## Part 2 Analysis:

### Listing 4:

Pairwise Pearson Correlation:

```
print(dataset.corr(method = 'pearson'))
```

```
             Ri     Na     Mg     Al     Si     K      Ca         Ba     Fe
Ri  1.000e+00 -0.192 -0.122 -0.407 -0.542 -0.290  0.810 -3.860e-04  0.143
Na -1.919e-01  1.000 -0.274  0.157 -0.070 -0.266 -0.275  3.266e-01 -0.241
Mg -1.223e-01 -0.274  1.000 -0.482 -0.166  0.005 -0.444 -4.923e-01  0.083
Al -4.073e-01  0.157 -0.482  1.000 -0.006  0.326 -0.260  4.794e-01 -0.074
Si -5.421e-01 -0.070 -0.166 -0.006  1.000 -0.193 -0.209 -1.022e-01 -0.094
K  -2.898e-01 -0.266  0.005  0.326 -0.193  1.000 -0.318 -4.262e-02 -0.008
Ca  8.104e-01 -0.275 -0.444 -0.260 -0.209 -0.318  1.000 -1.128e-01  0.125
Ba -3.860e-04  0.327 -0.492  0.479 -0.102 -0.043 -0.113  1.000e+00 -0.059
Fe  1.430e-01 -0.241  0.083 -0.074 -0.094 -0.008  0.125 -5.869e-02  1.000
```

Skew for Each Attribute:

```
print(dataset.skew())
```

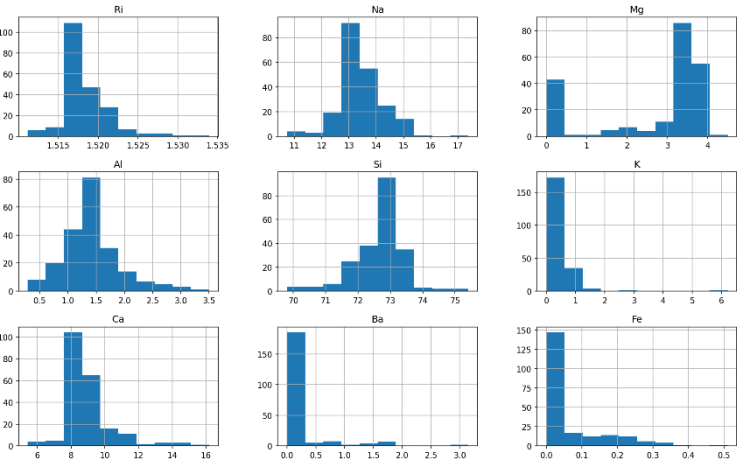
```
Ri     1.625
Na     0.454
Mg    -1.153
Al     0.907
Si    -0.730
K      6.552
Ca     2.047
Ba     3.416
Fe     1.754
```

Histograms and Density Plots:

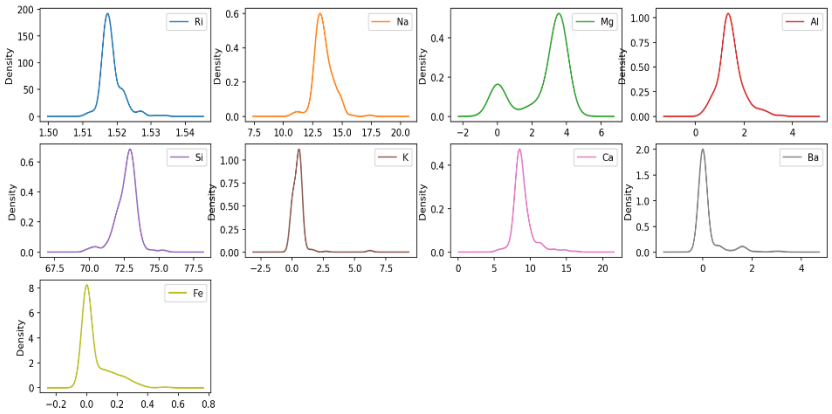
```
print(dataset.hist())
pyplot.figsize = (8,8)
pyplot.savefig('histograms.png', dpi=300)
pyplot.show()
```

```
dataset.plot(kind='density' , subplots=True, layout=(4,4), sharex=False, figsize=(14,14))
pyplot.show()
```

Histograms:

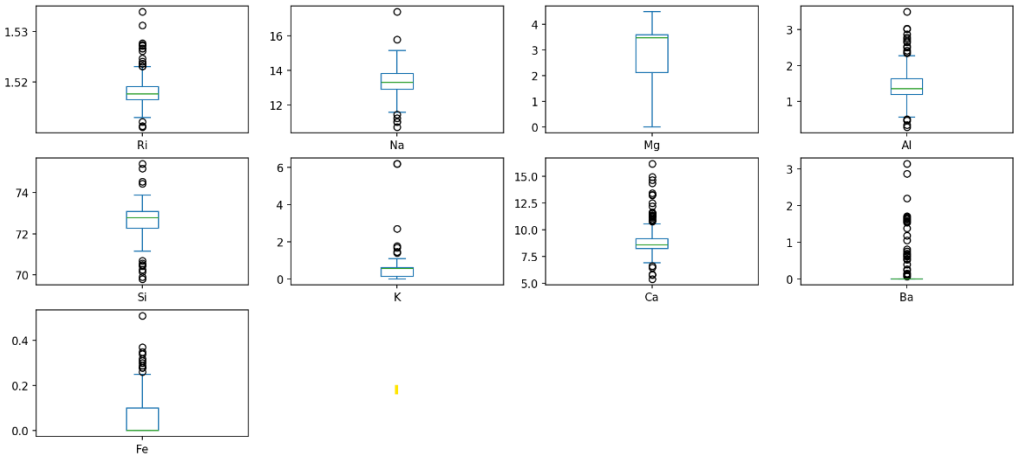



Density:




Box and Whisker Plots

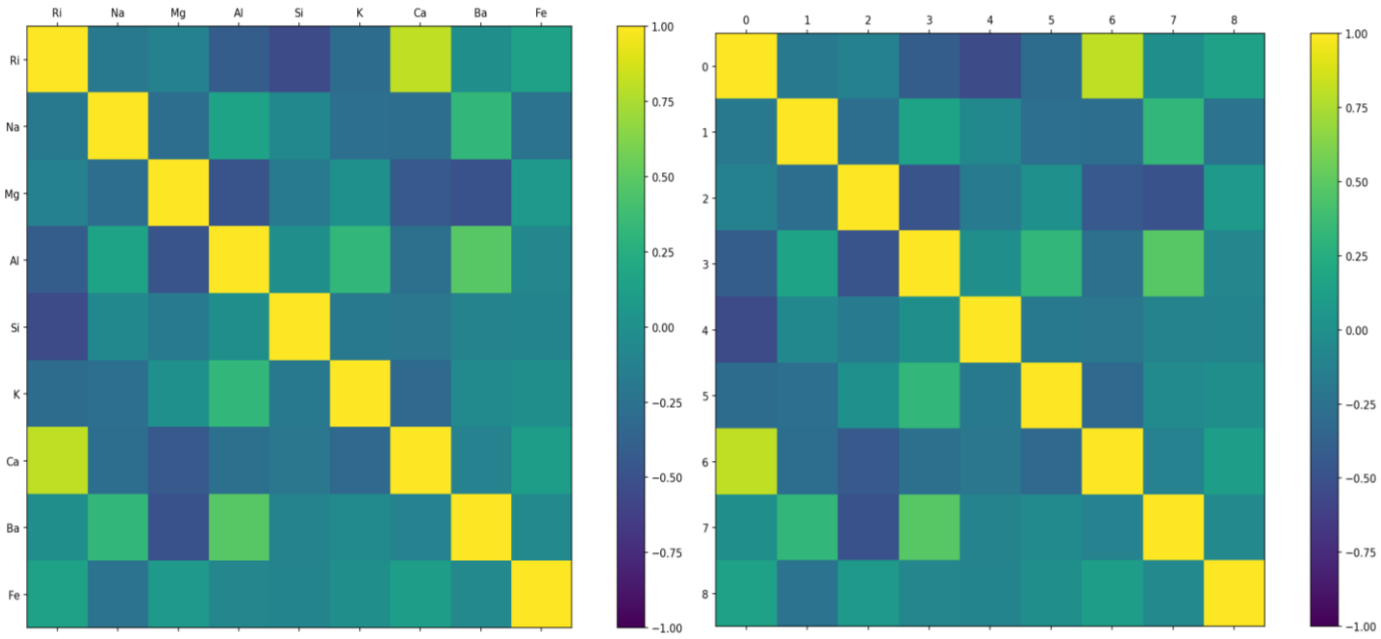
```
dataset.plot(kind='box', subplots=True, layout=(4,4), sharex=False, figsize=(14,14))
pyplot.show()
```



Correlation Matrix Plot & Correlation Matrix Generic Plot

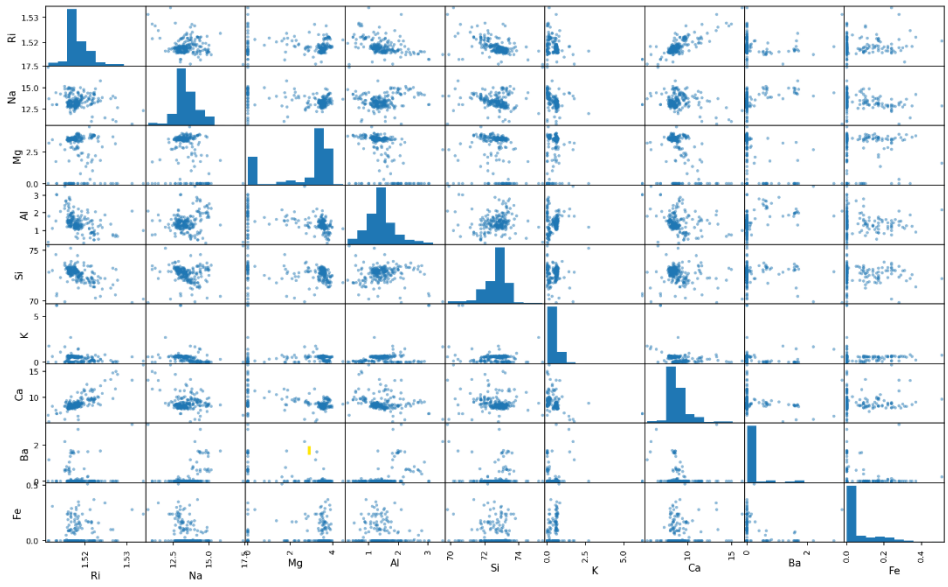
```
fig = pyplot.figure(figsize=(10,8))
ax = fig.add_subplot(111)
cax = ax.matshow(dataset.corr(), vmin=-1, vmax=1)
fig.colorbar(cax)
ticks = numpy.arange(0, 9, 1)
ax.set_xticks(ticks)
ax.set_yticks(ticks)
ax.set_xticklabels(columns_after_drop)
ax.set_yticklabels(columns_after_drop)
pyplot.show()
```

```
fig = pyplot.figure()
ax = fig.add_subplot(111)
cax = ax.matshow(dataset.corr(), vmin=-1, vmax=1)
fig.colorbar(cax)
pyplot.show()
```




Scatter Matrix

```
scatter_matrix(dataset)
pyplot.figure(figsize=(20,18))
pyplot.show()
```



Listing 4 analysis:

Our analysis of the chemical properties of the glass samples revealed several important patterns that provide deeper insight into the relationships between different chemical elements. Starting from the Pearson correlation analysis, we were able to demonstrate that Calcium (Ca) and Rhodium (Rh) exhibit a strong positive correlation, which suggests that higher levels of Calcium tend to be associated with increased Rhodium presence in the glass. This finding is important because it hints at a potential interaction between these elements in glass production, possibly influencing the overall strength or optical properties of the material. On the other hand, Aluminium (Al) showed a negative correlation with Rhodium, indicating that higher concentrations of Aluminium generally tend to decrease the amount of Rhodium.

Similarly, Potassium (K) and Sodium (Na) also demonstrated negative correlations with Rhodium. Elements such as Barium (Ba) and Iron (Fe) had minimal correlations with Rhodium, suggesting that they have little to no influence on its presence, making them less relevant in determining Rhodium levels in these glass samples.

When examining the skewness of each chemical attribute, Potassium (K) was highly skewed with a value of 6.552, indicating that a small number of glass samples contained disproportionately high levels of Potassium. This skewness could indicate that certain glass types are enriched with Potassium, potentially for specific functional purposes, such as thermal resistance and durability. Barium (Ba) and Calcium (Ca) also displayed positive skewness, meaning a few glass samples had considerably higher concentrations of these elements compared to the majority. This suggests the presence of specialized glass compositions, possibly engineered for industrial uses. In contrast, Sodium (Na) and Aluminium (Al) exhibited more balanced distributions, though still slightly skewed, while Magnesium (Mg) showed a negative skew, implying that most glass samples had lower Magnesium concentrations. This distribution indicates that Magnesium may not be a critical element in the majority of the glass types analyzed.

The histograms of the data confirmed these patterns of skewness. Potassium, Barium, and Calcium stood out with notably skewed distributions, suggesting the presence of unique glass formulations enriched in these elements. Sodium and Aluminium appeared closer to a normal distribution but were still slightly skewed. Silicon (Si) and Iron (Fe), on the other hand, showed signs of a bimodal distribution, indicating the possible existence of distinct subtypes of glass with differing levels of these elements. Additionally, we noticed many outliers in the data for Potassium and Barium, indicating that certain glass samples contain much higher levels of these elements, possibly as a result of customized compositions for specialized applications.

The correlation matrix provided a comprehensive view of how the chemical elements are interrelated. The strong positive correlation between Calcium (Ca) and the refractive index (Ri) stood out, suggesting that Calcium plays a significant role in influencing the optical properties of the glass. On the other hand, Silicon (Si) and the refractive index had a near-zero correlation, indicating that the primary glass-forming element does not substantially affect the refractive index in these samples.

The scatter matrix plots reinforced these observations, with clear positive trends between elements like Calcium and the refractive index, while others, such as Silicon and Iron, showed much more random relationships. Outliers were very noticeable in the scatter plots between Barium and Iron, which might represent glass samples with unique or rare compositions. The overall randomness in some scatter plots, especially between elements like Silicon and Iron, suggests that certain elements operate independently, with minimal direct influence on one another’s concentrations.

## Listing 5

Rescaling Data

```
array = dataset.values
X = array[:, 0:9]
scaler = MinMaxScaler(feature_range=(0, 1))
rescaledX = scaler.fit_transform(X)
set_printoptions(precision=3)
print(rescaledX[0:5, :])
```

```
[[0.433 0.438 1.    0.252 0.352 0.01  0.309 0.    0.   ]
 [0.284 0.475 0.802 0.333 0.521 0.077 0.223 0.    0.   ]
 [0.221 0.421 0.791 0.389 0.568 0.063 0.218 0.    0.   ]
 [0.286 0.373 0.822 0.312 0.5   0.092 0.259 0.    0.   ]
 [0.275 0.382 0.806 0.296 0.584 0.089 0.245 0.    0.   ]]
```

Standardize Data

```
X = array[:, 0:9]
scaler_standard = StandardScaler().fit(X)
rescaled_standardX = scaler_standard.transform(X)
set_printoptions(precision=3)
print(rescaled_standardX[0:5, :])
```

```
[[ 0.873  0.285  1.255 -0.692 -1.127 -0.672 -0.146 -0.353 -0.586]
 [-0.249  0.592  0.636 -0.17   0.102 -0.026 -0.794 -0.353 -0.586]
 [-0.721  0.15   0.601  0.191  0.439 -0.165 -0.829 -0.353 -0.586]
 [-0.233 -0.243  0.699 -0.311 -0.053  0.112 -0.519 -0.353 -0.586]
 [-0.312 -0.169  0.65  -0.411  0.555  0.081 -0.625 -0.353 -0.586]]
```

Normalize Data

```
X = array[:, 0:9]
scaler = Normalizer().fit(X)
normalizedX = scaler.transform(X)
set_printoptions(precision=3)
print(normalizedX[0:5, :])
```

```
[[2.062e-02 1.850e-01 6.088e-02 1.492e-02 9.733e-01 8.136e-04 1.186e-01
  0.000e+00 0.000e+00]
 [2.035e-02 1.863e-01 4.827e-02 1.824e-02 9.753e-01 6.437e-03 1.050e-01
  0.000e+00 0.000e+00]
 [2.028e-02 1.810e-01 4.749e-02 2.060e-02 9.764e-01 5.217e-03 1.041e-01
  0.000e+00 0.000e+00]
 [2.040e-02 1.776e-01 4.961e-02 1.734e-02 9.762e-01 7.663e-03 1.105e-01
  0.000e+00 0.000e+00]
 [2.028e-02 1.773e-01 4.838e-02 1.657e-02 9.766e-01 7.350e-03 1.078e-01
  0.000e+00 0.000e+00]]
```

Binarize Data

```
binarizer = Binarizer(threshold=0.0).fit(X)
binaryX = binarizer.transform(X)
set_printoptions(precision=3)
print(binaryX[0:5, :])
```

```
[[1. 1. 1. 1. 1. 1. 1. 0. 0.]
 [1. 1. 1. 1. 1. 1. 1. 0. 0.]
 [1. 1. 1. 1. 1. 1. 1. 0. 0.]
 [1. 1. 1. 1. 1. 1. 1. 0. 0.]
 [1. 1. 1. 1. 1. 1. 1. 0. 0.]]
```

### Listing 5 Analysis

In this second part of the analysis, we applied various transformation techniques to the dataset to better understand the properties of the glass samples and improve the quality of our analysis. The first method used was the MinMaxScaler, which scaled all feature values between 0 and 1. This rescaling was crucial in identifying which features exhibited the most variability. Elements like Sodium (Na), Magnesium (Mg), and Aluminum (Al) displayed significant variation across samples, indicating that these features might provide more meaningful insights and should be prioritized in subsequent analyses. In contrast, elements such as Barium (Ba) and Iron (Fe) showed minimal variability, suggesting that they may not contribute as much to understanding the trends in the data. Magnesium, in particular, frequently showed values close to its maximum, highlighting its importance in this dataset. This could mean that Magnesium plays a key role in certain glass formulations, potentially influencing physical properties such as strength and durability.

After rescaling, we used the StandardScaler to standardize the dataset. This technique transformed the features to have a mean of 0 and a standard deviation of 1, allowing us to more easily compare samples on a relative scale. One example can be seen in the refractive index, where a value of 0.0873 in one sample indicated a significantly higher value relative to other samples. On the opposite end, Aluminum consistently showed negative values in some samples, suggesting that its concentration was lower than average. These relative differences are important because they reflect variations that could impact the glass's physical and chemical properties, such as its viscosity, or its clarity. For example, the presence of Magnesium or Aluminum typically improves the durability and resilience of glass, so understanding how these elements vary from the mean provides insights into the performance of different glass samples under various conditions.

In the next part of our analysis, we focused on normalizing the data. This step ensures that all features are on the same scale, making it easier to compare them, even if their original ranges are quite different. During normalization, Potassium (K) stood out with values that were much higher than other elements, suggesting that it plays a major role in the composition of this glass. This could mean that Potassium affects important properties. Magnesium (Mg), though present in more moderate amounts, also emerged as a key element, likely contributing to the overall strength and structure of the glass. On the other hand, Sodium (Na) and Silicon (Si) were found at lower levels, implying they might not be as critical in this specific type of glass. One interesting observation was the lack of Calcium (Ca) in many samples. Since Calcium often affects color and opacity, its absence might suggest that these glasses are designed to be especially clear, free from impurities, and perhaps more focused on optical quality.

Lastly, we used a Binarizer to simplify the data and categorize whether certain chemical elements were present or absent in each glass sample. The Binarizer works by turning values greater than a threshold (usually 0.0) into 1, while those below it become 0. This method makes it easy to see which elements are consistently found across the samples and which ones are missing. It provided a straightforward way to spot patterns in the glass compositions. For example, Potassium and Magnesium showed up regularly in many samples, while the absence of Calcium in others confirmed our earlier thoughts that these glasses might have been made to be particularly clear or transparent. This binary approach helped us quickly understand how different chemicals contribute to the unique properties of each glass sample.