**Problem Statement :**

While researching, she found a raw dataset of over 30k datapoints (Raw\_Skills\_Dataset.csv) which contain technical skills and a lot of jargon mixed in. Can you help her develop a code that can clean this dataset and extract Technical (Hard) skills?

**Solution:**

Step1: Importing the pandas for cleaning the dataset

Step2: Opening the Raw\_skills\_Dataset and Example\_tecnical\_skills by using *open()* function and reading both the csv file line by line for comparing the dataset and extracting the similar column by using *readlines()* function

Step3: creating the for loop for iteration along the column as:

*for i in fileone:*

*if i != filetwo[x]:*

*outFile.write(filetwo[x])*

*x += 1*

Step4: Assigning the updated csv file to dataframe named technology\_skills using pd.

Step5: Checking if there is any null value using

*technology\_skills.isna().any()*

O/P is False means no null value present

Step6: checking for any duplicated value in dataset using *technology\_skills.duplicate(),* if any removed using *technology\_skills.drop\_duplicates(subset=None, keep='first', inplace=False, ignore\_index=False)*

Step6: Now the clean data is exported to a new dataset