

Out[]:

	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	upperExt-2	T-lowe
count	324.000000	319.000000	307.000000	308.000000	323.000000	322.000000	322.0
mean	20.635370	14.347937	87.464456	1237.837614	258.164483	356.904295	324.0
std	3.070036	1.499095	7.995012	100.593735	87.987452	9.209290	7.6
min	12.170000	9.983000	68.645000	0.000000	0.000000	339.168000	284.6
25%	18.382500	13.358000	81.823000	1193.215250	213.527000	350.241250	321.4
50%	20.845000	14.308000	86.739000	1273.138500	271.792000	356.843000	325.6
75%	23.032500	15.517000	92.372000	1289.196000	321.680000	362.242250	329.1
max	27.600000	16.958000	121.717000	1351.240000	419.014000	399.135000	337.0

8 rows × 22 columns

◀	▶
---	---

In []:

```
#show first five row
df.head()
```

Out[]:

	Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	upperExt-2	lowerExt-2
0	31-00:00	23.10	16.520	121.717	1177.607	169.805	358.282	329.5
1	31-01:00	27.60	16.810	79.022	1328.360	341.327	351.050	329.0
2	31-02:00	23.19	16.709	79.562	1329.407	239.161	350.022	329.2
3	31-03:00	23.60	16.478	81.011	1334.877	213.527	350.938	331.1
4	31-04:00	22.90	15.618	93.244	1334.168	243.131	351.640	332.7

5 rows × 23 columns

◀	▶
---	---

In []:

```
#show last five row
df.tail()
```

Out[]:

	Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	upperExt-2	T-lowe
319	10-16:00	23.75	12.667	93.450	1178.252	276.955	347.286	310.0
320	9-19:00	19.80	12.558	94.352	1184.119	297.071	399.135	310.0
321	9-20:00	23.01	12.550	90.842	1188.517	289.826	373.633	310.0

322	9-21:00	24.32	13.083	88.910	1192.879	318.006	364.081	30
323	9-22:00	25.75	13.417	85.451	1186.342	248.312	356.289	31

5 rows × 23 columns

In []:

```
#summary of data
df.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
 17  WeakLiquorF      323 non-null    float64
 18  BlackFlow-2       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
```

In []:

```
#drop duplicate items
df=df.drop_duplicates()
df
```

Out[]:

	Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	upperExt-2	lowe
0	31-00:00	23.10	16.520	121.717	1177.607	169.805	358.282	32
1	31-01:00	27.60	16.810	79.022	1328.360	341.327	351.050	32
2	31-02:00	23.19	16.709	79.562	1329.407	239.161	350.022	32
3	31-03:00	23.60	16.478	81.011	1334.877	213.527	350.938	33
4	31-04:00	22.90	15.618	93.244	1334.168	243.131	351.640	33

...
298	12-09:00	20.90	15.167	84.640	1283.706	339.440	354.803	31	
299	12-10:00	24.98	NaN	85.034	1278.345	368.564	357.723	32	
300	12-11:00	21.00	NaN	88.013	1307.722	278.842	357.438	32	
301	12-12:00	21.40	NaN	85.490	1255.986	273.484	361.365	32	
307	31-05:00	20.89	14.308	94.172	1327.832	251.120	351.263	33	

301 rows × 23 columns

In []:

```
#identify missing values
df.isnull()
```

Out[]:

	Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	upperExt-2	lowerExt-2
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...
298	False	False	False	False	False	False	False	False
299	False	False	True	False	False	False	False	False
300	False	False	True	False	False	False	False	False
301	False	False	True	False	False	False	False	False
307	False	False	False	False	False	False	False	False

301 rows × 23 columns

In []:

```
#sum of missing values per column
df.isnull().sum()
```

Out[]:

Observation	0
Y-Kappa	0
ChipRate	4
BF-CMratio	14
BlowFlow	13

```

ChipLevel4      1
T-upperExt-2   1
T-lowerExt-2   1
UCZAA          24
WhiteFlow-4    1
AAWhiteSt-4    141
AA-Wood-4     1
ChipMoisture-4 1
SteamFlow-4    1
Lower-HeatT-3   1
Upper-HeatT-3   1
ChipMass-4    1
WeakLiquorF    1
BlackFlow-2    1
WeakWashF     1
SteamHeatF-3   1
T-Top-Chips-4 1
SulphidityL-4  141
dtype: int64

```

In []:

```
#sum of missing values across the entire DataFrame
df.isnull().sum().sum()
```

Out[]:

In []:

```
#identify non-missing values
df.notnull()
```

Out[]:

	Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	upperExt-2	T-lowe
0	True	True	True	True	True	True	True	True
1	True	True	True	True	True	True	True	True
2	True	True	True	True	True	True	True	True
3	True	True	True	True	True	True	True	True
4	True	True	True	True	True	True	True	True
...
298	True	True	True	True	True	True	True	True
299	True	True	False	True	True	True	True	True
300	True	True	False	True	True	True	True	True
301	True	True	False	True	True	True	True	True
307	True	True	True	True	True	True	True	True

301 rows × 23 columns

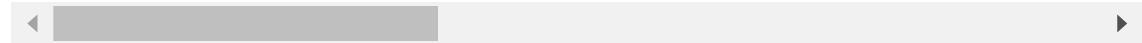
In []:

```
#fill missing values
df2=df.fillna(value=0)
df2
```

Out[]:

	Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	upperExt-2	lowerExt-2
0	31-00:00	23.10	16.520	121.717	1177.607	169.805	358.282	32
1	31-01:00	27.60	16.810	79.022	1328.360	341.327	351.050	32
2	31-02:00	23.19	16.709	79.562	1329.407	239.161	350.022	32
3	31-03:00	23.60	16.478	81.011	1334.877	213.527	350.938	33
4	31-04:00	22.90	15.618	93.244	1334.168	243.131	351.640	33
...
298	12-09:00	20.90	15.167	84.640	1283.706	339.440	354.803	31
299	12-10:00	24.98	0.000	85.034	1278.345	368.564	357.723	32
300	12-11:00	21.00	0.000	88.013	1307.722	278.842	357.438	32
301	12-12:00	21.40	0.000	85.490	1255.986	273.484	361.365	32
307	31-05:00	20.89	14.308	94.172	1327.832	251.120	351.263	33

301 rows × 23 columns



In []:

```
#sum of missing values across the entire DataFrame
df2.isnull().sum().sum()
```

Out[]: 0

In []:

```
#forward fill
df3=df.fillna(method='ffill')
df3
```

C:\Users\hp\AppData\Local\Temp\ipykernel_10468\667304635.py:2: FutureWarning: DataFrame.fillna with 'method' is deprecated and will raise in a future version.
Use obj.ffill() or obj.bfill() instead.
df3=df.fillna(method='ffill')

Out[]:

	Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	upperExt-2	lowerExt-2
0	31-00:00	23.10	16.520	121.717	1177.607	169.805	358.282	32
1	31-01:00	27.60	16.810	79.022	1328.360	341.327	351.050	32
2	31-02:00	23.19	16.709	79.562	1329.407	239.161	350.022	32

3	31-03:00	23.60	16.4/8	81.011	1334.877	213.527	350.938	33
4	31-04:00	22.90	15.618	93.244	1334.168	243.131	351.640	33
...
298	12-09:00	20.90	15.167	84.640	1283.706	339.440	354.803	31
299	12-10:00	24.98	15.167	85.034	1278.345	368.564	357.723	32
300	12-11:00	21.00	15.167	88.013	1307.722	278.842	357.438	32
301	12-12:00	21.40	15.167	85.490	1255.986	273.484	361.365	32
307	31-05:00	20.89	14.308	94.172	1327.832	251.120	351.263	33

301 rows × 23 columns

```
#backward fill
df4=df.fillna(method='bfill')
df4
```

C:\Users\hp\AppData\Local\Temp\ipykernel_10468\1750887326.py:2: FutureWarning:
 DataFrame.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
 df4=df.fillna(method='bfill')

Out[]:

	Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	upperExt-2	lowerExt-2
0	31-00:00	23.10	16.520	121.717	1177.607	169.805	358.282	32
1	31-01:00	27.60	16.810	79.022	1328.360	341.327	351.050	32
2	31-02:00	23.19	16.709	79.562	1329.407	239.161	350.022	32
3	31-03:00	23.60	16.478	81.011	1334.877	213.527	350.938	33
4	31-04:00	22.90	15.618	93.244	1334.168	243.131	351.640	33
...
298	12-09:00	20.90	15.167	84.640	1283.706	339.440	354.803	31
299	12-10:00	24.98	14.308	85.034	1278.345	368.564	357.723	32
300	12-11:00	21.00	14.308	88.013	1307.722	278.842	357.438	32
301	12-12:00	21.40	14.308	85.490	1255.986	273.484	361.365	32
307	31-05:00	20.89	14.308	94.172	1327.832	251.120	351.263	33

301 rows × 23 columns

```
#detect the outliers
df2.columns
```

```
Out[ ]: 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')
```

```
In [ ]: #drop particular outliers
df2.drop(['Observation'], axis=1, inplace=True)
df2.columns
```

```
Out[ ]: 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')
```

```
In [ ]: #calculate the Interquartile Range(IQR).
Q1=df2.quantile(0.25)
Q3=df2.quantile(0.75)
IQR=Q3-Q1
print(IQR)
```

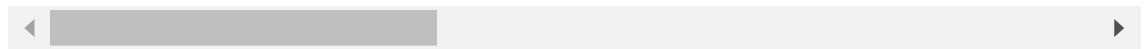
Y-Kappa	4.550
ChipRate	2.233
BF-CMratio	10.912
BlowFlow	96.766
ChipLevel4	105.868
T-upperExt-2	11.994
T-lowerExt-2	7.609
UCZAA	0.152
WhiteFlow-4	100.098
AAWhiteSt-4	6.143
AA-Wood-4	1.486
ChipMoisture-4	2.186
SteamFlow-4	8.840
Lower-HeatT-3	8.585
Upper-HeatT-3	7.852
ChipMass-4	19.347
WeakLiquorF	180.613
BlackFlow-2	280.829
WeakWashF	267.219
SteamHeatF-3	6.903
T-Top-Chips-4	2.044
SulphidityL-4	30.420
dtype: float64	

```
In [ ]: #drop outliers from the dataframe
df2=df2[-((df2<(Q1-1.5*IQR))|(df2>(Q3+1.5*IQR))).any(axis=1)]
df2
```

```
Out[ ]:      Y-    ChipRate      BF-      BlowFlow      ChipLevel4      T-    T-
           _____  _____  _____
           Y-    ChipRate      BF-      BlowFlow      ChipLevel4      upperExt-  lowerExt-  UCZAA
```

	Kappa	BF-CMratio	BlowFlow	ChipLevel4	T-2	2
1	27.60	16.810	79.022	1328.360	341.327	351.050
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	116.275	346.208
...
276	22.70	15.517	83.008	1288.010	306.886	350.155
296	20.50	13.358	97.662	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	339.440	354.803
307	20.89	14.308	94.172	1327.832	251.120	351.263

226 rows × 22 columns



In []:

```
#descriptive statistics function
df2.describe()
```

Out[]:

	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	upperExt-2	T-2	lowerExt-2
count	226.000000	226.000000	226.000000	226.000000	226.000000	226.000000	226.000000	226.000000
mean	20.690487	14.673491	85.882181	1255.288916	264.664912	356.861681	325.3	300.0
std	2.982916	1.297369	7.033155	47.896055	74.345135	7.466897	5.5	5.0
min	12.480000	10.833000	68.645000	1084.083000	61.783000	340.222000	310.4	280.0
25%	18.457500	13.850000	80.984000	1221.926000	220.356000	350.704250	322.3	300.0
50%	20.775000	14.729000	84.967000	1280.291500	270.965000	357.560500	326.5	300.0
75%	23.010000	15.708000	91.178750	1289.254000	322.492000	361.555000	329.2	300.0
max	27.600000	16.958000	108.104000	1351.240000	419.014000	375.047000	337.0	300.0

8 rows × 22 columns

