

```
import pandas as pd
```

```
df = pd.read_csv('01.Data Cleaning and Preprocessing.csv')
```

```
# Set display options to show all rows and columns
pd.set_option('display.max_rows', None)
```

```
pd.set_option('display.max_columns', None)
```

```
# Display the entire DataFrame
print(df)
```

	Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	\
0	31-00:00	23.10	16.520000	121.717000	1177.607000	169.805000	
1	31-01:00	27.60	16.810000	79.022000	1328.360000	341.327000	
2	31-02:00	23.19	16.709000	79.562000	1329.407000	239.161000	
3	31-03:00	23.60	16.478000	81.011000	1334.877000	213.527000	
4	31-04:00	22.90	15.618000	93.244000	1334.168000	243.131000	
5	1-08:00	14.23	15.350000	85.518000	1171.604000	198.538000	
6	1-09:00	13.49	13.700000	98.186000	1243.688000	116.275000	
7	31-06:00	22.65	14.100000	91.887000	1307.852000	288.989000	
8	31-07:00	22.50	14.233000	97.249000	1346.900000	330.325000	
9	31-08:00	24.70	13.850000	96.208000	1334.892000	362.511000	
10	31-09:00	22.60	13.875000	95.720000	1351.240000	383.731000	
11	31-10:00	24.40	14.117000	85.998000	1330.104000	394.234000	
12	31-11:00	26.62	15.467000	84.447000	1334.255000	386.971000	
13	31-12:00	27.20	16.083000	82.839000	1332.331000	366.855000	
14	31-13:00	24.05	16.675000	77.025000	1284.386000	246.336000	
15	31-14:00	25.40	16.425000	72.924000	1197.775000	118.821000	
16	31-15:00	24.10	15.800000	79.387000	1203.375000	52.241000	
17	31-16:00	17.40	15.158000	87.057000	1218.797000	6.476000	
18	31-17:00	17.80	14.000000	92.772000	1204.489000	61.783000	
19	31-18:00	20.00	12.983000	100.073000	1206.709000	191.648000	
20	31-19:00	20.30	12.058000	89.523000	1180.205000	300.596000	
21	31-20:00	20.80	13.183000	84.397000	1187.186000	349.159000	
22	31-21:00	23.00	14.067000	88.383000	1191.691000	374.687000	
23	31-22:00	22.70	13.483000	89.578000	1180.190000	353.574000	
24	31-23:00	21.40	12.875000	92.564000	1191.757000	229.014000	
25	1-00:00	20.70	12.573000	96.141000	1208.756000	272.985000	
26	1-01:00	18.90	12.542000	94.741000	1188.212000	302.073000	
27	1-02:00	21.35	12.467000	94.280000	1194.216000	296.608000	
28	1-03:00	21.10	12.667000	86.083000	1182.201000	295.825000	
29	1-04:00	20.14	13.733000	87.464456	1237.837614	274.481000	
30	1-05:00	24.15	14.533000	72.015000	1173.241000	267.057000	
31	1-06:00	16.10	16.292000	68.645000	1094.883000	245.962000	
32	1-07:00	12.48	15.950000	77.543000	1190.289000	189.957000	
33	1-08:00	14.23	15.350000	85.518000	1171.604000	198.538000	
34	1-09:00	13.49	13.700000	98.186000	1243.688000	116.275000	
35	1-10:00	12.17	12.667000	114.570000	1253.584000	188.043000	
36	1-11:00	12.81	10.942000	114.567000	1241.142000	215.485000	
37	1-12:00	14.53	10.833000	93.121000	1202.818000	339.119000	
38	1-13:00	18.45	12.917000	84.483000	1175.728000	391.599000	
39	1-14:00	24.30	13.917000	87.835000	1190.889000	390.140000	
40	1-15:00	27.10	13.558000	83.117000	1175.417000	289.256000	
41	1-16:00	24.70	14.142000	80.809000	1171.734000	104.205000	
42	1-17:00	22.30	14.500000	84.884000	1189.789000	133.151000	
43	1-18:00	23.60	14.017000	84.943000	1189.913000	224.176000	
44	1-19:00	23.00	14.008000	80.199000	1188.278000	194.959000	
45	1-20:00	22.00	14.817000	76.696000	1220.750000	228.285000	
46	1-21:00	21.50	15.917000	75.129000	1231.493000	279.928000	
47	1-22:00	20.70	16.208000	75.314000	1220.714000	250.234000	
48	1-23:00	19.20	15.736000	80.975000	1274.247000	197.078000	
49	2-00:00	17.80	15.708000	81.625000	1282.194000	186.236000	
50	2-01:00	17.00	15.067000	87.395000	1284.704000	265.526000	
51	2-02:00	18.08	14.700000	87.350000	1284.040000	296.951000	
52	2-03:00	19.53	14.700000	86.950000	1291.208000	311.740000	
53	2-04:00	20.43	14.850000	85.819000	1288.719000	329.008000	
54	2-05:00	20.80	15.017000	80.738000	1283.069000	311.598000	
55	2-06:00	23.01	15.892000	84.424000	1275.502000	307.254000	
56	2-07:00	22.38	15.108000	76.453000	1283.128000	279.875000	

```
print(df.columns)
```

```
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')
```

```
# Display the first few rows of the DataFrame
print("Initial DataFrame:")
print(df.head())
```

```
Initial DataFrame:
  Observation  Y-Kappa  ChipRate  BF-CMratio  BlowFlow  ChipLevel4  \
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

  T-upperExt-2  T-lowerExt-2  UCZAA  WhiteFlow-4  AAWhiteSt-4  \
0      358.282      329.545    1.443      599.253         NaN
1      351.050      329.067    1.549      537.201         6.076
2      350.022      329.260    1.600      549.611         NaN
3      350.938      331.142    1.604      623.362         6.054
4      351.640      332.709     NaN      638.672         6.110

  AA-Wood-4  ChipMoisture-4  SteamFlow-4  Lower-HeatT-3  Upper-HeatT-3  \
0      16.471         46.011      67.122      329.432      303.099
1      16.543         43.954      60.012      330.823      304.879
2      16.559         44.495      61.304      329.140      303.383
3      16.562         45.592      68.496      328.875      302.254
4      16.677         45.512      70.022      328.352      300.954

  ChipMass-4  WeakLiquorF  BlackFlow-2  WeakWashF  SteamHeatF-3  \
0      175.964     1127.197     1319.039     257.325       54.612
1      163.202     665.975     1297.317     241.182       46.603
2      164.013     677.534     1327.072     237.272       51.795
3      181.487     767.853     1324.461     239.478       54.846
4      183.929     888.448     1343.424     215.372       54.186

  T-Top-Chips-4  SulphidityL-4
0      252.077         NaN
1      251.406      29.11
2      251.335         NaN
3      250.312      29.02
4      249.916      29.01
```

```
# Display the first few rows of the DataFrame
print("Initial DataFrame:")
print(df.tail())
```

```
Initial DataFrame:
  Observation  Y-Kappa  ChipRate  BF-CMratio  BlowFlow  ChipLevel4  \
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    12.550     90.842    1188.517     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  AAWhiteSt-4  \
319      347.286      310.970    1.523      513.956         6.068
320      399.135      319.576    1.451      570.058         6.190
321      373.633      314.591    1.457      549.306     -999.000
322      364.081      308.559    1.523      504.852         6.128
323      356.289      310.482    1.474      497.375     -999.000

  AA-Wood-4  ChipMoisture-4  SteamFlow-4  Lower-HeatT-3  \
319      17.165         45.954      61.141      330.117
320      16.911         44.198      67.667      330.848
321      16.858         44.079      66.446      330.226
322      16.821         43.734      61.054      327.346
323      16.822         43.627      58.247      328.092

  Upper-HeatT-3  ChipMass-4  WeakLiquorF  BlackFlow-2  WeakWashF  \
319      304.006      148.174     1027.201     1357.271     381.643
320      304.616      165.178     906.962     1311.177      25.494
321      304.686      160.841     887.125     1319.226         0.638
322      304.363      147.589     804.423     1320.225         0.000
323      304.093      144.218     828.328     1320.848         1.276

  SteamHeatF-3  T-Top-Chips-4  SulphidityL-4
319         45.264      252.947         30.86
320         50.528      252.092         30.70
321         45.549      252.438     -999.00
322         43.725      253.176         31.13
323         43.840      253.216     -999.00
```

```
# Filter rows where 'Y-Kappa' is greater than 25
filtered_df = df[df['Y-Kappa'] > 25]
```

```
print("\nFiltered DataFrame (Y-Kappa > 25):")
print(filtered_df.head())
```



Filtered DataFrame (Y-Kappa > 25):

	Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	\
1	31-01:00	27.60	16.810	79.022	1328.360	341.327	
12	31-11:00	26.62	15.467	84.447	1334.255	386.971	
13	31-12:00	27.20	16.083	82.839	1332.331	366.855	
15	31-14:00	25.40	16.425	72.924	1197.775	118.821	
40	1-15:00	27.10	13.558	83.117	1175.417	289.256	

	T-upperExt-2	T-lowerExt-2	UCZAA	WhiteFlow-4	AAWhiteSt-4	\
1	351.050	329.067	1.549	537.201	6.076	
12	349.392	321.021	1.428	531.250	NaN	
13	350.094	327.439	1.486	527.893	6.190	
15	350.765	329.799	1.635	585.011	6.197	
40	339.168	318.386	1.360	480.184	NaN	

	AA-Wood-4	ChipMoisture-4	SteamFlow-4	Lower-HeatT-3	Upper-HeatT-3	\
1	16.543	43.954	60.012	330.823	304.879	
12	16.668	44.580	59.407	330.284	303.248	
13	16.669	44.541	60.271	330.023	302.883	
15	16.794	45.547	65.474	329.773	302.884	
40	18.353	43.745	48.568	318.228	294.850	

	ChipMass-4	WeakLiquorF	BlackFlow-2	WeakWashF	SteamHeatF-3	\
1	163.202	665.975	1297.317	241.182	46.603	
12	156.797	799.947	1299.782	118.901	46.597	
13	160.562	771.158	1299.974	153.647	47.175	
15	175.646	756.154	1300.037	401.418	54.628	
40	131.537	744.659	996.046	118.899	41.985	

	T-Top-Chips-4	SulphidityL-4
1	251.406	29.11
12	251.721	NaN
13	251.767	30.18
15	251.009	30.41
40	253.450	NaN

```
# Identify columns with missing values
missing_values = df.isnull().sum()
print("\nColumns with missing values:")
print(missing_values[missing_values > 0])
```



Columns with missing values:

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

```
# Fill missing values with the mean for columns with fewer missing values
columns_to_fill = missing_values[(missing_values < 50) & (df[missing_values.index].dtypes != 'object')].index # Filter out columns with object dtype
df[columns_to_fill] = df[columns_to_fill].fillna(df[columns_to_fill].mean())
```

```
# Fill columns with many missing values with a placeholder
columns_with_many_missing = missing_values[missing_values >= 50].index
df[columns_with_many_missing] = df[columns_with_many_missing].fillna(-999)
```

```
print("\nDataFrame after handling missing values:")
print(df.head())
```



DataFrame after handling missing values:

	Observation	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	\
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	

	T-upperExt-2	T-lowerExt-2	UCZAA	WhiteFlow-4	AAWhiteSt-4	\
0	358.282	329.545	1.44300	599.253	-999.000	
1	351.050	329.067	1.54900	537.201	6.076	
2	350.022	329.260	1.60000	549.611	-999.000	
3	350.938	331.142	1.60400	623.362	6.054	
4	351.640	332.709	1.49201	638.672	6.110	

	AA-Wood-4	ChipMoisture-4	SteamFlow-4	Lower-HeatT-3	Upper-HeatT-3	\
0	16.471	46.011	67.122	329.432	303.099	
1	16.543	43.954	60.012	330.823	304.879	
2	16.559	44.495	61.304	329.140	303.383	
3	16.562	45.592	68.496	328.875	302.254	
4	16.677	45.512	70.022	328.352	300.954	

	ChipMass-4	WeakLiquorF	BlackFlow-2	WeakWashF	SteamHeatF-3	\
0	175.964	1127.197	1319.039	257.325	54.612	
1	163.202	665.975	1297.317	241.182	46.603	
2	164.013	677.534	1327.072	237.272	51.795	
3	181.487	767.853	1324.461	239.478	54.846	
4	183.929	888.448	1343.424	215.372	54.186	

	T-Top-Chips-4	SulphidityL-4
0	252.077	-999.00
1	251.406	29.11
2	251.335	-999.00
3	250.312	29.02
4	249.916	29.01

```
# Calculate summary statistics for numerical columns
```

```
summary_statistics = df.describe()
print("\nSummary Statistics:")
print(summary_statistics)
```



Summary Statistics:

	Y-Kappa	ChipRate	BF-CMratio	BlowFlow	ChipLevel4	\
count	324.000000	324.000000	324.000000	324.000000	324.000000	
mean	20.635370	14.347937	87.464456	1237.837614	258.164483	
std	3.070036	1.487447	7.781774	98.070606	87.851143	
min	12.170000	9.983000	68.645000	0.000000	0.000000	
25%	18.382500	13.364750	82.156750	1194.525750	213.527000	
50%	20.845000	14.347937	87.253500	1254.658500	271.605500	
75%	23.032500	15.498250	92.123250	1288.628750	321.285000	
max	27.600000	16.958000	121.717000	1351.240000	419.014000	

	T-upperExt-2	T-lowerExt-2	UCZAA	WhiteFlow-4	AAWhiteSt-4	\
count	324.000000	324.000000	324.000000	324.000000	324.000000	
mean	356.904295	324.02018	1.492010	591.73226	-462.304657	
std	9.180734	7.59777	0.101741	66.91253	502.185878	
min	339.168000	284.63300	1.182000	405.11100	-999.000000	
25%	350.291750	321.48600	1.436000	541.00225	-999.000000	
50%	356.901648	325.63850	1.492010	592.71700	6.009500	
75%	362.104750	329.14700	1.555250	639.45775	6.140000	
max	399.135000	337.01200	1.747000	731.39400	6.340000	

	AA-Wood-4	ChipMoisture-4	SteamFlow-4	Lower-HeatT-3	\
count	324.000000	324.000000	324.000000	324.000000	
mean	17.835885	46.781969	66.668285	325.567820	
std	0.911975	1.478568	5.699744	4.595568	
min	16.471000	42.822000	48.568000	318.051000	
25%	17.105750	45.695000	62.539000	321.386500	
50%	17.759000	46.835500	67.408000	325.225500	
75%	18.560500	47.909000	71.519500	329.835750	
max	19.582000	50.658000	76.147000	333.854000	

Upper-HeatT-3 ChipMass-4 WeakLiquorF BlackFlow-2 WeakWashF \

count	324.000000	324.000000	324.000000	324.000000	324.000000
mean	300.525699	162.222322	873.828941	1175.917016	263.543068
std	4.554318	14.138751	121.884406	148.870958	163.413391
min	293.312000	113.922000	486.938000	838.948000	0.000000
25%	296.513750	153.089750	792.150250	1047.260500	134.688500
50%	299.184500	163.601000	865.457500	1152.547500	266.983500
75%	304.228250	172.555000	965.152250	1318.991750	405.411500
max	311.146000	189.268000	1226.277000	1395.767000	715.715000

	SteamHeatF-3	T-Top-Chips-4	SulphidityL-4
count	324.000000	324.000000	324.000000
mean	49.696907	251.240087	-449.345003
std	4.537794	1.281444	514.312479
min	35.510000	248.359000	-999.000000
25%	46.411250	250.312000	-999.000000
50%	50.192500	251.375000	29.236000
75%	53.284750	252.319750	30.402500
max	63.332000	254.122000	32.840000

Start coding or [generate](#) with AI.