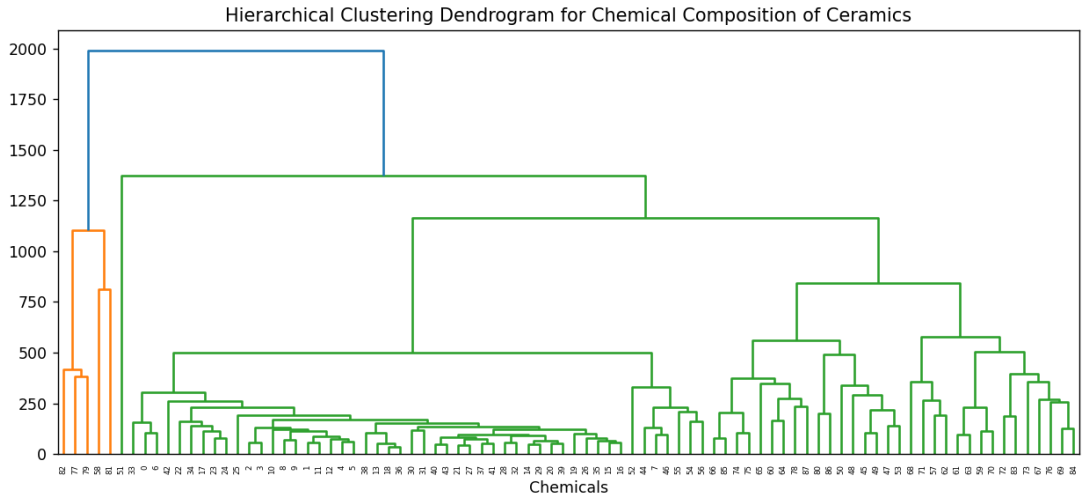
(1) Background (1/4 - 1/2 page) Describe your data, how you obtained it, and why it is interesting. Introduce all important concepts and background information. (2) Methods (1/2 - 3/4 page) Describe your methodology: give flowcharts, diagrams or formulas where appropriate. Describe evaluation strategy. (3) Results (1/4 - 1/2 page) Organize and present all your results and findings. Make sure it is easy to understand what your results are. (4) Conclusions (1/8 - 1/4 page) Summarize what the conclusions are and how they derive from the results. (5) References List books, scientific papers, web sites etc. that you referenced in the body of your manuscript.

My data has been taken from the UCI Machine Learning Repository. The data I chose for this assignment analyzes the chemical composition of ceramics and I have decided to show which certain chemical composition values are similar and can be clustered together. I thought this data would be an interesting choice since I enjoy learning about chemistry and it is interesting to see which chemicals can be grouped together and be considered similar in terms of behavior.

My strategy to take on this category was to start off by preprocessing my data by dropping all non numerical features. I added the following line “`X = data.drop(['Ceramic Name', 'Part'], axis=1)`”. This sets the variable X to equal the new value after dropping the Ceramic Name and Part features. After dropping these values, we are only working with numerical values, which is easier to cluster. After that I made a dendrogram of hierarchical clustering. I also added X values to my chemical values so that this can be plotted for k means clustering.

In my results in the dendrogram,



it has been clustered into two groups.