Business Analyst Role LLC

Task 1:

1. Clean and Preprocess the data:
   1. Importing the data.
   2. Finding null values
   3. FINDING top 5 and bottom 5 values
   4. By using shape function I got the number of rows and columns
   5. Identified the null values. – 16%
   6. Replaced all null values with mean values
2. Explotary analysis.
   1. Convert label encoding for classifier values.
   2. Standardization : convert unit free variable.
   3. Finding min and max value of all variables.
   4. Construct a boxplot & histogram to find outliers. (we did not remove).
   5. Data partition.
   6. Model fitting : Logistic model : KNeighborsClassifier.
   7. Cross validation: Principal component analysis.
   8. Reducing unnecessary columns- to avoid multicollinearity issues.

Task 2:

1. Cleaning the information:

a. Reading the text document.

b. Making a lowercase conversion.

d) Eliminating punctuation.

c. Splitting apart a text into words.

b. Eliminating stop words from the list of tokenized words.

1. Analysis

1. NLP Emotion Algorithm  
 a. Examine whether the term in emotion.txt is also included in the final word list by opening the emotion.txt file.   
 b. Clear each line by looping through it.   
 c. Use split to extract the word and emotion.   
 d. Add the feeling to emotion\_list if the term is present.

e. Finally count each emotion in the emotion list.