Name: Bhavya Narang

Roll No: 2019462

I am so sorry I haven't made functions for each task but they are written in a separate cell in jupyter notebook as I saw it at the very end that input will be text files :/

- Task 1
  - Method: regex1 for all vowels, they should start with aeiou or AEIOU and then can have any number of alphabets.
  - Assumption: word can only contain a-z and A-Z.

- Task 2
  - Method: used regex as all alphabets are capital hence only A-Z are allowed multiple times.
  - Assumption: word can only contain a-z and A-Z.
  - Formula for calculating percentage: (Total Capital words per class)/(Total words in a class(using word tokenizer))*100.

- Task 3
  - Method : Regex for both phone number and email
  - Assumption: email should be of the form string1@string2.string3 where any of the three string can contain characters a-z A-Z 0-9 and underscore hence I have used \w for this.
  - Assumption: phone number can be 10-12 digits long and should only contain numbers from 0 to 9 and nothing else.
  - Formula for percentage: (emails in messages of given class)/(total number of messages of the given class)*100.
  - Similar for phone number.

- Task 4:
  - Assumption: monetary values will be of the form SymbolNumber or NumberSymbol and should have no space in between and can contain commas to separate numbers. Symbols are: dollar and euro.

    - Method: regex containing digits and commas, either followed or preceded by the symbol.
    - Formula for percentage: (monetary values in messages of given class)/(total number of messages of the given class)*100.

- Task 5
    - Method: tokenized using inbuilt method in NLTK (EMOTICON_RE.findall)
    - Assumption: emojis should be only those that are present in nltk.
- Task 6
    - Assumption: clitics will be of the form: [a-zA-Z]'[a-zA-Z]
    - Method: regex using the above.
- Task 7
    - Method: used nltk word tokenizer to get the first word of each message and converted both the search and first word to lowercase to finally compare.
    - Assumption: Words should not be like: "Hello!" and expectation should not be "Hello."
- Task 8
    - Method: used sentence tokenizer in nltk after removing all multiple spaces and converting them to single space. Then repeated a similar process like task 7.
    - Assumption: Here after the word ends we can have ! or . or ? as we can have these things in the end.
- Task 9
    -