

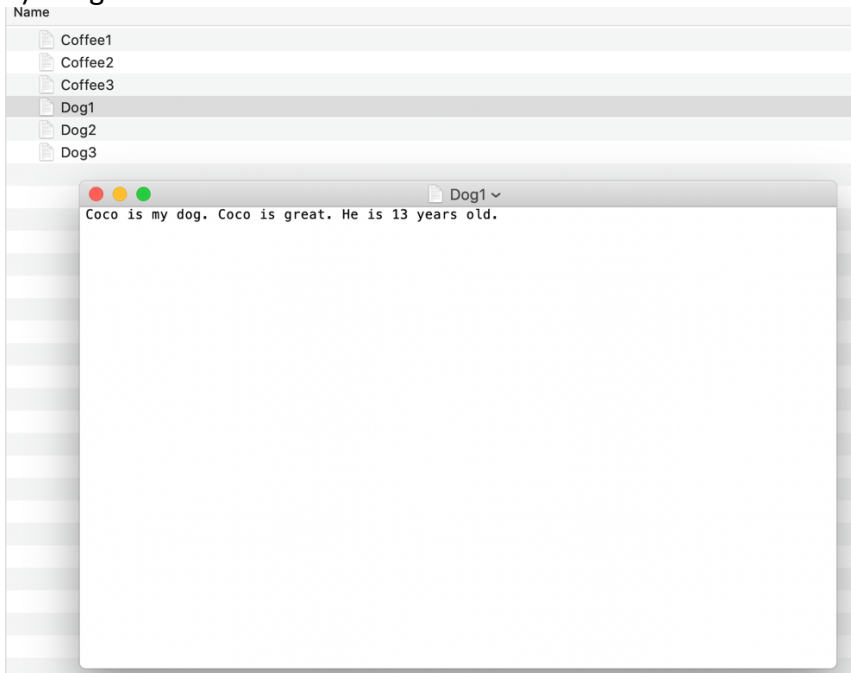
Sana Khan

Homework 1 & 2

1) Creation of corpus

MyCorpus				
Q Search				
Name		Date Modified	Size	Kind
Coffee1		Jan 14, 2024 at 4:27 PM	50 bytes	Plain Text
Coffee2		Jan 14, 2024 at 4:28 PM	107 bytes	Plain Text
Coffee3		Jan 14, 2024 at 4:29 PM	86 bytes	Plain Text
Dog1		Jan 14, 2024 at 4:09 PM	50 bytes	Plain Text
Dog2		Jan 14, 2024 at 4:10 PM	65 bytes	Plain Text
Dog3		Jan 14, 2024 at 4:26 PM	88 bytes	Plain Text

2) Dog1 file contents



3) Dataframe with vocabulary words

```
<class 'scipy.sparse._csr.csr_matrix'>
['13' 'americano' 'breakfast' 'chemex' 'chicken' 'chips' 'coco' 'coffee'
 'cream' 'doesn't' 'dog' 'drink' 'eat' 'everyday' 'favorite' 'great' 'home'
 'love' 'loves' 'make' 'old' 'outside' 'pourover' 'sit' 'sugar' 'sunshine'
 'teeth' 'use' 'warm' 'way' 'weather' 'years']
 13 americano breakfast chemex chicken chips coco coffee cream \
0 1 0 0 0 0 0 2 0 0
1 0 0 0 0 1 1 2 0 0
2 0 0 0 0 0 0 1 0 0
3 0 0 0 0 0 0 0 1 0
4 0 1 1 0 0 0 2 1
5 0 0 1 1 0 0 2 0

 doesn't ... pourover sit sugar sunshine teeth use warm way weather \
0 0 ... 0 0 0 0 0 0 0 0 0
1 1 ... 0 0 0 0 1 0 0 0 0
2 0 ... 0 1 0 1 0 0 1 0 1
3 0 ... 0 0 0 0 0 0 0 0 0
4 0 ... 0 0 1 0 0 0 0 0 0
5 0 ... 2 0 0 0 0 1 0 1 0

 years
0 1
1 0
2 0
3 0
4 0
5 0
[6 rows x 32 columns]
```

4) CSV File

	13	americano	breakfast	chemex	chicken	chips	coco	coffee	cream
0	1	0	0	0	0	0	2	0	0
1	0	0	0	0	1	1	2	0	0
2	0	0	0	0	0	0	1	0	0
3	0	0	0	0	0	0	0	1	0
4	0	1	1	0	0	0	0	2	1
5	0	0	0	1	0	0	0	2	0