

ADVANCED DATA MINING

ASSIGNMENT REPORT



**SUMANTH POBALA**

**UMID 71652304**

**DATA SCIENCE SEMESTER 1**

**DATA EXPLORATION ASSIGNMENT**

1. **Finding Anagrams using python** 
   1. Firstly we will import the file to fname and define a function to initialize and all the words and define other functions for creating dictionary and finding anagrams and get all anagrams from dictionary and we will store all anagrams in an attribute called allanas and create a text file and store all the anagrams in it.
   2. We can use my\_key to find the anagram of a specific word where word will be searched in anagram dictionary.

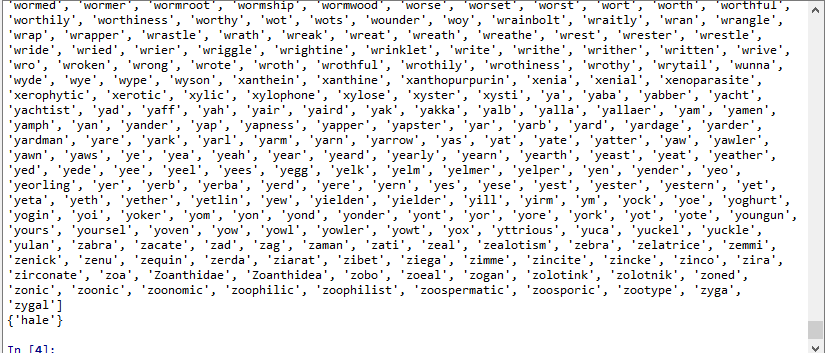


Fig.1 Anagrams and anagram of heal

* 1. Sorting the words based on the length, for doing that we will import the defualtdict which is located in collections and store output into ‘result’.

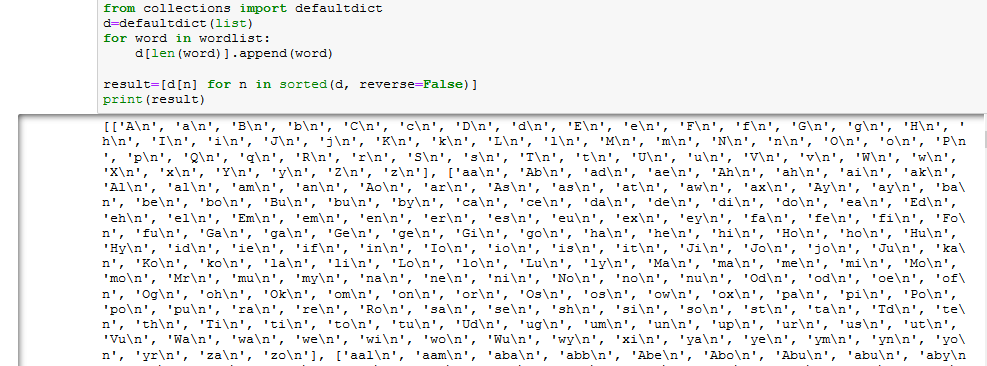


Fig 1.1. Sorting Anagrams

* 1. Creating the sublists based on the indexing and strip the data using the strip function

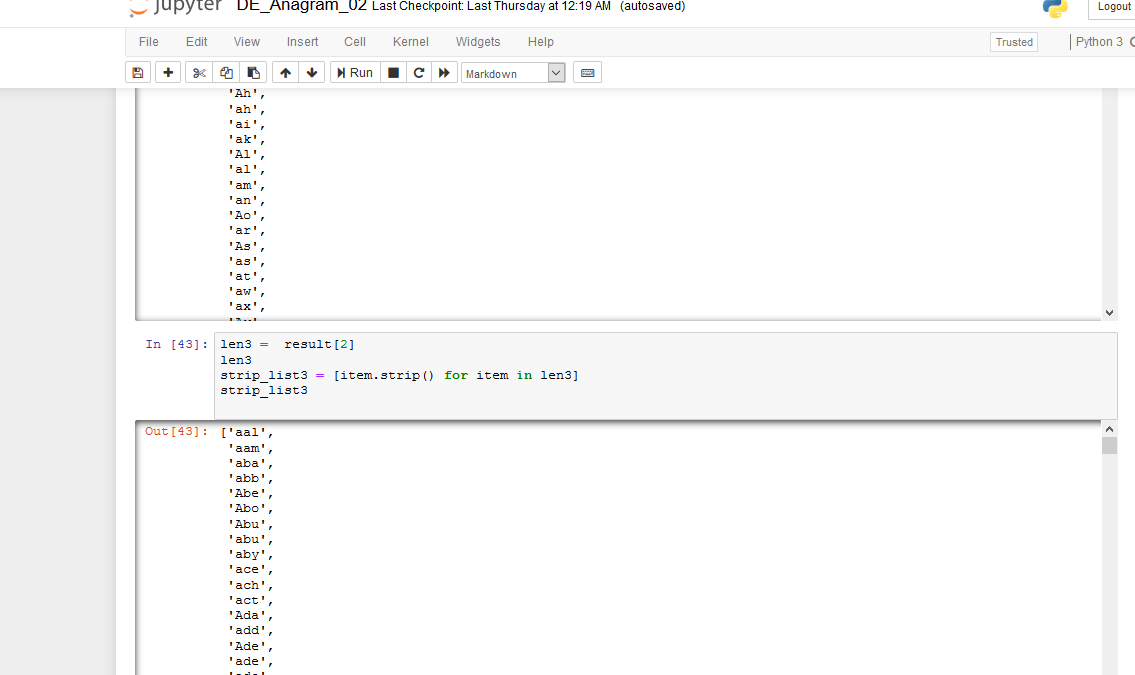


Fig 1.2. Dividing into sublists (Before finding anagrams)

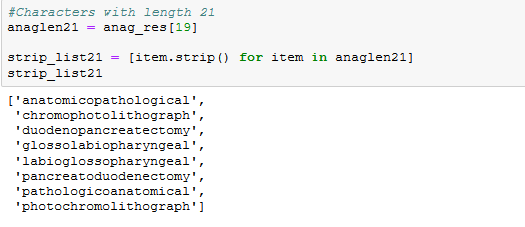


Fig 1.3. Fig 1.2. Dividing into sublists (After finding anagrams)

1. **Lengths of anagrams and sublists after finding anagrams**
   1. Length of words before finding anagrams : 235886 words
   2. Total anagrams found : 27400
   3. Length of anagrams of word length 2: 68
   4. Length of anagrams of word length 3: 682
   5. Length of anagrams of word length 4: 2341
   6. Length of anagrams of word length 5: 3560
   7. Length of anagrams of word length 6: 5137
   8. Length of anagrams of word length 7: 4700
   9. Length of anagrams of word length 8: 3983
   10. Length of anagrams of word length 9: 3011
   11. Length of anagrams of word length 10: 1837
   12. Length of anagrams of word length 11: 930
   13. Length of anagrams of word length 12: 519
   14. Length of anagrams of word length 13: 243
   15. Length of anagrams of word length 14: 140
   16. Length of anagrams of word length 15: 86
   17. Length of anagrams of word length 16: 70
   18. Length of anagrams of word length 17: 42
   19. Length of anagrams of word length 18: **20**
   20. Length of anagrams of word length 19: **14**
   21. Length of anagrams of word length 20: **6**
   22. Length of anagrams of word length 21: **8**
   23. Length of anagrams of word length 22: **4**