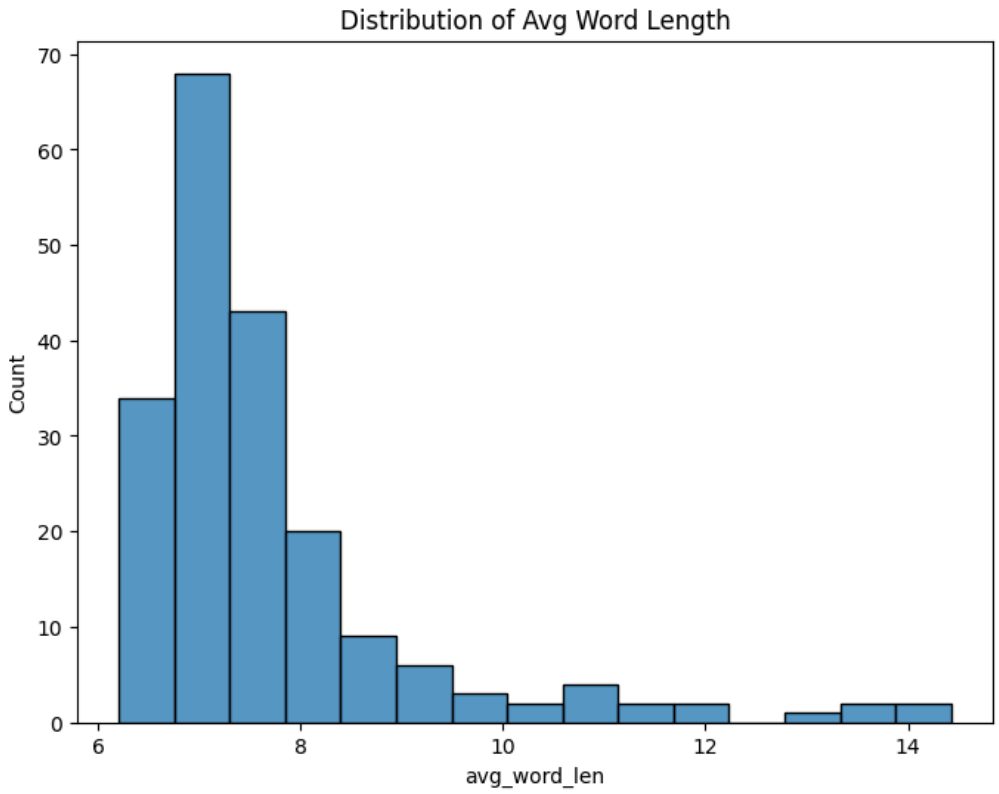
Week 10

Name: Sri Harsha, Nimmagadda
Email: harshachaitanya27@gmail.com
College: Oregon State University
Specialisation: NLP

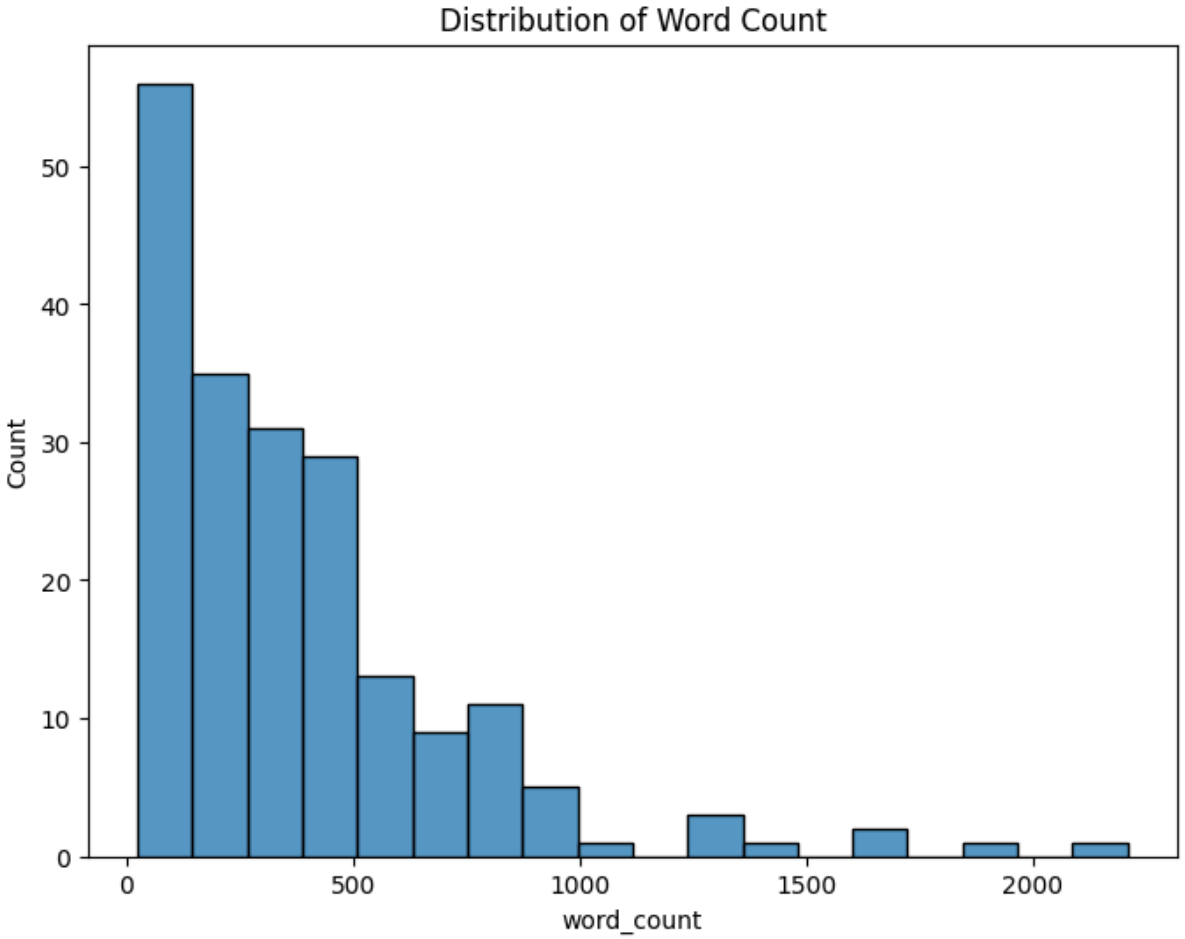
Problem Statement: The challenge arises when HR or hiring managers review resumes, as they often encounter an overwhelming amount of irrelevant information that requires careful examination. This tedious process makes it challenging and time-intensive for them to pinpoint the most qualified candidates. To tackle this issue, a proposed solution involves leveraging Named Entity Recognition (NER) within Natural Language Processing (NLP). This advanced technology can autonomously recognize and categorise key details in resumes, such as the candidate's name, educational history, work experience, and skills. Implementing NER streamlines the candidate shortlisting process, significantly enhancing efficiency for HR professionals and reducing their time and effort investment.

Github link - https://github.com/harshachaitanya27/DataGlacier/tree/main/week10

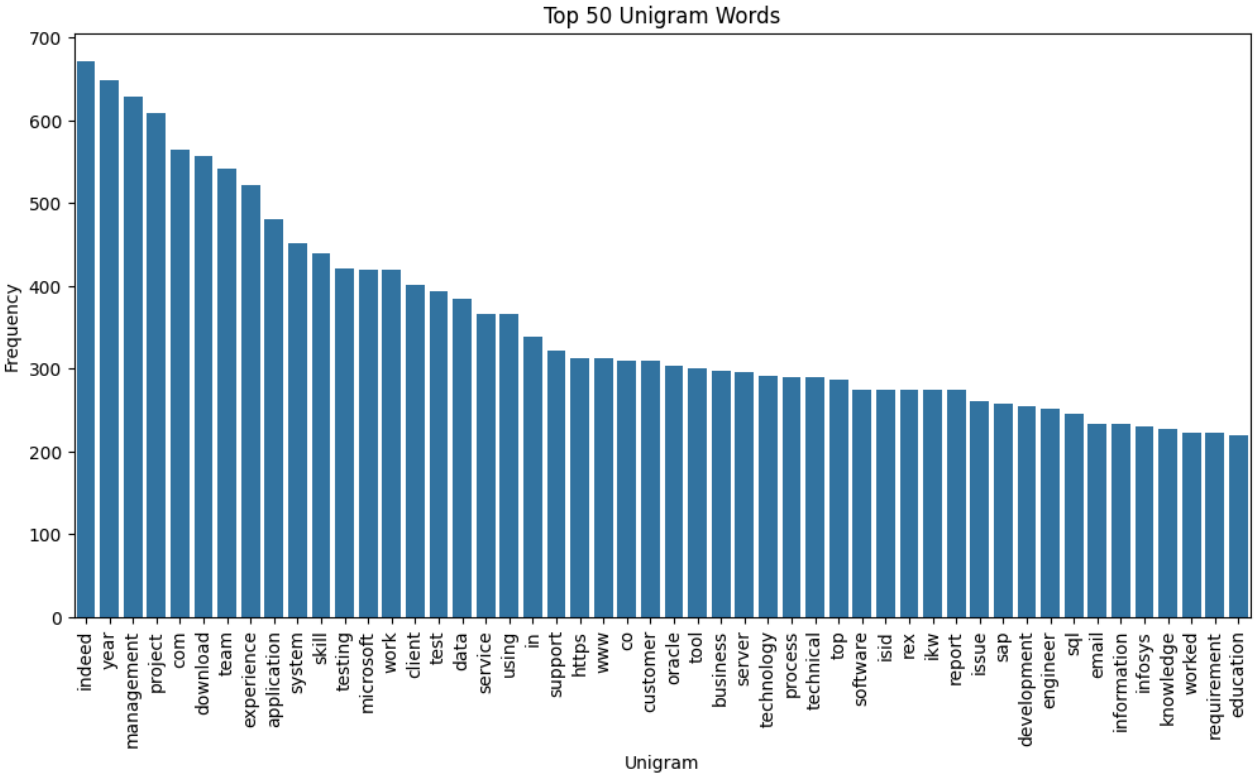
EDA:



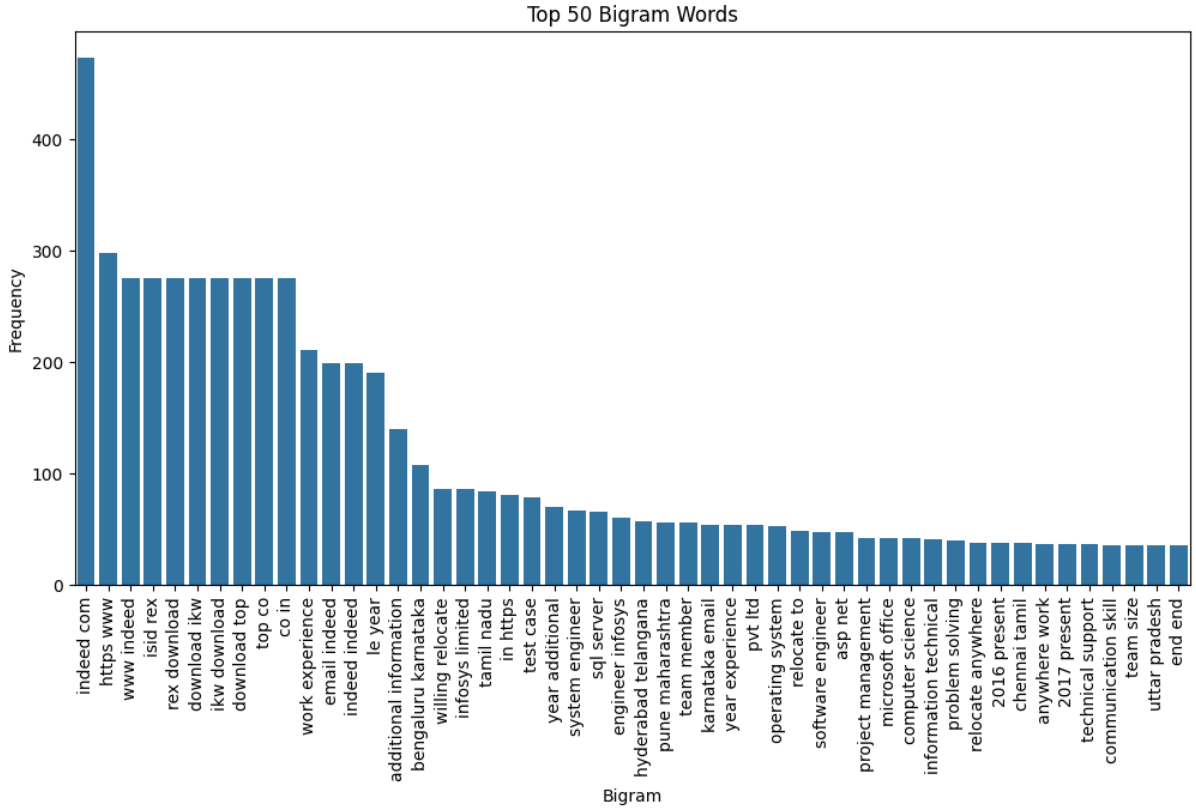
- The plot unmistakably indicates that the average word length is 7 words.



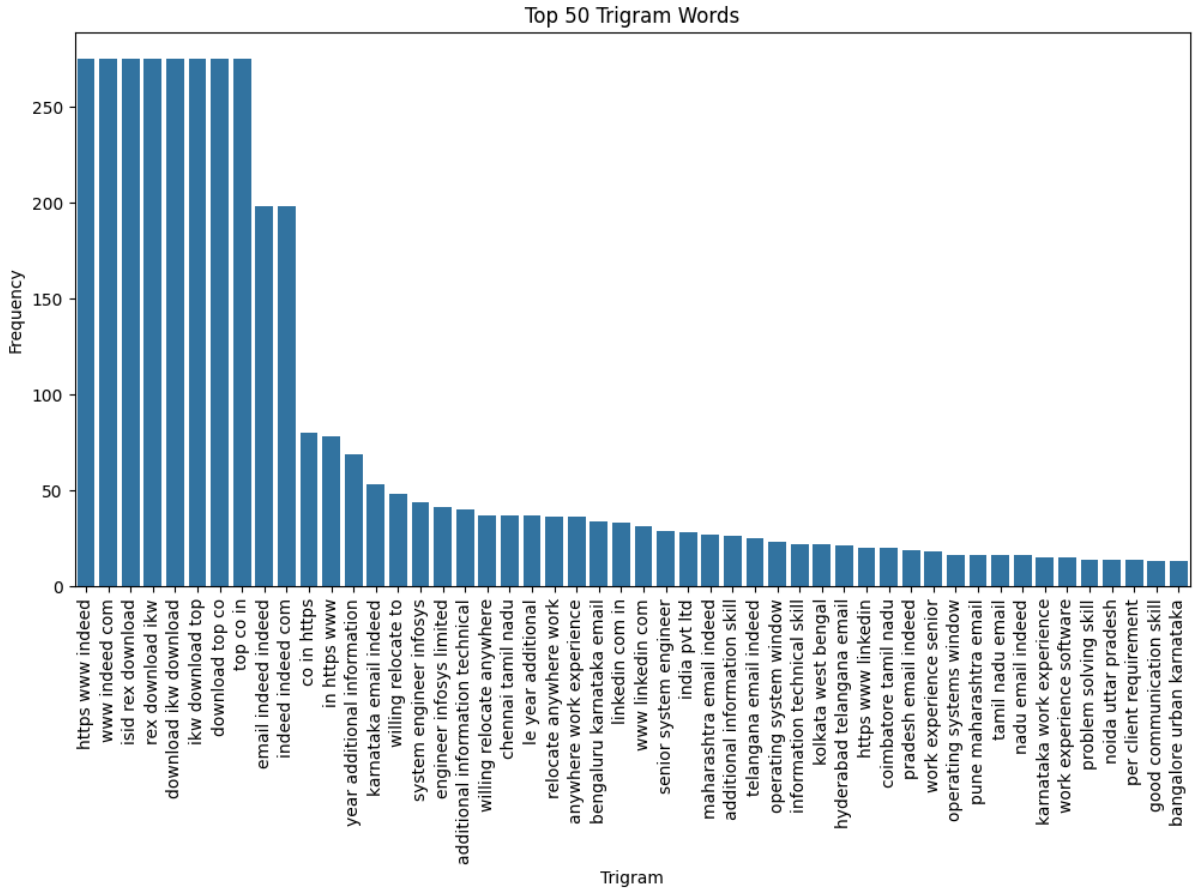
- Based on the graph provided, it can be inferred that the majority of resumes have fewer than 500 words.



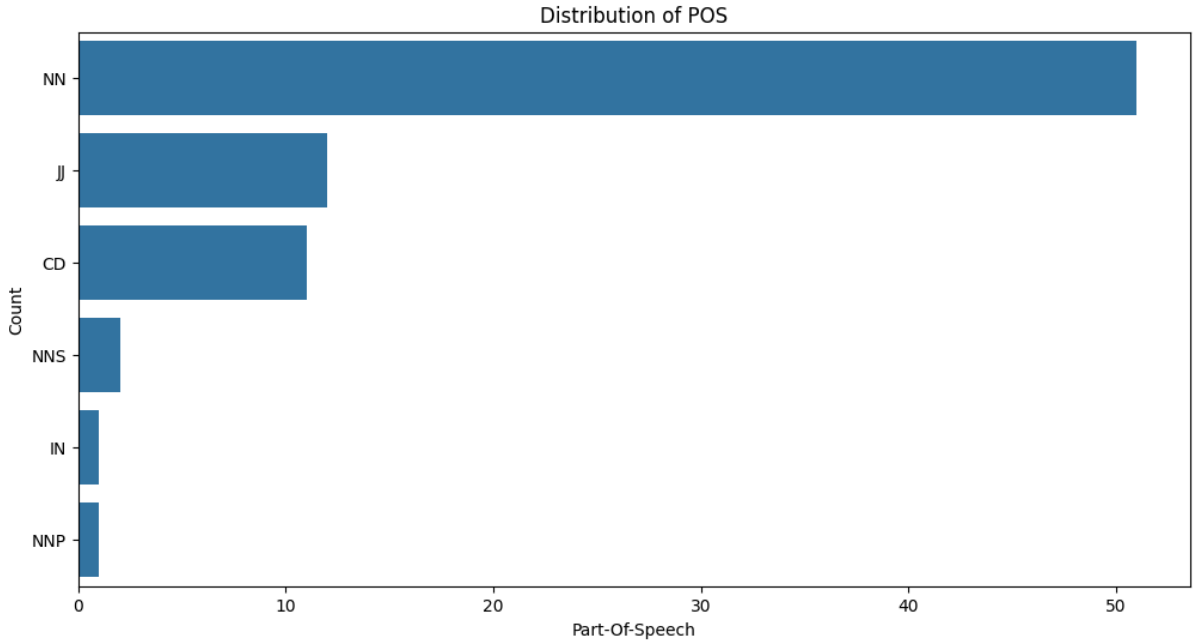
- The most common words are 'Indeed,' 'Year,' and 'Management,' ranking as the top three in frequency.



- The most frequently occurring word pairs are 'Work experience,' 'Email indeed,' and 'Additional information,' which could be valuable.



- The top three frequently occurring words together are 'Year additional information,' 'Karnataka email indeed,' and 'Willing relocate to,' which could be beneficial.



- The most commonly used is 'Singular noun,' followed by 'Adjective or numeral,' 'Numeral (cardinal),' 'Proper plural noun,' 'Singular proper noun,' and 'Preposition or conjunction, subordinating' in that order.

Final Recommendation:
➢ I plotted the word count distribution and found that my resume, along with most others, contained at most 500 words.

➢ Next, I wanted to analyze the frequency of each word occurring in the dataset. I created the distribution of Unigram words, Bigram words, and even Trigram words to examine the most frequent words individually, in pairs, and three words at a time, respectively.

➢ I conducted the distribution of Part of Speech to discover that the majority of words in my resume were singular nouns.