
0 2020-06-17 12:28:00                    0  2021-02-26 02:23:07.263777
1 2020-11-27 16:39:12  2021-02-02 15:11:57         2021-02-02 15:11:57
2 2020-12-09 11:59:27                    0  2021-02-26 02:23:07.263777
3 2020-09-24 12:54:06  2021-02-04 16:49:55         2021-02-04 16:49:55
4 2020-11-10 15:52:13                    0  2021-02-26 02:23:07.263777


0 253 days 13:55:07.263777  31.989049
1          66 days 22:32:45  39.619439
2  78 days 14:23:40.263777  32.536619
3         133 days 03:55:49  32.536619
4 107 days 10:30:54.263777  32.536619
```

0.0.1 1 Seq

1.1 Birth Date count, Brokers Account count

[7]:
```
# 2Se urity
# df_DB_fltr = df_any.groupby(by=['    ', '        ']).count().sort_values(['  ␣
↪     '])
# uni
df_DB_fltr = df_any[df_any['     '] == 0].groupby(by=['    ']).nunique().
↪sort_values(['         '])

df_DB_fltr['    '] = df_DB_fltr.index
df_DB_fltr
# df_DB_fltr[df_DB_fltr['        '] > 1] # 17271
```

[7]:
```
                                                  \

User_0                   1                1       1       1           1
User_5216                1                1       1       1           1
User_5215                1                1       1       1           1
User_5214                1                1       1       1           1
User_5213                1                1       1       1           1
…                    …                …     …     …        …
User_9881                1               13      13      13           1
User_10219               1               14       1      14           1
User_6293                1               15      15      15           1
User_88                  1               15      15      15           1
User_76                  1               16      16      16           1

                 2

User_0                1       1        1      User_0
User_5216             1       1        1   User_5216
User_5215             1       1        1   User_5215
User_5214             1       1        1   User_5214
```

```
User_5213           1       1        1   User_5213
…                …     …      …         …
User_9881           1      13        1   User_9881
User_10219          1      14        1  User_10219
User_6293           1      15        1   User_6293
User_88             1      15        1     User_88
User_76             1      16        1     User_76

[8410 rows x 9 columns]
```

[8]:

```
# df_any.sort_values(by=['    '])
#                  (    ).                ,
df_DB_fltr.describe().transpose()
```

[8]:

```
                      count       mean        std  min  25%  50%  75%   max
              8410.0  1.000000  0.000000  1.0  1.0  1.0  1.0   1.0
             8410.0  1.479548  1.127236  1.0  1.0  1.0  2.0  16.0
                 8410.0  1.468371  1.106639  1.0  1.0  1.0  2.0  16.0
                 8410.0  1.479310  1.127126  1.0  1.0  1.0  2.0  16.0
               8410.0  1.000000  0.000000  1.0  1.0  1.0  1.0   1.0
     2         8410.0  1.000000  0.000000  1.0  1.0  1.0  1.0   1.0
                 8410.0  1.479310  1.127126  1.0  1.0  1.0  2.0  16.0
                8410.0  1.000000  0.000000  1.0  1.0  1.0  1.0   1.0
```

1.2 Broker accounts count

[9]:

```
# Brokers Account count
df_DB_fltr[df_DB_fltr['         '] > 3].shape[0]/df_DB_fltr.shape[0]*100,\
# df_DB_fltr[df_DB_fltr['         '] == 3]
```

[9]:

```
(5.053507728894173,)
```

[10]:

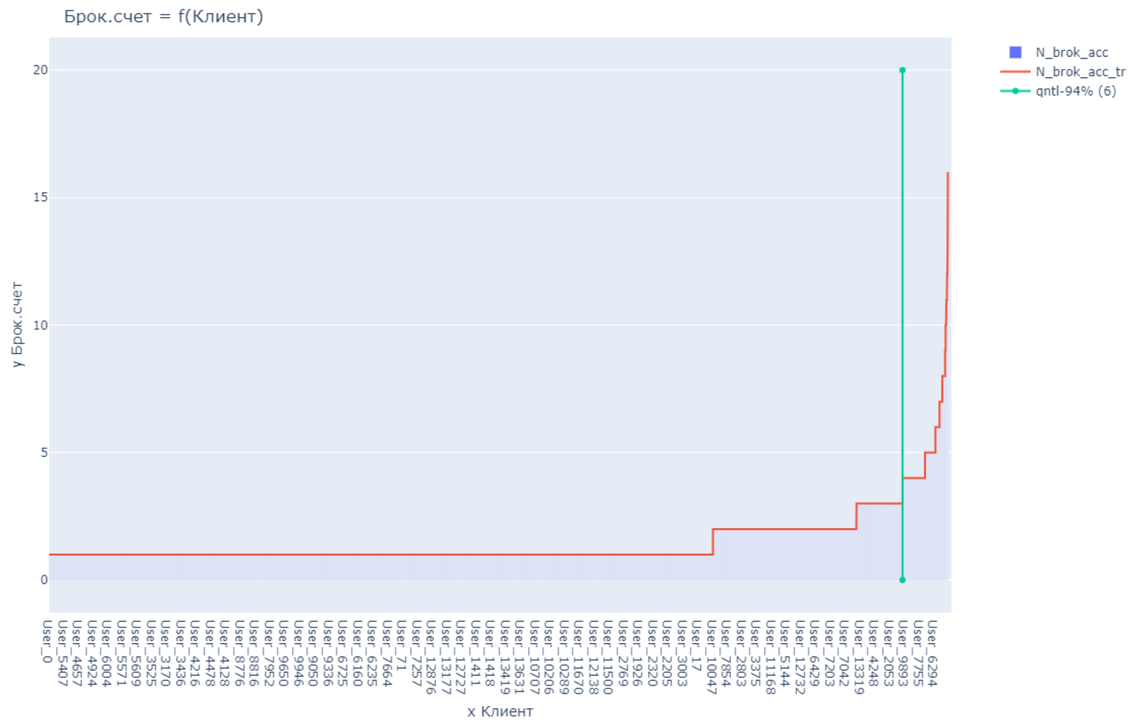
```
#
from plotly.subplots import make_subplots
import plotly.graph_objects as go

# asmpt_1 = X_Y[X_Y['    ']==X_Y['    '].quantile(.7)][0:1].index.values[0]

fig = make_subplots(rows=1, cols=1)
fig.add_trace(go.Bar(x=df_DB_fltr.index, y=df_DB_fltr['         '].values,␣
 ↪name='N_brok_acc'), row=1, col=1)
fig.add_trace(go.Scatter(x=df_DB_fltr.index, y=df_DB_fltr['         '].values,␣
 ↪name='N_brok_acc_tr'), row=1, col=1)
fig.add_trace(go.Scatter(x=['User_3427']*2, y=[0,20], name='qntl-94% (6)'))

fig.update_layout(legend_orientation="v",
                  legend=dict(x=1.2, xanchor="right"),
                  title="   .    = f(    )",
                  xaxis_title="x     ",
                  yaxis_title="y    .  ",
                  margin=dict(l=0, r=0, t=30, b=0)
                 )
fig.update_layout(
    autosize=False,
    width=900,
    height=700,)

fig.show()
```



[11]:
```
df_any.describe()
```

[11]:

```
count                            17271  17271.000000
mean   169 days 02:10:26.326163926     34.663181
std    102 days 04:46:58.340673475      3.167708
min                0 days 00:00:35      5.891855
25%        90 days 10:39:01.763777     31.865845
50%       143 days 09:11:07.263777     34.721424
75%       230 days 12:01:07.763777     37.462012
max       420 days 13:20:35.263777     39.991786
```

```
[22]: # 2Se urity
      # df_BA_fltr = df_DB_fltr[df_DB_fltr['        '] == 1].sort_values(['         '])
      # df_BA_fltr = df_BA_fltr.rename(columns={'       ':' ', '         ':' ', '   ':
       ↪' ',
      #                            '   ':' ', '     ':' ', '     2':'2', '    ':' '
      #                          })
      # np.max(df_BA_fltr)
```

```
[23]: # df_any_fltr = df_BA_fltr[['  ']].join(df_any.set_index('   '), on=['   '],␣
       ↪how='left') #lsuffix='_left', rsuffix='_right'
      # df_any_fltr
```

```
[ ]:
```

0.0.2 2 Law

2.1 Age < 18

```
[24]: #      < 18
      # min Age - 5.889117
```

```
[25]: # 2Lawyer
      df_any[df_any['    '] < 18]
```

```
[25]:                                                      \
      11306   User_6066    2015-04-07  Broker_account_7044  Strategy_59


                                                                2   \
      11306 2020-08-07 16:48:07         0  2021-02-26 01:12:03.088653



      11306 202 days 08:23:56.088653  5.891855
```

```
[ ]:
```

0.0.3 3 Product

3.1 Low Service Usage

```
[26]: df_any_fltr=\
      df_any[df_any['    '] >= 18][['    ','         ', '         ', '    ', '    ',
                                                '     ', '     2', '    ', '    ']].
       ↪reset_index(drop=True)
      df_any_fltr.head()
      # df_any_fltr3['    '] = df_any_fltr3.index
```

```
[26]:                                                 \
      0    User_0     1989-03-02      Broker_account_0    Strategy_0
      1    User_1     1981-07-15      Broker_account_1    Strategy_0
      2   User_10     1988-08-14     Broker_account_10    Strategy_2
```

```
3  User_10    1988-08-14  Broker_account_14023  Strategy_238
4  User_10    1988-08-14  Broker_account_14024  Strategy_236

                                                          2  \
0 2020-06-17 12:28:00                    0  2021-02-26 01:12:03.088653
1 2020-11-27 16:39:12  2021-02-02 15:11:57         2021-02-02 15:11:57
2 2020-12-09 11:59:27                    0  2021-02-26 01:12:03.088653
3 2020-09-24 12:54:06  2021-02-04 16:49:55         2021-02-04 16:49:55
4 2020-11-10 15:52:13                    0  2021-02-26 01:12:03.088653



0 253 days 12:44:03.088653  31.989049
1         66 days 22:32:45  39.619439
2  78 days 13:12:36.088653  32.536619
3        133 days 03:55:49  32.536619
4 107 days 09:19:50.088653  32.536619
```

[27]:
```
# 2Prod - Service
# df_any_fltr3.groupby(by=['    ', '    ']).count()#.sort_values(['         '])



#
df_any_fltr.groupby(by=['   ']).count().sort_values(['   '])
```

[27]:
```
                                                    \

Strategy_27         1              1                1       1          1
Strategy_306        1              1                1       1          1
Strategy_178        1              1                1       1          1
Strategy_308        1              1                1       1          1
Strategy_313        1              1                1       1          1
…              …            …              …     …        …
Strategy_362      644            644              644     644        644
Strategy_18       717            717              717     717        717
Strategy_17      1398           1398             1398    1398       1398
Strategy_16      1408           1408             1408    1408       1408
Strategy_59      1502           1502             1502    1502       1502

                   2

Strategy_27             1       1        1
Strategy_306            1       1        1
Strategy_178            1       1        1
Strategy_308            1       1        1
Strategy_313            1       1        1
…                   …     …      …
Strategy_362          644     644      644
```

```
Strategy_18          717     717      717
Strategy_17         1398    1398     1398
Strategy_16         1408    1408     1408
Strategy_59         1502    1502     1502

[331 rows x 8 columns]
```

[28]:

```
#
X_Y = df_any_fltr.groupby(by=['    ']).count().sort_values(['    '])
```

[30]:

```
# import pandas_profiling
# pandas_profiling.ProfileReport(X_Y[['    ']])
# X_Y['    '].describe().transpose(), len(df_any_fltr3['    '].unique()),␣
 ↪len(df_any_fltr3['    '].unique())
```

[32]:

```
X_Y['    '].quantile(.7),\
X_Y['    '].quantile(.7)/len(df_any_fltr['    '].unique())*100,\
X_Y['    '].quantile(.8),\
X_Y['    '].quantile(.8)/len(df_any_fltr['    '].unique())*100
```

[32]:

```
(15.0, 4.531722054380665, 36.0, 10.876132930513595)
```

[33]:

```
# X_Y[X_Y['    ']==X_Y['    '].quantile(.7)].index
# X_Y[X_Y['    ']==X_Y['    '].quantile(.8)].index
```

[35]:

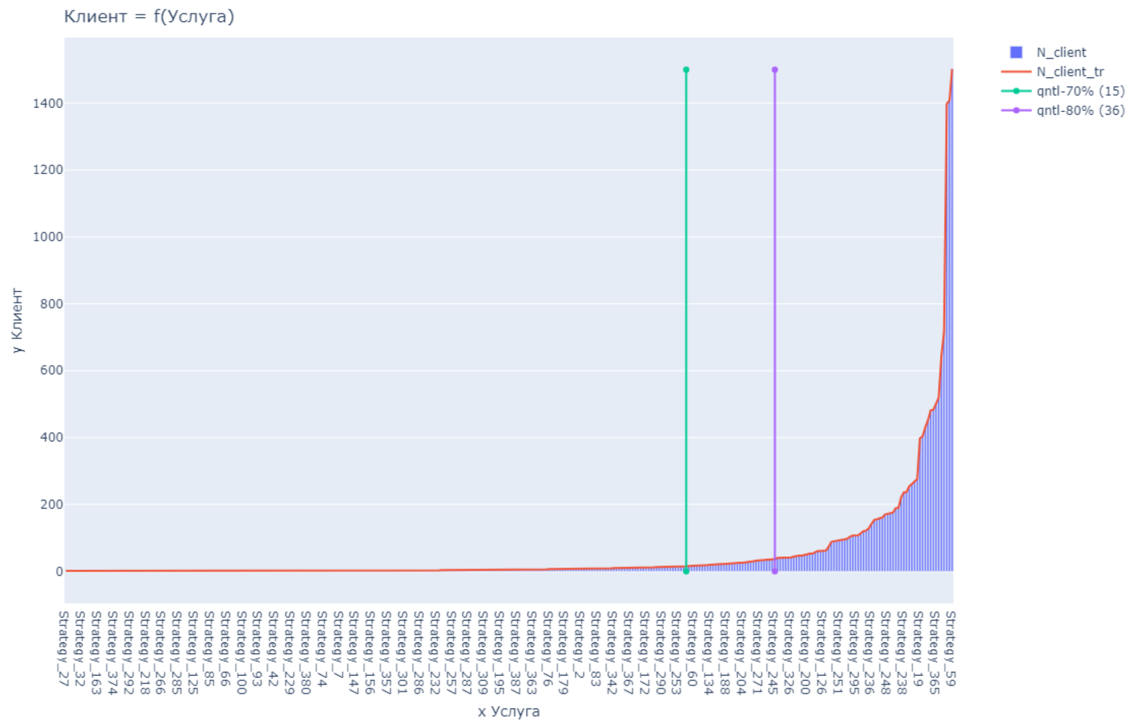
```
#
from plotly.subplots import make_subplots
import plotly.graph_objects as go

# asmpt_1 = X_Y[X_Y['    ']==X_Y['    '].quantile(.7)][0:1].index.values[0]
# asmpt_2 = X_Y[X_Y['    ']==X_Y['    '].quantile(.8)][0:1].index.values[0]

fig = make_subplots(rows=1, cols=1)
fig.add_trace(go.Bar(x=X_Y.index, y=X_Y['    '].values, name='N_client'), row=1,␣
 ↪col=1)
fig.add_trace(go.Scatter(x=X_Y.index, y=X_Y['    '].values, name='N_client_tr'),␣
 ↪row=1, col=1, )
fig.add_trace(go.Scatter(x=['Strategy_304']*2, y=[0,1500], name='qntl-70%␣
 ↪(15)'))
fig.add_trace(go.Scatter(x=['Strategy_245']*2, y=[0,1500], name='qntl-80%␣
 ↪(36)'))

fig.update_layout(legend_orientation="v",
                  legend=dict(x=1.2, xanchor="right"),
                  title="    = f(    )",
                  xaxis_title="x    ",
                  yaxis_title="y    ",
                  margin=dict(l=0, r=0, t=30, b=0)
                 )
fig.update_layout(
    autosize=False,
    width=900,
    height=700,)

fig.show()
```



[ ]:

```
```

3.2 Period Service Usage

[36]:

```
#
X_Y = df_any_fltr.sort_values(by=['    '])

# X_Y['    '].describe().transpose(), len(df_any_fltr3['    '].unique()),␣
↪len(df_any_fltr3['    '].unique())
f = lambda x: [x.days for x in df_any_fltr['    ']]
X_Y['    '] = f(5)
X_Y = X_Y.sort_values(by=['    '])
X_Y.head()
```

[36]:

```
                                                    \
7845      User_3783    1984-04-09   Broker_account_4236   Strategy_16
11876      User_641    1984-04-10    Broker_account_709   Strategy_17
6972      User_3286    1984-04-05   Broker_account_6811   Strategy_59
946      User_10417    1991-04-13  Broker_account_15105  Strategy_264
13131      User_749    1989-05-15    Broker_account_826   Strategy_17

                                                              2  \
7845  2020-02-25 17:45:51                    0  2021-02-26 01:12:03.088653
11876 2020-11-20 18:35:19                    0  2021-02-26 01:12:03.088653
6972  2020-06-30 22:59:51                    0  2021-02-26 01:12:03.088653
946   2020-11-20 18:34:50                    0  2021-02-26 01:12:03.088653
13131 2020-05-26 14:00:12  2021-01-21 16:03:10         2021-01-21 16:03:10



7845        0  36.884326
11876       0  39.882272
6972        0  36.895277
946         0  29.875428
13131       0  31.786448
```

[37]:

```
X_Y.describe()
```

[37]:

```
count  17270.000000  17270.000000
mean     168.599884     34.664847
std      102.191257      3.160224
min        0.000000     29.281314
25%       90.000000     31.865845
50%      143.000000     34.721424
75%      230.000000     37.462012
max      420.000000     39.991786
```

[38]:

```
X_Y_p = X_Y.groupby(by=['    ']).mean().sort_values(['    '])
X_Y_p.head()
```

[38]:

```


Strategy_280    10.0  31.950719
Strategy_306    11.0  34.559890
Strategy_293    13.0  36.457221
Strategy_180    26.0  33.459274
Strategy_34     34.5  31.596167
```

[39]:

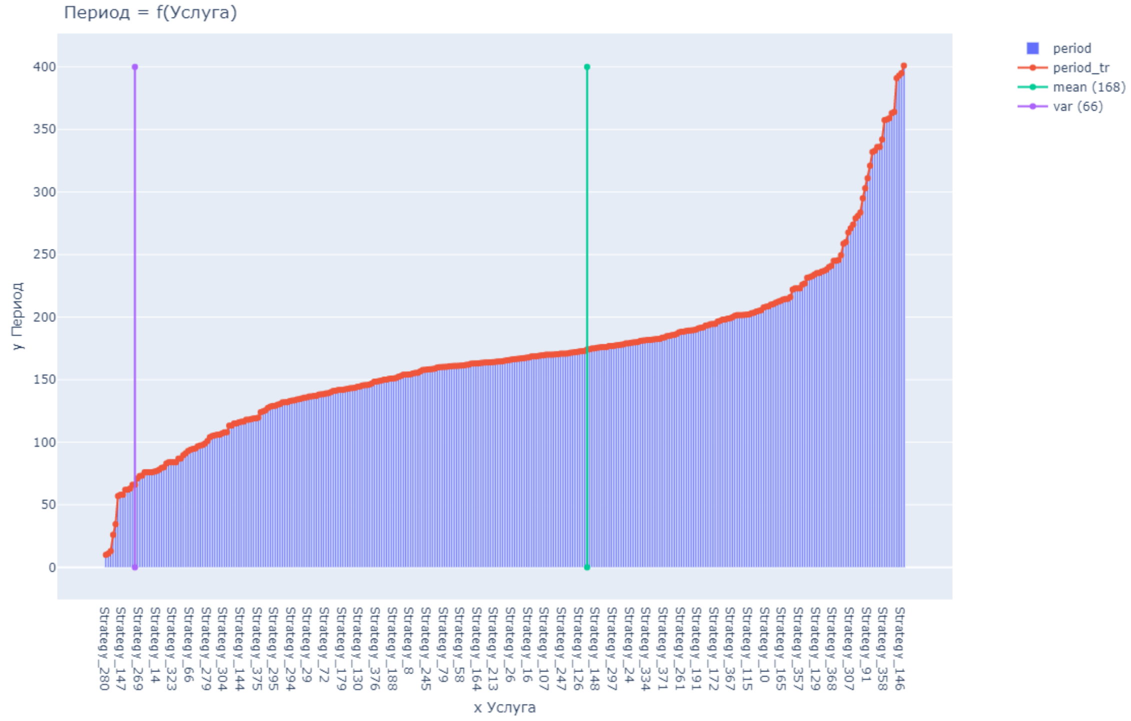
```
#
from plotly.subplots import make_subplots
import plotly.graph_objects as go
```

```
# asmpt_1 = X_Y[X_Y['    ']==X_Y['    '].quantile(.7)][0:1].index.values[0]
# asmpt_2 = X_Y[X_Y['    ']==X_Y['    '].quantile(.8)][0:1].index.values[0]

fig = make_subplots(rows=1, cols=1)
fig.add_trace(go.Bar(x=X_Y_p.index, y=X_Y_p['    '].values, name='period'))
fig.add_trace(go.Scatter(x=X_Y_p.index, y=X_Y_p['    '].values,␣
 ↪name='period_tr', mode = 'lines+markers'))
fig.add_trace(go.Scatter(x=['Strategy_249']*2, y=[0,400], name='mean (168)'))
fig.add_trace(go.Scatter(x=['Strategy_121']*2, y=[0,400], name='var (66)'))

fig.update_layout(legend_orientation="v",
                  legend=dict(x=1.2, xanchor="right"),
                  title="      = f(    )",
                  xaxis_title="x     ",
                  yaxis_title="y     ",
                  margin=dict(l=0, r=0, t=30, b=0)
                 )
fig.update_layout(
    autosize=False,
    width=900,
    height=700,)

fig.show()
```



[ ]:

```

```

0.0.4 4 Market

4.1 Low Service Usage

[40]:

```
# df_any_open = df_any[df_any['     '] == 0]
# df_any_close = df_any[df_any['     '] != 0]
# X_Y[X_Y['     '] ==0]
X_Y_cl = X_Y.groupby(by=['   ']).count().sort_values(['   '])
X_Y_cl.head()
```

[40]:

```
                                                       \

User_0                 1                1       1       1          1
User_5129              1                1       1       1          1
User_5127              1                1       1       1          1
User_5126              1                1       1       1          1
User_5125              1                1       1       1          1

                2

User_0               1       1        1
User_5129            1       1        1
User_5127            1       1        1
User_5126            1       1        1
User_5125            1       1        1
```

[41]:

```
X_Y_cl[X_Y_cl['   ']<3].shape[0]/X_Y_cl.shape[0]*100
```

[41]:

```
85.87072390735536
```

[42]:

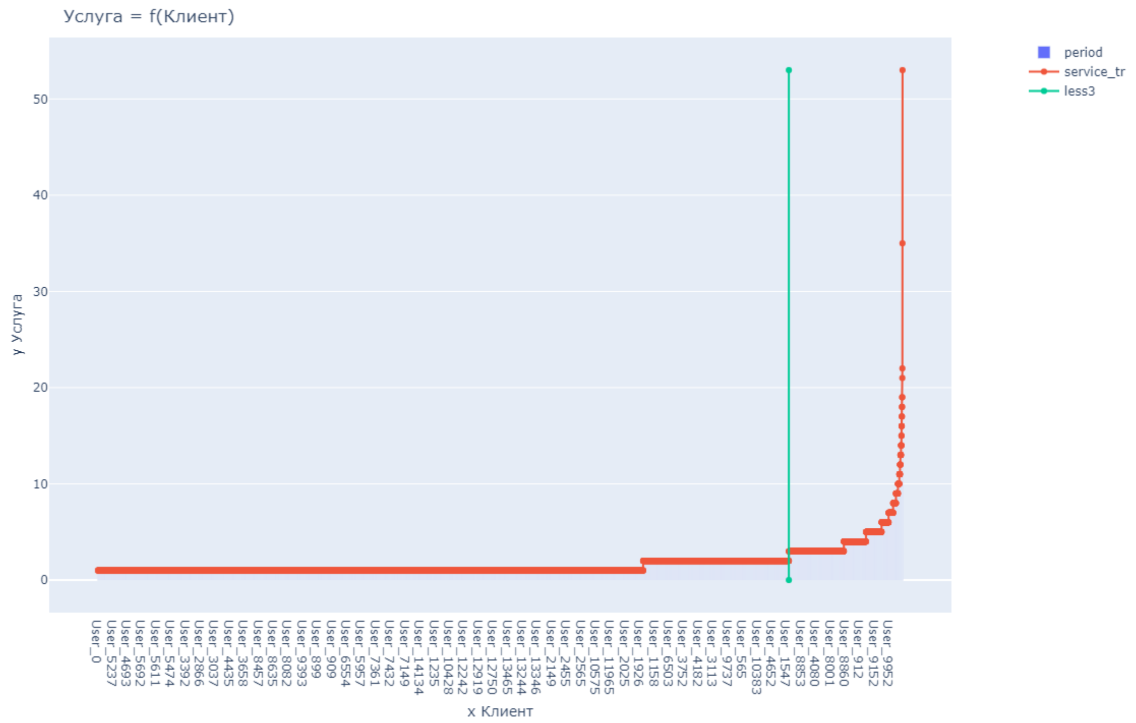
```
#
from plotly.subplots import make_subplots
import plotly.graph_objects as go

# asmpt_1 = X_Y[X_Y['   ']==X_Y['   '].quantile(.7)][0:1].index.values[0]
# asmpt_2 = X_Y[X_Y['   ']==X_Y['   '].quantile(.8)][0:1].index.values[0]

fig = make_subplots(rows=1, cols=1)
fig.add_trace(go.Bar(x=X_Y_cl.index, y=X_Y_cl['   '].values, name='period'))
fig.add_trace(go.Scatter(x=X_Y_cl.index, y=X_Y_cl['   '].values,␣
 ↪name='service_tr', mode = 'lines+markers'))
fig.add_trace(go.Scatter(x=['User_1535']*2, y=[0,53], name='less3'))
# fig.add_trace(go.Scatter(x=['Strategy_121']*2, y=[0,50], name='var (66)'))

fig.update_layout(legend_orientation="v",
                  legend=dict(x=1.2, xanchor="right"),
                  title="    = f(   )",
                  xaxis_title="x     ",
                  yaxis_title="y     ",
                  margin=dict(l=0, r=0, t=30, b=0)
                 )
fig.update_layout(
    autosize=False,
    width=900,
    height=700,)

fig.show()
```



[ ]:

```
```

[344]:

```
#                 (          ,                          )
df_any_close = df_any[df_any['     '] != 0]
df_any_close
```

[344]:

```
                                                    \
1         User_1   1981-07-15      Broker_account_1    Strategy_0
3        User_10   1988-08-14  Broker_account_14023  Strategy_238
11    User_10000   1989-08-01  Broker_account_14041  Strategy_244
13    User_10001   1984-07-15  Broker_account_14042  Strategy_237
14    User_10001   1984-07-15  Broker_account_14043  Strategy_251
…            …            …                     …             …
```

```
17256   User_9994    1983-07-09  Broker_account_14028  Strategy_239
17257   User_9994    1983-07-09  Broker_account_14029  Strategy_236
17258   User_9994    1983-07-09  Broker_account_14030  Strategy_247
17260   User_9995    1984-05-15  Broker_account_14032  Strategy_246
17270   User_9999    1984-09-07  Broker_account_14037  Strategy_258

                                                       2  \
1     2020-11-27 16:39:12  2021-02-02 15:11:57  2021-02-02 15:11:57
3     2020-09-24 12:54:06  2021-02-04 16:49:55  2021-02-04 16:49:55
11    2020-11-09 19:05:27  2020-12-24 13:24:11  2020-12-24 13:24:11
13    2020-12-07 20:49:10  2020-12-26 23:17:55  2020-12-26 23:17:55
14    2020-11-05 14:40:50  2020-12-17 00:47:36  2020-12-17 00:47:36
…                   …                  …                  …
17256 2020-11-24 20:27:42  2020-12-21 16:15:32  2020-12-21 16:15:32
17257 2020-11-27 16:50:04  2021-01-19 11:01:44  2021-01-19 11:01:44
17258 2020-11-30 14:02:33  2021-01-19 11:01:07  2021-01-19 11:01:07
17260 2020-08-04 13:13:33  2021-01-19 13:41:35  2021-01-19 13:41:35
17270 2020-09-16 10:37:06  2020-12-07 20:49:09  2020-12-07 20:49:09


1      66 days 22:32:45  39.619439
3     133 days 03:55:49  32.536619
11     44 days 18:18:44  31.572895
13     19 days 02:28:45  36.618754
14     41 days 10:06:46  36.618754
…                  …        …
17256  26 days 19:47:50  37.637235
17257  52 days 18:11:40  37.637235
17258  49 days 20:58:34  37.637235
17260 168 days 00:28:02  36.785763
17270  82 days 10:12:03  36.470910

[4828 rows x 9 columns]
```

```
[ ]:
```