EDA_Buencafe

June 30, 2022

1 Lyophilized Buencafé

The following is a first approach to the databases obtained from Buencafé.

1.1 Concentrator 1 (C1)

Initially the data and its corresponding type will be read

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Visualization and loading of the data delivered for concentrator 1

The three datasets that were delivered by the BuenCafe team are loaded and after this the exploratory analysis begins.

- To know the relationship between dataset 1 and dataset 2 and compare their initial data since they have the same start date.
- A concentration of null data is evidenced in the first samples of the datasets.
- Complex names are identified in each of the columns.

[2]:
```
db1 = pd.read_csv('Data/TableFloatConc1-1.csv', delimiter=',')
db1.head(5)
```

[2]:
```
                          Time  \
0  24/02/2022 12:00:27,945 a.m.
1  24/02/2022 12:01:47,954 a.m.
2  24/02/2022 12:04:47,914 a.m.
3  24/02/2022 12:05:17,937 a.m.
4  24/02/2022 12:05:47,918 a.m.

  [CONCENTRACION1]_11VM1Recristalizador1Corriente_AI.Valor  \
0                                                  NaN
1                                               22,70
2                                               22,70
3                                               22,70
4                                               22,70
```

```
  [CONCENTRACION1]_21VM1Recristalizador2Corriente_AI.Valor  \
0                                                  NaN
1                                                  NaN
2                                                  NaN
3                                                  NaN
4                                                  NaN

  [CONCENTRACION1]_31VM1Recristalizador3Corriente_AI.Valor  \
0                                                  NaN
1                                                  NaN
2                                                36,45
3                                                36,45
4                                                36,45

  [CONCENTRACION1]_11V1Recrist1Temp_TT1231_AI.Valor  \
0                                                  NaN
1                                                  NaN
2                                                  NaN
3                                                  NaN
4                                                  NaN

  [CONCENTRACION1]_21V1Recrist2Temp_TT2231_AI.Valor  \
0                                                  NaN
1                                                  NaN
2                                                  NaN
3                                                  NaN
4                                                  NaN

  [CONCENTRACION1]_31V1Recrist3Temp_TT3231_AI.Valor  \
0                                                  NaN
1                                                  NaN
2                                                  NaN
3                                                  NaN
4                                                  NaN

  [CONCENTRACION1]_16V01Recibidor1Temp_TT1120_AI.Valor  \
0                                                  NaN
1                                                  NaN
2                                                  NaN
3                                                  NaN
4                                                  NaN

  [CONCENTRACION1]_26V01Recibidor2Temp_TT2120_AI.Valor  \
0                                                  NaN
1                                                  NaN
2                                                  NaN
3                                                  NaN
```

```
4                                                -33,23

  [CONCENTRACION1]_36V01Recibidor3Temp_TT3120_AI.Valor  \
0                                                  NaN
1                                                  NaN
2                                                  NaN
3                                               -39,51
4                                               -39,51

  [CONCENTRACION1]_44V01TAPTemp_TT4401_AI.Valor  \
0                                           NaN
1                                           NaN
2                                           NaN
3                                           NaN
4                                           NaN

  [CONCENTRACION1]_11V1Recrist1Presion_PT1105_AI.Valor  \
0                                                  NaN
1                                                  NaN
2                                                  NaN
3                                                  NaN
4                                                  NaN

  [CONCENTRACION1]_21V1Recrist2Presion_PT2105_AI.Valor  \
0                                                  NaN
1                                                  NaN
2                                                  NaN
3                                                  NaN
4                                                  NaN

  [CONCENTRACION1]_31V1Recrist3Presion_PT3105_AI.Valor
0                                               2,37
1                                               2,37
2                                               2,37
3                                               2,37
4                                               2,37
```

[3]:
```
db2 = pd.read_csv('Data/TableFloatConc1-2.csv', delimiter=',')
db2.head(5)
```

[3]:
```
                          Time  \
0  24/02/2022 12:00:17,908 a.m.
1  24/02/2022 12:02:27,929 a.m.
2  24/02/2022 12:02:47,958 a.m.
3  24/02/2022 12:21:27,911 a.m.
4  24/02/2022 12:26:27,927 a.m.
```

	[CONCENTRACION1]_16V01Recibidor1Presion_PT1108_AI.Valor	[CONCENTRACION1]_26V01Recibidor2Presion_PT2108_AI.Valor	[CONCENTRACION1]_36V01Recibidor3Presion_PT3108_AI.Valor	[CONCENTRACION1]_12EM1Cristalizador1Corriente_AI.Valor	[CONCENTRACION1]_12EM2Cristalizador2Corriente_AI.Valor	[CONCENTRACION1]_12EM3Cristalizador3Corriente_AI.Valor	[CONCENTRACION1]_22EM1Cristalizador4Corriente_AI.Valor
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	7,36	NaN	NaN	NaN
2	NaN	NaN	NaN	7,36	6,01	NaN	NaN
3	NaN	NaN	NaN	7,36	6,01	NaN	NaN
4	NaN	NaN	NaN	7,36	6,01	NaN	NaN

```
4                                                   NaN

  [CONCENTRACION1]_22EM2Cristalizador5Corriente_AI.Valor  \
0                                                   NaN
1                                                   NaN
2                                                   NaN
3                                                   NaN
4                                                   NaN

  [CONCENTRACION1]_22EM3Cristalizador6Corriente_AI.Valor  \
0                                                   NaN
1                                                   NaN
2                                                   NaN
3                                                   NaN
4                                                   NaN

  [CONCENTRACION1]_32EM1Cristalizador7Corriente_AI.Valor  \
0                                                   NaN
1                                                   NaN
2                                                   NaN
3                                                   NaN
4                                                   NaN

  [CONCENTRACION1]_32EM2Cristalizador8Corriente_AI.Valor  \
0                                                   NaN
1                                                   NaN
2                                                  7,74
3                                                  7,74
4                                                  7,74

  [CONCENTRACION1]_13V01Colum1Conduct_CT1319_AI.Valor  \
0                                                   NaN
1                                                   NaN
2                                                   NaN
3                                                   NaN
4                                                   NaN

  [CONCENTRACION1]_23V01Colum2Conduct_CT2319_AI.Valor  \
0                                                 61,81
1                                                 61,81
2                                                 61,81
3                                                 61,81
4                                                308,21

  [CONCENTRACION1]_33V01Colum3Conduct_CT3319_AI.Valor
0                                                   NaN
1                                                   NaN
```

```
2                                                    NaN
3                                                 515,45
4                                                 515,45
```

[4]:
```
db1.describe()
```

[4]:
```
                              Time  \
count                        41187
unique                       39995
top     12/05/2022 1:36:29,249 a.m.
freq                             6

       [CONCENTRACION1]_11VM1Recristalizador1Corriente_AI.Valor  \
count                                                  41186
unique                                                   886
top                                                     0,03
freq                                                    1793

       [CONCENTRACION1]_21VM1Recristalizador2Corriente_AI.Valor  \
count                                                  41182
unique                                                   835
top                                                    -0,36
freq                                                    1075

       [CONCENTRACION1]_31VM1Recristalizador3Corriente_AI.Valor  \
count                                                  41185
unique                                                   987
top                                                     0,05
freq                                                    1648

       [CONCENTRACION1]_11V1Recrist1Temp_TT1231_AI.Valor  \
count                                              41173
unique                                               537
top                                                25,02
freq                                                 881

       [CONCENTRACION1]_21V1Recrist2Temp_TT2231_AI.Valor  \
count                                              41179
unique                                               473
top                                                26,82
freq                                                1828

       [CONCENTRACION1]_31V1Recrist3Temp_TT3231_AI.Valor  \
count                                              41178
unique                                               529
top                                                26,82
freq                                                1536
```

```
       [CONCENTRACION1]_16V01Recibidor1Temp_TT1120_AI.Valor  \
count                                                41172
unique                                                1288
top                                                 -26,68
freq                                                   260

       [CONCENTRACION1]_26V01Recibidor2Temp_TT2120_AI.Valor  \
count                                                41183
unique                                                1174
top                                                 -34,75
freq                                                   280

       [CONCENTRACION1]_36V01Recibidor3Temp_TT3120_AI.Valor  \
count                                                41184
unique                                                1346
top                                                 -35,92
freq                                                   221

       [CONCENTRACION1]_44V01TAPTemp_TT4401_AI.Valor  \
count                                          41175
unique                                          1129
top                                            24,99
freq                                            2678

       [CONCENTRACION1]_11V1Recrist1Presion_PT1105_AI.Valor  \
count                                                41176
unique                                                 239
top                                                   1,64
freq                                                   717

       [CONCENTRACION1]_21V1Recrist2Presion_PT2105_AI.Valor  \
count                                                41169
unique                                                 284
top                                                   2,09
freq                                                   576

       [CONCENTRACION1]_31V1Recrist3Presion_PT3105_AI.Valor
count                                                41187
unique                                                 341
top                                                   1,89
freq                                                   466
```

[5]:
```
db2.describe()
```

[5]:
```
                              Time  \
count                        43395
```

```
unique                         41646
top     19/05/2022 7:40:19,223 a.m.
freq                               7

       [CONCENTRACION1]_16V01Recibidor1Presion_PT1108_AI.Valor  \
count                                                 43372
unique                                                  384
top                                                    1,96
freq                                                   1076

       [CONCENTRACION1]_26V01Recibidor2Presion_PT2108_AI.Valor  \
count                                                 43370
unique                                                    1
top                                                  -3,21
freq                                                  43370

       [CONCENTRACION1]_36V01Recibidor3Presion_PT3108_AI.Valor  \
count                                                 43390
unique                                                  296
top                                                    0,66
freq                                                   1167

       [CONCENTRACION1]_12EM1Cristalizador1Corriente_AI.Valor  \
count                                                 43394
unique                                                  399
top                                                    0,01
freq                                                   2423

       [CONCENTRACION1]_12EM2Cristalizador2Corriente_AI.Valor  \
count                                                 43393
unique                                                  528
top                                                    0,02
freq                                                   2639

       [CONCENTRACION1]_12EM3Cristalizador3Corriente_AI.Valor  \
count                                                 43374
unique                                                  629
top                                                    0,04
freq                                                   2557

       [CONCENTRACION1]_22EM1Cristalizador4Corriente_AI.Valor  \
count                                                 43380
unique                                                  183
top                                                    0,06
freq                                                   2259

       [CONCENTRACION1]_22EM2Cristalizador5Corriente_AI.Valor  \
```

```
count                                                 43378
unique                                                  475
top                                                    0,04
freq                                                   2687

       [CONCENTRACION1]_22EM3Cristalizador6Corriente_AI.Valor  \
count                                                 43382
unique                                                  377
top                                                    0,04
freq                                                   3793

       [CONCENTRACION1]_32EM1Cristalizador7Corriente_AI.Valor  \
count                                                 43388
unique                                                  355
top                                                    0,04
freq                                                   1887

       [CONCENTRACION1]_32EM2Cristalizador8Corriente_AI.Valor  \
count                                                 43393
unique                                                  293
top                                                    0,01
freq                                                   2204

       [CONCENTRACION1]_13V01Colum1Conduct_CT1319_AI.Valor  \
count                                                 43385
unique                                                 2706
top                                                  515,24
freq                                                    624

       [CONCENTRACION1]_23V01Colum2Conduct_CT2319_AI.Valor  \
count                                                 43395
unique                                                 2593
top                                                  515,46
freq                                                    560

       [CONCENTRACION1]_33V01Colum3Conduct_CT3319_AI.Valor
count                                                 43392
unique                                                 2172
top                                                  515,48
freq                                                   2144
```

1.1.1 Convert data types to the corresponding type

Change column names Due to the long and confusing column names in the databases that were identified in the first analysis, the names are changed to more precise identifiers to facilitate their handling.

[6]:
```
db1.isnull().mean()
```

[6]:
```
Time                                                  0.000000
[CONCENTRACION1]_11VM1Recristalizador1Corriente_AI.Valor    0.000024
[CONCENTRACION1]_21VM1Recristalizador2Corriente_AI.Valor    0.000121
[CONCENTRACION1]_31VM1Recristalizador3Corriente_AI.Valor    0.000049
[CONCENTRACION1]_11V1Recrist1Temp_TT1231_AI.Valor           0.000340
[CONCENTRACION1]_21V1Recrist2Temp_TT2231_AI.Valor           0.000194
[CONCENTRACION1]_31V1Recrist3Temp_TT3231_AI.Valor           0.000219
[CONCENTRACION1]_16V01Recibidor1Temp_TT1120_AI.Valor        0.000364
[CONCENTRACION1]_26V01Recibidor2Temp_TT2120_AI.Valor        0.000097
[CONCENTRACION1]_36V01Recibidor3Temp_TT3120_AI.Valor        0.000073
[CONCENTRACION1]_44V01TAPTemp_TT4401_AI.Valor               0.000291
[CONCENTRACION1]_11V1Recrist1Presion_PT1105_AI.Valor        0.000267
[CONCENTRACION1]_21V1Recrist2Presion_PT2105_AI.Valor        0.000437
[CONCENTRACION1]_31V1Recrist3Presion_PT3105_AI.Valor        0.000000
dtype: float64
```

[7]:
```
db2.isnull().mean()*100
```

[7]:
```
Time                                                  0.000000
[CONCENTRACION1]_16V01Recibidor1Presion_PT1108_AI.Valor    0.053001
[CONCENTRACION1]_26V01Recibidor2Presion_PT2108_AI.Valor    0.057610
[CONCENTRACION1]_36V01Recibidor3Presion_PT3108_AI.Valor    0.011522
[CONCENTRACION1]_12EM1Cristalizador1Corriente_AI.Valor     0.002304
[CONCENTRACION1]_12EM2Cristalizador2Corriente_AI.Valor     0.004609
[CONCENTRACION1]_12EM3Cristalizador3Corriente_AI.Valor     0.048393
[CONCENTRACION1]_22EM1Cristalizador4Corriente_AI.Valor     0.034566
[CONCENTRACION1]_22EM2Cristalizador5Corriente_AI.Valor     0.039175
[CONCENTRACION1]_22EM3Cristalizador6Corriente_AI.Valor     0.029957
[CONCENTRACION1]_32EM1Cristalizador7Corriente_AI.Valor     0.016131
[CONCENTRACION1]_32EM2Cristalizador8Corriente_AI.Valor     0.004609
[CONCENTRACION1]_13V01Colum1Conduct_CT1319_AI.Valor        0.023044
[CONCENTRACION1]_23V01Colum2Conduct_CT2319_AI.Valor        0.000000
[CONCENTRACION1]_33V01Colum3Conduct_CT3319_AI.Valor        0.006913
dtype: float64
```

1.1.2 Null values

After processing the null values of each column in each dataset, you can see that the average value is less than 0.1%.

[8]:
```
```

```
db1.columns =['Time', 'Corriente_recristalizador1',␣
 ↪'Corriente_recristalizador2', 'Corriente_recristalizador3',␣
 ↪'Temperatura_recristalizador1', 'Temperatura_recristalizador2',␣
 ↪'Temperatura_recristalizador3','Temperatura_recibidor1',␣
 ↪'Temperatura_recibidor2', 'Temperatura_recibidor3',␣
 ↪'Temperatura_Tap','Presión_recristalizador1', 'Presión_recristalizador2',␣
 ↪'Presión_recristalizador3']
db1.columns
```

[8]:
```
Index(['Time', 'Corriente_recristalizador1', 'Corriente_recristalizador2',
       'Corriente_recristalizador3', 'Temperatura_recristalizador1',
       'Temperatura_recristalizador2', 'Temperatura_recristalizador3',
       'Temperatura_recibidor1', 'Temperatura_recibidor2',
       'Temperatura_recibidor3', 'Temperatura_Tap', 'Presión_recristalizador1',
       'Presión_recristalizador2', 'Presión_recristalizador3'],
      dtype='object')
```

[9]:
```
db2.columns = ['Time', 'Presión_recibidor1', 'Presión_recibidor2',
               'Presión_recibidor3', 'Corriente_cristalizador1',
               'Corriente_cristalizador2', 'Corriente_cristalizador3',
               'Corriente_cristalizador4', 'Corriente_cristalizador5',
               'Corriente_cristalizador6', 'Corriente_cristalizador7',
               'Corriente_cristalizador8', 'Conductividad_columna1',
               'Conductividad_columna2', 'Conductividad_columna3']
db2.columns
```

[9]:
```
Index(['Time', 'Presión_recibidor1', 'Presión_recibidor2',
       'Presión_recibidor3', 'Corriente_cristalizador1',
       'Corriente_cristalizador2', 'Corriente_cristalizador3',
       'Corriente_cristalizador4', 'Corriente_cristalizador5',
       'Corriente_cristalizador6', 'Corriente_cristalizador7',
       'Corriente_cristalizador8', 'Conductividad_columna1',
       'Conductividad_columna2', 'Conductividad_columna3'],
      dtype='object')
```

Removing rows with null values At the suggestion of the Buencafe freeze-dried coffee team, we proceeded to eliminate the rows with null values because it is considered that the information on the operation of the C1 is reduced and can lead to erroneous conclusions. Additionally, they tell us that at the beginning of the database the machine was in the process of recalibration and therefore, not all the required data is stored.

[10]:
```
db1.Temperatura_recristalizador3[9:].isnull().sum()
```

[10]:
```
0
```

[11]:
```
db1.Temperatura_recristalizador3.isnull()[db1.Temperatura_recristalizador3.
 ↪isnull()]
```

[11]:
```
0    True
1    True
2    True
3    True
4    True
5    True
6    True
7    True
8    True
Name: Temperatura_recristalizador3, dtype: bool
```

[12]:
```
db1 = db1.dropna()
db2 = db2.dropna()
```

Change the variable type After adjusting the datasets db1 and db2 we proceed to identify the variables and their data type, from this it is evident that they are stored as objects which are composed of numerical and dates.

All objects are cast to be handled as number and date type in order to clean the data and obtain a better analysis of the information.

[13]:
```
print(db1.info())
print(db2.info())
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 41169 entries, 18 to 41186
Data columns (total 14 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Time                            41169 non-null  object
 1   Corriente_recristalizador1      41169 non-null  object
 2   Corriente_recristalizador2      41169 non-null  object
 3   Corriente_recristalizador3      41169 non-null  object
 4   Temperatura_recristalizador1    41169 non-null  object
 5   Temperatura_recristalizador2    41169 non-null  object
 6   Temperatura_recristalizador3    41169 non-null  object
 7   Temperatura_recibidor1          41169 non-null  object
 8   Temperatura_recibidor2          41169 non-null  object
 9   Temperatura_recibidor3          41169 non-null  object
 10  Temperatura_Tap                 41169 non-null  object
 11  Presión_recristalizador1        41169 non-null  object
 12  Presión_recristalizador2        41169 non-null  object
 13  Presión_recristalizador3        41169 non-null  object
dtypes: object(14)
memory usage: 4.7+ MB
None
<class 'pandas.core.frame.DataFrame'>
Int64Index: 43370 entries, 25 to 43394
```

```
Data columns (total 15 columns):
 #   Column                    Non-Null Count  Dtype 
---  ------                    --------------  ----- 
 0   Time                      43370 non-null  object
 1   Presión_recibidor1        43370 non-null  object
 2   Presión_recibidor2        43370 non-null  object
 3   Presión_recibidor3        43370 non-null  object
 4   Corriente_cristalizador1  43370 non-null  object
 5   Corriente_cristalizador2  43370 non-null  object
 6   Corriente_cristalizador3  43370 non-null  object
 7   Corriente_cristalizador4  43370 non-null  object
 8   Corriente_cristalizador5  43370 non-null  object
 9   Corriente_cristalizador6  43370 non-null  object
 10  Corriente_cristalizador7  43370 non-null  object
 11  Corriente_cristalizador8  43370 non-null  object
 12  Conductividad_columna1    43370 non-null  object
 13  Conductividad_columna2    43370 non-null  object
 14  Conductividad_columna3    43370 non-null  object
dtypes: object(15)
memory usage: 5.3+ MB
None
```

1.2 Variables Date / doubles

In the records of the time column of db1, there is an additional hour after the initial date. Therefore, it was decided to separate this information into two variables, in order to identify with the BuenCafe team if the second date has any relevance.

[14]:
```
db1.columns[1:14]
for i in db1.columns[1:14]:
  db1[i]=db1[i].astype("string").str.replace(",", ".").astype(float)

db1['Time2'] = db1['Time'].astype("string").str.split(',',expand=True)[0]
db1['Hour_unknow'] = db1['Time'].astype("string").str.split(',',expand=True)[1]
db1['Time2'] = pd.to_datetime(db1['Time2'])

db1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 41169 entries, 18 to 41186
Data columns (total 16 columns):
 #   Column                        Non-Null Count  Dtype  
---  ------                        --------------  -----  
 0   Time                          41169 non-null  object 
 1   Corriente_recristalizador1    41169 non-null  float64
 2   Corriente_recristalizador2    41169 non-null  float64
 3   Corriente_recristalizador3    41169 non-null  float64
 4   Temperatura_recristalizador1  41169 non-null  float64
```

```
 5   Temperatura_recristalizador2  41169 non-null  float64
 6   Temperatura_recristalizador3  41169 non-null  float64
 7   Temperatura_recibidor1        41169 non-null  float64
 8   Temperatura_recibidor2        41169 non-null  float64
 9   Temperatura_recibidor3        41169 non-null  float64
 10  Temperatura_Tap               41169 non-null  float64
 11  Presión_recristalizador1      41169 non-null  float64
 12  Presión_recristalizador2      41169 non-null  float64
 13  Presión_recristalizador3      41169 non-null  float64
 14  Time2                         41169 non-null  datetime64[ns]
 15  Hour_unknow                   41169 non-null  string
dtypes: datetime64[ns](1), float64(13), object(1), string(1)
memory usage: 5.3+ MB
```

1.3 Variables Date / doubles

The same adjustment is made to the dates for db2.

[15]:
```
db2.columns[1:]
for i in db2.columns[1:]:
  db2[i]=db2[i].astype("string").str.replace(",", ".").astype(float)

db2['Time2'] = db2['Time'].astype("string").str.split(',',expand=True)[0]
db2['Hour_unknow'] = db2['Time'].astype("string").str.split(',',expand=True)[1]
db2['Time2'] = pd.to_datetime(db2['Time2'])

db2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 43370 entries, 25 to 43394
Data columns (total 17 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Time                      43370 non-null  object
 1   Presión_recibidor1        43370 non-null  float64
 2   Presión_recibidor2        43370 non-null  float64
 3   Presión_recibidor3        43370 non-null  float64
 4   Corriente_cristalizador1  43370 non-null  float64
 5   Corriente_cristalizador2  43370 non-null  float64
 6   Corriente_cristalizador3  43370 non-null  float64
 7   Corriente_cristalizador4  43370 non-null  float64
 8   Corriente_cristalizador5  43370 non-null  float64
 9   Corriente_cristalizador6  43370 non-null  float64
 10  Corriente_cristalizador7  43370 non-null  float64
 11  Corriente_cristalizador8  43370 non-null  float64
 12  Conductividad_columna1    43370 non-null  float64
 13  Conductividad_columna2    43370 non-null  float64
 14  Conductividad_columna3    43370 non-null  float64
```

```
 15  Time2                    43370 non-null  datetime64[ns]
 16  Hour_unknow              43370 non-null  string
dtypes: datetime64[ns](1), float64(14), object(1), string(1)
memory usage: 6.0+ MB
```

We take the hours of each dataset. It may be useful to know the hours in which energy is best used in C1, so we adjust the format in each dataset to later make comparisons in time ranges.

[16]:
```
db1["Hour"] = pd.to_datetime(db1['Time2'], format = "%H:%M").dt.hour

db1.head(5)

db2["Hour"] = pd.to_datetime(db2['Time2'], format = "%H:%M").dt.hour
#hourly_accidents = df.groupby("HOUR")["TIME"].size()
```

1.3.1 DB1 graphs

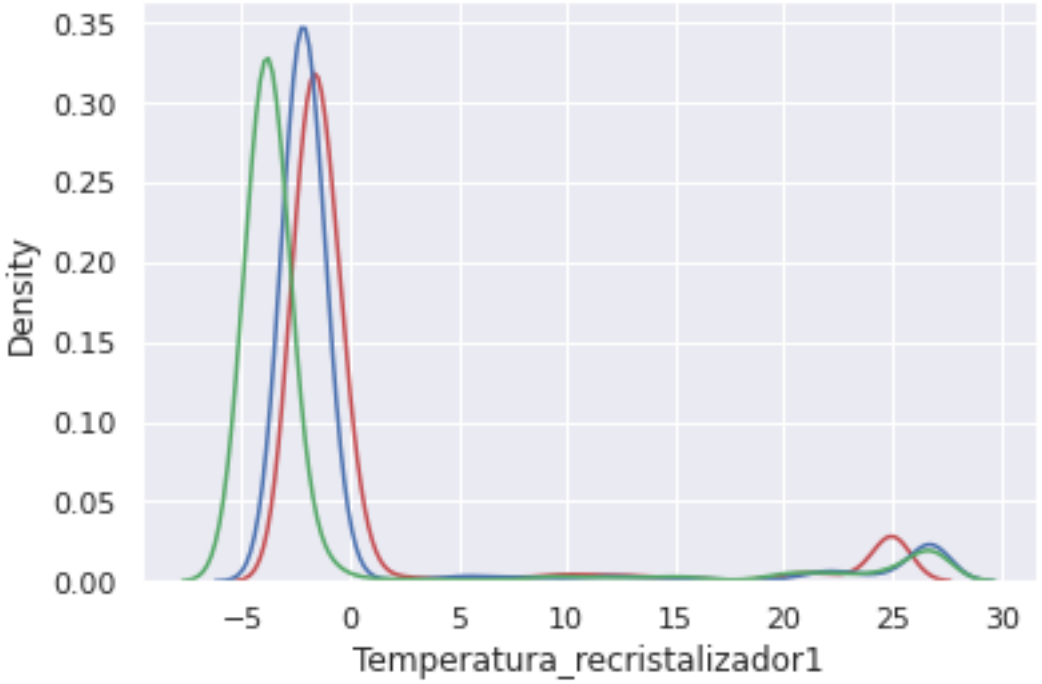
Taking the previous information, it is necessary to plot dataset 1, in order to contrast the data. Since they are variables with elements in common, important differences in density can be appreciated, which can lead us to identify improvement alternatives as points to analyze.

[17]:
```
sns.set(style="darkgrid")
fig = sns.kdeplot(db1.Corriente_recristalizador1, color="r")
fig = sns.kdeplot(db1.Corriente_recristalizador2, color="b")
fig = sns.kdeplot(db1.Corriente_recristalizador3, color="g")
plt.show()
```

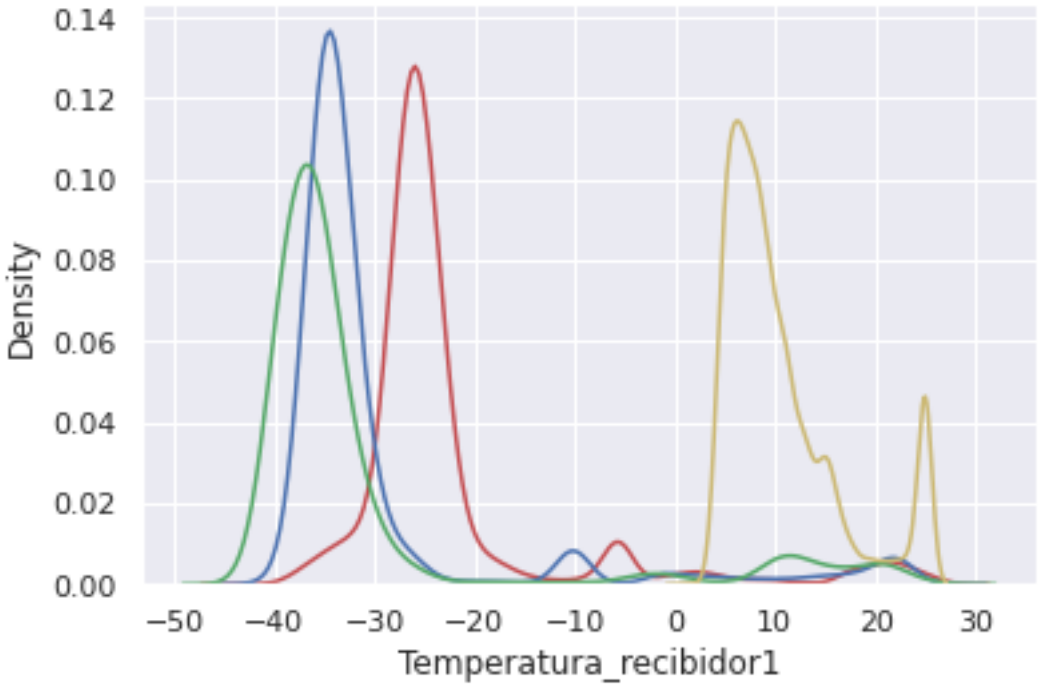
1.4 Graphical Analysis

Although most of the data are between 20 and 45 values corresponding to the correinte_recristalizador1, there are also high values in density with a current equal to 0.

[18]:
```
sns.set(style="darkgrid")
fig = sns.kdeplot(db1.Temperatura_recristalizador1, color="r")
fig = sns.kdeplot(db1.Temperatura_recristalizador2, color="b")
fig = sns.kdeplot(db1.Temperatura_recristalizador3, color="g")
plt.show()
```



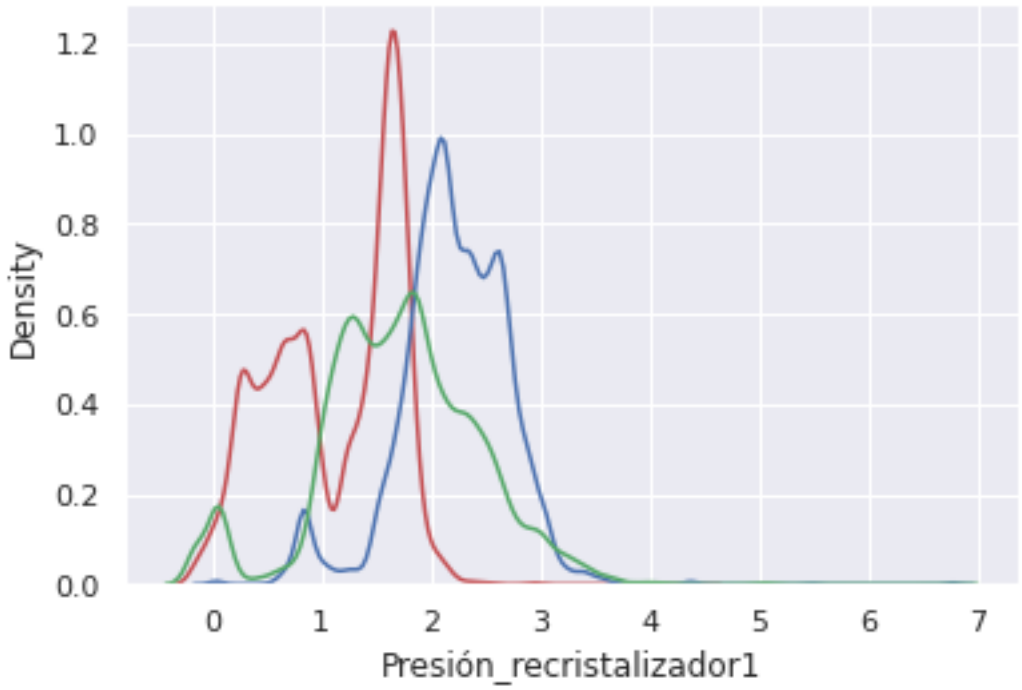
[19]:
```
sns.set(style="darkgrid")
fig = sns.kdeplot(db1.Temperatura_recibidor1, color="r")
fig = sns.kdeplot(db1.Temperatura_recibidor2, color="b")
fig = sns.kdeplot(db1.Temperatura_recibidor3, color="g")
fig = sns.kdeplot(db1.Temperatura_Tap, color="y")
plt.show()
```



1.5 Temperature Analysis

Most of the temperatures identified with the records of dataset1 are negative values. dataset1 records are negative values, this helps us to reaffirm that during the process the ice crystals are maintained in order to separate the coffee extract. The ice crystals are maintained in order to separate the coffee extract leaving the water in the form of ice for later extraction.

[20]:
```
sns.set(style="darkgrid")
fig = sns.kdeplot(db1.Presión_recristalizador1, color="r")
fig = sns.kdeplot(db1.Presión_recristalizador2, color="b")
fig = sns.kdeplot(db1.Presión_recristalizador3, color="g")
plt.show()
```



1.6 Pressure analysis

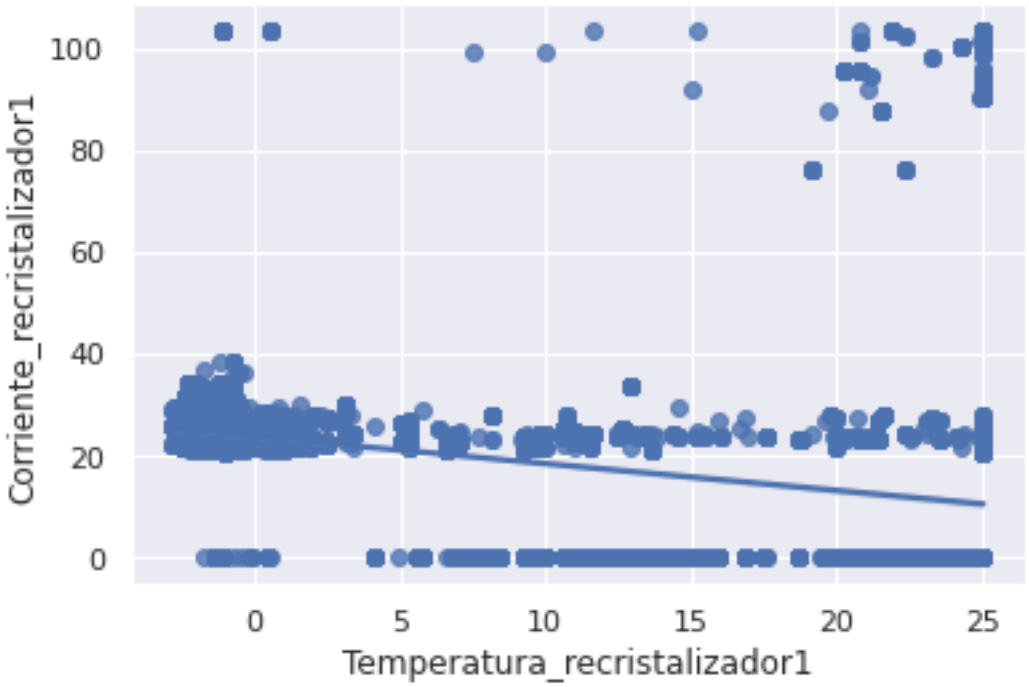
A critical point that is evident in the graph is at a value of 2 for the recrystallizer pressure axis and on the y-axis at a value of 0.6.

the highest density expressed in the graph is above the value of 2.

most of the records of recrystallizer 1 are in the lower limit of 2, while for recrystallizer 2 they are in the upper limit.

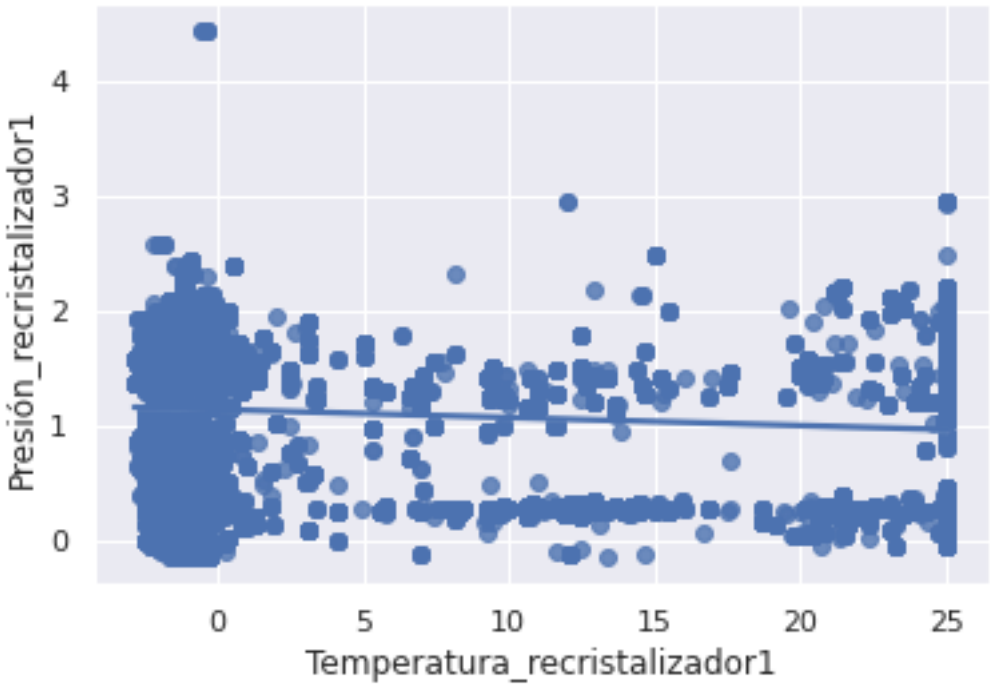
[21]:
```
sns.regplot(x=db1["Temperatura_recristalizador1"],␣
↪y=db1["Corriente_recristalizador1"])
```

[21]:
```
<AxesSubplot:xlabel='Temperatura_recristalizador1',
ylabel='Corriente_recristalizador1'>
```



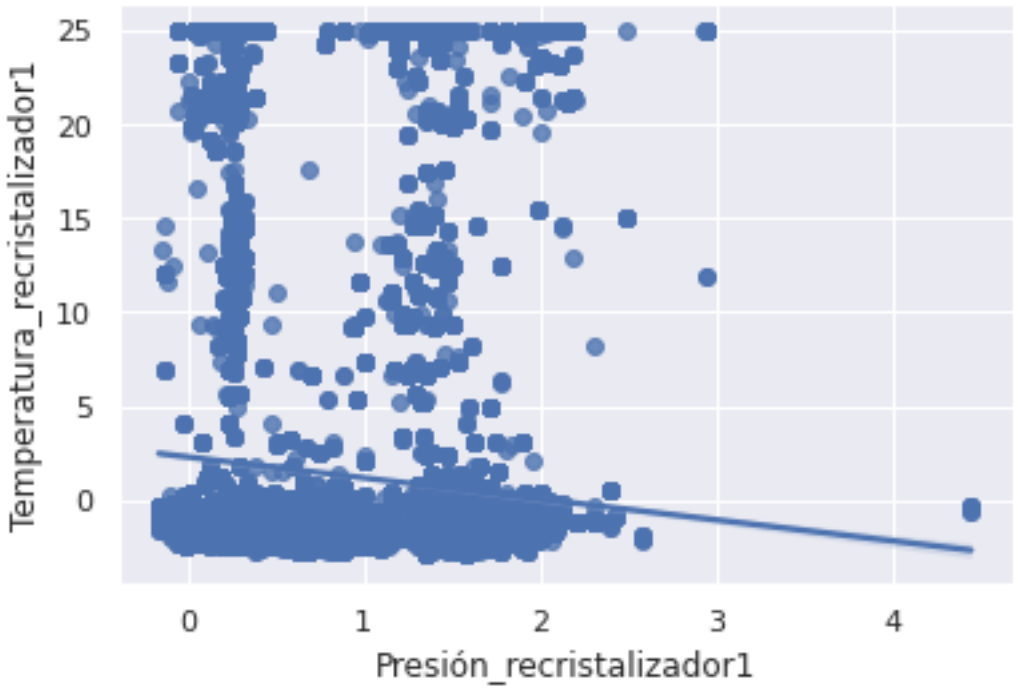
[22]:
```
sns.regplot(x=db1["Temperatura_recristalizador1"],␣
↪y=db1["Presión_recristalizador1"])
```

[22]:
```
<AxesSubplot:xlabel='Temperatura_recristalizador1',
ylabel='Presión_recristalizador1'>
```



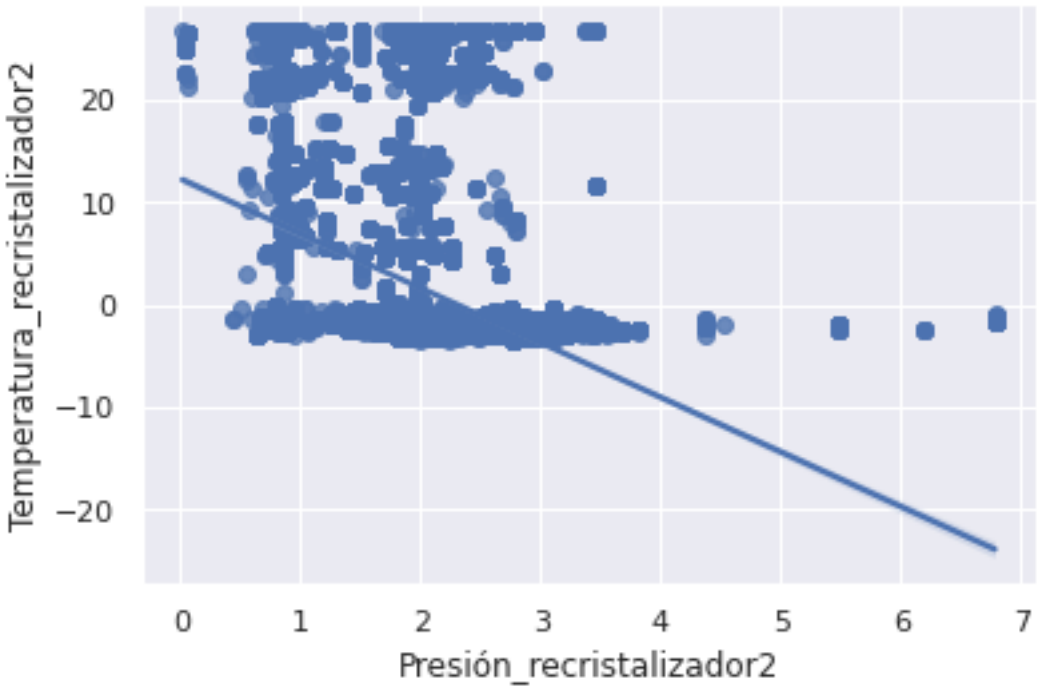
[23]:
```
sns.regplot(x=db1["Presión_recristalizador1"],␣
↪y=db1["Temperatura_recristalizador1"])
```

[23]:
```
<AxesSubplot:xlabel='Presión_recristalizador1',
ylabel='Temperatura_recristalizador1'>
```



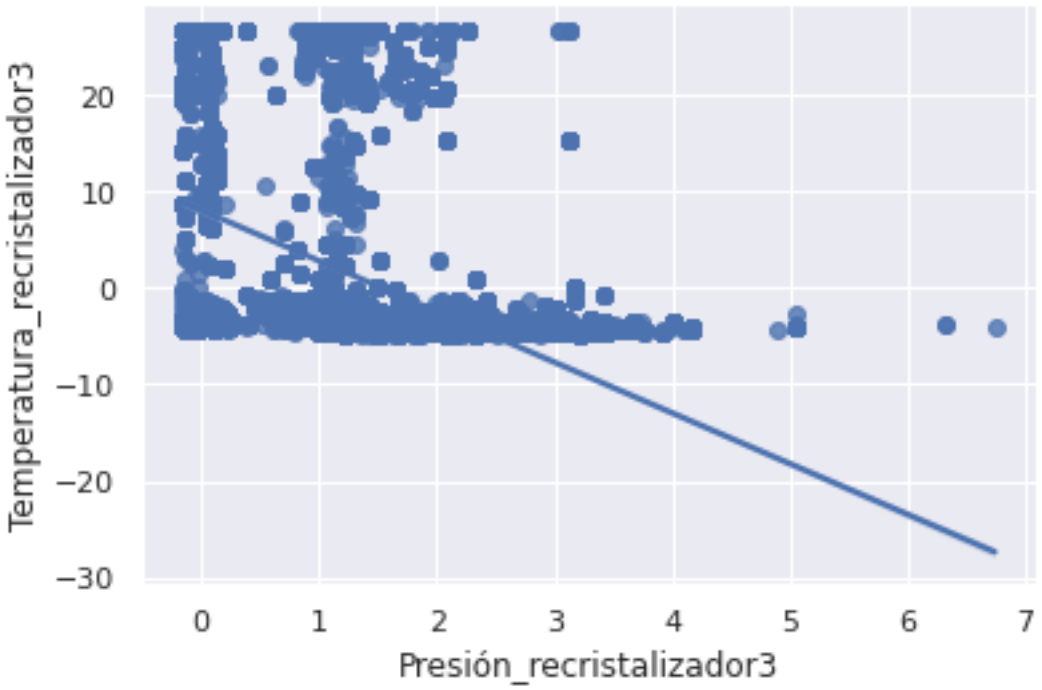
[24]:
```
sns.regplot(x=db1["Presión_recristalizador2"],␣
↪y=db1["Temperatura_recristalizador2"])
```

[24]:
```
<AxesSubplot:xlabel='Presión_recristalizador2',
ylabel='Temperatura_recristalizador2'>
```



[25]:
```
sns.regplot(x=db1["Presión_recristalizador3"],␣
↪y=db1["Temperatura_recristalizador3"])
```

[25]:
```
<AxesSubplot:xlabel='Presión_recristalizador3',
ylabel='Temperatura_recristalizador3'>
```

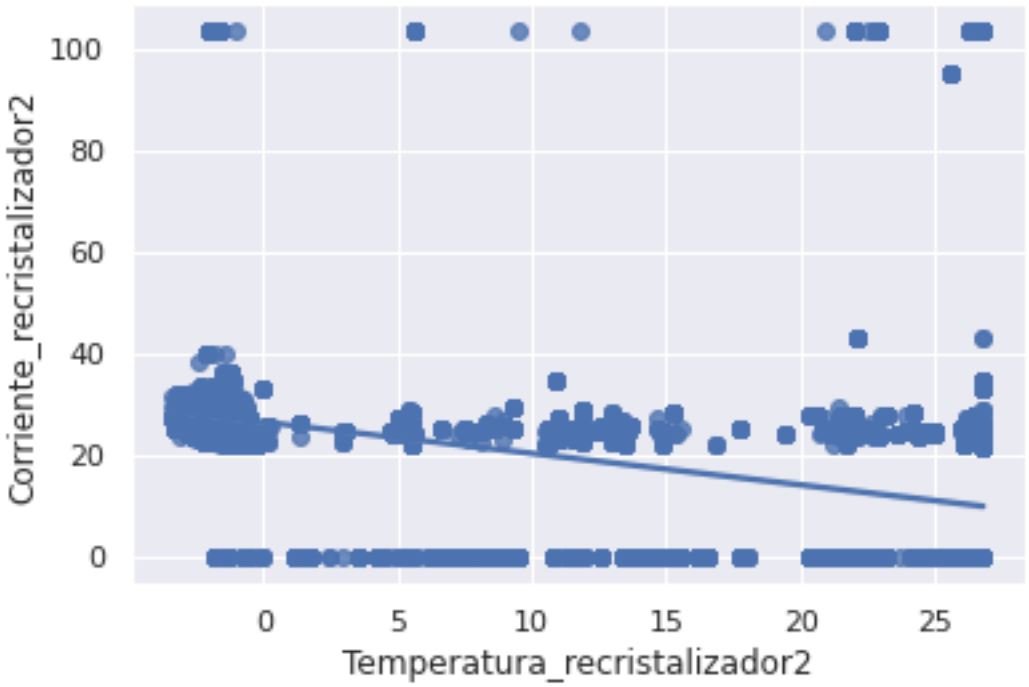


1.7 Pressure / Temperature Graph Analysis

- More linear values are evident in recrystallizer 3 as pressure increases and temperature decreases to values less than zero.

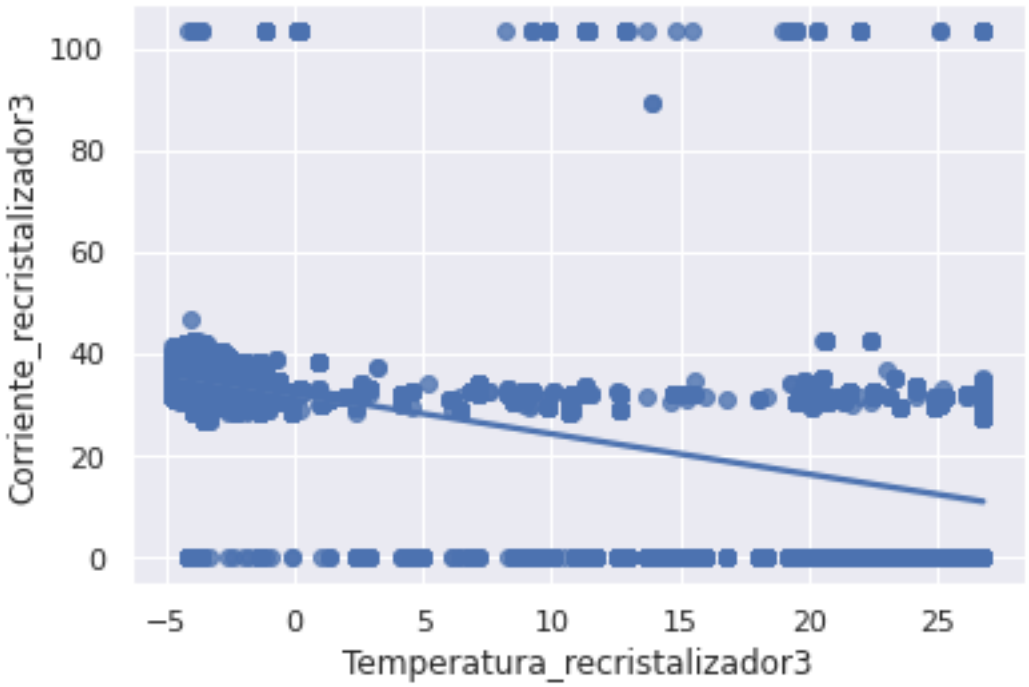
[26]:
```
sns.regplot(x=db1["Temperatura_recristalizador2"],␣
↪y=db1["Corriente_recristalizador2"])
```

[26]:
```
<AxesSubplot:xlabel='Temperatura_recristalizador2',
ylabel='Corriente_recristalizador2'>
```



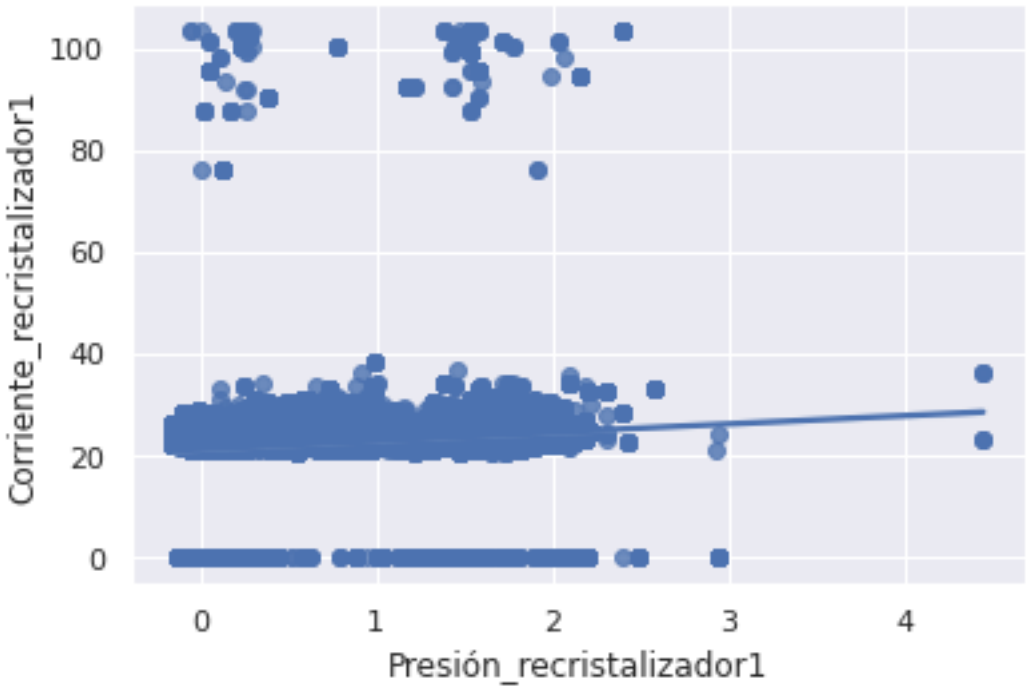
[27]:
```
sns.regplot(x=db1["Temperatura_recristalizador3"],␣
↪y=db1["Corriente_recristalizador3"])
```

[27]:
```
<AxesSubplot:xlabel='Temperatura_recristalizador3',
ylabel='Corriente_recristalizador3'>
```



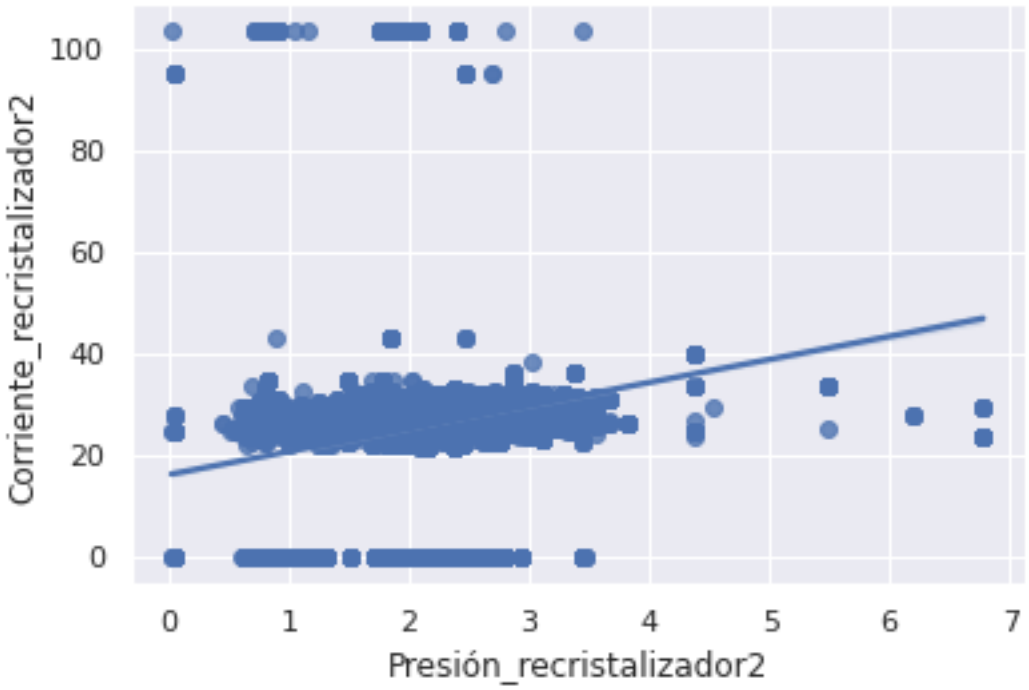
[28]:
```
sns.regplot(x=db1["Presión_recristalizador1"],␣
↪y=db1["Corriente_recristalizador1"])
```

[28]:
```
<AxesSubplot:xlabel='Presión_recristalizador1',
ylabel='Corriente_recristalizador1'>
```



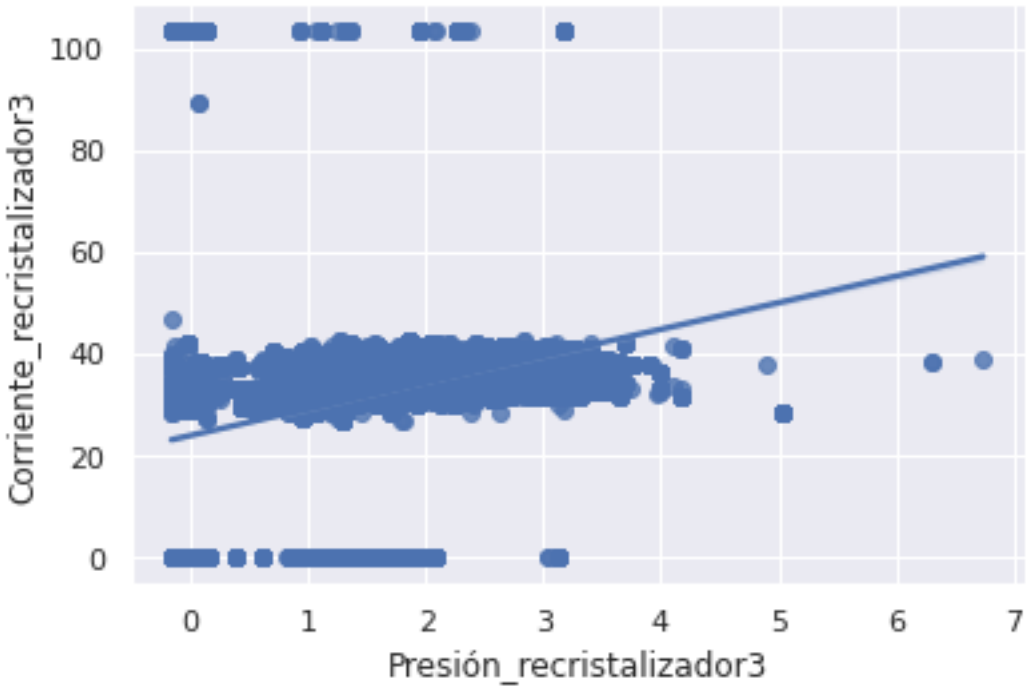
[29]:
```
sns.regplot(x=db1["Presión_recristalizador2"],␣
↪y=db1["Corriente_recristalizador2"])
```

[29]:
```
<AxesSubplot:xlabel='Presión_recristalizador2',
ylabel='Corriente_recristalizador2'>
```



[30]:
```
sns.regplot(x=db1["Presión_recristalizador3"],␣
↪y=db1["Corriente_recristalizador3"])
```

[30]:
```
<AxesSubplot:xlabel='Presión_recristalizador3',
ylabel='Corriente_recristalizador3'>
```



2 DB2 Graphs

2.0.1 DB2 graphs

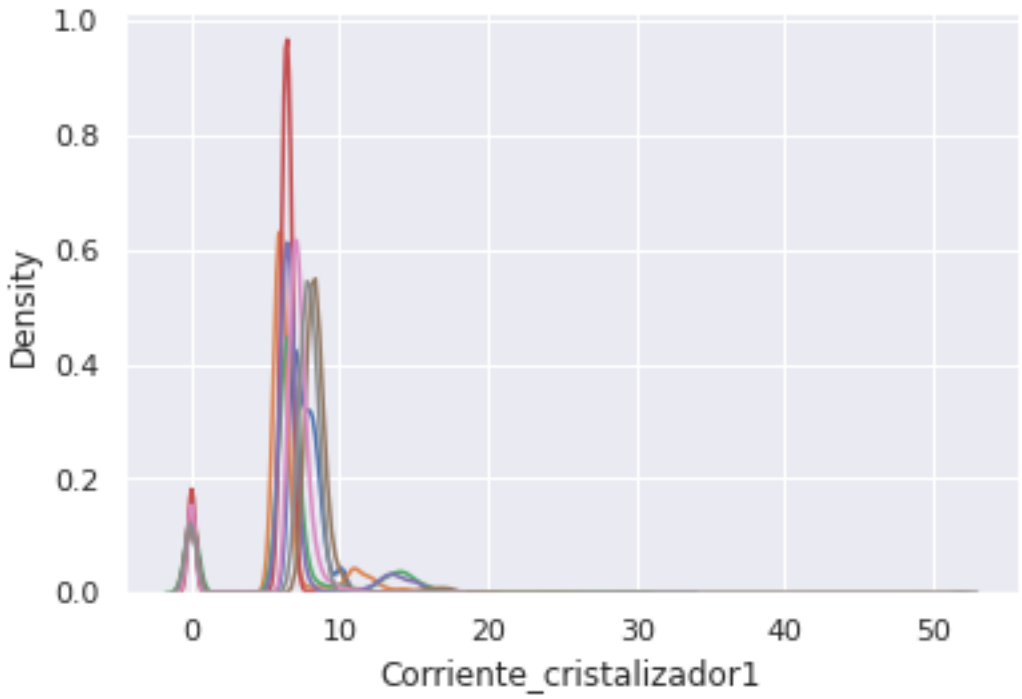
Taking the previous information, it is necessary to plot dataset 1, in order to contrast the data. Since they are variables with elements in common, important differences in density can be appreciated, which can lead us to identify improvement alternatives as points to analyze.

[31]:
```
sns.set(style="darkgrid")
fig = sns.kdeplot(db2.Presión_recibidor1, color="r")
#fig = sns.kdeplot(db2.Presión_recibidor2, color="b")
fig = sns.kdeplot(db2.Presión_recibidor3, color="g")
plt.show()
```

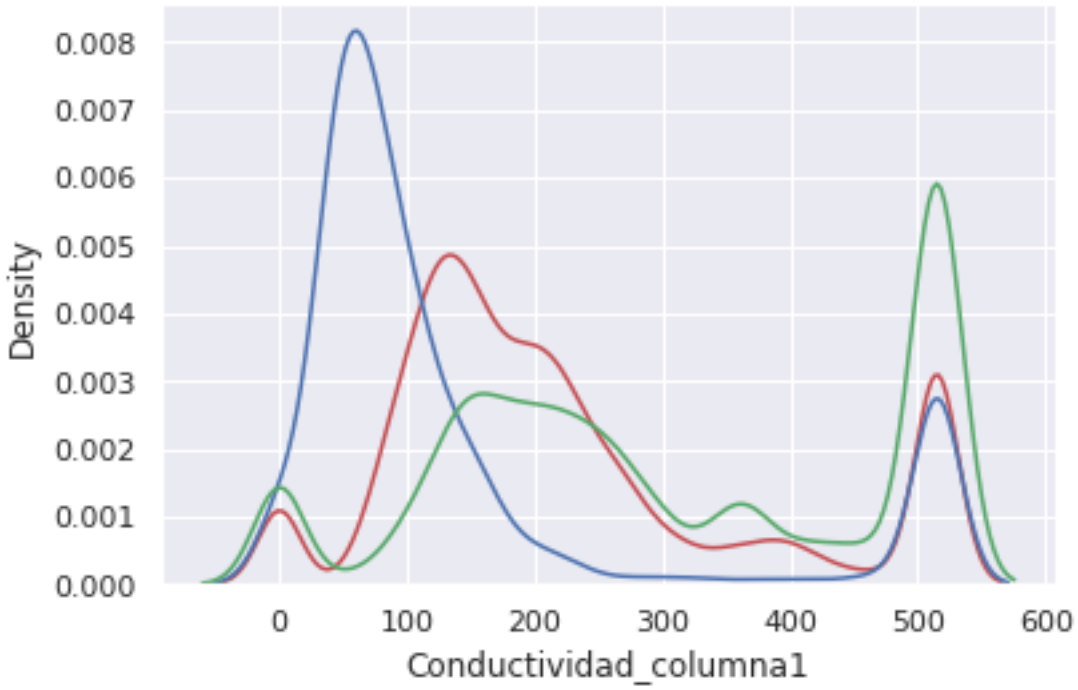
[32]:
```
db2.Presión_recibidor2.unique()
```

[32]:
```
array([-3.21])
```

[33]:
```
sns.set(style="darkgrid")
fig = sns.kdeplot(db2.Corriente_cristalizador1)
fig = sns.kdeplot(db2.Corriente_cristalizador2)
fig = sns.kdeplot(db2.Corriente_cristalizador3)
fig = sns.kdeplot(db2.Corriente_cristalizador4)
fig = sns.kdeplot(db2.Corriente_cristalizador5)
fig = sns.kdeplot(db2.Corriente_cristalizador6)
fig = sns.kdeplot(db2.Corriente_cristalizador7)
fig = sns.kdeplot(db2.Corriente_cristalizador8)
current_palette = sns.color_palette()
#sns.palplot(current_palette)
plt.show(current_palette)
```



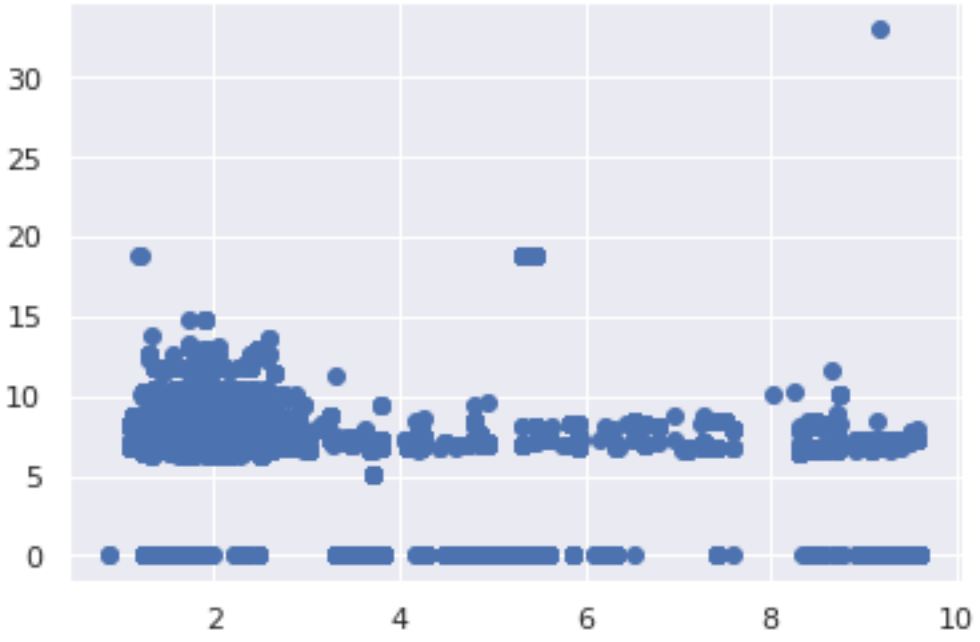
[34]:
```
sns.set(style="darkgrid")
fig = sns.kdeplot(db2.Conductividad_columna1, color="r")
fig = sns.kdeplot(db2.Conductividad_columna2, color="b")
fig = sns.kdeplot(db2.Conductividad_columna3, color="g")
plt.show()
```



2.1 Pressure / Stream graph analysis

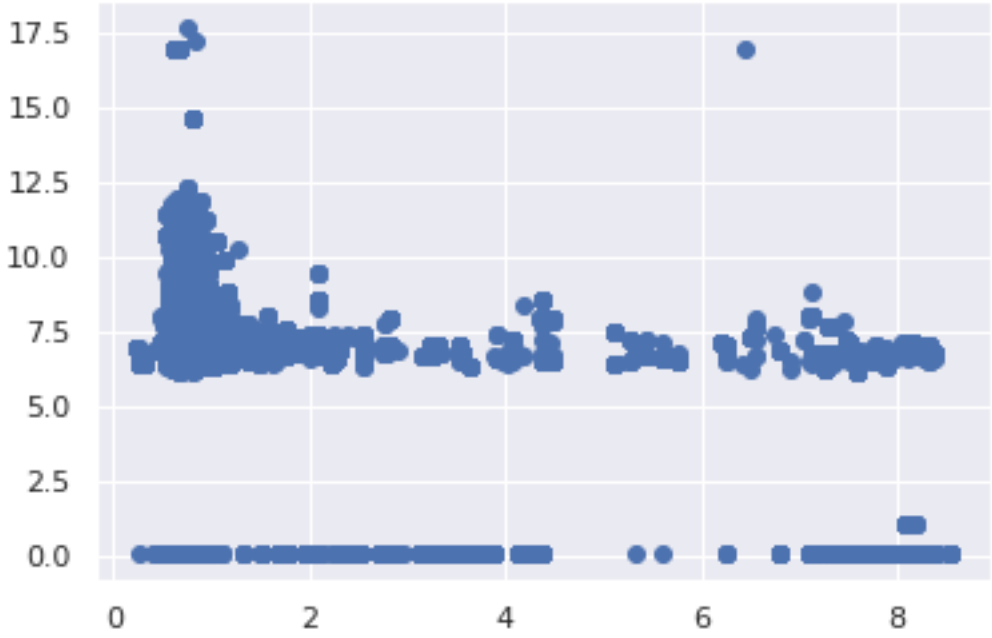
[35]:
```
plt.scatter(db2.Presión_recibidor1, db2.Corriente_cristalizador1 )
```

[35]:
```
<matplotlib.collections.PathCollection at 0x7f17f7266a60>
```



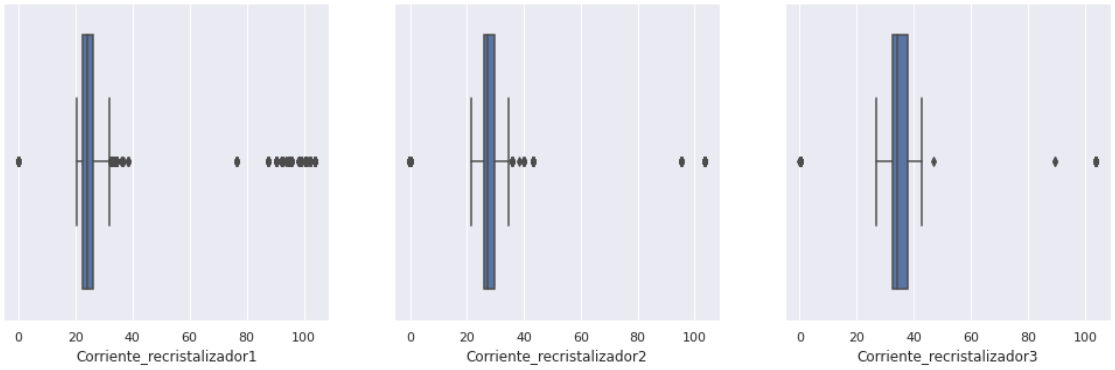
[36]:
```
plt.scatter(db2.Presión_recibidor3, db2.Corriente_cristalizador7 )
```

[36]:
```
<matplotlib.collections.PathCollection at 0x7f17f65a5190>
```



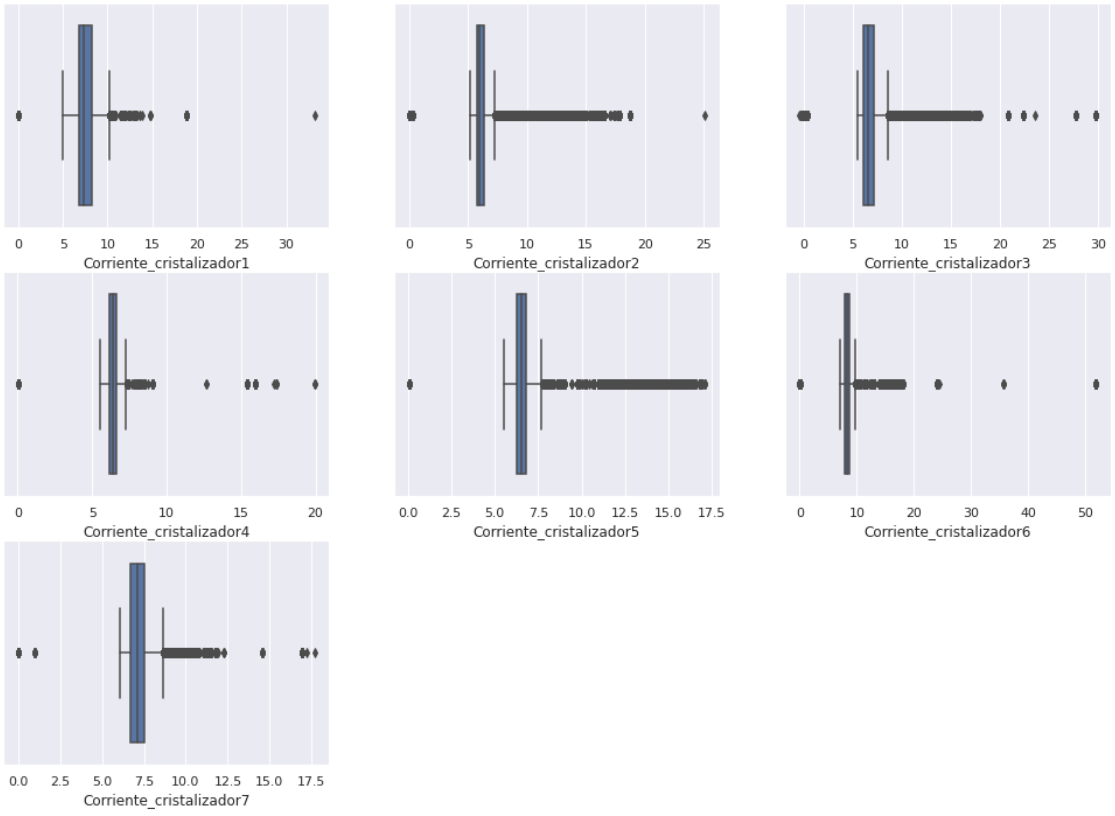
2.2 Graficas boxplot

[37]:
```
plt.figure(figsize=(17,5))
plt.subplot(1,3,1)
sns.boxplot(x=db1.Corriente_recristalizador1)
plt.subplot(1,3,2)
sns.boxplot(x=db1.Corriente_recristalizador2)
plt.subplot(1,3,3)
sns.boxplot(x=db1.Corriente_recristalizador3);
```



For all recrystallizers, 50% of the data are between 20Amp and 40Amp. The first recrystallizer has several outliers, unlike 2 and 3. The current range of recrystallizer 3 varies with respect to 1 and 2 (30Amp - 40Amp), why is this? Does the latter consume more power?

[38]:
```
plt.figure(figsize=(17,12))
plt.subplot(3,3,1)
sns.boxplot(x=db2.Corriente_cristalizador1)
plt.subplot(3,3,2)
sns.boxplot(x=db2.Corriente_cristalizador2)
plt.subplot(3,3,3)
sns.boxplot(x=db2.Corriente_cristalizador3)
plt.subplot(3,3,4)
sns.boxplot(x=db2.Corriente_cristalizador4)
plt.subplot(3,3,5)
sns.boxplot(x=db2.Corriente_cristalizador5)
plt.subplot(3,3,6)
sns.boxplot(x=db2.Corriente_cristalizador6)
plt.subplot(3,3,7)
sns.boxplot(x=db2.Corriente_cristalizador7);
```



For all crystallizers 50% of the data are between 5Amp and 10 Amp. For crystallizers 2, 3 and 5 the outliers are more. The current of crystallizer 6 reaches higher values than any other crystallizer, what happened at that time? Why did it show a higher power? Crystallizers 4,5 and 7 are the ones that consume less energy.

[39]:
```
# without regression
# sns.pairplot(db1, kind="scatter")
# plt.show()
```

3 How are we going to evaluate the efficiency of the machine?

Could it be with the current you use for the production process?