```
import pandas as pd
```

```
data = pd.read_csv('/content/01.Data Cleaning and Preprocessing.csv')
```

```
data.shape
```

```
(324, 23)
```

```
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
 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
```

```
data.describe()
```

	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	T-upperExt-2	lowerEx
count	324.000000	319.000000	307.000000	308.000000	323.000000	322.000000	322.000(
mean	20.635370	14.347937	87.464456	1237.837614	258.164483	356.904295	324.020
std	3.070036	1.499095	7.995012	100.593735	87.987452	9.209290	7.621
min	12.170000	9.983000	68.645000	0.000000	0.000000	339.168000	284.633(
25%	18.382500	13.358000	81.823000	1193.215250	213.527000	350.241250	321.420(
50%	20.845000	14.308000	86.739000	1273.138500	271.792000	356.843000	325.669(
75%	23.032500	15.517000	92.372000	1289.196000	321.680000	362.242250	329.175(
max	27.600000	16.958000	121.717000	1351.240000	419.014000	399.135000	337.012(

8 rows × 22 columns

```
data.notnull().sum()
```

```
Observation        324
Y-Kappa            324
ChipRate           319
BF-CMratio         307
BlowFlow           308
ChipLevel4         323
T-upperExt-2       322
T-lowerExt-2       322
UCZAA              299
WhiteFlow-4        323
AAWhiteSt-4        173
AA-Wood-4          323
ChipMoisture-4     323
SteamFlow-4        323
Lower-HeatT-3      322
Upper-HeatT-3      322
```

```
ChipMass-4          323
WeakLiquorF         323
BlackFlow-2         322
WeakWashF           323
SteamHeatF-3        322
T-Top-Chips-4       323
SulphidityL-4       173
dtype: int64
```

```
data.isnull()
```

	Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	T-upperExt-2	T lowerExt
0	False	False	False	False	False	False	False	Fals
1	False	False	False	False	False	False	False	Fals
2	False	False	False	False	False	False	False	Fals
3	False	False	False	False	False	False	False	Fals
4	False	False	False	False	False	False	False	Fals
...	...	...	...	...	...	...	...	..
319	False	False	False	False	False	False	False	Fals
320	False	False	False	False	False	False	False	Fals
321	False	False	False	False	False	False	False	Fals
322	False	False	False	False	False	False	False	Fals
323	False	False	False	False	False	False	False	Fals

324 rows × 23 columns

```
data.isnull().sum()
```

```
Observation          0
Y-Kappa              0
ChipRate             5
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
```

```
data.isnull().sum().sum()
```

```
386
```

```
data1 = data.fillna(value=0)
data1
```

	Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	T-upperExt-2	T lowerExt
0	31-00:00	23.10	16.520	121.717	1177.607	169.805	358.282	329.54
1	31-01:00	27.60	16.810	79.022	1328.360	341.327	351.050	329.06
2	31-02:00	23.19	16.709	79.562	1329.407	239.161	350.022	329.26
3	31-03:00	23.60	16.478	81.011	1334.877	213.527	350.938	331.14
4	31-04:00	22.90	15.618	93.244	1334.168	243.131	351.640	332.70
...	...	...	...	...	...	...	...	..
319	10-16:00	23.75	12.667	93.450	1178.252	276.955	347.286	310.97
320	9-19:00	19.80	12.558	94.352	1184.119	297.071	399.135	319.57
321	9-20:00	23.01	12.550	90.842	1188.517	289.826	373.633	314.59
322	9-21:00	24.32	13.083	88.910	1192.879	318.006	364.081	308.55
323	9-22:00	25.75	13.417	85.451	1186.342	248.312	356.289	310.48

324 rows × 23 columns

```
data1.isnull().sum()
```

```
Observation        0
Y-Kappa            0
ChipRate           0
BF-CMratio         0
BlowFlow           0
ChipLevel4         0
T-upperExt-2       0
T-lowerExt-2       0
UCZAA              0
WhiteFlow-4        0
AAWhiteSt-4        0
AA-Wood-4          0
ChipMoisture-4     0
SteamFlow-4        0
Lower-HeatT-3      0
Upper-HeatT-3      0
ChipMass-4         0
WeakLiquorF        0
BlackFlow-2        0
WeakWashF          0
SteamHeatF-3       0
T-Top-Chips-4      0
SulphidityL-4      0
dtype: int64
```

```
data2 = data.fillna(method = 'bfill')
data2
```

	Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	T-upperExt-2	T lowerExt
0	31-00:00	23.10	16.520	121.717	1177.607	169.805	358.282	329.54
1	31-01:00	27.60	16.810	79.022	1328.360	341.327	351.050	329.06
2	31-02:00	23.19	16.709	79.562	1329.407	239.161	350.022	329.26
3	31-03:00	23.60	16.478	81.011	1334.877	213.527	350.938	331.14
4	31-04:00	22.90	15.618	93.244	1334.168	243.131	351.640	332.70
...	...	...	...	...	...	...	...	..
319	10-16:00	23.75	12.667	93.450	1178.252	276.955	347.286	310.97
320	9-19:00	19.80	12.558	94.352	1184.119	297.071	399.135	319.57
321	9-20:00	23.01	12.550	90.842	1188.517	289.826	373.633	314.59
322	9-21:00	24.32	13.083	88.910	1192.879	318.006	364.081	308.55
323	9-22:00	25.75	13.417	85.451	1186.342	248.312	356.289	310.48

324 rows × 23 columns

```
data3 = data.fillna(method = 'pad')
data3
```

	Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	T-upperExt-2	T lowerExt
0	31-00:00	23.10	16.520	121.717	1177.607	169.805	358.282	329.54
1	31-01:00	27.60	16.810	79.022	1328.360	341.327	351.050	329.06
2	31-02:00	23.19	16.709	79.562	1329.407	239.161	350.022	329.26
3	31-03:00	23.60	16.478	81.011	1334.877	213.527	350.938	331.14
4	31-04:00	22.90	15.618	93.244	1334.168	243.131	351.640	332.70
...	...	...	...	...	...	...	...	..
319	10-16:00	23.75	12.667	93.450	1178.252	276.955	347.286	310.97
320	9-19:00	19.80	12.558	94.352	1184.119	297.071	399.135	319.57
321	9-20:00	23.01	12.550	90.842	1188.517	289.826	373.633	314.59
322	9-21:00	24.32	13.083	88.910	1192.879	318.006	364.081	308.55
323	9-22:00	25.75	13.417	85.451	1186.342	248.312	356.289	310.48

324 rows × 23 columns

```
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats


data1.drop(['Observation'],axis = 1, inplace = True)
data1.columns
```

```
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')
```

```
Q1 = data1.quantile(0.25)
Q3 =  data1.quantile(0.75)
Q = Q3 - Q1
Q
```

```
Y-Kappa              4.65000
ChipRate             2.25625
BF-CMratio          11.11225
BlowFlow            98.43375
ChipLevel4         107.92275
T-upperExt-2        11.96500
T-lowerExt-2         7.82875
UCZAA                0.13925
WhiteFlow-4         98.59525
AAWhiteSt-4          6.14000
AA-Wood-4            1.45900
ChipMoisture-4       2.22000
SteamFlow-4          9.04675
Lower-HeatT-3        8.46750
Upper-HeatT-3        7.77050
ChipMass-4          19.70375
WeakLiquorF        174.05550
BlackFlow-2        276.51675
WeakWashF          271.44325
SteamHeatF-3         6.94975
T-Top-Chips-4        2.01025
SulphidityL-4       30.40250
dtype: float64
```

```
data1 = data1[~((data1 < Q1 - 1.5 * Q) | (data1 > Q3 + 1.5 * Q)).any(axis=1)]

data1
```

	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	T-upperExt-2	T-lowerExt-2	UCZAA	Whi
1	27.60	16.810	79.022	1328.360	341.327	351.050	329.067	1.549	
2	23.19	16.709	79.562	1329.407	239.161	350.022	329.260	1.600	
3	23.60	16.478	81.011	1334.877	213.527	350.938	331.142	1.604	
5	14.23	15.350	85.518	1171.604	198.538	344.014	325.195	1.436	
6	13.49	13.700	98.186	1243.688	116.275	346.208	326.982	1.434	
...	...	...	...	...	...	...	...	...	
317	17.80	16.625	78.367	1276.082	202.744	360.127	329.266	1.488	
318	18.20	16.283	83.508	1288.104	234.284	359.412	328.670	1.534	
319	23.75	12.667	93.450	1178.252	276.955	347.286	310.970	1.523	
321	23.01	12.550	90.842	1188.517	289.826	373.633	314.591	1.457	
323	25.75	13.417	85.451	1186.342	248.312	356.289	310.482	1.474	

241 rows × 22 columns