NAME –HRISHIKESH MAGADUM

EMAIL- hrishikeshmagadum09@gmail.com

APPROACH

The sample data set contained unnecessary jargon and we required to remove it to get the technical skills. This was done using Data Cleaning method which is one of the important steps in Data Science and analytics. Data cleaning involves removing missing values, converting the data into meaningful and suitable form, removing irrelevant data and extracting data which is useful for further analysis. This data contained uncleaned strings with the skill set as a substring of it. Data cleaning was performed using Jupyter Notebooks.

Libraries imported-

- Pandas
- Numpy
- Seaborn
- search

First the dataset was imported. The dataset containing raw skills was our main concern and hence we performed data cleaning on it specifically. I defined a method called clean string which takes a string as an input and removes all unnecessary characters present in it like extra spaces, exclamation marks etc. and removes it from a string to give a string containing meaningful text which is our motive. The cleaned string is returned. Using a simple loop we perform this operation on the entire data frame. We also get some null strings which we remove in the further code and then save the extracted skill set in a new csv file.