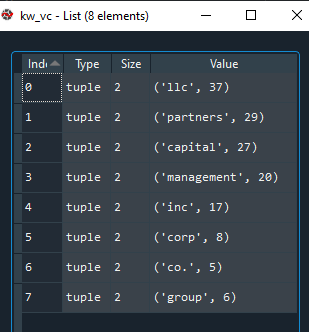
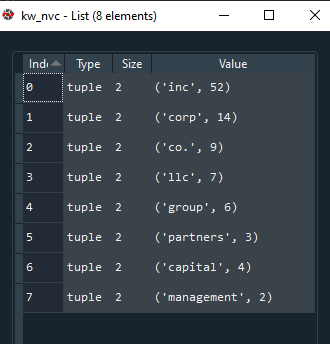
**PROBLEM 2 (ANALYSIS):**

* In order to determine a list of keywords that would help to decide whether a company is a venture capital firm or not, we have written a code (from line 141 to 184) that extracts the list of keywords from the dataset. We have then created a comparative keywords list for VC and Non-VC by considering the 5 most frequent keywords in VC and Non-VC separately.
* Fig(1) below refers to comparative keywords list of Non-VC firm whereas fig(2) refers to that of VC firm. From these figures, we can choose the keywords that are predominantly used in one of the firm and not in the other.
* By following the above condition, the list of keywords will be [‘inc’, ‘corp’, ‘co.’, ‘llc’, ‘partners’, ‘capital’, ‘management’]. With the help of these keywords, we can determine whether a company is VC or Non-VC firm.



  
 Fig(1) Fig(2)

**RESEARCH MATERIAL:**

* <https://medium.com/@datamonsters/text-preprocessing-in-python-steps-tools-and-examples-bf025f872908>
* <https://www.kdnuggets.com/2019/04/text-preprocessing-nlp-machine-learning.html>