1. **How many tokens and types (unique tokens) in each document, and in the entire corpus (collections)?**

Ans: Total tokens in corpus: 24165

Total unique tokens in corpus: 4105

2. **How many tokens and types (unique tokens) in each document after removing all the stop words?**

Ans: Total tokens in corpus (after stop word removal): 15526

Total unique tokens in corpus (after stop word removal): 3646

3. **How many terms (size of vocabulary) are left in each document after stemming/ lemmatization and what's the total vocabulary size for the entire corpus?**

Ans: Total terms in corpus (after stem): 15526

Total unique terms in corpus (after stem): 2878