

Q1. How many observations are in this dataset?

A: 214.

How many attributes (including the target variable: types of glasses) are in this dataset?

A: 10. (after removing 'ID number' attributes.) 'ID number' is not a helpful attribute to do data analysis.

Also, show the first 10 observations in this dataset, but please remove the first id column and add the title row before showing.

```
first 10 observations:
  RI      Na      Mg      Al      Si      K      Ca      Ba      Fe  Type of glass
0  1.52101  13.64  4.49  1.10  71.78  0.06  8.75  0.0  0.00      1
1  1.51761  13.89  3.60  1.36  72.73  0.48  7.83  0.0  0.00      1
2  1.51618  13.53  3.55  1.54  72.99  0.39  7.78  0.0  0.00      1
3  1.51766  13.21  3.69  1.29  72.61  0.57  8.22  0.0  0.00      1
4  1.51742  13.27  3.62  1.24  73.08  0.55  8.07  0.0  0.00      1
5  1.51596  12.79  3.61  1.62  72.97  0.64  8.07  0.0  0.26      1
6  1.51743  13.30  3.60  1.14  73.09  0.58  8.17  0.0  0.00      1
7  1.51756  13.15  3.61  1.05  73.24  0.57  8.24  0.0  0.00      1
8  1.51918  14.04  3.58  1.37  72.08  0.56  8.30  0.0  0.00      1
9  1.51755  13.00  3.60  1.36  72.99  0.57  8.40  0.0  0.11      1
```

Q2. Is there any missing value in any column?

A: No

Q3. Among all 8 kinds of metal oxide, which of them has the highest mean?

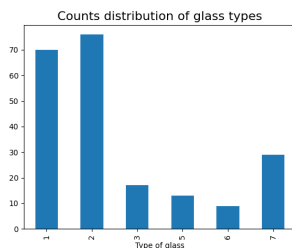
A: Si (Silicon)

Which of them has the smallest standard deviation?

A: Fe (Iron)

Q4. Which kind of plot is more suitable for drawing the target variable (glass types), histogram or bar chart? Briefly explain your thoughts and draw it out. Also, what do you see from the plot?

A: Bar chart. Because bar chart can clearly show counts distribution of glass types.



Q5. Draw out the distribution plots, boxplots for each variable, as well as plots that can demonstrate the pairwise correlation between each 2 variables. Briefly explain what you see in these plots.

A: From the pairplot, I can see each variable value distribution, and find which variables value distribution are quite different of different glass types, and maybe these variables are useful to be used in machine learning and help classify different types of glass.

Boxplot grouped by Type of glass

