9 - Integrated Project 2

January 3, 2024

Prepare a prototype of a machine learning model for Zyfra. The company develops efficiency solutions for heavy industry. The model should predict the amount of gold recovered from gold ore. You have the data on extraction and purification. The model will help to optimize the production and eliminate unprofitable parameters.

Mined ore undergoes primary processing to get the ore mixture or rougher feed, which is the raw material for flotation (also known as the rougher process). After flotation, the material is sent to two-stage purification.

[ ]:
```python
import pandas as pd
from sklearn.metrics import mean_absolute_error
import numpy as np
import matplotlib.pyplot as plt
```

[ ]:
```python
gold_train = pd.read_csv('/datasets/gold_recovery_train.csv')
gold_test = pd.read_csv('/datasets/gold_recovery_test.csv')
gold_full = pd.read_csv('/datasets/gold_recovery_full.csv')
```

[ ]:
```python
gold_train
```

[ ]:
```
                      date  final.output.concentrate_ag  \
0      2016-01-15 00:00:00                     6.055403
1      2016-01-15 01:00:00                     6.029369
2      2016-01-15 02:00:00                     6.055926
3      2016-01-15 03:00:00                     6.047977
4      2016-01-15 04:00:00                     6.148599
…                        …                          …
16855  2018-08-18 06:59:59                     3.224920
16856  2018-08-18 07:59:59                     3.195978
16857  2018-08-18 08:59:59                     3.109998
16858  2018-08-18 09:59:59                     3.367241
16859  2018-08-18 10:59:59                     3.598375

       final.output.concentrate_pb  final.output.concentrate_sol  \
0                         9.889648                      5.507324
1                         9.968944                      5.257781
2                        10.213995                      5.383759
3                         9.977019                      4.858634
4                        10.142511                      4.939416
```

```
…                         …                               …
16855             11.356233                        6.803482
16856             11.349355                        6.862249
16857             11.434366                        6.886013
16858             11.625587                        6.799433
16859             11.737832                        6.717509

       final.output.concentrate_au  final.output.recovery  \
0                        42.192020              70.541216
1                        42.701629              69.266198
2                        42.657501              68.116445
3                        42.689819              68.347543
4                        42.774141              66.927016
…                                …                      …
16855                    46.713954              73.755150
16856                    46.866780              69.049291
16857                    46.795691              67.002189
16858                    46.408188              65.523246
16859                    46.299438              70.281454

       final.output.tail_ag  final.output.tail_pb  final.output.tail_sol  \
0                 10.411962              0.895447              16.904297
1                 10.462676              0.927452              16.634514
2                 10.507046              0.953716              16.208849
3                 10.422762              0.883763              16.532835
4                 10.360302              0.792826              16.525686
…                         …                     …                      …
16855              8.769645              3.141541              10.403181
16856              8.897321              3.130493              10.549470
16857              8.529606              2.911418              11.115147
16858              8.777171              2.819214              10.463847
16859              8.406690              2.517518              10.652193

       final.output.tail_au  …  secondary_cleaner.state.floatbank4_a_air  \
0                  2.143149  …                                 14.016835
1                  2.224930  …                                 13.992281
2                  2.257889  …                                 14.015015
3                  2.146849  …                                 14.036510
4                  2.055292  …                                 14.027298
…                         …  …                                       …
16855              1.529220  …                                 23.031497
16856              1.612542  …                                 22.960095
16857              1.596616  …                                 23.015718
16858              1.602879  …                                 23.024963
16859              1.389434  …                                 23.018622

       secondary_cleaner.state.floatbank4_a_level  \
```

```
0                                       -502.488007
1                                       -505.503262
2                                       -502.520901
3                                       -500.857308
4                                       -499.838632
…                                             …
16855                                   -501.167942
16856                                   -501.612783
16857                                   -501.711599
16858                                   -501.153409
16859                                   -500.492702

       secondary_cleaner.state.floatbank4_b_air  \
0                                       12.099931
1                                       11.950531
2                                       11.912783
3                                       11.999550
4                                       11.953070
…                                           …
16855                                   20.007571
16856                                   20.035660
16857                                   19.951231
16858                                   20.054122
16859                                   20.020205

       secondary_cleaner.state.floatbank4_b_level  \
0                                       -504.715942
1                                       -501.331529
2                                       -501.133383
3                                       -501.193686
4                                       -501.053894
…                                             …
16855                                   -499.740028
16856                                   -500.251357
16857                                   -499.857027
16858                                   -500.314711
16859                                   -500.220296

       secondary_cleaner.state.floatbank5_a_air  \
0                                        9.925633
1                                       10.039245
2                                       10.070913
3                                        9.970366
4                                        9.925709
…                                           …
16855                                   18.006038
16856                                   17.998535
```

```
16857                                       18.019543
16858                                       17.979515
16859                                       17.963512

       secondary_cleaner.state.floatbank5_a_level  \
0                                     -498.310211
1                                     -500.169983
2                                     -500.129135
3                                     -499.201640
4                                     -501.686727
…                                               …
16855                                 -499.834374
16856                                 -500.395178
16857                                 -500.451156
16858                                 -499.272871
16859                                 -499.939490

       secondary_cleaner.state.floatbank5_b_air  \
0                                       8.079666
1                                       7.984757
2                                       8.013877
3                                       7.977324
4                                       7.894242
…                                             …
16855                                  13.001114
16856                                  12.954048
16857                                  13.023431
16858                                  12.992404
16859                                  12.990306

       secondary_cleaner.state.floatbank5_b_level  \
0                                     -500.470978
1                                     -500.582168
2                                     -500.517572
3                                     -500.255908
4                                     -500.356035
…                                               …
16855                                 -500.155694
16856                                 -499.895163
16857                                 -499.914391
16858                                 -499.976268
16859                                 -500.080993

       secondary_cleaner.state.floatbank6_a_air  \
0                                      14.151341
1                                      13.998353
2                                      14.028663
```

```
3                                    14.005551
4                                    13.996647
…                                        …
16855                                20.007840
16856                                19.968498
16857                                19.990885
16858                                20.013986
16859                                19.990336

       secondary_cleaner.state.floatbank6_a_level
0                                    -605.841980
1                                    -599.787184
2                                    -601.427363
3                                    -599.996129
4                                    -601.496691
…                                          …
16855                                -501.296428
16856                                -501.041608
16857                                -501.518452
16858                                -500.625471
16859                                -499.191575

[16860 rows x 87 columns]
```

[ ]:
```python
gold_train['date'] = pd.to_datetime(gold_train['date'])
```

[ ]:
```python
gold_test
```

[ ]:
```
                     date  primary_cleaner.input.sulfate  \
0     2016-09-01 00:59:59                     210.800909
1     2016-09-01 01:59:59                     215.392455
2     2016-09-01 02:59:59                     215.259946
3     2016-09-01 03:59:59                     215.336236
4     2016-09-01 04:59:59                     199.099327
…                     …                            …
5851  2017-12-31 19:59:59                     173.957757
5852  2017-12-31 20:59:59                     172.910270
5853  2017-12-31 21:59:59                     171.135718
5854  2017-12-31 22:59:59                     179.697158
5855  2017-12-31 23:59:59                     181.556856

      primary_cleaner.input.depressant  primary_cleaner.input.feed_size  \
0                            14.993118                         8.080000
1                            14.987471                         8.080000
2                            12.884934                         7.786667
3                            12.006805                         7.640000
4                            10.682530                         7.530000
```

```
…                            …                                 …
5851                 15.963399                          8.070000
5852                 16.002605                          8.070000
5853                 15.993669                          8.070000
5854                 15.438979                          8.070000
5855                 14.995850                          8.070000

      primary_cleaner.input.xanthate  primary_cleaner.state.floatbank8_a_air  \
0                           1.005021                             1398.981301
1                           0.990469                             1398.777912
2                           0.996043                             1398.493666
3                           0.863514                             1399.618111
4                           0.805575                             1401.268123
…                               …                                     …
5851                        0.896701                             1401.930554
5852                        0.896519                             1447.075722
5853                        1.165996                             1498.836182
5854                        1.501068                             1498.466243
5855                        1.623454                             1498.096303

      primary_cleaner.state.floatbank8_a_level  \
0                                  -500.225577
1                                  -500.057435
2                                  -500.868360
3                                  -498.863574
4                                  -500.808305
…                                         …
5851                               -499.728848
5852                               -494.716823
5853                               -501.770403
5854                               -500.483984
5855                               -499.796922

      primary_cleaner.state.floatbank8_b_air  \
0                                1399.144926
1                                1398.055362
2                                1398.860436
3                                1397.440120
4                                1398.128818
…                                       …
5851                             1401.441445
5852                             1448.851892
5853                             1499.572353
5854                             1497.986986
5855                             1501.743791

      primary_cleaner.state.floatbank8_b_level  \
```

```
0                                     -499.919735
1                                     -499.778182
2                                     -499.764529
3                                     -499.211024
4                                     -499.504543
…                                             …
5851                                  -499.193423
5852                                  -465.963026
5853                                  -495.516347
5854                                  -519.200340
5855                                  -505.146931

      primary_cleaner.state.floatbank8_c_air  …  \
0                                1400.102998  …
1                                1396.151033  …
2                                1398.075709  …
3                                1400.129303  …
4                                1402.172226  …
…                                        …  …
5851                             1399.810313  …
5852                             1443.890424  …
5853                             1502.749213  …
5854                             1496.569047  …
5855                             1499.535978  …

      secondary_cleaner.state.floatbank4_a_air  \
0                                    12.023554
1                                    12.058140
2                                    11.962366
3                                    12.033091
4                                    12.025367
…                                          …
5851                                 13.995957
5852                                 16.749781
5853                                 19.994130
5854                                 19.958760
5855                                 20.034715

      secondary_cleaner.state.floatbank4_a_level  \
0                                    -497.795834
1                                    -498.695773
2                                    -498.767484
3                                    -498.350935
4                                    -500.786497
…                                            …
5851                                 -500.157454
5852                                 -496.031539
```

```
5853                                 -499.791312
5854                                 -499.958750
5855                                 -500.728588

      secondary_cleaner.state.floatbank4_b_air  \
0                                     8.016656
1                                     8.130979
2                                     8.096893
3                                     8.074946
4                                     8.054678
…                                          …
5851                                 12.069155
5852                                 13.365371
5853                                 15.101425
5854                                 15.026853
5855                                 14.914199

      secondary_cleaner.state.floatbank4_b_level  \
0                                    -501.289139
1                                    -499.634209
2                                    -500.827423
3                                    -499.474407
4                                    -500.397500
…                                            …
5851                                 -499.673279
5852                                 -499.122723
5853                                 -499.936252
5854                                 -499.723143
5855                                 -499.948518

      secondary_cleaner.state.floatbank5_a_air  \
0                                     7.946562
1                                     7.958270
2                                     8.071056
3                                     7.897085
4                                     8.107890
…                                          …
5851                                  7.977259
5852                                  9.288553
5853                                 10.989181
5854                                 11.011607
5855                                 10.986607

      secondary_cleaner.state.floatbank5_a_level  \
0                                    -432.317850
1                                    -525.839648
2                                    -500.801673
```

```
3                                         -500.868509
4                                         -509.526725
…                                                 …
5851                                      -499.516126
5852                                      -496.892967
5853                                      -498.347898
5854                                      -499.985046
5855                                      -500.658027

      secondary_cleaner.state.floatbank5_b_air  \
0                                       4.872511
1                                       4.878850
2                                       4.905125
3                                       4.931400
4                                       4.957674
…                                            …
5851                                    5.933319
5852                                    7.372897
5853                                    9.020944
5854                                    9.009783
5855                                    8.989497

      secondary_cleaner.state.floatbank5_b_level  \
0                                      -500.037437
1                                      -500.162375
2                                      -499.828510
3                                      -499.963623
4                                      -500.360026
…                                               …
5851                                   -499.965973
5852                                   -499.942956
5853                                   -500.040448
5854                                   -499.937902
5855                                   -500.337588

      secondary_cleaner.state.floatbank6_a_air  \
0                                      26.705889
1                                      25.019940
2                                      24.994862
3                                      24.948919
4                                      25.003331
…                                            …
5851                                    8.987171
5852                                    8.986832
5853                                    8.982038
5854                                    9.012660
5855                                    8.988632
```

```
      secondary_cleaner.state.floatbank6_a_level
0                                  -499.709414
1                                  -499.819438
2                                  -500.622559
3                                  -498.709987
4                                  -500.856333
…                                          …
5851                               -499.755909
5852                               -499.903761
5853                               -497.789882
5854                               -500.154284
5855                               -500.764937

[5856 rows x 53 columns]
```

[ ]:
```
gold_test['date'] = pd.to_datetime(gold_test['date'])
```

[ ]:
```
gold_full
```

[ ]:
```
                     date  final.output.concentrate_ag  \
0     2016-01-15 00:00:00                     6.055403
1     2016-01-15 01:00:00                     6.029369
2     2016-01-15 02:00:00                     6.055926
3     2016-01-15 03:00:00                     6.047977
4     2016-01-15 04:00:00                     6.148599
…                     …                          …
22711 2018-08-18 06:59:59                     3.224920
22712 2018-08-18 07:59:59                     3.195978
22713 2018-08-18 08:59:59                     3.109998
22714 2018-08-18 09:59:59                     3.367241
22715 2018-08-18 10:59:59                     3.598375

       final.output.concentrate_pb  final.output.concentrate_sol  \
0                         9.889648                      5.507324
1                         9.968944                      5.257781
2                        10.213995                      5.383759
3                         9.977019                      4.858634
4                        10.142511                      4.939416
…                              …                           …
22711                    11.356233                      6.803482
22712                    11.349355                      6.862249
22713                    11.434366                      6.886013
22714                    11.625587                      6.799433
22715                    11.737832                      6.717509

       final.output.concentrate_au  final.output.recovery  \
```

```
0                        42.192020            70.541216
1                        42.701629            69.266198
2                        42.657501            68.116445
3                        42.689819            68.347543
4                        42.774141            66.927016
…                            …                   …
22711                    46.713954            73.755150
22712                    46.866780            69.049291
22713                    46.795691            67.002189
22714                    46.408188            65.523246
22715                    46.299438            70.281454

       final.output.tail_ag  final.output.tail_pb  final.output.tail_sol  \
0                 10.411962              0.895447              16.904297
1                 10.462676              0.927452              16.634514
2                 10.507046              0.953716              16.208849
3                 10.422762              0.883763              16.532835
4                 10.360302              0.792826              16.525686
…                     …                     …                     …
22711              8.769645              3.141541              10.403181
22712              8.897321              3.130493              10.549470
22713              8.529606              2.911418              11.115147
22714              8.777171              2.819214              10.463847
22715              8.406690              2.517518              10.652193

       final.output.tail_au  …  secondary_cleaner.state.floatbank4_a_air  \
0                  2.143149  …                                 14.016835
1                  2.224930  …                                 13.992281
2                  2.257889  …                                 14.015015
3                  2.146849  …                                 14.036510
4                  2.055292  …                                 14.027298
…                     …   …                                       …
22711              1.529220  …                                 23.031497
22712              1.612542  …                                 22.960095
22713              1.596616  …                                 23.015718
22714              1.602879  …                                 23.024963
22715              1.389434  …                                 23.018622

       secondary_cleaner.state.floatbank4_a_level  \
0                                     -502.488007
1                                     -505.503262
2                                     -502.520901
3                                     -500.857308
4                                     -499.838632
…                                           …
22711                                 -501.167942
22712                                 -501.612783
```

```
22713                                   -501.711599
22714                                   -501.153409
22715                                   -500.492702

       secondary_cleaner.state.floatbank4_b_air  \
0                                     12.099931
1                                     11.950531
2                                     11.912783
3                                     11.999550
4                                     11.953070
…                                         …
22711                                 20.007571
22712                                 20.035660
22713                                 19.951231
22714                                 20.054122
22715                                 20.020205

       secondary_cleaner.state.floatbank4_b_level  \
0                                      -504.715942
1                                      -501.331529
2                                      -501.133383
3                                      -501.193686
4                                      -501.053894
…                                              …
22711                                  -499.740028
22712                                  -500.251357
22713                                  -499.857027
22714                                  -500.314711
22715                                  -500.220296

       secondary_cleaner.state.floatbank5_a_air  \
0                                      9.925633
1                                     10.039245
2                                     10.070913
3                                      9.970366
4                                      9.925709
…                                         …
22711                                 18.006038
22712                                 17.998535
22713                                 18.019543
22714                                 17.979515
22715                                 17.963512

       secondary_cleaner.state.floatbank5_a_level  \
0                                      -498.310211
1                                      -500.169983
2                                      -500.129135
```

```
3                                        -499.201640
4                                        -501.686727
…                                                  …
22711                                    -499.834374
22712                                    -500.395178
22713                                    -500.451156
22714                                    -499.272871
22715                                    -499.939490

       secondary_cleaner.state.floatbank5_b_air  \
0                                      8.079666
1                                      7.984757
2                                      8.013877
3                                      7.977324
4                                      7.894242
…                                             …
22711                                 13.001114
22712                                 12.954048
22713                                 13.023431
22714                                 12.992404
22715                                 12.990306

       secondary_cleaner.state.floatbank5_b_level  \
0                                        -500.470978
1                                        -500.582168
2                                        -500.517572
3                                        -500.255908
4                                        -500.356035
…                                                  …
22711                                    -500.155694
22712                                    -499.895163
22713                                    -499.914391
22714                                    -499.976268
22715                                    -500.080993

       secondary_cleaner.state.floatbank6_a_air  \
0                                     14.151341
1                                     13.998353
2                                     14.028663
3                                     14.005551
4                                     13.996647
…                                             …
22711                                 20.007840
22712                                 19.968498
22713                                 19.990885
22714                                 20.013986
22715                                 19.990336
```

```
       secondary_cleaner.state.floatbank6_a_level
0                                     -605.841980
1                                     -599.787184
2                                     -601.427363
3                                     -599.996129
4                                     -601.496691
…                                               …
22711                                 -501.296428
22712                                 -501.041608
22713                                 -501.518452
22714                                 -500.625471
22715                                 -499.191575

[22716 rows x 87 columns]
```

[ ]:
```
gold_full['date'] = pd.to_datetime(gold_full['date'])
```

[ ]:
```
col_mapping_dict = {c[0]:c[1] for c in enumerate(gold_train.columns)}
col_mapping_dict
```

[ ]:
```
{0: 'date',
 1: 'final.output.concentrate_ag',
 2: 'final.output.concentrate_pb',
 3: 'final.output.concentrate_sol',
 4: 'final.output.concentrate_au',
 5: 'final.output.recovery',
 6: 'final.output.tail_ag',
 7: 'final.output.tail_pb',
 8: 'final.output.tail_sol',
 9: 'final.output.tail_au',
 10: 'primary_cleaner.input.sulfate',
 11: 'primary_cleaner.input.depressant',
 12: 'primary_cleaner.input.feed_size',
 13: 'primary_cleaner.input.xanthate',
 14: 'primary_cleaner.output.concentrate_ag',
 15: 'primary_cleaner.output.concentrate_pb',
 16: 'primary_cleaner.output.concentrate_sol',
 17: 'primary_cleaner.output.concentrate_au',
 18: 'primary_cleaner.output.tail_ag',
 19: 'primary_cleaner.output.tail_pb',
 20: 'primary_cleaner.output.tail_sol',
 21: 'primary_cleaner.output.tail_au',
 22: 'primary_cleaner.state.floatbank8_a_air',
 23: 'primary_cleaner.state.floatbank8_a_level',
 24: 'primary_cleaner.state.floatbank8_b_air',
 25: 'primary_cleaner.state.floatbank8_b_level',
```

```
26: 'primary_cleaner.state.floatbank8_c_air',
27: 'primary_cleaner.state.floatbank8_c_level',
28: 'primary_cleaner.state.floatbank8_d_air',
29: 'primary_cleaner.state.floatbank8_d_level',
30: 'rougher.calculation.sulfate_to_au_concentrate',
31: 'rougher.calculation.floatbank10_sulfate_to_au_feed',
32: 'rougher.calculation.floatbank11_sulfate_to_au_feed',
33: 'rougher.calculation.au_pb_ratio',
34: 'rougher.input.feed_ag',
35: 'rougher.input.feed_pb',
36: 'rougher.input.feed_rate',
37: 'rougher.input.feed_size',
38: 'rougher.input.feed_sol',
39: 'rougher.input.feed_au',
40: 'rougher.input.floatbank10_sulfate',
41: 'rougher.input.floatbank10_xanthate',
42: 'rougher.input.floatbank11_sulfate',
43: 'rougher.input.floatbank11_xanthate',
44: 'rougher.output.concentrate_ag',
45: 'rougher.output.concentrate_pb',
46: 'rougher.output.concentrate_sol',
47: 'rougher.output.concentrate_au',
48: 'rougher.output.recovery',
49: 'rougher.output.tail_ag',
50: 'rougher.output.tail_pb',
51: 'rougher.output.tail_sol',
52: 'rougher.output.tail_au',
53: 'rougher.state.floatbank10_a_air',
54: 'rougher.state.floatbank10_a_level',
55: 'rougher.state.floatbank10_b_air',
56: 'rougher.state.floatbank10_b_level',
57: 'rougher.state.floatbank10_c_air',
58: 'rougher.state.floatbank10_c_level',
59: 'rougher.state.floatbank10_d_air',
60: 'rougher.state.floatbank10_d_level',
61: 'rougher.state.floatbank10_e_air',
62: 'rougher.state.floatbank10_e_level',
63: 'rougher.state.floatbank10_f_air',
64: 'rougher.state.floatbank10_f_level',
65: 'secondary_cleaner.output.tail_ag',
66: 'secondary_cleaner.output.tail_pb',
67: 'secondary_cleaner.output.tail_sol',
68: 'secondary_cleaner.output.tail_au',
69: 'secondary_cleaner.state.floatbank2_a_air',
70: 'secondary_cleaner.state.floatbank2_a_level',
71: 'secondary_cleaner.state.floatbank2_b_air',
72: 'secondary_cleaner.state.floatbank2_b_level',
```

```
 73: 'secondary_cleaner.state.floatbank3_a_air',
 74: 'secondary_cleaner.state.floatbank3_a_level',
 75: 'secondary_cleaner.state.floatbank3_b_air',
 76: 'secondary_cleaner.state.floatbank3_b_level',
 77: 'secondary_cleaner.state.floatbank4_a_air',
 78: 'secondary_cleaner.state.floatbank4_a_level',
 79: 'secondary_cleaner.state.floatbank4_b_air',
 80: 'secondary_cleaner.state.floatbank4_b_level',
 81: 'secondary_cleaner.state.floatbank5_a_air',
 82: 'secondary_cleaner.state.floatbank5_a_level',
 83: 'secondary_cleaner.state.floatbank5_b_air',
 84: 'secondary_cleaner.state.floatbank5_b_level',
 85: 'secondary_cleaner.state.floatbank6_a_air',
 86: 'secondary_cleaner.state.floatbank6_a_level'}
```

[ ]:
```
gold_train.isna().sum().sort_values(ascending=False)/len(gold_train)
```

[ ]:
```
rougher.output.recovery                               0.152610
rougher.output.tail_ag                                0.133452
rougher.output.tail_sol                               0.133393
rougher.output.tail_au                                0.133393
secondary_cleaner.output.tail_sol                     0.117794
                                                        …
rougher.calculation.sulfate_to_au_concentrate         0.001601
rougher.calculation.floatbank10_sulfate_to_au_feed    0.001601
rougher.calculation.floatbank11_sulfate_to_au_feed    0.001601
primary_cleaner.input.feed_size                       0.000000
date                                                  0.000000
Length: 87, dtype: float64
```

[ ]:
```
gold_test.isna().sum().sort_values(ascending=False)/len(gold_test)
```

[ ]:
```
rougher.input.floatbank11_xanthate            0.060280
primary_cleaner.input.sulfate                 0.051571
primary_cleaner.input.depressant              0.048497
rougher.input.floatbank10_sulfate             0.043887
primary_cleaner.input.xanthate                0.028347
rougher.input.floatbank10_xanthate            0.021004
rougher.input.feed_sol                        0.011441
rougher.input.floatbank11_sulfate             0.009392
rougher.input.feed_rate                       0.006831
secondary_cleaner.state.floatbank3_a_air      0.005806
secondary_cleaner.state.floatbank2_b_air      0.003928
rougher.input.feed_size                       0.003757
secondary_cleaner.state.floatbank2_a_air      0.003415
rougher.state.floatbank10_e_air               0.002903
rougher.state.floatbank10_d_air               0.002903
```

```
rougher.state.floatbank10_a_air              0.002903
rougher.state.floatbank10_b_air              0.002903
rougher.state.floatbank10_c_air              0.002903
rougher.state.floatbank10_f_air              0.002903
primary_cleaner.state.floatbank8_a_air       0.002732
primary_cleaner.state.floatbank8_a_level     0.002732
rougher.input.feed_au                        0.002732
primary_cleaner.state.floatbank8_d_air       0.002732
primary_cleaner.state.floatbank8_b_air       0.002732
primary_cleaner.state.floatbank8_d_level     0.002732
primary_cleaner.state.floatbank8_b_level     0.002732
rougher.input.feed_pb                        0.002732
primary_cleaner.state.floatbank8_c_air       0.002732
rougher.input.feed_ag                        0.002732
primary_cleaner.state.floatbank8_c_level     0.002732
secondary_cleaner.state.floatbank6_a_level   0.002732
rougher.state.floatbank10_b_level            0.002732
rougher.state.floatbank10_a_level            0.002732
secondary_cleaner.state.floatbank6_a_air     0.002732
secondary_cleaner.state.floatbank5_b_level   0.002732
secondary_cleaner.state.floatbank5_b_air     0.002732
secondary_cleaner.state.floatbank5_a_level   0.002732
secondary_cleaner.state.floatbank5_a_air     0.002732
secondary_cleaner.state.floatbank4_b_level   0.002732
secondary_cleaner.state.floatbank4_b_air     0.002732
secondary_cleaner.state.floatbank4_a_level   0.002732
secondary_cleaner.state.floatbank4_a_air     0.002732
secondary_cleaner.state.floatbank3_b_level   0.002732
secondary_cleaner.state.floatbank3_b_air     0.002732
secondary_cleaner.state.floatbank3_a_level   0.002732
secondary_cleaner.state.floatbank2_b_level   0.002732
secondary_cleaner.state.floatbank2_a_level   0.002732
rougher.state.floatbank10_f_level            0.002732
rougher.state.floatbank10_e_level            0.002732
rougher.state.floatbank10_d_level            0.002732
rougher.state.floatbank10_c_level            0.002732
primary_cleaner.input.feed_size              0.000000
date                                         0.000000
dtype: float64
```

[ ]:
```
gold_full.isna().sum().sort_values(ascending=False)/len(gold_train)
```

[ ]:
```
rougher.output.recovery                    0.184994
rougher.output.tail_ag                     0.162337
rougher.output.tail_sol                    0.162278
rougher.output.tail_au                     0.162278
rougher.input.floatbank11_xanthate         0.133867
```

```
                                                     …
primary_cleaner.state.floatbank8_b_level    0.002550
primary_cleaner.state.floatbank8_c_level    0.002550
primary_cleaner.state.floatbank8_d_level    0.002550
primary_cleaner.input.feed_size             0.000000
date                                        0.000000
Length: 87, dtype: float64
```

[ ]:
```python
gold_test.fillna(method='ffill', inplace=True)
gold_test
```

[ ]:
```
                     date  primary_cleaner.input.sulfate  \
0     2016-09-01 00:59:59                     210.800909
1     2016-09-01 01:59:59                     215.392455
2     2016-09-01 02:59:59                     215.259946
3     2016-09-01 03:59:59                     215.336236
4     2016-09-01 04:59:59                     199.099327
…                     …                            …
5851 2017-12-31 19:59:59                     173.957757
5852 2017-12-31 20:59:59                     172.910270
5853 2017-12-31 21:59:59                     171.135718
5854 2017-12-31 22:59:59                     179.697158
5855 2017-12-31 23:59:59                     181.556856

      primary_cleaner.input.depressant  primary_cleaner.input.feed_size  \
0                            14.993118                         8.080000
1                            14.987471                         8.080000
2                            12.884934                         7.786667
3                            12.006805                         7.640000
4                            10.682530                         7.530000
…                                  …                              …
5851                         15.963399                         8.070000
5852                         16.002605                         8.070000
5853                         15.993669                         8.070000
5854                         15.438979                         8.070000
5855                         14.995850                         8.070000

      primary_cleaner.input.xanthate  primary_cleaner.state.floatbank8_a_air  \
0                           1.005021                             1398.981301
1                           0.990469                             1398.777912
2                           0.996043                             1398.493666
3                           0.863514                             1399.618111
4                           0.805575                             1401.268123
…                                …                                   …
5851                        0.896701                             1401.930554
5852                        0.896519                             1447.075722
5853                        1.165996                             1498.836182
```

```
5854                          1.501068                                1498.466243
5855                          1.623454                                1498.096303

      primary_cleaner.state.floatbank8_a_level  \
0                                  -500.225577
1                                  -500.057435
2                                  -500.868360
3                                  -498.863574
4                                  -500.808305
…                                          …
5851                               -499.728848
5852                               -494.716823
5853                               -501.770403
5854                               -500.483984
5855                               -499.796922

      primary_cleaner.state.floatbank8_b_air  \
0                                1399.144926
1                                1398.055362
2                                1398.860436
3                                1397.440120
4                                1398.128818
…                                        …
5851                             1401.441445
5852                             1448.851892
5853                             1499.572353
5854                             1497.986986
5855                             1501.743791

      primary_cleaner.state.floatbank8_b_level  \
0                                  -499.919735
1                                  -499.778182
2                                  -499.764529
3                                  -499.211024
4                                  -499.504543
…                                          …
5851                               -499.193423
5852                               -465.963026
5853                               -495.516347
5854                               -519.200340
5855                               -505.146931

      primary_cleaner.state.floatbank8_c_air  …  \
0                                1400.102998  …
1                                1396.151033  …
2                                1398.075709  …
3                                1400.129303  …
```

```
4                                  1402.172226  …
…                                          …   …
5851                               1399.810313  …
5852                               1443.890424  …
5853                               1502.749213  …
5854                               1496.569047  …
5855                               1499.535978  …

      secondary_cleaner.state.floatbank4_a_air  \
0                                    12.023554
1                                    12.058140
2                                    11.962366
3                                    12.033091
4                                    12.025367
…                                          …
5851                                 13.995957
5852                                 16.749781
5853                                 19.994130
5854                                 19.958760
5855                                 20.034715

      secondary_cleaner.state.floatbank4_a_level  \
0                                    -497.795834
1                                    -498.695773
2                                    -498.767484
3                                    -498.350935
4                                    -500.786497
…                                            …
5851                                 -500.157454
5852                                 -496.031539
5853                                 -499.791312
5854                                 -499.958750
5855                                 -500.728588

      secondary_cleaner.state.floatbank4_b_air  \
0                                     8.016656
1                                     8.130979
2                                     8.096893
3                                     8.074946
4                                     8.054678
…                                          …
5851                                 12.069155
5852                                 13.365371
5853                                 15.101425
5854                                 15.026853
5855                                 14.914199
```

```
      secondary_cleaner.state.floatbank4_b_level  \
0                                    -501.289139
1                                    -499.634209
2                                    -500.827423
3                                    -499.474407
4                                    -500.397500
…                                            …
5851                                 -499.673279
5852                                 -499.122723
5853                                 -499.936252
5854                                 -499.723143
5855                                 -499.948518

      secondary_cleaner.state.floatbank5_a_air  \
0                                     7.946562
1                                     7.958270
2                                     8.071056
3                                     7.897085
4                                     8.107890
…                                          …
5851                                  7.977259
5852                                  9.288553
5853                                 10.989181
5854                                 11.011607
5855                                 10.986607

      secondary_cleaner.state.floatbank5_a_level  \
0                                    -432.317850
1                                    -525.839648
2                                    -500.801673
3                                    -500.868509
4                                    -509.526725
…                                            …
5851                                 -499.516126
5852                                 -496.892967
5853                                 -498.347898
5854                                 -499.985046
5855                                 -500.658027

      secondary_cleaner.state.floatbank5_b_air  \
0                                     4.872511
1                                     4.878850
2                                     4.905125
3                                     4.931400
4                                     4.957674
…                                          …
5851                                  5.933319
```

```
5852                                  7.372897
5853                                  9.020944
5854                                  9.009783
5855                                  8.989497

      secondary_cleaner.state.floatbank5_b_level  \
0                                  -500.037437
1                                  -500.162375
2                                  -499.828510
3                                  -499.963623
4                                  -500.360026
…                                          …
5851                               -499.965973
5852                               -499.942956
5853                               -500.040448
5854                               -499.937902
5855                               -500.337588

      secondary_cleaner.state.floatbank6_a_air  \
0                                  26.705889
1                                  25.019940
2                                  24.994862
3                                  24.948919
4                                  25.003331
…                                        …
5851                                8.987171
5852                                8.986832
5853                                8.982038
5854                                9.012660
5855                                8.988632

      secondary_cleaner.state.floatbank6_a_level
0                                  -499.709414
1                                  -499.819438
2                                  -500.622559
3                                  -498.709987
4                                  -500.856333
…                                          …
5851                               -499.755909
5852                               -499.903761
5853                               -497.789882
5854                               -500.154284
5855                               -500.764937

[5856 rows x 53 columns]
```

When we look at NA values in the 3 sets, we see that some rows have almost 20% of values missing. We are told that parameters that are close in time tend to be similar, so I have used forward fill to fill these. In order to manually compute the recovery statistic below, I will drop NA values to avoid dividing by 0.

[ ]:
```
gold_train.dropna(inplace=True)
c = gold_train['rougher.output.concentrate_au']
f = gold_train['rougher.input.feed_au']
t = gold_train['rougher.output.tail_au']
recovery = ((c*(f-t))/(f*(c-t))) *100
recovery
```

[ ]:
```
0        87.107763
1        86.843261
2        86.842308
3        87.226430
4        86.688794
           …
16855    89.574376
16856    87.724007
16857    88.890579
16858    89.858126
16859    89.514960
Length: 11017, dtype: float64
```

[ ]:
```
recovery.shape
```

[ ]:
```
(11017,)
```

[ ]:
```
gold_train['rougher.output.recovery'].shape
```

[ ]:
```
(11017,)
```

[ ]:
```
mean_absolute_error(gold_train['rougher.output.recovery'], recovery)
```

[ ]:
```
9.555596961987514e-15
```

[ ]:
```
sum(np.abs(recovery-gold_train['rougher.output.recovery']))/len(recovery)
```

[ ]:
```
9.555596961987514e-15
```

MAE is confirmed to be reasonably small (very close to zero) for the manually calculated recovery values. This tells us that this equation is a good model to predict the output of the process.

[ ]:
```
col_mapping_dict = {c[0]:c[1] for c in enumerate(gold_test.columns)}
col_mapping_dict
```

[ ]:
```
{0: 'date',
 1: 'primary_cleaner.input.sulfate',
 2: 'primary_cleaner.input.depressant',
 3: 'primary_cleaner.input.feed_size',
```

```
4: 'primary_cleaner.input.xanthate',
5: 'primary_cleaner.state.floatbank8_a_air',
6: 'primary_cleaner.state.floatbank8_a_level',
7: 'primary_cleaner.state.floatbank8_b_air',
8: 'primary_cleaner.state.floatbank8_b_level',
9: 'primary_cleaner.state.floatbank8_c_air',
10: 'primary_cleaner.state.floatbank8_c_level',
11: 'primary_cleaner.state.floatbank8_d_air',
12: 'primary_cleaner.state.floatbank8_d_level',
13: 'rougher.input.feed_ag',
14: 'rougher.input.feed_pb',
15: 'rougher.input.feed_rate',
16: 'rougher.input.feed_size',
17: 'rougher.input.feed_sol',
18: 'rougher.input.feed_au',
19: 'rougher.input.floatbank10_sulfate',
20: 'rougher.input.floatbank10_xanthate',
21: 'rougher.input.floatbank11_sulfate',
22: 'rougher.input.floatbank11_xanthate',
23: 'rougher.state.floatbank10_a_air',
24: 'rougher.state.floatbank10_a_level',
25: 'rougher.state.floatbank10_b_air',
26: 'rougher.state.floatbank10_b_level',
27: 'rougher.state.floatbank10_c_air',
28: 'rougher.state.floatbank10_c_level',
29: 'rougher.state.floatbank10_d_air',
30: 'rougher.state.floatbank10_d_level',
31: 'rougher.state.floatbank10_e_air',
32: 'rougher.state.floatbank10_e_level',
33: 'rougher.state.floatbank10_f_air',
34: 'rougher.state.floatbank10_f_level',
35: 'secondary_cleaner.state.floatbank2_a_air',
36: 'secondary_cleaner.state.floatbank2_a_level',
37: 'secondary_cleaner.state.floatbank2_b_air',
38: 'secondary_cleaner.state.floatbank2_b_level',
39: 'secondary_cleaner.state.floatbank3_a_air',
40: 'secondary_cleaner.state.floatbank3_a_level',
41: 'secondary_cleaner.state.floatbank3_b_air',
42: 'secondary_cleaner.state.floatbank3_b_level',
43: 'secondary_cleaner.state.floatbank4_a_air',
44: 'secondary_cleaner.state.floatbank4_a_level',
45: 'secondary_cleaner.state.floatbank4_b_air',
46: 'secondary_cleaner.state.floatbank4_b_level',
47: 'secondary_cleaner.state.floatbank5_a_air',
48: 'secondary_cleaner.state.floatbank5_a_level',
49: 'secondary_cleaner.state.floatbank5_b_air',
50: 'secondary_cleaner.state.floatbank5_b_level',
```

```
51: 'secondary_cleaner.state.floatbank6_a_air',
52: 'secondary_cleaner.state.floatbank6_a_level'}
```

[ ]:
```
col_mapping_dict = {c[0]:c[1] for c in enumerate(gold_full.columns)}
col_mapping_dict
```

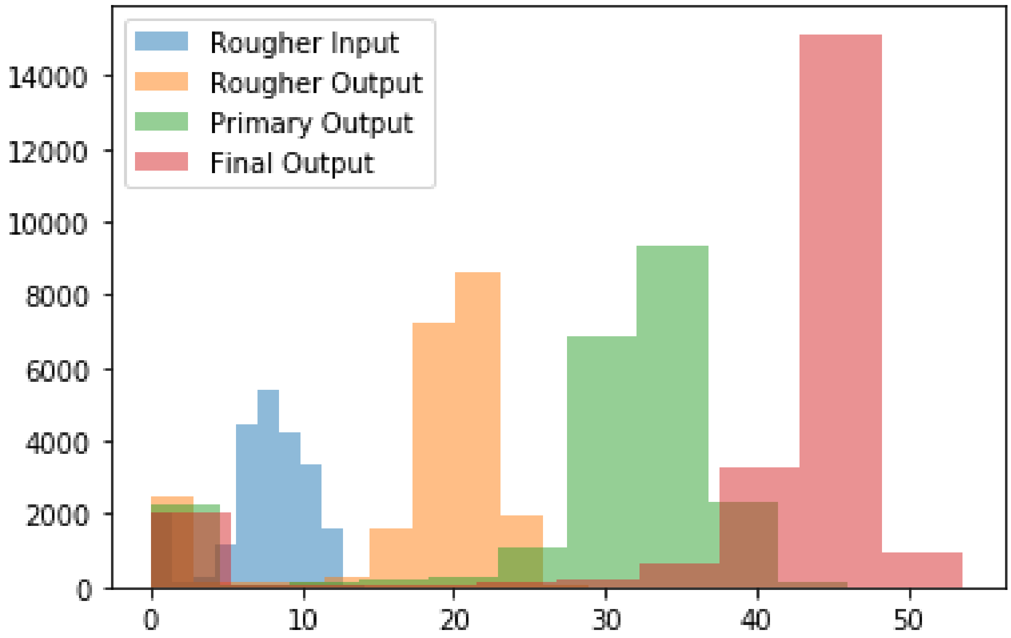
[ ]:
```
{0: 'date',
 1: 'final.output.concentrate_ag',
 2: 'final.output.concentrate_pb',
 3: 'final.output.concentrate_sol',
 4: 'final.output.concentrate_au',
 5: 'final.output.recovery',
 6: 'final.output.tail_ag',
 7: 'final.output.tail_pb',
 8: 'final.output.tail_sol',
 9: 'final.output.tail_au',
 10: 'primary_cleaner.input.sulfate',
 11: 'primary_cleaner.input.depressant',
 12: 'primary_cleaner.input.feed_size',
 13: 'primary_cleaner.input.xanthate',
 14: 'primary_cleaner.output.concentrate_ag',
 15: 'primary_cleaner.output.concentrate_pb',
 16: 'primary_cleaner.output.concentrate_sol',
 17: 'primary_cleaner.output.concentrate_au',
 18: 'primary_cleaner.output.tail_ag',
 19: 'primary_cleaner.output.tail_pb',
 20: 'primary_cleaner.output.tail_sol',
 21: 'primary_cleaner.output.tail_au',
 22: 'primary_cleaner.state.floatbank8_a_air',
 23: 'primary_cleaner.state.floatbank8_a_level',
 24: 'primary_cleaner.state.floatbank8_b_air',
 25: 'primary_cleaner.state.floatbank8_b_level',
 26: 'primary_cleaner.state.floatbank8_c_air',
 27: 'primary_cleaner.state.floatbank8_c_level',
 28: 'primary_cleaner.state.floatbank8_d_air',
 29: 'primary_cleaner.state.floatbank8_d_level',
 30: 'rougher.calculation.sulfate_to_au_concentrate',
 31: 'rougher.calculation.floatbank10_sulfate_to_au_feed',
 32: 'rougher.calculation.floatbank11_sulfate_to_au_feed',
 33: 'rougher.calculation.au_pb_ratio',
 34: 'rougher.input.feed_ag',
 35: 'rougher.input.feed_pb',
 36: 'rougher.input.feed_rate',
 37: 'rougher.input.feed_size',
 38: 'rougher.input.feed_sol',
 39: 'rougher.input.feed_au',
 40: 'rougher.input.floatbank10_sulfate',
```

```
41: 'rougher.input.floatbank10_xanthate',
42: 'rougher.input.floatbank11_sulfate',
43: 'rougher.input.floatbank11_xanthate',
44: 'rougher.output.concentrate_ag',
45: 'rougher.output.concentrate_pb',
46: 'rougher.output.concentrate_sol',
47: 'rougher.output.concentrate_au',
48: 'rougher.output.recovery',
49: 'rougher.output.tail_ag',
50: 'rougher.output.tail_pb',
51: 'rougher.output.tail_sol',
52: 'rougher.output.tail_au',
53: 'rougher.state.floatbank10_a_air',
54: 'rougher.state.floatbank10_a_level',
55: 'rougher.state.floatbank10_b_air',
56: 'rougher.state.floatbank10_b_level',
57: 'rougher.state.floatbank10_c_air',
58: 'rougher.state.floatbank10_c_level',
59: 'rougher.state.floatbank10_d_air',
60: 'rougher.state.floatbank10_d_level',
61: 'rougher.state.floatbank10_e_air',
62: 'rougher.state.floatbank10_e_level',
63: 'rougher.state.floatbank10_f_air',
64: 'rougher.state.floatbank10_f_level',
65: 'secondary_cleaner.output.tail_ag',
66: 'secondary_cleaner.output.tail_pb',
67: 'secondary_cleaner.output.tail_sol',
68: 'secondary_cleaner.output.tail_au',
69: 'secondary_cleaner.state.floatbank2_a_air',
70: 'secondary_cleaner.state.floatbank2_a_level',
71: 'secondary_cleaner.state.floatbank2_b_air',
72: 'secondary_cleaner.state.floatbank2_b_level',
73: 'secondary_cleaner.state.floatbank3_a_air',
74: 'secondary_cleaner.state.floatbank3_a_level',
75: 'secondary_cleaner.state.floatbank3_b_air',
76: 'secondary_cleaner.state.floatbank3_b_level',
77: 'secondary_cleaner.state.floatbank4_a_air',
78: 'secondary_cleaner.state.floatbank4_a_level',
79: 'secondary_cleaner.state.floatbank4_b_air',
80: 'secondary_cleaner.state.floatbank4_b_level',
81: 'secondary_cleaner.state.floatbank5_a_air',
82: 'secondary_cleaner.state.floatbank5_a_level',
83: 'secondary_cleaner.state.floatbank5_b_air',
84: 'secondary_cleaner.state.floatbank5_b_level',
85: 'secondary_cleaner.state.floatbank6_a_air',
86: 'secondary_cleaner.state.floatbank6_a_level'}
```

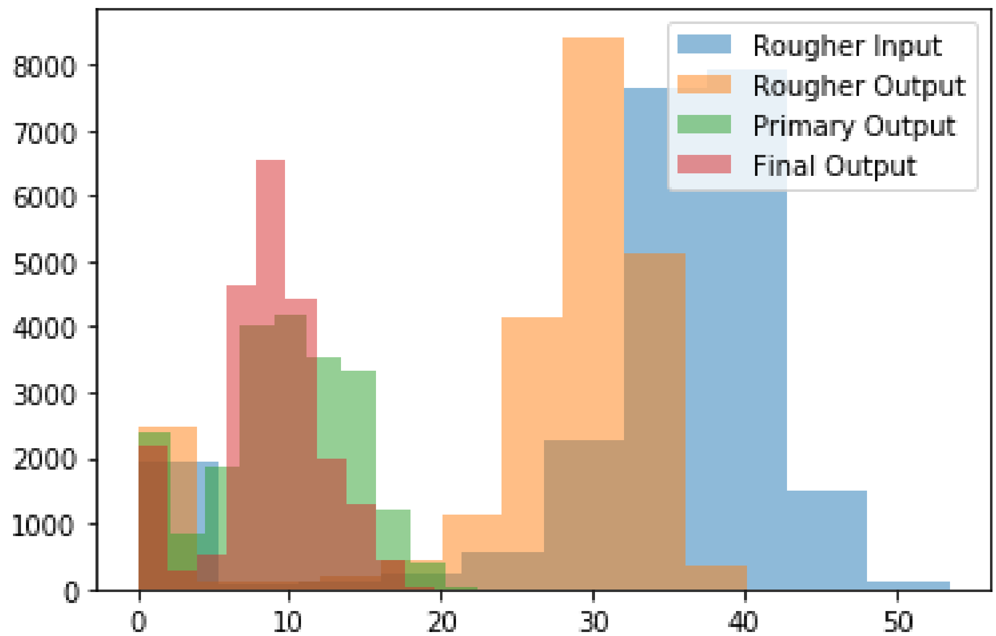
The test dataset doesn't include any of the final states, calculations, or the outputs of any of the stages. Which makes sense, because we would want the model to learn to predict outcomes from what goes in to each stage. The test dataset would only need the inputs, while we use the calculations, and actual outputs, to verify the accuracy of the model.

```
[ ]: au_conc = gold_full.filter(['rougher.input.feed_au', 'rougher.output.
      ↪concentrate_au', 'primary_cleaner.output.concentrate_au', 'final.output.
      ↪concentrate_au'])
     sol_conc = gold_full.filter(['rougher.input.feed_sol', 'rougher.output.
      ↪concentrate_sol', 'primary_cleaner.output.concentrate_sol', 'final.output.
      ↪concentrate_sol'])
     pb_conc = gold_full.filter(['rougher.input.feed_pb', 'rougher.output.
      ↪concentrate_pb', 'primary_cleaner.output.concentrate_pb', 'final.output.
      ↪concentrate_pb'])
     ag_conc = gold_full.filter(['rougher.input.feed_ag', 'rougher.output.
      ↪concentrate_ag', 'primary_cleaner.output.concentrate_ag', 'final.output.
      ↪concentrate_ag'])
```

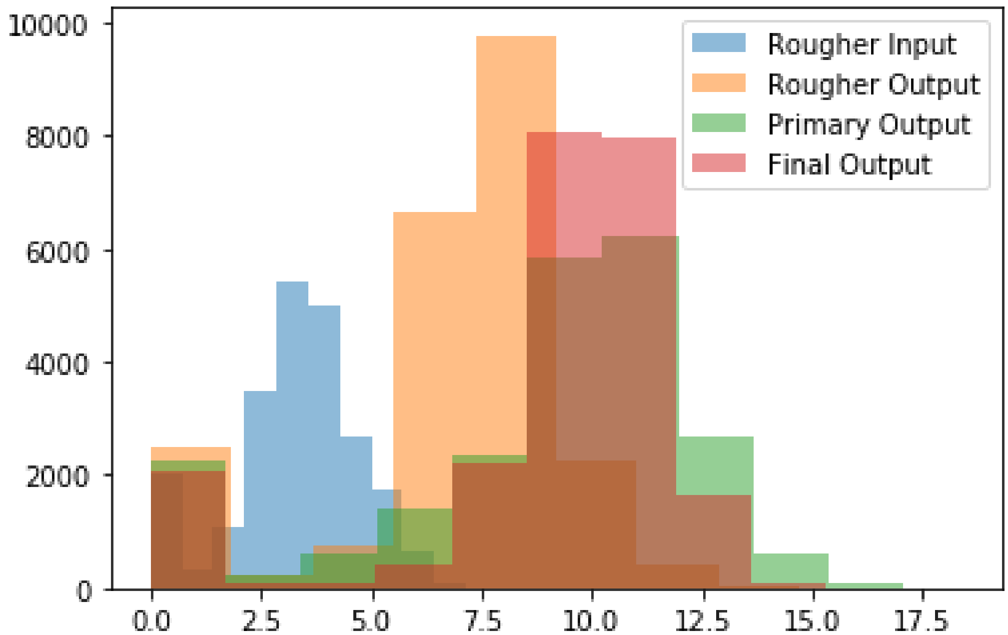
```
[ ]: plt.hist(au_conc['rougher.input.feed_au'], alpha=0.5, label='Rougher Input')
     plt.hist(au_conc['rougher.output.concentrate_au'], alpha=0.5, label='Rougher␣
      ↪Output')
     plt.hist(au_conc['primary_cleaner.output.concentrate_au'], alpha=0.5,␣
      ↪label='Primary Output')
     plt.hist(au_conc['final.output.concentrate_au'], alpha=0.5, label='Final␣
      ↪Output')
     plt.legend(loc='upper left')
     plt.show()
```



[ ]:
```
plt.hist(sol_conc['rougher.input.feed_sol'], alpha=0.5, label='Rougher Input')
plt.hist(sol_conc['rougher.output.concentrate_sol'], alpha=0.5, label='Rougher␣
    ↪Output')
plt.hist(sol_conc['primary_cleaner.output.concentrate_sol'], alpha=0.5,␣
    ↪label='Primary Output')
plt.hist(sol_conc['final.output.concentrate_sol'], alpha=0.5, label='Final␣
    ↪Output')
plt.legend(loc='upper right')
plt.show()
```

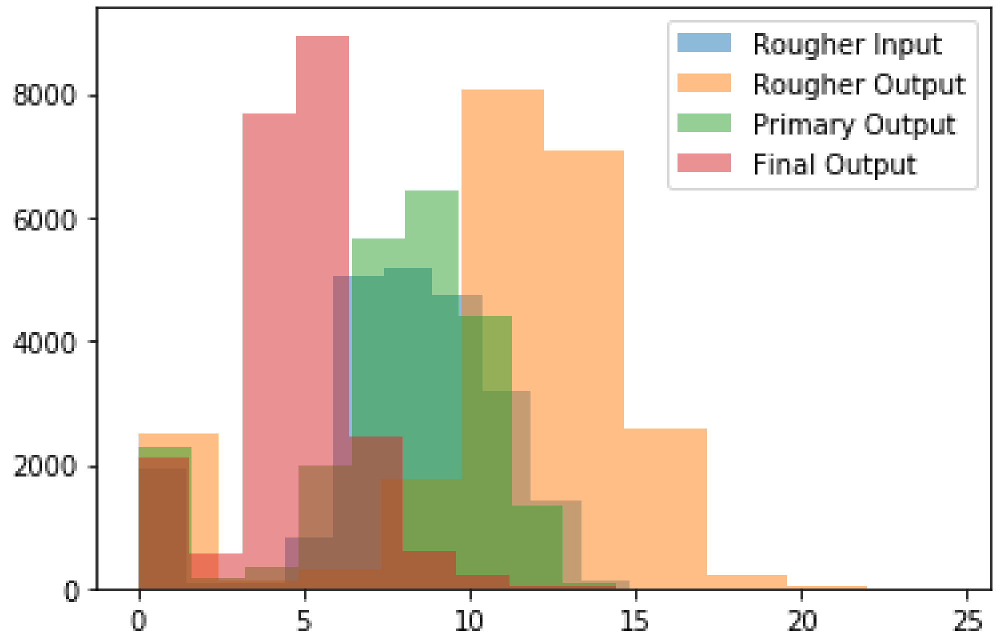


[ ]:
```
plt.hist(pb_conc['rougher.input.feed_pb'], alpha=0.5, label='Rougher Input')
plt.hist(pb_conc['rougher.output.concentrate_pb'], alpha=0.5, label='Rougher␣
    ↪Output')
plt.hist(pb_conc['primary_cleaner.output.concentrate_pb'], alpha=0.5,␣
    ↪label='Primary Output')
plt.hist(pb_conc['final.output.concentrate_pb'], alpha=0.5, label='Final␣
    ↪Output')
plt.legend(loc='upper right')
plt.show()
```



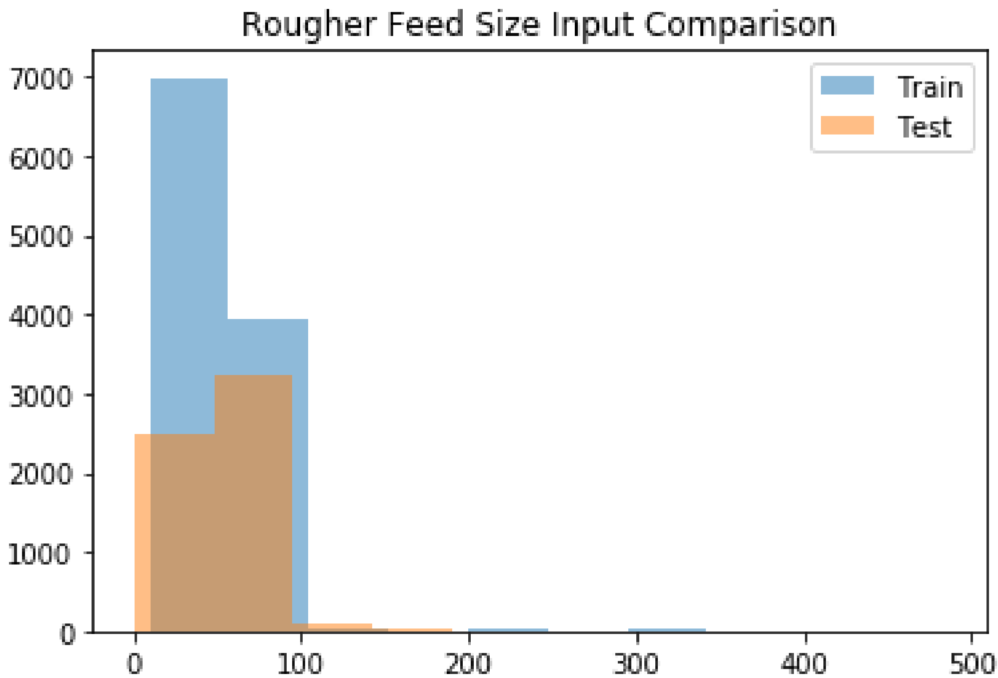
[ ]:

```
plt.hist(ag_conc['rougher.input.feed_ag'], alpha=0.5, label='Rougher Input')
plt.hist(ag_conc['rougher.output.concentrate_ag'], alpha=0.5, label='Rougher␣
↪Output')
plt.hist(ag_conc['primary_cleaner.output.concentrate_ag'], alpha=0.5,␣
↪label='Primary Output')
plt.hist(ag_conc['final.output.concentrate_ag'], alpha=0.5, label='Final␣
↪Output')
plt.legend(loc='upper right')
plt.show()
```

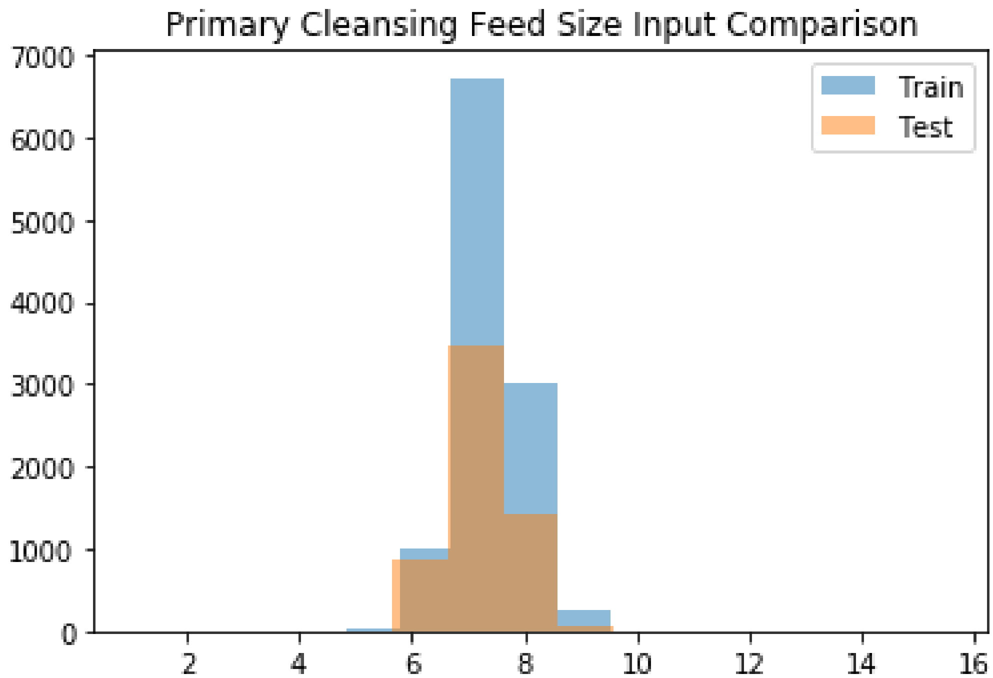


When we look at how concentrations change throughout the process, the most important to observe is the Au (gold) concentration. We can see that the concentration grows as the ore moves through the process, which aligns with the goal to purify and improve the concentration of gold with each step. The other minerals do not show as clear a progress, but they do not mimic the Au process. This tells us that the process is prioritizing the purification of Au, and that each different step assists with reducing a different contaminant.

[ ]:
```
plt.hist(gold_train['rougher.input.feed_size'], alpha=0.5, label='Train')
plt.hist(gold_test['rougher.input.feed_size'], alpha=0.5, label='Test')
plt.title('Rougher Feed Size Input Comparison')
plt.legend(loc='upper right')
plt.show()
```

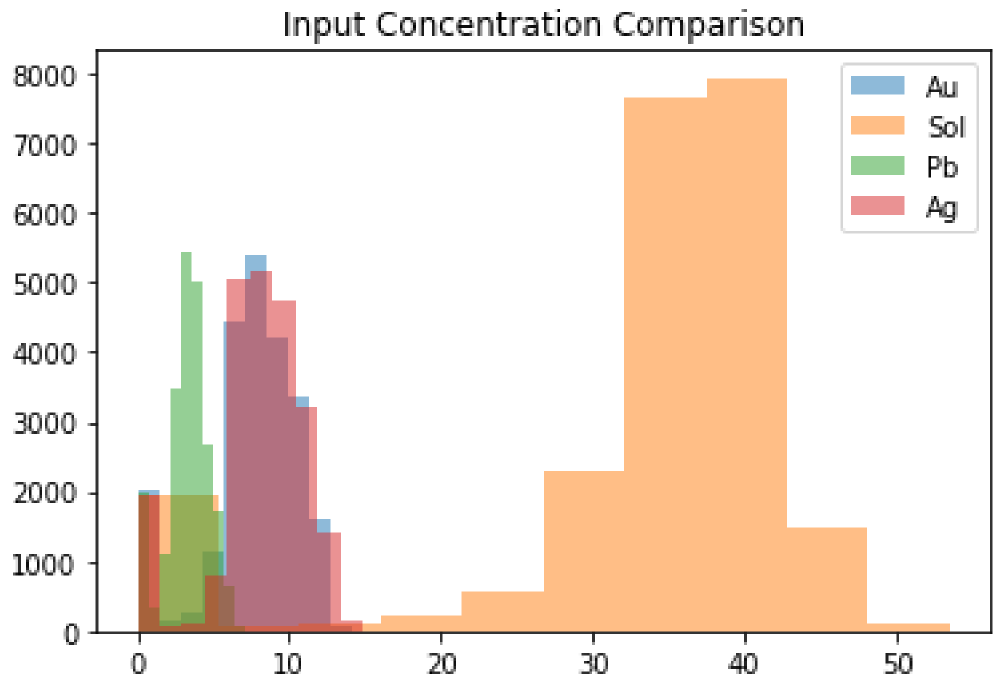


[ ]:
```
plt.hist(gold_train['primary_cleaner.input.feed_size'], alpha=0.5,␣
 ↪label='Train')
plt.hist(gold_test['primary_cleaner.input.feed_size'], alpha=0.5, label='Test')
plt.title('Primary Cleansing Feed Size Input Comparison')
plt.legend(loc='upper right')
plt.show()
```



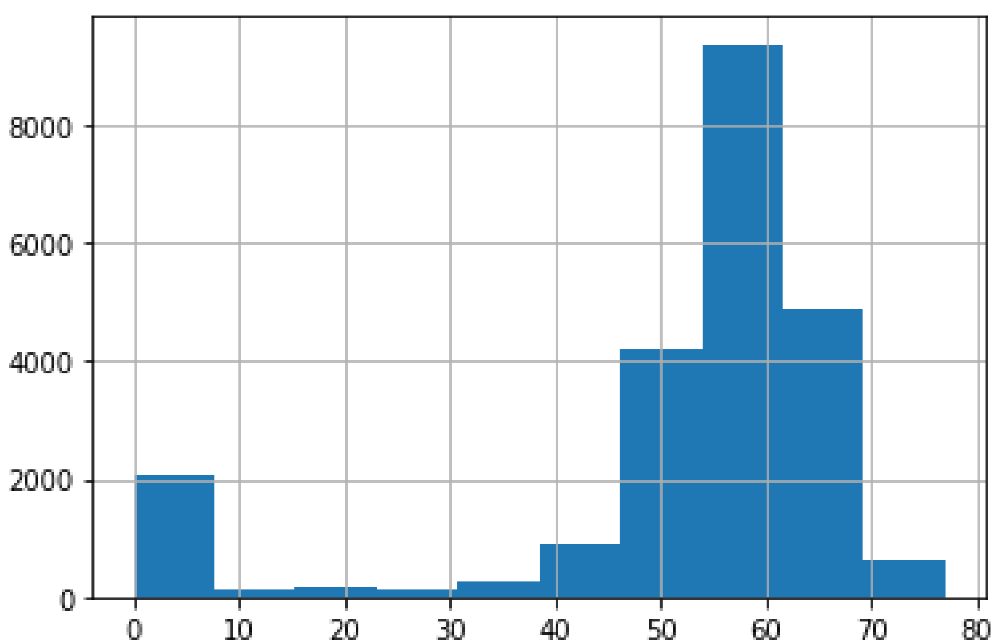
Feed size has approximately similar distributions between train and test sets, so models should be accurate.

[ ]:
```
plt.hist(gold_full['rougher.input.feed_au'], alpha=0.5, label='Au')
plt.hist(gold_full['rougher.input.feed_sol'], alpha=0.5, label='Sol')
plt.hist(gold_full['rougher.input.feed_pb'], alpha=0.5, label='Pb')
plt.hist(gold_full['rougher.input.feed_ag'], alpha=0.5, label='Ag')
plt.title('Input Concentration Comparison')
plt.legend(loc='upper right')
plt.show()
```



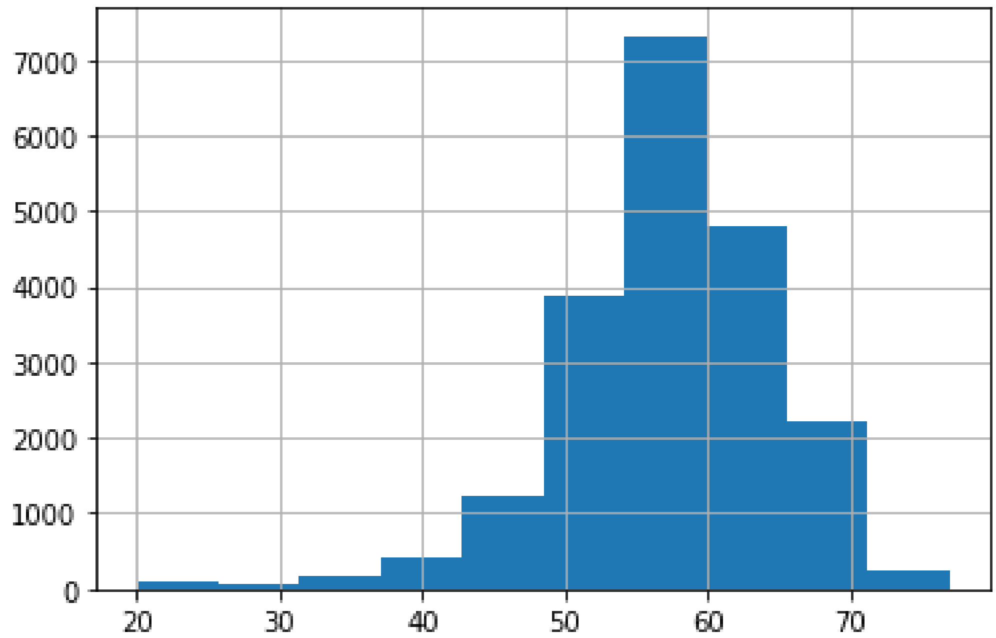
[ ]:
```
rougher_input = ['rougher.input.feed_au', 'rougher.input.feed_sol', 'rougher.
 ↪input.feed_pb', 'rougher.input.feed_ag']
gold_full[rougher_input].sum(1).hist()
```

[ ]: <matplotlib.axes._subplots.AxesSubplot at 0x7f443d4d6410>



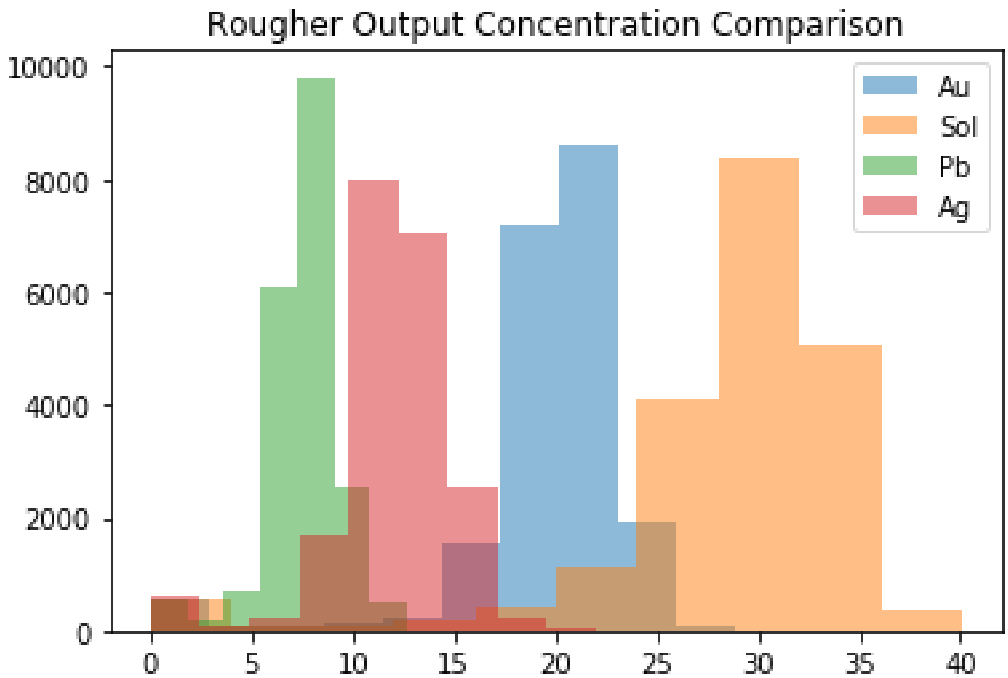
```
[ ]: gold_full['rougher_input'] = gold_full[rougher_input].sum(1)
     gold_full.drop(gold_full[gold_full['rougher_input'] <= 20].index, inplace=True)
     gold_full['rougher_input'].hist()
```

```
[ ]: <matplotlib.axes._subplots.AxesSubplot at 0x7f443d54f2d0>
```



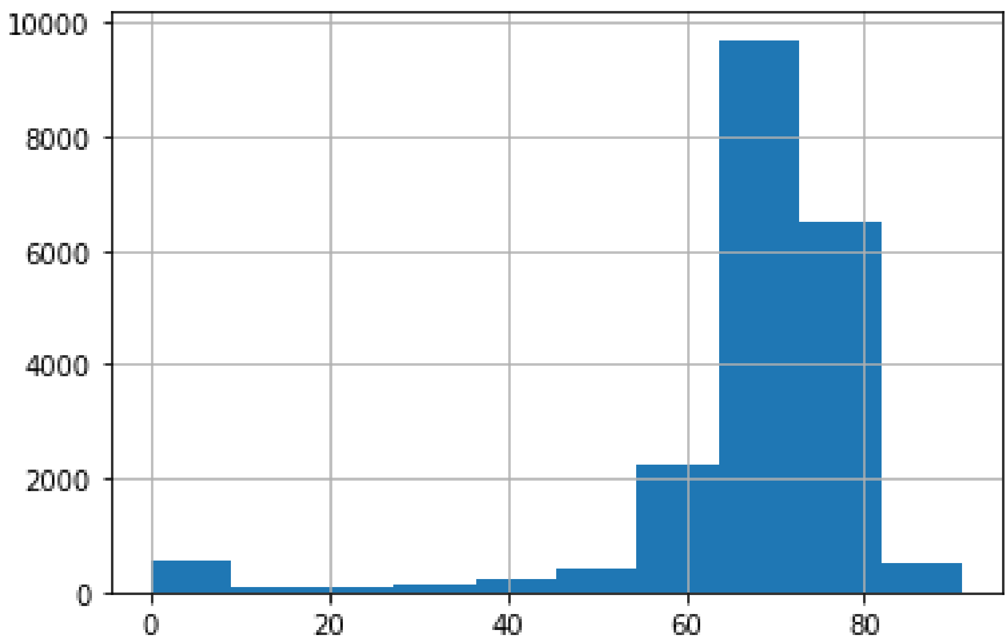

```
[ ]: gold_train['rougher_input'] = gold_train[rougher_input].sum(1)
     gold_train.drop(gold_train[gold_train['rougher_input'] <= 20].index,␣
      ↪inplace=True)
```

```
[ ]: plt.hist(gold_full['rougher.output.concentrate_au'], alpha=0.5, label='Au')
     plt.hist(gold_full['rougher.output.concentrate_sol'], alpha=0.5, label='Sol')
     plt.hist(gold_full['rougher.output.concentrate_pb'], alpha=0.5, label='Pb')
     plt.hist(gold_full['rougher.output.concentrate_ag'], alpha=0.5, label='Ag')
     plt.title('Rougher Output Concentration Comparison')
     plt.legend(loc='upper right')
     plt.show()
```



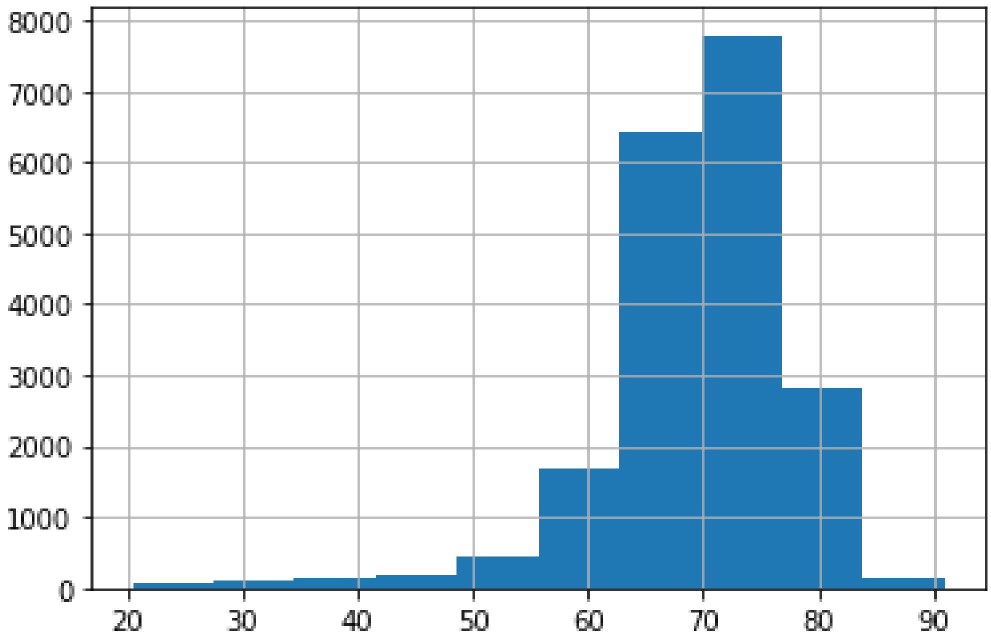
[ ]:
```
rougher_output = ['rougher.output.concentrate_au', 'rougher.output.
↪concentrate_sol', 'rougher.output.concentrate_pb', 'rougher.output.
↪concentrate_ag']
gold_full[rougher_output].sum(1).hist()
```

[ ]: <matplotlib.axes._subplots.AxesSubplot at 0x7f443db39850>

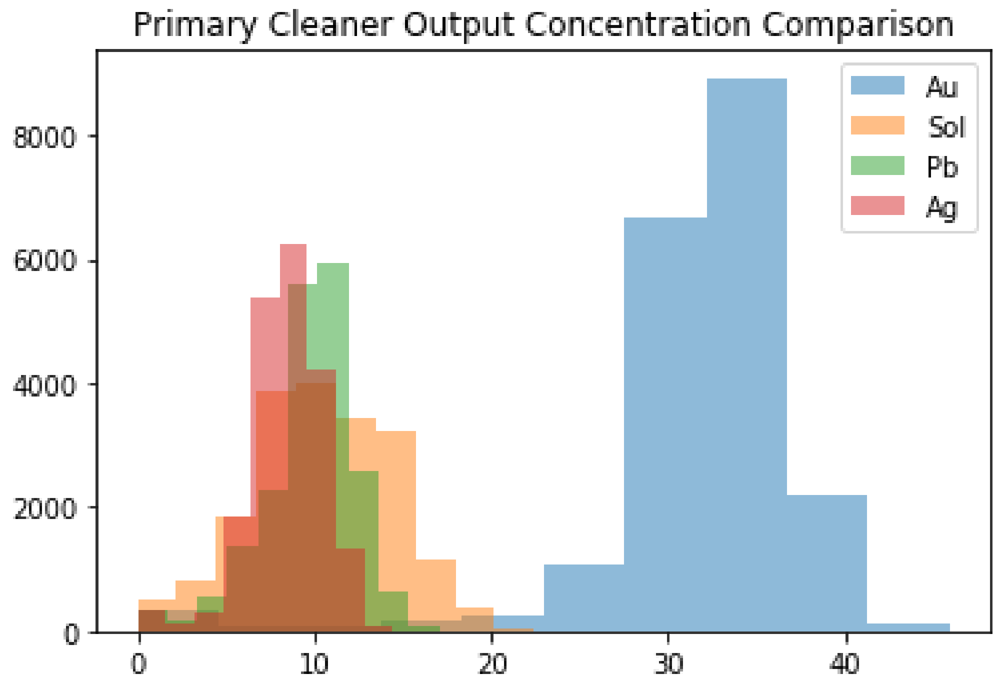



```
[ ]: gold_full['rougher_output'] = gold_full[rougher_output].sum(1)
     gold_full.drop(gold_full[gold_full['rougher_output'] <= 20].index, inplace=True)
     gold_train['rougher_output'] = gold_train[rougher_output].sum(1)
     gold_train.drop(gold_train[gold_train['rougher_output'] <= 20].index,␣
      ↪inplace=True)
     gold_full['rougher_output'].hist()
```

```
[ ]: <matplotlib.axes._subplots.AxesSubplot at 0x7f443e1da4d0>
```

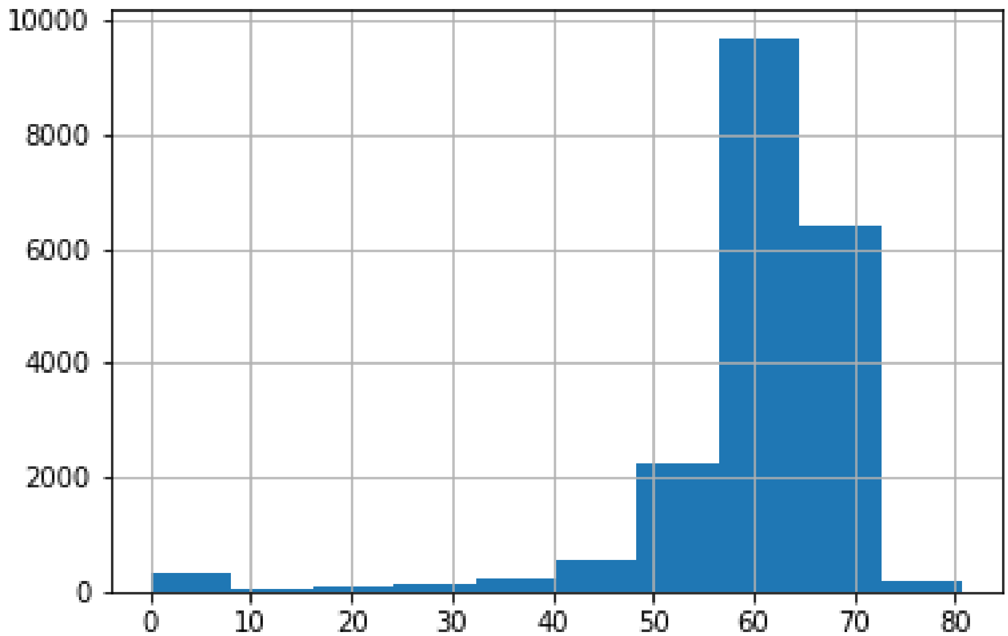


[ ]:
```
plt.hist(gold_full['primary_cleaner.output.concentrate_au'], alpha=0.5,␣
↪label='Au')
plt.hist(gold_full['primary_cleaner.output.concentrate_sol'], alpha=0.5,␣
↪label='Sol')
plt.hist(gold_full['primary_cleaner.output.concentrate_pb'], alpha=0.5,␣
↪label='Pb')
plt.hist(gold_full['primary_cleaner.output.concentrate_ag'], alpha=0.5,␣
↪label='Ag')
plt.title('Primary Cleaner Output Concentration Comparison')
plt.legend(loc='upper right')
plt.show()
```



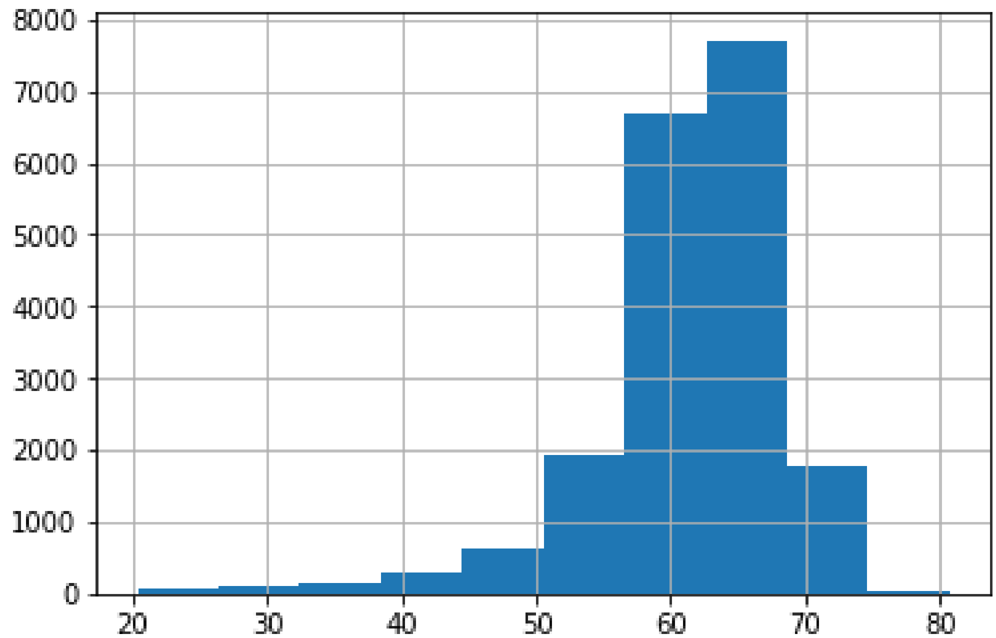

```
[ ]: primary_cleaner_output = ['primary_cleaner.output.concentrate_au',␣
      ↪'primary_cleaner.output.concentrate_sol', 'primary_cleaner.output.
      ↪concentrate_pb', 'primary_cleaner.output.concentrate_ag']
     gold_full[primary_cleaner_output].sum(1).hist()
```

```
[ ]: <matplotlib.axes._subplots.AxesSubplot at 0x7f443e004ed0>
```

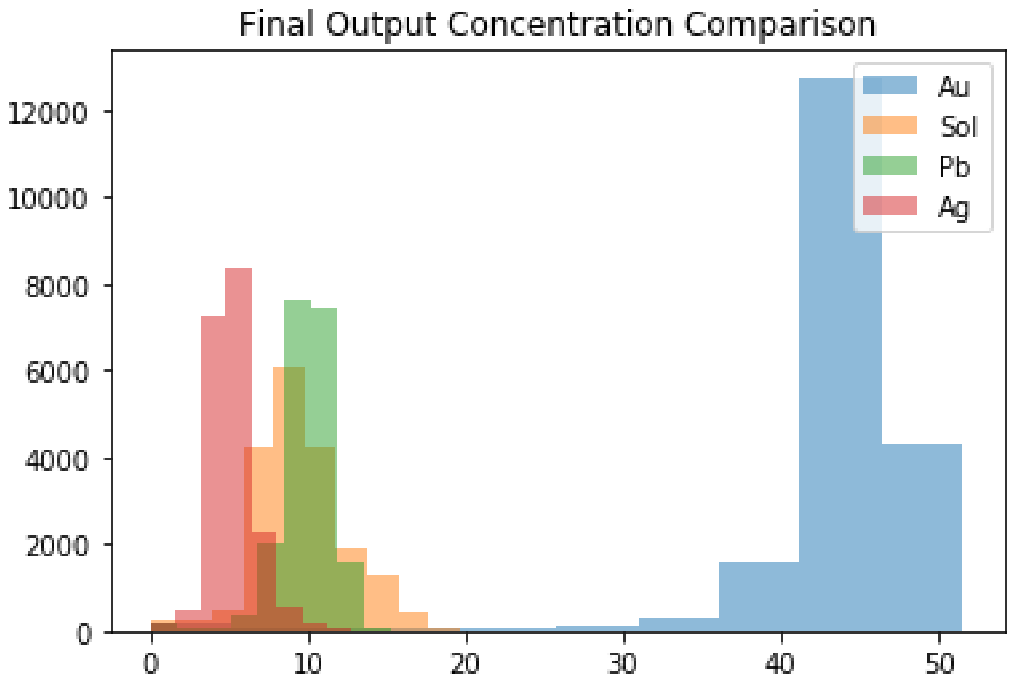


[ ]:
```
gold_full['primary_cleaner_output'] = gold_full[primary_cleaner_output].sum(1)
gold_full.drop(gold_full[gold_full['primary_cleaner_output'] <= 20].index, inplace=True)
gold_train['primary_cleaner_output'] = gold_train[primary_cleaner_output].sum(1)
gold_train.drop(gold_train[gold_train['primary_cleaner_output'] <= 20].index, inplace=True)
gold_full['primary_cleaner_output'].hist()
```

[ ]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f443d49c5d0>
```

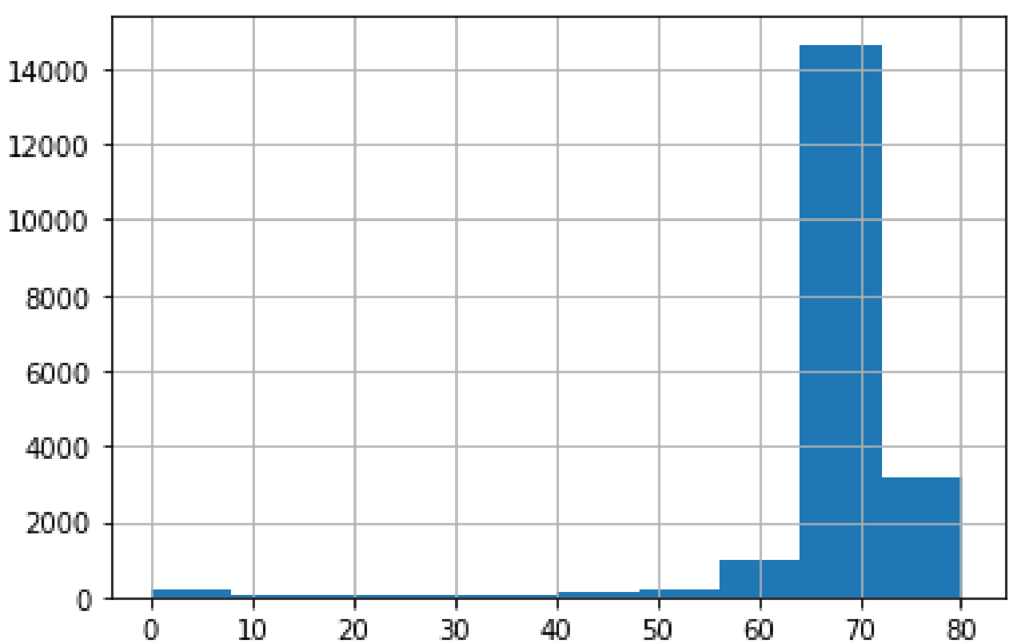


[ ]:
```
plt.hist(gold_full['final.output.concentrate_au'], alpha=0.5, label='Au')
plt.hist(gold_full['final.output.concentrate_sol'], alpha=0.5, label='Sol')
plt.hist(gold_full['final.output.concentrate_pb'], alpha=0.5, label='Pb')
plt.hist(gold_full['final.output.concentrate_ag'], alpha=0.5, label='Ag')
plt.title('Final Output Concentration Comparison')
plt.legend(loc='upper right')
plt.show()
```



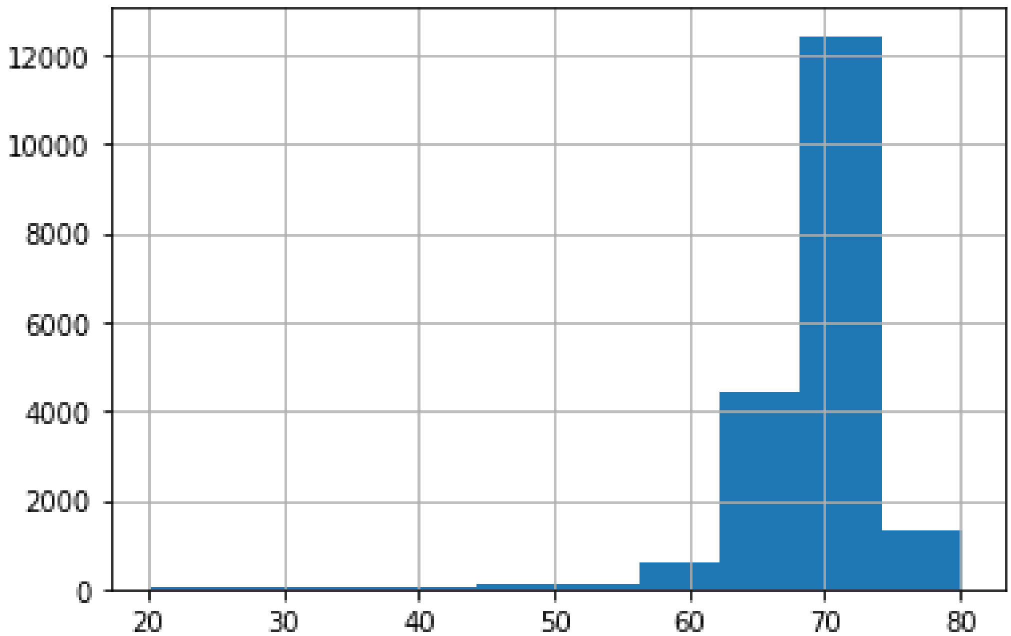
[ ]:
```
final_output = ['final.output.concentrate_au', 'final.output.concentrate_sol',␣
↪'final.output.concentrate_pb', 'final.output.concentrate_ag']
gold_full[final_output].sum(1).hist()
```

[ ]: <matplotlib.axes._subplots.AxesSubplot at 0x7f443db5ffd0>



[ ]:
```
gold_full['final_output'] = gold_full[final_output].sum(1)
gold_full.drop(gold_full[gold_full['final_output'] <= 20].index, inplace=True)
gold_full['final_output'].hist()
```

[ ]: <matplotlib.axes._subplots.AxesSubplot at 0x7f443e558110>



It does appear that there are some outliers of 0 in all of our concentrations, these may just be missing values or erroneous values. They have been dropped from the full dataset, and also from the train dataset to avoid mistraining the model.

[ ]:
```
def smape(y_true, y_pred):
    return (1/(len(y_true)))*np.sum((np.abs(y_true-y_pred))/(np.abs(y_true)+np.
 ↪abs(y_pred))/2)*100
```

[ ]:
```
def finalsmape(smape_rough, smape_final):
    return .25* smape_rough + .75* smape_final
```

[ ]:
```
smape_rough = smape(gold_train['rougher.output.recovery'], recovery)
```

[ ]:
```
c = gold_train['final.output.concentrate_au']
f = gold_train['rougher.output.concentrate_au']
t = gold_train['final.output.tail_au']
recovery = ((c*(f-t))/(f*(c-t))) *100
smape_final = smape(gold_train['final.output.recovery'], recovery)
```

[ ]:
```
finalsmape(smape_rough, smape_final)
```

[ ]: 5.856398012142628

[ ]:
```
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import train_test_split

features = gold_train[gold_test.columns].drop(['date'], axis=1)
target = gold_train[['rougher.output.recovery', 'final.output.recovery']]
features_train, features_test, target_train, target_test =␣
↪train_test_split(features, target, test_size=.25, random_state=42)
gold_test_target = gold_test.merge(gold_full, on='date', how='inner')[['date',␣
↪'rougher.output.recovery', 'final.output.recovery']]
gold_test = gold_test.merge(gold_test_target, on='date', how='inner')
gold_test_target = gold_test_target.drop(['date'], axis=1)
gold_test = gold_test.drop(['date', 'rougher.output.recovery', 'final.output.
↪recovery'], axis=1)
```

[ ]:
```
gold_test_target
```

[ ]:

	rougher.output.recovery	final.output.recovery
0	89.993421	70.273583
1	88.089657	68.910432
2	88.412756	68.143213
3	87.360133	67.776393
4	83.236367	61.467078
…	…	…
5154	95.172585	68.919891
5155	94.575036	68.440582
5156	93.018138	67.092759
5157	92.599042	68.061186
5158	91.177695	71.699976

[5159 rows x 2 columns]

[ ]:
```
gold_test
```

[ ]:

	primary_cleaner.input.sulfate	primary_cleaner.input.depressant	\
0	210.800909	14.993118	
1	215.392455	14.987471	
2	215.259946	12.884934	
3	215.336236	12.006805	
4	199.099327	10.682530	
…	…	…	

```
5154                    173.957757                               15.963399
5155                    172.910270                               16.002605
5156                    171.135718                               15.993669
5157                    179.697158                               15.438979
5158                    181.556856                               14.995850

      primary_cleaner.input.feed_size  primary_cleaner.input.xanthate  \
0                            8.080000                        1.005021
1                            8.080000                        0.990469
2                            7.786667                        0.996043
3                            7.640000                        0.863514
4                            7.530000                        0.805575
…                                 …                               …
5154                         8.070000                        0.896701
5155                         8.070000                        0.896519
5156                         8.070000                        1.165996
5157                         8.070000                        1.501068
5158                         8.070000                        1.623454

      primary_cleaner.state.floatbank8_a_air  \
0                                1398.981301
1                                1398.777912
2                                1398.493666
3                                1399.618111
4                                1401.268123
…                                        …
5154                             1401.930554
5155                             1447.075722
5156                             1498.836182
5157                             1498.466243
5158                             1498.096303

      primary_cleaner.state.floatbank8_a_level  \
0                                  -500.225577
1                                  -500.057435
2                                  -500.868360
3                                  -498.863574
4                                  -500.808305
…                                          …
5154                               -499.728848
5155                               -494.716823
5156                               -501.770403
5157                               -500.483984
5158                               -499.796922

      primary_cleaner.state.floatbank8_b_air  \
0                                1399.144926
```

```
1                                  1398.055362
2                                  1398.860436
3                                  1397.440120
4                                  1398.128818
…                                        …
5154                               1401.441445
5155                               1448.851892
5156                               1499.572353
5157                               1497.986986
5158                               1501.743791

      primary_cleaner.state.floatbank8_b_level  \
0                                  -499.919735
1                                  -499.778182
2                                  -499.764529
3                                  -499.211024
4                                  -499.504543
…                                        …
5154                               -499.193423
5155                               -465.963026
5156                               -495.516347
5157                               -519.200340
5158                               -505.146931

      primary_cleaner.state.floatbank8_c_air  \
0                                  1400.102998
1                                  1396.151033
2                                  1398.075709
3                                  1400.129303
4                                  1402.172226
…                                        …
5154                               1399.810313
5155                               1443.890424
5156                               1502.749213
5157                               1496.569047
5158                               1499.535978

      primary_cleaner.state.floatbank8_c_level  …  \
0                                  -500.704369  …
1                                  -499.240168  …
2                                  -502.151509  …
3                                  -498.355873  …
4                                  -500.810606  …
…                                        …  …
5154                               -499.599127  …
5155                               -503.587739  …
5156                               -520.667442  …
```

```
5157                              -487.479567  …
5158                              -492.428226  …

      secondary_cleaner.state.floatbank4_a_air  \
0                                    12.023554
1                                    12.058140
2                                    11.962366
3                                    12.033091
4                                    12.025367
…                                          …
5154                                 13.995957
5155                                 16.749781
5156                                 19.994130
5157                                 19.958760
5158                                 20.034715

      secondary_cleaner.state.floatbank4_a_level  \
0                                    -497.795834
1                                    -498.695773
2                                    -498.767484
3                                    -498.350935
4                                    -500.786497
…                                            …
5154                                 -500.157454
5155                                 -496.031539
5156                                 -499.791312
5157                                 -499.958750
5158                                 -500.728588

      secondary_cleaner.state.floatbank4_b_air  \
0                                     8.016656
1                                     8.130979
2                                     8.096893
3                                     8.074946
4                                     8.054678
…                                          …
5154                                 12.069155
5155                                 13.365371
5156                                 15.101425
5157                                 15.026853
5158                                 14.914199

      secondary_cleaner.state.floatbank4_b_level  \
0                                    -501.289139
1                                    -499.634209
2                                    -500.827423
3                                    -499.474407
```

```
4                                           -500.397500
…                                                     …
5154                                        -499.673279
5155                                        -499.122723
5156                                        -499.936252
5157                                        -499.723143
5158                                        -499.948518

      secondary_cleaner.state.floatbank5_a_air  \
0                                         7.946562
1                                         7.958270
2                                         8.071056
3                                         7.897085
4                                         8.107890
…                                                …
5154                                      7.977259
5155                                      9.288553
5156                                     10.989181
5157                                     11.011607
5158                                     10.986607

      secondary_cleaner.state.floatbank5_a_level  \
0                                         -432.317850
1                                         -525.839648
2                                         -500.801673
3                                         -500.868509
4                                         -509.526725
…                                                   …
5154                                      -499.516126
5155                                      -496.892967
5156                                      -498.347898
5157                                      -499.985046
5158                                      -500.658027

      secondary_cleaner.state.floatbank5_b_air  \
0                                         4.872511
1                                         4.878850
2                                         4.905125
3                                         4.931400
4                                         4.957674
…                                                …
5154                                      5.933319
5155                                      7.372897
5156                                      9.020944
5157                                      9.009783
5158                                      8.989497
```

```
      secondary_cleaner.state.floatbank5_b_level  \
0                                     -500.037437
1                                     -500.162375
2                                     -499.828510
3                                     -499.963623
4                                     -500.360026
…                                             …
5154                                  -499.965973
5155                                  -499.942956
5156                                  -500.040448
5157                                  -499.937902
5158                                  -500.337588

      secondary_cleaner.state.floatbank6_a_air  \
0                                     26.705889
1                                     25.019940
2                                     24.994862
3                                     24.948919
4                                     25.003331
…                                           …
5154                                   8.987171
5155                                   8.986832
5156                                   8.982038
5157                                   9.012660
5158                                   8.988632

      secondary_cleaner.state.floatbank6_a_level
0                                     -499.709414
1                                     -499.819438
2                                     -500.622559
3                                     -498.709987
4                                     -500.856333
…                                             …
5154                                  -499.755909
5155                                  -499.903761
5156                                  -497.789882
5157                                  -500.154284
5158                                  -500.764937

[5159 rows x 52 columns]
```

[ ]:
```
LinReg = LinearRegression()
LinReg.fit(features_train, target_train)
LinRegPredict=pd.DataFrame(LinReg.predict(features_test))
smape_rough = smape(target_test['rougher.output.recovery'], LinRegPredict[0])
smape_final = smape(target_test['final.output.recovery'], LinRegPredict[1])
finalsmape(smape_rough, smape_final)
```

[ ]: 0.45784446552154245

```
[ ]: for estim in range(5, 51, 5):
         for depth in range(5, 20, 5):
             RanFor = RandomForestRegressor(n_estimators=estim, max_depth=depth,␣
      ↪random_state=12345)
             RanFor.fit(features_train, target_train)
             RanForPredict = RanFor.predict(features_test)
             smape_rough = smape(target_test['rougher.output.recovery'],␣
      ↪RanForPredict[:,0])
             smape_final = smape(target_test['final.output.recovery'],␣
      ↪RanForPredict[:,1])
             total_smape = finalsmape(smape_rough, smape_final)
             print('N-estimator:', estim, ' | Depth:', depth, ' | sMAPE:',␣
      ↪total_smape)
```

```
N-estimator: 5  | Depth: 5  | sMAPE: 1.5571076204997591
N-estimator: 5  | Depth: 10  | sMAPE: 1.3897854182543772
N-estimator: 5  | Depth: 15  | sMAPE: 1.3426019169488914
N-estimator: 10  | Depth: 5  | sMAPE: 1.5629710859071269
N-estimator: 10  | Depth: 10  | sMAPE: 1.3481799513693564
N-estimator: 10  | Depth: 15  | sMAPE: 1.2718503085106792
N-estimator: 15  | Depth: 5  | sMAPE: 1.5604005124628415
N-estimator: 15  | Depth: 10  | sMAPE: 1.3357094442175526
N-estimator: 15  | Depth: 15  | sMAPE: 1.249945280817667
N-estimator: 20  | Depth: 5  | sMAPE: 1.5558131352357407
N-estimator: 20  | Depth: 10  | sMAPE: 1.3267885572538463
N-estimator: 20  | Depth: 15  | sMAPE: 1.2380285277364578
N-estimator: 25  | Depth: 5  | sMAPE: 1.557714292473178
N-estimator: 25  | Depth: 10  | sMAPE: 1.3207788988866975
N-estimator: 25  | Depth: 15  | sMAPE: 1.231689094585287
N-estimator: 30  | Depth: 5  | sMAPE: 1.5550407275282083
N-estimator: 30  | Depth: 10  | sMAPE: 1.319476847625792
N-estimator: 30  | Depth: 15  | sMAPE: 1.2261083665833772
N-estimator: 35  | Depth: 5  | sMAPE: 1.5552075683746347
N-estimator: 35  | Depth: 10  | sMAPE: 1.3158867301953852
N-estimator: 35  | Depth: 15  | sMAPE: 1.2195801835854312
N-estimator: 40  | Depth: 5  | sMAPE: 1.5545822464269303
N-estimator: 40  | Depth: 10  | sMAPE: 1.3123101134919433
N-estimator: 40  | Depth: 15  | sMAPE: 1.2141682166341456
N-estimator: 45  | Depth: 5  | sMAPE: 1.5522834570870234
N-estimator: 45  | Depth: 10  | sMAPE: 1.3081021581822214
N-estimator: 45  | Depth: 15  | sMAPE: 1.2085760211054908
N-estimator: 50  | Depth: 5  | sMAPE: 1.5537979650960074
N-estimator: 50  | Depth: 10  | sMAPE: 1.3062765161898446
N-estimator: 50  | Depth: 15  | sMAPE: 1.2050041919216232
```

```
[ ]: for depth in range(1, 41):
        DecTree = DecisionTreeRegressor(max_depth=depth, random_state=12345)
        DecTree.fit(features_train, target_train)
        DecTreePredict = DecTree.predict(features_test)
        smape_rough = smape(target_test['rougher.output.recovery'], DecTreePredict[:
     ↪,0])
        smape_final = smape(target_test['final.output.recovery'], DecTreePredict[:
     ↪,1])
        total_smape = finalsmape(smape_rough, smape_final)
        print('Depth:', depth, ' | sMAPE:', total_smape)
```

```
Depth: 1  | sMAPE: 1.8904854399935058
Depth: 2  | sMAPE: 1.8007765642156657
Depth: 3  | sMAPE: 1.7316619050914737
Depth: 4  | sMAPE: 1.655258829244849
Depth: 5  | sMAPE: 1.6214803590320255
Depth: 6  | sMAPE: 1.6341874432713217
Depth: 7  | sMAPE: 1.5934930915221932
Depth: 8  | sMAPE: 1.6114083163084576
Depth: 9  | sMAPE: 1.5856983317329798
Depth: 10  | sMAPE: 1.5492788264586281
Depth: 11  | sMAPE: 1.5828456042477665
Depth: 12  | sMAPE: 1.5423959809095495
Depth: 13  | sMAPE: 1.5581503908115497
Depth: 14  | sMAPE: 1.5890104319485339
Depth: 15  | sMAPE: 1.5866599430750707
Depth: 16  | sMAPE: 1.6389366497653821
Depth: 17  | sMAPE: 1.650925478624972
Depth: 18  | sMAPE: 1.6623329328419585
Depth: 19  | sMAPE: 1.732813872177719
Depth: 20  | sMAPE: 1.699985476513528
Depth: 21  | sMAPE: 1.6744488210163637
Depth: 22  | sMAPE: 1.7137702353462976
Depth: 23  | sMAPE: 1.732146692691994
Depth: 24  | sMAPE: 1.7144717037420194
Depth: 25  | sMAPE: 1.6658849011075436
Depth: 26  | sMAPE: 1.72046869054606
Depth: 27  | sMAPE: 1.7129876439659473
Depth: 28  | sMAPE: 1.6876205312917438
Depth: 29  | sMAPE: 1.7046304481807646
Depth: 30  | sMAPE: 1.6822253034790204
Depth: 31  | sMAPE: 1.702510290888545
Depth: 32  | sMAPE: 1.7166626469323076
Depth: 33  | sMAPE: 1.6733371055703437
Depth: 34  | sMAPE: 1.692050236901453
Depth: 35  | sMAPE: 1.7371661831129872
Depth: 36  | sMAPE: 1.7264300080688697
Depth: 37  | sMAPE: 1.7264300080688697
```

```
Depth: 38  | sMAPE: 1.7264300080688697
Depth: 39  | sMAPE: 1.7264300080688697
Depth: 40  | sMAPE: 1.7264300080688697
```

[ ]:
```
mean_absolute_error(target_test, LinReg.predict(features_test))
```

[ ]:
```
3.924412492800389
```

[ ]:
```
predictions0 = pd.Series(target_test.iloc[:,0].mean(), index=target_test.index)
predictions1 = pd.Series(target_test.iloc[:,1].mean(), index=target_test.index)
predictions = pd.concat([predictions0, predictions1], axis=1)
predictions
```

[ ]:
```
               0         1
1888   84.257823  66.75486
2913   84.257823  66.75486
2862   84.257823  66.75486
9329   84.257823  66.75486
324    84.257823  66.75486
…            …         …
7592   84.257823  66.75486
3981   84.257823  66.75486
5625   84.257823  66.75486
12172  84.257823  66.75486
1454   84.257823  66.75486

[2663 rows x 2 columns]
```

[ ]:
```
mean_absolute_error(target_test, predictions)
```

[ ]:
```
5.790588358203398
```

[ ]:
```
smape_rough = smape(target_test['rougher.output.recovery'], predictions.iloc[:
  ↪,0])
smape_final = smape(target_test['final.output.recovery'], predictions.iloc[:,1])
finalsmape(smape_rough, smape_final)
```

[ ]:
```
2.1674310014318103
```

When we compare sMAPE values across various methods of determining recovery, we can see that LinearRegressor gives us the best predictions, with the lowest sMAPE score. With the provided equation to determine recovery as a baseline, our sMAPE score was 5.85, much higher than the score of 0.45 that LinearRegressor got. Estimating with the mean recoveries as a baseline, the MAE is 5.79, so we can see that LinearRegressor is more effective than the baseline prediction. sMAPE with the baseline predictions is also 2.1, which is still higher than LinearRegressor.

[ ]:
```
linreg_valid = LinReg.predict(gold_test)
```

[ ]:
```
smape_rough = smape(gold_test_target['rougher.output.recovery'], linreg_valid[:
 ↪,0])
smape_final = smape(gold_test_target['final.output.recovery'], linreg_valid[:
 ↪,1])
finalsmape(smape_rough, smape_final)
```

[ ]:
```
1.793065917135099
```

Final sMAPE value with the chosen model is 1.79, which is still a very good score, and lower than many of the trained models.