Solution: Starting with reading the dataset and dropping all the null values from dataset. Then I drop all the duplicates from our dataset, which almost reduce the number by half in our dataset

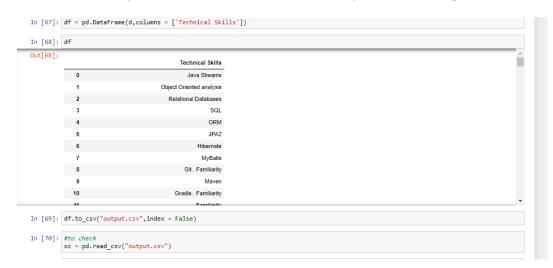
```
In [1]: import pandas as pd
         df = pd.read_csv("Raw_Skills_Dataset.csv")
         df.shape
Out[1]: (34116, 1)
         df.head()
In [2]:
Out[2]:
              RAW DATA
          0
                What ifs
          1
                seniority
               familiarity
            functionalities
               Lambdas
In [3]: df.dropna(inplace = True)
         df.drop duplicates(inplace = True)
In [4]: df.shape
Out[4]: (15677, 1)
```

Then I used some regex to find all the required data, and replace all the numerical data with empty sting from our dataset.

```
In [5]: y = df['RAW DATA'].str.findall(r'(^[A-Z].*)')
In [6]: y.replace(to_replace ="\d+",
                         value ="")
Out[6]: 0
                                                        [What ifs]
                                                                []
        2
                                                                []
        3
                                                         [Lambdas]
        34106
                                           [Leadership qualities]
                                                       [Actalent]
        34108
        34109
                 [Self-motivated, enthusiastic and strong drive]
        34110
        34111
        Name: RAW DATA, Length: 15677, dtype: object
```

After that I did the basic things, like removing empty list from our dataset and some simple cleaning tasks. And convert it to pandas data frame

Then I convert our panda's data frame to a CSV file as required in our assignment



Final to check our CSV file

