Instructions

1. Follow the instructions in each question carefully.
2. Only two files should be uploaded in canvas without zipping them. One is ipynb file and other one html output of the ipynb file. No other files should be uploaded
3. Any assignment submitted using other python IDEs are not considered for grading.
4. If there are any issues in accessing the links to datasets, you can search for the same dataset from any repositories and use them.
5. Incorrect Assignment Set submitted will not be considered.

NLP Assignment 2

Set A

Link to the Dataset:

https://drive.google.com/file/d/1x0oiWyLUns9002jTDj2CzIE6yqbglN__/view?usp=sharing

Note: Use first 10000 rows of dataset from the original dataset given

Description of Data:

This is the Amazon Fine food review dataset.Each record consists of the following attributes:
The column or features in the dataset:

- Id
- ProductId — unique identifier for the product
- UserId — unqiue identifier for the user
- ProfileName
- HelpfulnessNumerator — number of users who found the review helpful
- HelpfulnessDenominator — number of users who indicated whether they found the review helpful or not
- Score — rating between 1 and 5
- Time — timestamp for the review
- Summary — brief summary of the review
- Text — text of the review

Task 1 : Load the Amazon Fine Food Review dataset, select the first 10,000 rows, and perform a comprehensive Exploratory Data Analysis (EDA) followed by necessary preprocessing steps. ***(2 Mark)***

Task 2 : Implementing Parsing Algorithms **[1+ 3 +3+1]**

1. Display the POS tagging on the first 2 rows of 'Text'

2. Plot the dependency parser for any two random sentences from the entire corpus/dataset that has at least 10 words in the sentence. Make sure that dependency parser looks good and should visually understandable. (3 Mark)

3. Plot the Statistical parser for any two random sentences from the entire corpus/dataset that has at least 10 words in the sentence. Make sure that Statistical parser looks good and should visually understandable. (3 Mark)

4. Compare the efficiency of each approach in terms of the number of operations and time taken.