**Steps for all three tasks step by step in details-**

**TASK 1 POS TAG :**

1) Importing the necessary packages nltk,numpy,

Pandas and matplotlib

2) Reading the given corpus

3) Checking and exploring the corpus,using

shape,columns,head() and describe()

4) Performing word tokenization on the ‘Section\_Text’

Of the corpus using word\_tokenize

5) Using nltk.pos\_tag to perform the POS tagging operations

6) Using nltk.FreqDist to check the frequency distribution

of the respective POS Tags

7) Creating patterns to be used

8) Using nltk.RegexpTagger regex to get the pattern for taggers

9) Storing the frequencies in a new dataframe and then

plotting the histogram for all of them

**TASK 2 SEMANTIC SIMILARITY :**

1) Importing the necessary packages nltk,numpy,pandas,

Matplotlib,sklearn Count vectizer and TfIdf Vectorizer

And distance from scipy.spatial

2) Reading the given corpus

3) Checking and exploring the corpus,using

shape,columns,head() and describe()

4) Dropping Section\_ID column as it is not required

5) Calculating similarity of two senetnces,first by using

count vectorizer method

6) For this first we convert the text from Section\_text to

Vector,by fitting and transforming using count vectorizer

7) Using distance,imported from scipy.spacial calculating the cosine distance

8) Printing out the similarity in terms of percentage

9) Similarly, following the procedure using TfIdf vectorizer

10) Printing the similarity of two senetences in terms of percentage

**TASK 3 TEXT SUMMARIZATION :**

1) Importing the necessary packages nltk, pandas,

Genism(summarize) and sumy (summarizers,stopwords,

LsaSummarizer, LuhnSummarizer, PlaintextParser,

Stemmer,LexRankSummarizer)

2) Taking the text to be summarized, a text based on

the famous "Cockroach Theory" by Sundar Pichai

3) Checking and exploring the corpus,using

shape,columns,head() and describe()

4) Using PlaintextParser imported from sumy for effective

parsing the text

5) Summarizing the text with different summarizers-

Lexrank summarizer, lsa summarizer, luhn summarize

And genism summarizer

6) Lastly calling all the functions to print the summaries