Data Analysis with Python-Task2

March 14, 2024

```
import pandas as pd
[6]: data = pd.read_csv('C:\\Users\\SANDEEP\\Downloads\\01.Data Cleaning and_
       ⇔Preprocessing.csv')
     data
[7]:
[7]:
         Observation
                        Y-Kappa
                                  ChipRate
                                             BF-CMratio
                                                          BlowFlow
                                                                     ChipLevel4
                                                121.717
     0
             31-00:00
                          23.10
                                    16.520
                                                          1177.607
                                                                         169.805
                          27.60
     1
             31-01:00
                                    16.810
                                                 79.022
                                                          1328.360
                                                                         341.327
     2
                          23.19
                                    16.709
                                                 79.562
                                                          1329.407
                                                                         239.161
             31-02:00
     3
             31-03:00
                          23.60
                                    16.478
                                                 81.011
                                                          1334.877
                                                                         213.527
     4
             31-04:00
                          22.90
                                                 93.244
                                                          1334.168
                                                                         243.131
                                    15.618
     319
             10-16:00
                          23.75
                                    12.667
                                                 93.450
                                                          1178.252
                                                                         276.955
     320
              9-19:00
                          19.80
                                    12.558
                                                 94.352
                                                          1184.119
                                                                         297.071
     321
              9-20:00
                          23.01
                                                 90.842
                                                          1188.517
                                    12.550
                                                                         289.826
     322
              9-21:00
                          24.32
                                    13.083
                                                 88.910
                                                          1192.879
                                                                         318.006
     323
              9-22:00
                          25.75
                                    13.417
                                                 85.451
                                                          1186.342
                                                                         248.312
           T-upperExt-2
                           T-lowerExt-2
                                             UCZAA
                                                    WhiteFlow-4
                                                                       SteamFlow-4
     0
                 358.282
                                   329.545
                                             1.443
                                                          599.253
                                                                              67.122
     1
                 351.050
                                   329.067
                                             1.549
                                                          537.201
                                                                              60.012
     2
                 350.022
                                             1.600
                                                          549.611
                                   329.260
                                                                              61.304
     3
                 350.938
                                   331.142
                                             1.604
                                                          623.362
                                                                              68.496
     4
                 351.640
                                   332.709
                                               NaN
                                                          638.672
                                                                              70.022
     319
                 347.286
                                   310.970
                                             1.523
                                                          513.956
                                                                              61.141
     320
                                                                              67.667
                 399.135
                                   319.576
                                             1.451
                                                          570.058
     321
                 373.633
                                   314.591
                                             1.457
                                                          549.306
                                                                              66.446
     322
                 364.081
                                   308.559
                                             1.523
                                                          504.852
                                                                              61.054
     323
                 356.289
                                   310.482
                                             1.474
                                                          497.375
                                                                              58.247
                                             ChipMass-4
          Lower-HeatT-3
                           Upper-HeatT-3
                                                           WeakLiquorF
                                                                          BlackFlow-2
     0
                 329.432
                                   303.099
                                                 175.964
                                                                1127.197
                                                                               1319.039
     1
                 330.823
                                   304.879
                                                 163.202
                                                                 665.975
                                                                               1297.317
     2
                 329.140
                                   303.383
                                                 164.013
                                                                 677.534
                                                                               1327.072
                 328.875
                                   302.254
                                                 181.487
                                                                767.853
                                                                               1324.461
```

4	328.3	52 300.9	954 183.929	888.448	1343.424
	•••	•••	•••	•••	•••
319	330.1	17 304.0	148.174	1027.201	1357.271
320	330.8	48 304.6	165.178	906.962	1311.177
321	330.2	26 304.6	386 160.841	887.125	1319.226
322	327.3	46 304.3	363 147.589	804.423	1320.225
323	328.0	92 304.0	93 144.218	828.328	1320.848
	WeakWashF	SteamHeatF-3	T-Top-Chips-4	SulphidityL-4	
0	257.325	54.612	252.077	NaN	
1	241.182	46.603	251.406	29.11	
2	237.272	51.795	251.335	NaN	
3	239.478	54.846	250.312	29.02	
4	215.372	54.186	249.916	29.01	
	***	•••	•••	•••	
319	381.643	45.264	252.947	30.86	
320	25.494	50.528	252.092	30.70	
321	0.638	45.549	252.438	NaN	
322	0.000	43.725	253.176	31.13	
323	1.276	43.840	253.216	NaN	

[324 rows x 23 columns]

[8]: # Now we will deal with the missing #info() gives us both datatypes and the sum_of null values
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 324 entries, 0 to 323
Data columns (total 23 columns):

#	Column	Non-Null Count	Dtype
0	Observation	324 non-null	object
1	Y-Kappa	324 non-null	float64
2	ChipRate	319 non-null	float64
3	BF-CMratio	307 non-null	float64
4	BlowFlow	308 non-null	float64
5	ChipLevel4	323 non-null	float64
6	T-upperExt-2	322 non-null	float64
7	T-lowerExt-2	322 non-null	float64
8	UCZAA	299 non-null	float64
9	WhiteFlow-4	323 non-null	float64
10	AAWhiteSt-4	173 non-null	float64
11	AA-Wood-4	323 non-null	float64
12	ChipMoisture-4	323 non-null	float64
13	SteamFlow-4	323 non-null	float64
14	Lower-HeatT-3	322 non-null	float64
15	Upper-HeatT-3	322 non-null	float64

```
16 ChipMass-4
                            323 non-null
                                            float64
          WeakLiquorF
                            323 non-null
                                            float64
      17
         BlackFlow-2
      18
                            322 non-null
                                            float64
      19
         WeakWashF
                            323 non-null
                                            float64
      20 SteamHeatF-3
                            322 non-null
                                            float64
      21 T-Top-Chips-4
                            323 non-null
                                            float64
      22 SulphidityL-4
                            173 non-null
                                            float64
     dtypes: float64(22), object(1)
     memory usage: 58.3+ KB
 [9]: # It will give us the sum of null values in corresponding column of our data
      data.isnull().sum()
 [9]: Observation
                           0
      Y-Kappa
                           0
                           5
      ChipRate
      BF-CMratio
                          17
      BlowFlow
                          16
      ChipLevel4
                           1
      T-upperExt-2
                           2
      T-lowerExt-2
                           2
     UCZAA
                          25
      WhiteFlow-4
                           1
      AAWhiteSt-4
                         151
      AA-Wood-4
                           1
      ChipMoisture-4
                           1
      SteamFlow-4
                           1
      Lower-HeatT-3
                           2
      Upper-HeatT-3
                           2
      ChipMass-4
                           1
      WeakLiquorF
                           1
      BlackFlow-2
                           2
      WeakWashF
                           1
      SteamHeatF-3
                           2
      T-Top-Chips-4
                           1
      SulphidityL-4
                         151
      dtype: int64
[10]: # It will give us the total null values present in our data
      data.isnull().sum().sum()
[10]: 386
[11]: # Describe() is very useful if we are trying to get a simple statistical report
       ⇔for the data
```

data.describe()

```
[11]:
                             ChipRate
                                       BF-CMratio
                                                        BlowFlow
                                                                  ChipLevel4
                 Y-Kappa
      count
             324.000000
                          319.000000
                                       307.000000
                                                     308.000000
                                                                    323.000000
                            14.347937
              20.635370
                                                    1237.837614
                                         87.464456
                                                                    258.164483
      mean
                                                                     87.987452
      std
                3.070036
                             1.499095
                                          7.995012
                                                      100.593735
      min
               12.170000
                             9.983000
                                         68.645000
                                                        0.000000
                                                                      0.00000
      25%
                                         81.823000
               18.382500
                            13.358000
                                                     1193.215250
                                                                    213.527000
      50%
               20.845000
                            14.308000
                                         86.739000
                                                     1273.138500
                                                                    271.792000
      75%
              23.032500
                            15.517000
                                         92.372000
                                                     1289.196000
                                                                    321.680000
               27.600000
                            16.958000
                                       121.717000
                                                    1351.240000
                                                                    419.014000
      max
                                                    UCZAA
              T-upperExt-2
                              T-lowerExt-2
                                                            WhiteFlow-4
                                                                           AAWhiteSt-4
                                                                                          \
                 322.000000
                                  322.000000
                                               299.000000
                                                              323.000000
      count
                                                                             173.000000
                 356.904295
                                  324.020180
                                                 1.492010
                                                              591.732260
                                                                               6.140410
      mean
      std
                   9.209290
                                    7.621402
                                                 0.105923
                                                               67.016351
                                                                               0.081609
      min
                 339.168000
                                  284.633000
                                                 1.182000
                                                              405.111000
                                                                               5.890000
                 350.241250
      25%
                                  321.420000
                                                 1.431500
                                                              540.989500
                                                                               6.089000
      50%
                 356.843000
                                  325.669000
                                                 1.498000
                                                              592.895000
                                                                               6.135000
                 362.242250
      75%
                                  329.175000
                                                 1.560500
                                                              639.480500
                                                                               6.199000
                 399.135000
                                  337.012000
                                                 1.747000
                                                              731.394000
                                                                               6.340000
      max
                                                Upper-HeatT-3
                 SteamFlow-4
                                Lower-HeatT-3
                                                                 ChipMass-4
                   323.000000
                                   322.000000
                                                     322.000000
                                                                   323.000000
      count
      mean
                    66.668285
                                   325.567820
                                                     300.525699
                                                                   162.222322
      std
                     5.708587
                                     4.609862
                                                       4.568484
                                                                    14.160688
      min
                    48.568000
                                   318.051000
                                                     293.312000
                                                                   113.922000
      25%
                    62.518000
                                   321.385500
                                                     296.513250
                                                                   153.032500
      50%
                                   324.741000
                    67.429000
                                                     299.126000
                                                                   163.690000
      75%
                    71.522000
                                   329.845250
                                                     304.244750
                                                                   172.555000
              •••
                                   333.854000
                    76.147000
                                                     311.146000
                                                                   189.268000
      max
              WeakLiquorF
                             BlackFlow-2
                                            WeakWashF
                                                         SteamHeatF-3
                                                                         T-Top-Chips-4
                323.000000
                               322.000000
                                            323.000000
                                                            322.000000
                                                                             323.000000
      count
                873.828941
                              1175.917016
                                            263.543068
                                                             49.696907
                                                                             251.240087
      mean
                122.073521
                               149.334010
      std
                                            163.666942
                                                              4.551909
                                                                                1.283432
      min
                486.938000
                               838.948000
                                              0.000000
                                                                             248.359000
                                                             35.510000
      25%
                792.019500
                              1044.817500
                                            134.649000
                                                             46.389750
                                                                             250.312000
      50%
                865.254000
                              1150.221500
                                            269.193000
                                                             50.277000
                                                                             251.380000
      75%
                965.286500
                              1319.021250
                                                             53.294250
                                                                             252.323500
                                            405.563000
               1226.277000
                                                                             254.122000
      max
                              1395.767000
                                            715.715000
                                                             63.332000
              SulphidityL-4
                  173.000000
      count
      mean
                   30.411671
      std
                    0.701317
      min
                   29.010000
      25%
                   29.970000
      50%
                   30.370000
```

75% 30.820000 max 32.840000

[8 rows x 22 columns]

```
[12]: # Dropping the duplicate values from our data
data = data.drop_duplicates()
data
```

[12]:		Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow C	hipLevel4 \	
	0	31-00:00	23.10	16.520	121.717	1177.607	169.805	
	1	31-01:00	27.60	16.810	79.022	1328.360	341.327	
	2	31-02:00	23.19	16.709	79.562	1329.407	239.161	
	3	31-03:00	23.60	16.478	81.011	1334.877	213.527	
	4	31-04:00	22.90	15.618	93.244	1334.168	243.131	
		•••	•••	•••		•••		
	298	12-09:00	20.90	15.167	84.640	1283.706	339.440	
	299	12-10:00	24.98	NaN	85.034	1278.345	368.564	
	300	12-11:00	21.00	NaN	88.013	1307.722	278.842	
	301	12-12:00	21.40	NaN	85.490	1255.986	273.484	
	307	31-05:00	20.89	14.308	94.172	1327.832	251.120	
		T-upperExt-2	T-low	erExt-2	UCZAA Whit	eFlow-4	SteamFlow-4 \	
	0	358.28		329.545	1.443	599.253	67.122	
	1	351.05	50	329.067	1.549	537.201	60.012	
	2	350.02	22	329.260	1.600	549.611	61.304	
	3	350.93	88	331.142	1.604	623.362	68.496	
	4	351.64	-0	332.709	NaN	638.672	70.022	
		•••			· · · · · · · · · · · · · · · · · · ·	•••	•••	
	298	354.80	3	311.041	1.635	532.419	65.561	
	299	357.72	23	321.387	NaN	520.365	65.729	
	300	357.43	88	323.757	NaN	553.070	65.795	
	301	361.36	55	322.689	NaN	590.199	71.456	
	307	351.26	3	332.485	1.522	631.514	71.286	
		Lower-HeatT-	-3 Upper	-HeatT-3	ChipMass-4	WeakLiquor	F BlackFlow-2	\
	0	329.43	32	303.099	175.964	1127.19	97 1319.039	
	1	330.82	23	304.879	163.202	665.9	75 1297.317	
	2	329.14	-0	303.383	164.013	677.5	1327.072	
	3	328.87	' 5	302.254	181.487	767.8	1324.461	
	4	328.35	52	300.954	183.929	888.4	1343.424	
	• •							
	298	332.92		307.626	145.299			
	299	332.52		307.169	151.544			
	300	331.26		306.400	157.954			
	301	333.03		308.732	174.069			
	307	328.69	19	300.706	180.229	903.60	05 1323.082	

	WeakWashF	SteamHeatF-3	T-Top-Chips-4	SulphidityL-4
0	257.325	54.612	252.077	NaN
1	241.182	46.603	251.406	29.11
2	237.272	51.795	251.335	NaN
3	239.478	54.846	250.312	29.02
4	215.372	54.186	249.916	29.01
	•••	•••	•••	•••
298	388.911	49.524	251.833	30.29
299	418.979	48.135	251.614	30.47
300	462.712	54.373	251.197	NaN
301	457.313	53.194	251.324	30.46
307	232.729	54.503	250.084	NaN

[301 rows x 23 columns]

```
[30]: # Replacing the null values with 0
data2 = data.fillna(value = 0)
data2
```

	Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4 \	
0	31-00:00	23.10	_	121.717	1177.607	169.805	
1	31-01:00	27.60	16.810	79.022	1328.360	341.327	
2	31-02:00	23.19	16.709	79.562	1329.407	239.161	
3	31-03:00	23.60	16.478	81.011	1334.877	213.527	
4	31-04:00	22.90	15.618	93.244	1334.168	243.131	
	***	•••	•••		••		
298	12-09:00	20.90	15.167	84.640	1283.706	339.440	
299	12-10:00	24.98	0.000	85.034	1278.345	368.564	
300	12-11:00	21.00	0.000	88.013	1307.722	278.842	
301	12-12:00	21.40	0.000	85.490	1255.986	273.484	
307	31-05:00	20.89	14.308	94.172	1327.832	251.120	
			erExt-2		eFlow-4	SteamFlow-4	\
0			329.545				
1	351.09	50	329.067	1.549	537.201	60.012	
2	350.02	22	329.260	1.600	549.611	61.304	
3	350.93	38	331.142	1.604	623.362	68.496	
4	351.64	40	332.709	0.000	638.672	70.022	
• •	***		•••	•	• •••	•••	
298	354.80	03	311.041	1.635	532.419	65.561	
299	357.72	23	321.387	0.000	520.365	65.729	
300	357.43	38	323.757	0.000	553.070	65.795	
301	361.36	65	322.689	0.000	590.199	71.456	
307	351.26	20	332.485	1.522	631.514	71.286	
	0 1 2 3 4 298 299 300 301 307 0 1 2 3 4 298 299 300 301 307	0 31-00:00 1 31-01:00 2 31-02:00 3 31-03:00 4 31-04:00 298 12-09:00 299 12-10:00 300 12-11:00 301 12-12:00 307 31-05:00 T-upperExt-2 0 358.28 1 351.08 2 350.02 3 350.93 4 351.64 298 354.86 299 357.72 300 357.43 301 361.36	0 31-00:00 23.10 1 31-01:00 27.60 2 31-02:00 23.19 3 31-03:00 23.60 4 31-04:00 22.90 298 12-09:00 20.90 299 12-10:00 24.98 300 12-11:00 21.00 301 12-12:00 21.40 307 31-05:00 20.89 T-upperExt-2 T-low 0 358.282 1 351.050 2 350.022 3 350.938 4 351.640 298 354.803 299 357.723 300 357.438 301 361.365	0 31-00:00 23.10 16.520 1 31-01:00 27.60 16.810 2 31-02:00 23.19 16.709 3 31-03:00 23.60 16.478 4 31-04:00 22.90 15.618 298 12-09:00 20.90 15.167 299 12-10:00 24.98 0.000 300 12-11:00 21.00 0.000 301 12-12:00 21.40 0.000 307 31-05:00 20.89 14.308 T-upperExt-2 T-lowerExt-2 0 358.282 329.545 1 351.050 329.067 2 350.022 329.260 3 350.938 331.142 4 351.640 332.709 298 354.803 311.041 299 357.723 321.387 300 357.438 323.757 301 361.365 322.689	0 31-00:00 23.10 16.520 121.717 1 31-01:00 27.60 16.810 79.022 2 31-02:00 23.19 16.709 79.562 3 31-03:00 23.60 16.478 81.011 4 31-04:00 22.90 15.618 93.244 298 12-09:00 20.90 15.167 84.640 299 12-10:00 24.98 0.000 85.034 300 12-11:00 21.00 0.000 85.490 301 12-12:00 21.40 0.000 85.490 307 31-05:00 20.89 14.308 94.172 T-upperExt-2 T-lowerExt-2 UCZAA Whit 0 358.282 329.545 1.443 1 351.050 329.260 1.600 3 350.938 331.142 1.604 4 351.640 332.709 0.000	0 31-00:00 23.10 16.520 121.717 1177.607 1 31-01:00 27.60 16.810 79.022 1328.360 2 31-02:00 23.19 16.709 79.562 1329.407 3 31-03:00 23.60 16.478 81.011 1334.877 4 31-04:00 22.90 15.618 93.244 1334.168 298 12-09:00 20.90 15.167 84.640 1283.706 299 12-10:00 24.98 0.000 85.034 1278.345 300 12-11:00 21.00 0.000 85.490 1255.986 307 31-05:00 20.89 14.308 94.172 1327.832 T-upperExt-2 T-lowerExt-2 UCZAA WhiteFlow-4 0 358.282 329.545 1.443 599.253 1 351.050 329.067 1.549 537.201 2 350.022 329.260 1.600 549.611 3	0 31-00:00 23.10 16.520 121.717 1177.607 169.805 1 31-01:00 27.60 16.810 79.022 1328.360 341.327 2 31-02:00 23.19 16.709 79.562 1329.407 239.161 3 31-03:00 23.60 16.478 81.011 1334.877 213.527 4 31-04:00 22.90 15.618 93.244 1334.168 243.131 298 12-09:00 20.90 15.167 84.640 1283.706 339.440 299 12-10:00 24.98 0.000 85.034 1278.345 368.564 300 12-11:00 21.00 0.000 85.490 1255.986 273.484 301 12-12:00 21.40 0.000 85.490 1255.986 273.484 307 31-05:00 20.89 14.308 94.172 1327.832 251.120 T-upperExt-2 T-lowerExt-2 UCZAA WhiteFlow-4 S

7.317 7.072 4.461 3.424 4.708 4.469 4.588 8.747 3.082
4.461 3.424 4.708 4.469 4.588 8.747
4.461 3.424 4.708 4.469 4.588 8.747
3.424 4.708 4.469 4.588 8.747
4.708 4.469 4.588 8.747
4.469 4.588 8.747
4.469 4.588 8.747
4.588 8.747
8.747
3.082
\
\
\
\
\
\
\
\
\
\
\
\
\

303.099 175.964

1127.197

1319.039

0

329.432

```
5
                            325.195 1.436
           344.014
                                                   628.245 ...
                                                                      65.225
7
                                                   625.549
                                                                      71.298
           352.321
                            331.162 1.468
9
           352.372
                            327.358 1.515
                                                   553.172
                                                                      64.249
. .
               •••
                                                   •••
290
           343.515
                            302.364 1.592
                                                   452.718
                                                                      55.963
292
           345.261
                            310.115 1.513
                                                   428.202
                                                                      52.494
294
           345.811
                            307.806 1.633
                                                   469.045
                                                                      60.307
296
           347.672
                            313.147 1.546
                                                   496.460
                                                                      60.119
298
                                                   532.419 ...
           354.803
                            311.041 1.635
                                                                      65.561
                                                    WeakLiquorF
     Lower-HeatT-3 Upper-HeatT-3
                                      ChipMass-4
                                                                  BlackFlow-2
1
           330.823
                            304.879
                                          163.202
                                                         665.975
                                                                       1297.317
3
           328.875
                            302.254
                                          181.487
                                                         767.853
                                                                       1324.461
5
           322.103
                            298.517
                                          165.814
                                                         826.243
                                                                        907.641
7
           329.662
                            301.539
                                          179.886
                                                         837.178
                                                                       1315.111
9
           332.264
                            305.419
                                          166.120
                                                         909.810
                                                                       1318.725
. .
               •••
290
                            308.789
                                          128.701
                                                         783.417
                                                                       1293.108
           330.842
292
           330.589
                            309.152
                                          122.011
                                                         816.020
                                                                       1294.891
294
           329.997
                            308.072
                                          137.719
                                                         865.661
                                                                       1292.216
296
           332.615
                            308.575
                                          141.076
                                                         997.904
                                                                       1334.703
298
           332.924
                                          145.299
                                                         832.906
                                                                       1344.708
                            307.626
     WeakWashF
                                  T-Top-Chips-4
                                                   SulphidityL-4
                 SteamHeatF-3
1
        241.182
                         46.603
                                         251.406
                                                            29.11
3
        239.478
                         54.846
                                         250.312
                                                            29.02
5
                                                            30.34
        595.875
                         52.807
                                         249.580
7
        234.047
                         53.805
                                         249.971
                                                            29.22
9
        180.375
                         48.842
                                         251.121
                                                            29.21
290
        390.049
                         41.596
                                         252.428
                                                            30.73
292
                         42.080
                                                            30.24
        391.607
                                         253.206
294
                         44.330
                                                            30.31
        391.645
                                         253.323
296
                         46.206
        389.497
                                         252.423
                                                            30.43
298
        388.911
                         49.524
                                         251.833
                                                            30.29
[131 rows x 23 columns]
```

```
[32]: data2.isnull().sum().sum()
```

[32]: 0

```
[33]: # Filling null values with next value data4 = data.fillna(method = 'bfill') data4
```

[33]:		Observation Y	Y-Kappa	${\tt ChipRate}$	BF-CMratio	BlowFlow Chip	Level4 \
	0	31-00:00	23.10	16.520	121.717	1177.607	169.805
	1	31-01:00	27.60	16.810	79.022	1328.360	341.327
	2	31-02:00	23.19	16.709	79.562		239.161
	3	31-03:00	23.60	16.478	81.011		213.527
	4	31-04:00	22.90	15.618	93.244	1334.168	243.131
	• •	•••	•••	•••		•••	
	298	12-09:00	20.90	15.167	84.640	1283.706	339.440
	299	12-10:00	24.98	14.308	85.034	1278.345	368.564
	300	12-11:00	21.00	14.308	88.013	1307.722	278.842
	301	12-12:00	21.40	14.308	85.490		273.484
	307	31-05:00	20.89	14.308	94.172		251.120
	301	31 03.00	20.03	14.500	34.172	1027.002	201.120
		T-upperExt-2	Т-1 оч	erExt-2	UCZAA Whit	eFlow-4 St	eamFlow-4 \
	Λ	358.282		329.545			67.122
	0						
	1	351.050		329.067		537.201	60.012
	2	350.022	2	329.260	1.600	549.611	61.304
	3	350.938	3	331.142	1.604	623.362	68.496
	4	351.640)	332.709	1.436	638.672	70.022
		•••					
	298	354.803	3	311.041	1.635	532.419	65.561
	299	357.723		321.387		520.365	65.729
	300	357.438		323.757		553.070	65.795
	301	361.36		322.689			71.456
	307	351.263)	332.485	1.522	631.514	71.286
		Lower-HeatT-3	3 Unner	-H02+T-3	ChipMass-4	WeakLiquorF	BlackFlow-2 \
	^				_	-	
	0	329.432		303.099	175.964		1319.039
	1	330.823		304.879	163.202		1297.317
	2	329.140		303.383	164.013		1327.072
	3	328.87		302.254	181.487		1324.461
	4	328.352	2	300.954	183.929	888.448	1343.424
		•••		•••	•••	•••	•••
	298	332.924	1	307.626	145.299	832.906	1344.708
	299	332.523	3	307.169	151.544	905.639	1344.469
	300	331.263		306.400	157.954		1344.588
	301	333.032		308.732			1348.747
	307	328.699		300.706			1323.082
	301	520.033	,	300.700	100.229	903.003	1323.002
		WeakWashF S	SteamHea	+F-3 T-T	Con-Chins-4	SulphidityL-4	
	0	257.325		4.612	252.077	29.11	
	1	241.182		6.603	251.406	29.11	
	2	237.272		1.795	251.335	29.02	
	3	239.478	5	4.846	250.312	29.02	2
	4	215.372	5	4.186	249.916	29.01	-
		•••		•••	•••	•••	
	298	388.911	4	9.524	251.833	30.29)

```
301
             457.313
                              53.194
                                             251.324
                                                               30.46
      307
             232.729
                              54.503
                                             250.084
                                                                 NaN
      [301 rows x 23 columns]
[34]: import numpy as np
      import matplotlib.pyplot as plt
      from scipy import stats
[35]: # Detect the outliers using IQR
      data2.columns
[35]: Index(['Observation', 'Y-Kappa', 'ChipRate', 'BF-CMratio', 'BlowFlow',
             'ChipLevel4 ', 'T-upperExt-2 ', 'T-lowerExt-2 ', 'UCZAA',
             'WhiteFlow-4 ', 'AAWhiteSt-4 ', 'AA-Wood-4 ', 'ChipMoisture-4 ',
             'SteamFlow-4', 'Lower-HeatT-3', 'Upper-HeatT-3', 'ChipMass-4',
             'WeakLiquorF ', 'BlackFlow-2 ', 'WeakWashF ', 'SteamHeatF-3 ',
             'T-Top-Chips-4', 'SulphidityL-4'],
            dtype='object')
[36]: data2.drop(['Observation'], axis = 1, inplace = True)
[37]: data2.columns
[37]: Index(['Y-Kappa', 'ChipRate', 'BF-CMratio', 'BlowFlow', 'ChipLevel4',
             'T-upperExt-2 ', 'T-lowerExt-2 ', 'UCZAA', 'WhiteFlow-4 ',
             'AAWhiteSt-4', 'AA-Wood-4', 'ChipMoisture-4', 'SteamFlow-4',
             'Lower-HeatT-3', 'Upper-HeatT-3', 'ChipMass-4', 'WeakLiquorF',
             'BlackFlow-2 ', 'WeakWashF ', 'SteamHeatF-3 ', 'T-Top-Chips-4 ',
             'SulphidityL-4 '],
            dtype='object')
[38]: Q1 = data2.quantile(0.25)
      Q3 = data2.quantile(0.75)
      IQR = Q3 - Q1
      print(IQR)
     Y-Kappa
                          4.550
     ChipRate
                          2.233
     BF-CMratio
                         10.912
     BlowFlow
                         96.766
     ChipLevel4
                        105.868
     T-upperExt-2
                         11.994
     T-lowerExt-2
                          7.609
     UCZAA
                          0.152
     WhiteFlow-4
                        100.098
```

299

300

418.979

462.712

48.135

54.373

251.614

251.197

30.47

30.46

```
AA-Wood-4
                            1.486
     ChipMoisture-4
                            2.186
     SteamFlow-4
                            8.840
     Lower-HeatT-3
                            8.585
     Upper-HeatT-3
                            7.852
     ChipMass-4
                           19.347
     WeakLiquorF
                          180.613
     BlackFlow-2
                          280.829
     WeakWashF
                          267.219
     SteamHeatF-3
                            6.903
     T-Top-Chips-4
                            2.044
     SulphidityL-4
                           30.420
     dtype: float64
[39]: data2 = data2[\sim((data2 < (Q1 - 1.5 * IQR)) | (data2 > (Q3 + 1.5 * IQR))).
       \rightarrowany(axis = 1)]
      data2
[39]:
                     ChipRate
                               BF-CMratio
                                            BlowFlow ChipLevel4
                                                                     T-upperExt-2
           Y-Kappa
              27.60
                       16.810
                                    79.022
                                                           341.327
                                                                           351.050
      1
                                            1328.360
      2
              23.19
                       16.709
                                    79.562
                                            1329.407
                                                           239.161
                                                                           350.022
      3
              23.60
                       16.478
                                    81.011
                                            1334.877
                                                           213.527
                                                                           350.938
      5
              14.23
                       15.350
                                    85.518
                                            1171.604
                                                           198.538
                                                                           344.014
      6
              13.49
                       13.700
                                    98.186
                                            1243.688
                                                                           346.208
                                                           116.275
      . .
               •••
      276
              22.70
                       15.517
                                    83.008
                                            1288.010
                                                           306.886
                                                                           350.155
             20.50
                                    97.662
      296
                       13.358
                                            1304.597
                                                           377.678
                                                                           347.672
      297
              20.40
                       14.233
                                    89.790
                                            1278.006
                                                           379.458
                                                                           354.290
      298
              20.90
                       15.167
                                    84.640
                                            1283.706
                                                                           354.803
                                                           339.440
                                    94.172 1327.832
      307
              20.89
                       14.308
                                                           251.120
                                                                           351.263
           T-lowerExt-2
                            UCZAA
                                    WhiteFlow-4
                                                   AAWhiteSt-4
                                                                     SteamFlow-4
      1
                   329.067
                            1.549
                                         537.201
                                                          6.076
                                                                           60.012
      2
                   329.260
                            1.600
                                         549.611
                                                          0.000
                                                                           61.304
      3
                   331.142
                            1.604
                                         623.362
                                                          6.054
                                                                           68.496
      5
                   325.195
                            1.436
                                         628.245
                                                          6.020
                                                                           65.225
                                                                           72.989
      6
                   326.982
                            1.434
                                         696.766
                                                          0.000
                                                          6.170
      276
                   322.485
                            1.590
                                         568.752
                                                                           67.678
                                                          6.340
      296
                   313.147
                                                                           60.119
                            1.546
                                         496.460
                                                          0.000
      297
                   315.558
                            1.515
                                         491.374
                                                                           60.424
      298
                            1.635
                                                          6.340
                   311.041
                                         532.419
                                                                           65.561
                                                          0.000
      307
                   332.485
                            1.522
                                         631.514
                                                                           71.286
           Lower-HeatT-3 Upper-HeatT-3
                                            ChipMass-4
                                                          WeakLiquorF
                                                                         BlackFlow-2
      1
                  330.823
                                   304.879
                                                 163.202
                                                                665.975
                                                                              1297.317
```

AAWhiteSt-4

6.143

```
2
                  329.140
                                   303.383
                                                 164.013
                                                                677.534
                                                                              1327.072
      3
                  328.875
                                   302.254
                                                 181.487
                                                                767.853
                                                                              1324.461
      5
                  322.103
                                   298.517
                                                 165.814
                                                                826.243
                                                                               907.641
      6
                  322.982
                                   296.080
                                                 182.018
                                                                784.281
                                                                               929.527
                      •••
      276
                  331.854
                                   309.346
                                                 160.061
                                                                910.013
                                                                              1381.389
      296
                  332.615
                                   308.575
                                                 141.076
                                                                997.904
                                                                              1334.703
      297
                  331.980
                                   308.078
                                                 140.301
                                                                975.016
                                                                              1344.835
      298
                                                                              1344.708
                  332.924
                                   307.626
                                                 145.299
                                                                832.906
      307
                  328.699
                                   300.706
                                                                903.605
                                                                              1323.082
                                                 180.229
           WeakWashF
                        SteamHeatF-3
                                        T-Top-Chips-4
                                                         SulphidityL-4
      1
              241.182
                                46.603
                                                251.406
                                                                   29.11
      2
              237.272
                                51.795
                                                251.335
                                                                    0.00
      3
                                54.846
                                                                   29.02
              239.478
                                                250.312
      5
              595.875
                                52.807
                                                249.580
                                                                   30.34
      6
                                                                    0.00
              201.272
                                58.118
                                                248.741
      . .
      276
              441.934
                                51.466
                                                252.216
                                                                   29.59
      296
              389.497
                                46.206
                                                252.423
                                                                   30.43
                                                                    0.00
      297
              388.676
                                47.803
                                                252.311
      298
                                49.524
                                                251.833
                                                                   30.29
              388.911
      307
              232.729
                                54.503
                                                250.084
                                                                    0.00
      [226 rows x 22 columns]
[40]: # Transforming dataset
      import scipy
      import sklearn
      from sklearn import preprocessing
      from sklearn.preprocessing import scale
[42]:
     data2.describe()
[42]:
                                                                  ChipLevel4
                 Y-Kappa
                            ChipRate
                                       BF-CMratio
                                                       BlowFlow
              226.000000
                          226.000000
                                       226.000000
                                                     226.000000
                                                                   226.000000
      count
      mean
               20.690487
                           14.673491
                                        85.882181
                                                    1255.288916
                                                                   264.664912
      std
               2.982916
                            1.297369
                                         7.033155
                                                      47.896055
                                                                    74.345135
      min
               12.480000
                           10.833000
                                        68.645000 1084.083000
                                                                    61.783000
      25%
                                                    1221.926000
               18.457500
                           13.850000
                                        80.984000
                                                                   220.356000
      50%
              20.775000
                           14.729000
                                        84.967000
                                                    1280.291500
                                                                   270.965000
      75%
              23.010000
                           15.708000
                                        91.178750
                                                    1289.254000
                                                                   322.492000
      max
              27.600000
                           16.958000
                                       108.104000
                                                    1351.240000
                                                                   419.014000
                                                           WhiteFlow-4
              T-upperExt-2
                             T-lowerExt-2
                                                    UCZAA
                                                                          AAWhiteSt-4
                                                                                         \
                 226.000000
                                  226.000000
                                              226.000000
                                                              226.000000
                                                                             226.000000
      count
                 356.861681
                                  325.341124
                                                 1.487146
                                                              603.242482
                                                                               3.098164
      mean
```

```
7.466897
                                    5.557537
                                                 0.108054
                                                               61.052197
                                                                               3.078138
      std
      min
                 340.222000
                                  310.421000
                                                 1.182000
                                                              468.841000
                                                                               0.000000
      25%
                 350.704250
                                  322.355500
                                                 1.429000
                                                              549.611000
                                                                               0.000000
      50%
                                                 1.492000
                                                              602.508000
                 357.560500
                                  326.508500
                                                                               5.904500
      75%
                 361.555000
                                  329.264500
                                                 1.556000
                                                              653.358500
                                                                               6.140000
                 375.047000
                                  337.012000
                                                 1.712000
      max
                                                              731.394000
                                                                               6.340000
                 SteamFlow-4
                                Lower-HeatT-3
                                                Upper-HeatT-3
                                                                 ChipMass-4
                                                                               \
                   226.000000
                                   226.000000
                                                    226.000000
                                                                  226.000000
      count
      mean
                    67.545478
                                   324.752212
                                                    299.655420
                                                                  164.220102
      std
                     4.914301
                                     4.526481
                                                      4.383788
                                                                   11.423269
             •••
                                   318.051000
      min
                    52.962000
                                                    293.312000
                                                                  133.878000
      25%
                    63.954000
                                   321.179500
                                                    296.338500
                                                                  156.091000
             •••
                                   322.380000
      50%
                    68.147000
                                                    297.636500
                                                                  164.333000
      75%
                    71.760750
                                                                  172.555000
                                   329.575000
                                                    303.777000
      max
                    75.974000
                                   333.223000
                                                    309.854000
                                                                  189.268000
             WeakLiquorF
                             BlackFlow-2
                                            WeakWashF
                                                         SteamHeatF-3
                                                                         T-Top-Chips-4
                226.000000
                               226.000000
                                            226.000000
                                                            226.000000
                                                                             226.000000
      count
      mean
                874.123035
                              1149.895257
                                            273.739403
                                                             49.810239
                                                                             251.177779
      std
                120.259977
                               150.321416
                                            163.452307
                                                              4.143153
                                                                               1.221296
                                              0.00000
                                                             38.283000
      min
                596.446000
                               838.948000
                                                                             248.359000
      25%
                784.366750
                              1014.977000
                                            149.331750
                                                             46.639750
                                                                             250.290750
      50%
                866.170000
                              1126.513500
                                            283.079500
                                                             50.128500
                                                                             251.233000
      75%
                968.683250
                              1302.847000
                                            414.599750
                                                             52.889250
                                                                             252.240000
               1132.181000
                              1392.868000
                                            715.715000
                                                             59.564000
                                                                             254.122000
      max
              SulphidityL-4
      count
                  226.000000
                   15.391987
      mean
      std
                   15.297984
      min
                    0.000000
      25%
                    0.000000
      50%
                   29.065000
      75%
                   30.437500
                   32.840000
      max
      [8 rows x 22 columns]
[43]: data2.matrix = data2.values.reshape(-1,1)
      scaled = preprocessing.MinMaxScaler(feature_range = (0,10))
```

C:\Users\Dell\AppData\Local\Temp\ipykernel_11856\2925766736.py:1: UserWarning:
Pandas doesn't allow columns to be created via a new attribute name - see
https://pandas.pydata.org/pandas-docs/stable/indexing.html#attribute-access
data2.matrix = data2.values.reshape(-1,1)

scaled_data =scaled.fit_transform(data2)

data2

[43]:		Y-Kappa	ChipRate	e BF-CMra	atio B	lowFlow	ChipLevel4	T-upperExt-2	\
	1	27.60	16.810			328.360	341.327	351.050	
	2	23.19	16.709	9 79	.562 13	329.407	239.161	350.022	
	3	23.60	16.478	81	.011 13	334.877	213.527	350.938	
	5	14.23	15.350			171.604	198.538	344.014	
	6	13.49	13.700			243.688	116.275	346.208	
	276	22.70	15.517	7 83	.008 1	288.010	306.886	350.155	
	296	20.50	13.358	3 97	.662 13	304.597	377.678	347.672	
	297	20.40	14.233		.790 1	278.006	379.458	354.290	
	298	20.90	15.167	7 84	.640 1	283.706	339.440	354.803	
	307	20.89	14.308			327.832	251.120	351.263	
		T-lowerE	+ O I	JCZAA Wh:	iteFlow [.]	A A A T.T	hiteSt-4	SteamFlow-4	\
	1			1.549	537.			60.012	\
	1								
	2			1.600	549.		0.000	61.304	
	3			1.604	623.		6.054	68.496	
	5			1.436	628.		6.020	65.225	
	6	3:		1.434	696.	766	0.000	72.989	
	 276	2		 1.590	 568.	750	 6.170	 67.678	
	296			1.546	496.		0.040	60.119	
	296 297								
				1.515	491.3			60.424	
	298			1.635	532.4		6.340	65.561	
	307	٥.	32.485	1.522	631.	014	0.000	71.286	
		Lower-He	atT-3 Uj	pper-Heat	Γ-3 C	nipMass-	4 WeakLiquo	rF BlackFlow-	-2 \
	1	330	0.823	304	. 879	163.2	02 665.	975 1297.3	317
	2	329	9.140	303	. 383	164.0	13 677.	534 1327.0	72
	3	32	8.875	302	. 254	181.4	87 767.	853 1324.4	161
	5	32:	2.103	298	.517	165.8	14 826.	243 907.6	641
	6	32:	2.982	296	.080	182.0	18 784.	281 929.5	527
			•••	•••		•••	•••	•••	
	276	33	1.854	309	.346	160.0	61 910.	013 1381.3	389
	296		2.615		. 575	141.0			
	297		1.980	308	.078	140.3	01 975.		
	298		2.924		. 626	145.2			
	307		8.699		.706	180.2			
			п а.			G1 : 4	G 3 1 1 1 1	T 4	
	4	WeakWash		nHeatF-3	1-10p	-Chips-4			
	1	241.13		46.603		251.40		9.11	
	2	237.2		51.795		251.33		0.00	
	3	239.4		54.846		250.31		9.02	
	5	595.8		52.807		249.58		0.34	
	6	201.2	72	58.118		248.74	1	0.00	
	 276	 441.9		 51.466		 252.21	•••	9.59	

296	389.497	46.206	252.423	30.43
297	388.676	47.803	252.311	0.00
298	388.911	49.524	251.833	30.29
307	232.729	54.503	250.084	0.00

[226 rows x 22 columns]

[]: