# Data Mining
## Assignment 1.1

Name: Ayush Kumar

Group: CSE-6B1

Reg. No.: 20214284

Q1:

```python
def checker(str):
  if len(str)==0:
    return False
  for ch in str:
    if not ch.isalpha():
      return False
  return True

sentence=input("Enter the data set: ")
wordsArr=sentence.strip().lower().split(" ")
freq={}
for i in range(len(wordsArr)):
  wordsArr[i]=wordsArr[i].replace("'","");
  wordsArr[i]=wordsArr[i].replace(".","");
  if checker(wordsArr[i]):
      freq.update({wordsArr[i]:0})
for x in wordsArr:
  if checker(x):
    freq.update({x:freq.get(x)+1})
ls=list(freq.items())
ls.sort()
print(ls)
```

```
Enter the data set: This is a very very short 'text'. It is only 2 sentences long'.
[('a', 1), ('is', 2), ('it', 1), ('long', 1), ('only', 1), ('sentences', 1), ('short', 1), ('text', 1), ('this', 1), ('very', 2)]
```

Q2:

```python
import nltk
nltk.download('stopwords')

stop_words = nltk.corpus.stopwords.words('english')

def remove_stop_words(text):
  words = text.split()
  filtered_words = [word for word in words if word not in
stop_words]
  return ' '.join(filtered_words)

text = input("Enter the sentence: ")
text=text.lower()
text_without_stop_words = remove_stop_words(text)
print(text_without_stop_words.split(" "))
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
Enter the sentence: this is a very very short text It is only 2 sentences long
['short', 'text', '2', 'sentences', 'long']
```