1. Cleaning steps: the element types of compound should be 2,3 or 4.

We need to find the index of certain compound so we can access the compound information even the library increase some day.

1 so I redo the data gain process, just left 'material\_id' and 'pretty\_formula'. Any further information can be easily accessed by finding certain compounds by its 'material\_id'.

I store the data set named 'third\_dataset\_09\_12\_2018.csv', and check it has the same length of data as I gained for the first time and second time.

Choose first 80 or 85 elements’ compound.