**Review**

How I approached the task?

1)I extracted the data from the pdf using the tabula software

2)I then imported the information into excel and cleansed the information as much as I could.

3) I then made a list of all the different hazards mentioned in the tables given and placed them in a column next to the data fields in the field key sheet

4) I then used python to associate the different field names to the phrases. The software iterated through each option in the columns. Then printed out the word category with the unique number and letter if that specific hazard matched the field name. I’ve attached the code that I used.

5) I then outputted this result into the Excel Spreadsheet

6) I then copied and pasted the notes column and aligned the relevant comments to the columns

The biggest problem that I faced was assigning the different categories to the chemicals. I realised that I couldn’t just use Excel. I was trying to use Excel, to complete this part of the task, which was arduous. So, I thought I could use some programming language to achieve this instead. I used python. This made the process quicker and more efficient. I was able to assign the hazard types to the correct columns for the 800+ chemicals.

Whilst I was using python, a problem that came up was trying to separate the different hazards for each chemical into different columns. To separate them into different columns, the hazards in each row needed to be separated by either a comma or a space. In my case, the hazards were separated by commas and spaces. Therefore, I had to consider both when separating the columns. However, with some work, I didn’t need to separate them into columns and could make it work without it.

I took a trial-and-error approach with the task given. After completion of the task, I believe I’ve developed a method or steps that I need to take that will enable me to complete such tasks in the future.