

## CMPT 459 Assignment 3

Huiyi Zou 301355563

### Question 1

Proof:

The itemset  $Y$  with length  $k$  contains  $k$  distinct items. Assuming  $Y = \{y_1, y_2, \dots, y_k\}$  and  $\text{sup}(Y) = n$ , which means that the number of occurrences of the itemset  $Y$  is  $n$ , for each item contained in the itemset, it must appear no less than  $n$  times. Thus,  $\text{sup}(y_1) \geq n$ ,  $\text{sup}(y_2) \geq n$ , ...,  $\text{sup}(y_k) \geq n$ . Since  $\text{sup}(Y) \leq$  any  $\text{sup}(y_i)$  where  $y_i$  in the itemset  $Y$ , we have  $\text{sup}(Y) \leq \min \{\text{sup}(y_1), \text{sup}(y_2), \dots, \text{sup}(y_k)\}$ . And  $x_1, x_2, \dots, x_k$  are the  $k$  most frequent items. Then we have  $\min \{\text{sup}(y_1), \text{sup}(y_2), \dots, \text{sup}(y_k)\} \leq \min \{\text{sup}(x_1), \text{sup}(x_2), \dots, \text{sup}(x_k)\}$ . Therefore,  $\text{sup}(Y) \leq \min \{\text{sup}(x_1), \text{sup}(x_2), \dots, \text{sup}(x_k)\}$ .

## Question 2

$sup(X) = sup(Y) = sup(X \cup Y)$  does not always hold.

Counterexample:

TID	Items
1	a, b, c
2	c, d, e
3	a, b, c, d, e

Here is a transaction database.

Assume  $X = \{a, b, c\}$  and  $Y = \{c, d, e\}$ , where  $sup(X) = sup(Y) = 2$ .

$X \cup Y = \{a, b, c, d, e\}$ , there is only one item containing  $\{a, b, c, d, e\}$ , then  $sup(X \cup Y) = 1$ .

### Question 3

1. We can use the Apriori algorithm. Firstly, get all patterns of length  $k = 1$  and their support with a minimum support, and then generate patterns of length  $k + 1$  from previous patterns and filter them with the minimum support. Repeat this step until we get all patterns of length up to 5. For different pattern length, we can easily sort those patterns by their support and get top 100.

The code is in the file “code.py”. I modified the code from assignment 1. I used the function `apriori()` from library `mlxtend`. I followed this website to apply the Apriori algorithm in python:

[http://rasbt.github.io/mlxtend/user\\_guide/frequent\\_patterns/apriori/](http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/apriori/)

I got every pattern of length up to 5 with their support for D1 and D2 by using the Apriori algorithm, where the minimum support for D1 was 10, and the minimum support for D2 was 25. The function `apriori()` returned a dataframe with columns “support” and “itemsets”. I added another column “length” to store the length of itemsets and sorted the support of itemset in descending order. I created 5 lists in a list `result[]`. For each pattern length  $k$ , `result[k - 1]` stored the top 100 length- $k$  patterns and their supports extracted from the dataframe.

## 2. The top 100 length-1 patterns for D1:

[[['the'], 391], [['to'], 269], [['and'], 253], [['\r'], 252], [['a'], 251], [['is'], 217], [['of'], 201], [['in'], 177], [['this'], 148], [['you'], 140], [['i'], 136], [['for'], 120], [['that'], 111], [['are'], 97], [['on'], 95], [['not'], 78], [['be'], 72], [['people'], 71], [['they'], 71], [['if'], 70], [['it'], 70], [['was'], 70], [['but'], 66], [['just'], 63], [['with'], 63], [['black'], 60], [['he'], 60], [['my'], 60], [['have'], 59], [['we'], 59], [['about'], 58], [['all'], 57], [['who'], 56], [['so'], 54], [['your'], 54], [['its'], 52], [['what'], 52], [['police'], 49], [['out'], 46], [['us'], 45], [['at'], 43], [['dont'], 43], [['from'], 43], [['how'], 43], [['their'], 42], [['as'], 41], [['amp'], 38], [['like'], 38], [['me'], 38], [['one'], 38], [['up'], 38], [['an'], 37], [['white'], 37], [['by'], 36], [['no'], 36], [['trump'], 36], [['anonymous'], 35], [['has'], 35], [['can'], 34], [['now'], 34], [['will'], 33], [['do'], 32], [['when'], 32], [['her'], 31], [['been'], 30], [['please'], 30], [['get'], 29], [['because'], 28], [['more'], 28], [['our'], 28], [['youranoncentral'], 27], [['his'], 26], [['man'], 26], [['go'], 25], [['im'], 25], [['or'], 25], [['protests'], 24], [['see'], 23], [['than'], 23], [['there'], 23], [['were'], 23], [['antifa'], 22], [['make'], 22], [['protest'], 22], [['should'], 22], [['these'], 22], [['had'], 21], [['him'], 21], [['say'], 21], [['them'], 21], [['america'], 20], [['didnt'], 20], [['right'], 20], [['fuck'], 19], [['other'], 19], [['she'], 19], [['want'], 19], [['after'], 18], [['into'], 18], [['know'], 18]]

## The top 100 length-2 patterns for D1:

[[['and', 'the'], 125], [['the', 'to'], 121], [['the', 'of'], 115], [['the', 'in'], 101], [['the', 'a'], 98], [['the', 'is'], 97], [['the', '\r'], 87], [['and', 'to'], 86], [['to', 'a'], 85], [['and', 'a'], 73], [['this', 'is'], 73], [['to', 'is'], 72], [['of', 'to'], 70], [['and', 'of'], 69], [['of', 'a'], 68], [['is', 'a'], 63], [['the', 'for'], 59], [['in', 'a'], 56], [['and', 'in'], 56], [['and', 'is'], 56], [['this', 'the'], 55], [['i', 'the'], 54], [['r', 'a'], 53], [['you', 'the'], 53], [['r', 'to'], 48], [['of', 'in'], 47], [['to', 'in'], 47], [['and', '\r'], 46], [['i', '\r'], 46], [['this', 'a'], 45], [['the', 'are'], 44], [['this', 'to'], 44], [['i', 'to'], 43], [['the', 'that'], 43], [['r', 'is'], 42], [['of', '\r'], 42], [['the', 'on'], 42], [['this', '\r'], 41], [['i', 'and'], 41], [['this', 'and'], 41], [['i', 'a'], 40], [['you', 'a'], 40], [['and', 'for'], 40], [['for', 'to'], 39], [['for', 'a'], 38], [['in', 'is'], 38], [['to', 'that'], 38], [['of', 'is'], 36], [['the', 'they'], 36], [['you', 'to'], 36], [['the', 'was'], 35], [['r', 'in'], 34], [['that', 'a'], 34], [['and', 'that'], 34], [['the', 'black'], 34], [['not', 'is'], 34], [['you', 'and'], 33], [['this', 'in'], 33], [['the', 'us'], 33], [['for', '\r'], 32], [['the', 'all'], 32], [['you', 'are'], 32], [['you', 'if'], 32], [['you', 'of'], 32], [['people', 'the'], 32], [['i', 'of'], 31], [['to', 'on'], 31], [['the', 'be'], 30], [['the', 'he'], 30], [['i', 'this'], 30], [['you', 'in'], 30], [['this', 'of'], 30], [['you', 'this'], 30], [['on', 'a'], 29], [['my', 'and'], 29], [['that', 'is'], 29], [['the', 'not'], 29], [['the', 'police'], 29], [['are', 'to'], 28], [['for', 'in'], 28], [['the', 'from'], 28], [['to', 'not'], 28], [['people', 'of'], 28], [['the', 'who'], 28], [['and', 'on'], 27], [['we', 'to'], 27], [['you', '\r'], 26], [['if', 'a'], 26], [['the', 'amp'], 26], [['and', 'are'], 26], [['are', 'is'], 26], [['is',

'he'], 26], [['i', 'in'], 26], [['not', 'in'], 26], [['you', 'is'], 26], [['trump', 'the'], 26],  
[['the', 'what'], 26], [['to', 'us'], 26], [['was', 'a'], 25], [['and', 'out'], 25]]

### The top 100 length-3 patterns for D1:

[[['and', 'of', 'the'], 41], [['the', 'of', 'a'], 40], [['and', 'in', 'the'], 37], [['and', 'to',  
'the'], 35], [['the', 'to', 'is'], 35], [['the', 'of', 'to'], 33], [['and', 'the', 'a'], 32], [['the',  
'of', 'in'], 32], [['the', 'to', 'a'], 31], [['the', 'r', 'of'], 28], [['the', 'in', 'a'], 28], [['the',  
'is', 'a'], 28], [['and', 'is', 'the'], 26], [['and', 'r', 'the'], 25], [['and', 'to', 'a'], 25], [['of',  
'to', 'a'], 25], [['the', 'to', 'in'], 25], [['and', 'of', 'a'], 24], [['and', 'to', 'is'], 24], [['this',  
'is', 'the'], 24], [['this', 'is', 'a'], 23], [['to', 'is', 'a'], 23], [['this', 'and', 'the'], 23],  
[['the', 'of', 'is'], 23], [['this', 'r', 'is'], 22], [['and', 'for', 'the'], 22], [['this', 'and', 'is'],  
22], [['and', 'is', 'a'], 21], [['the', 'this', 'a'], 21], [['i', 'and', 'the'], 21], [['the', 'r', 'in'],  
20], [['and', 'of', 'to'], 20], [['the', 'for', 'in'], 20], [['this', 'to', 'the'], 20],  
[['youranoncentral', 'the', 'us'], 20], [['i', 'the', 'r'], 18], [['this', 'to', 'a'], 18], [['you',  
'and', 'the'], 18], [['the', 'in', 'is'], 18], [['the', 'is', 'police'], 18], [['this', 'to', 'is'], 18],  
[['the', 'to', 'us'], 18], [['the', 'r', 'a'], 17], [['the', 'for', 'a'], 17], [['to', 'in', 'a'], 17],  
[['the', 'amp', 'is'], 17], [['and', 'to', 'in'], 17], [['they', 'and', 'the'], 17], [['the', 'for',  
'of'], 17], [['the', 'to', 'he'], 17], [['i', 'the', 'of'], 17], [['you', 'the', 'of'], 17], [['the',  
'to', 'that'], 17], [['i', 'r', 'a'], 16], [['of', 'in', 'a'], 16], [['and', 'black', 'the'], 16],  
[['you', 'the', 'are'], 16], [['the', 'to', 'be'], 16], [['the', 'is', 'he'], 16], [['the', 'not', 'is'],  
16], [['trump', 'the', 'is'], 16], [['this', 'of', 'the'], 16], [['youranoncentral', 'the', 'to'],  
16], [['r', 'to', 'a'], 15], [['i', 'r', 'to'], 15], [['i', 'to', 'a'], 15], [['and', 'are', 'the'], 15],  
[['and', 'of', 'out'], 15], [['the', 'and', 'that'], 15], [['i', 'the', 'in'], 15], [['the', 'is',  
'what'], 15], [['youranoncentral', 'is', 'the'], 15], [['people', 'of', 'the'], 15],  
[['trump', 'the', 'to'], 15], [['the', 'to', 'what'], 15], [['i', 'and', 'r'], 14], [['the', 'r', 'to'],  
14], [['the', 'was', 'a'], 14], [['you', 'to', 'a'], 14], [['the', 'amp', 'police'], 14], [['the',  
'to', 'amp'], 14], [['i', 'and', 'to'], 14], [['my', 'and', 'the'], 14], [['and', 'on', 'the'], 14],  
[['and', 'their', 'to'], 14], [['the', 'are', 'is'], 14], [['the', 'for', 'to'], 14], [['is', 'he',  
'police'], 14], [['the', 'this', 'in'], 14], [['is', 'us', 'police'], 14], [['youranoncentral',  
'is', 'police'], 14], [['the', 'us', 'is'], 14], [['to', 'is', 'what'], 14], [['youranoncentral',  
'us', 'is'], 14], [['the', 'of', 'not'], 14], [['the', 'of', 'out'], 14], [['of', 'to', 'that'], 14],  
[['the', 'to', 'police'], 14], [['the', 'us', 'police'], 14], [['youranoncentral', 'the',  
'police'], 14]]

### The top 100 length-4 patterns for D1:

[[['and', 'of', 'the', 'a'], 14], [['the', 'is', 'us', 'police'], 14], [['the', 'youranoncentral',  
'is', 'police'], 14], [['youranoncentral', 'is', 'us', 'police'], 14], [['youranoncentral',  
'is', 'us', 'the'], 14], [['youranoncentral', 'the', 'us', 'police'], 14],  
[['youranoncentral', 'the', 'to', 'us'], 14], [['the', 'of', 'in', 'a'], 13], [['the', 'is', 'amp',

'he'], 13], [['the', 'is', 'amp', 'police'], 13], [['is', 'amp', 'he', 'police'], 12], [['the', 'amp', 'he', 'police'], 12], [['the', 'is', 'he', 'police'], 12], [['trump', 'the', 'to', 'is'], 12], [['the', 'to', 'is', 'what'], 12], [['trump', 'youranoncentral', 'to', 'the'], 12], [['and', 'is', 'the', 'a'], 11], [['this', 'is', 'the', 'a'], 11], [['the', 'of', 'to', 'a'], 11], [['the', 'didnt', 'amp', 'is'], 11], [['didnt', 'to', 'amp', 'is'], 11], [['the', 'didnt', 'to', 'amp'], 11], [['the', 'to', 'amp', 'is'], 11], [['trump', 'the', 'amp', 'is'], 11], [['us', 'to', 'amp', 'what'], 11], [['this', 'is', 'and', 'the'], 11], [['and', 'of', 'to', 'the'], 11], [['the', 'didnt', 'to', 'he'], 11], [['the', 'didnt', 'to', 'is'], 11], [['the', 'didnt', 'to', 'police'], 11], [['the', 'to', 'is', 'he'], 11], [['the', 'to', 'is', 'police'], 11], [['allows', 'create', 'created', 'amp'], 10], [['allows', 'criminals', 'create', 'amp'], 10], [['allows', 'create', 'amp', 'didnt'], 10], [['allows', 'create', 'go\r', 'amp'], 10], [['allows', 'create', 'amp', 'he'], 10], [['allows', 'impunity', 'create', 'amp'], 10], [['allows', 'create', 'amp', 'is'], 10], [['allows', 'create', 'amp', 'police'], 10], [['allows', 'racists', 'create', 'amp'], 10], [['allows', 'create', 'system', 'amp'], 10], [['allows', 'the', 'create', 'amp'], 10], [['allows', 'create', 'to', 'amp'], 10], [['allows', 'create', 'amp', 'trump'], 10], [['amp', 'create', 'us', 'allows'], 10], [['allows', 'create', 'amp', 'what'], 10], [['allows', 'youranoncentral', 'create', 'amp'], 10], [['allows', 'criminals', 'created', 'amp'], 10], [['didnt', 'allows', 'created', 'amp'], 10], [['go\r', 'allows', 'created', 'amp'], 10], [['allows', 'created', 'amp', 'he'], 10], [['impunity', 'allows', 'created', 'amp'], 10], [['allows', 'created', 'amp', 'is'], 10], [['allows', 'created', 'amp', 'police'], 10], [['racists', 'allows', 'created', 'amp'], 10], [['allows', 'created', 'system', 'amp'], 10], [['the', 'allows', 'created', 'amp'], 10], [['allows', 'created', 'to', 'amp'], 10], [['trump', 'allows', 'created', 'amp'], 10], [['allows', 'amp', 'created', 'us'], 10], [['allows', 'created', 'amp', 'what'], 10], [['youranoncentral', 'allows', 'created', 'amp'], 10], [['allows', 'criminals', 'amp', 'didnt'], 10], [['allows', 'criminals', 'go\r', 'amp'], 10], [['allows', 'criminals', 'amp', 'he'], 10], [['allows', 'impunity', 'criminals', 'amp'], 10], [['allows', 'criminals', 'amp', 'is'], 10], [['allows', 'criminals', 'amp', 'police'], 10], [['allows', 'racists', 'criminals', 'amp'], 10], [['allows', 'criminals', 'system', 'amp'], 10], [['allows', 'the', 'criminals', 'amp'], 10], [['allows', 'criminals', 'to', 'amp'], 10], [['allows', 'criminals', 'amp', 'trump'], 10], [['amp', 'criminals', 'us', 'allows'], 10], [['allows', 'criminals', 'amp', 'what'], 10], [['allows', 'youranoncentral', 'criminals', 'amp'], 10], [['didnt', 'allows', 'go\r', 'amp'], 10], [['didnt', 'allows', 'amp', 'he'], 10], [['didnt', 'impunity', 'allows', 'amp'], 10], [['didnt', 'allows', 'amp', 'is'], 10], [['didnt', 'allows', 'amp', 'police'], 10], [['didnt', 'racists', 'allows', 'amp'], 10], [['didnt', 'allows', 'system', 'amp'], 10], [['didnt', 'the', 'allows', 'amp'], 10], [['didnt', 'allows', 'to', 'amp'], 10], [['didnt', 'allows', 'amp', 'trump'], 10], [['allows', 'amp', 'us', 'didnt'], 10], [['didnt', 'allows', 'amp', 'what'], 10], [['didnt', 'youranoncentral', 'allows', 'amp'], 10], [['go\r', 'allows', 'amp', 'he'], 10], [['go\r', 'impunity', 'allows', 'amp'], 10], [['allows', 'amp',

'go\r', 'is'], 10], [['go\r', 'allows', 'amp', 'police'], 10], [['go\r', 'racists', 'allows', 'amp'], 10], [['go\r', 'allows', 'system', 'amp'], 10], [['go\r', 'the', 'allows', 'amp'], 10], [['go\r', 'allows', 'to', 'amp'], 10], [['trump', 'allows', 'go\r', 'amp'], 10], [['allows', 'amp', 'go\r', 'us'], 10]]

### The top 100 length-5 patterns for D1:

[[['is', 'police', 'youranoncentral', 'the', 'us'], 14], [['is', 'police', 'the', 'amp', 'he'], 12], [['is', 'the', 'amp', 'to', 'didnt'], 11], [['criminals', 'created', 'create', 'amp', 'allows'], 10], [['didnt', 'created', 'create', 'amp', 'allows'], 10], [['created', 'create', 'amp', 'go\r', 'allows'], 10], [['created', 'create', 'amp', 'allows', 'he'], 10], [['impunity', 'created', 'create', 'amp', 'allows'], 10], [['created', 'is', 'create', 'amp', 'allows'], 10], [['created', 'police', 'create', 'amp', 'allows'], 10], [['racists', 'created', 'create', 'amp', 'allows'], 10], [['created', 'create', 'system', 'amp', 'allows'], 10], [['created', 'the', 'create', 'amp', 'allows'], 10], [['created', 'create', 'amp', 'to', 'allows'], 10], [['trump', 'created', 'create', 'amp', 'allows'], 10], [['created', 'create', 'amp', 'allows', 'us'], 10], [['created', 'create', 'amp', 'allows', 'what'], 10], [['created', 'youranoncentral', 'create', 'amp', 'allows'], 10], [['didnt', 'criminals', 'create', 'amp', 'allows'], 10], [['criminals', 'create', 'amp', 'go\r', 'allows'], 10], [['criminals', 'create', 'amp', 'allows', 'he'], 10], [['impunity', 'criminals', 'create', 'amp', 'allows'], 10], [['criminals', 'is', 'create', 'amp', 'allows'], 10], [['criminals', 'police', 'create', 'amp', 'allows'], 10], [['racists', 'criminals', 'create', 'amp', 'allows'], 10], [['criminals', 'create', 'system', 'amp', 'allows'], 10], [['criminals', 'the', 'create', 'amp', 'allows'], 10], [['criminals', 'create', 'amp', 'to', 'allows'], 10], [['trump', 'criminals', 'create', 'amp', 'allows'], 10], [['criminals', 'create', 'amp', 'allows', 'us'], 10], [['criminals', 'create', 'amp', 'allows', 'what'], 10], [['criminals', 'youranoncentral', 'create', 'amp', 'allows'], 10], [['didnt', 'create', 'amp', 'go\r', 'allows'], 10], [['didnt', 'create', 'amp', 'allows', 'he'], 10], [['impunity', 'didnt', 'create', 'amp', 'allows'], 10], [['allows', 'is', 'create', 'amp', 'didnt'], 10], [['didnt', 'police', 'create', 'amp', 'allows'], 10], [['racists', 'didnt', 'create', 'amp', 'allows'], 10], [['didnt', 'create', 'system', 'amp', 'allows'], 10], [['didnt', 'the', 'create', 'amp', 'allows'], 10], [['didnt', 'create', 'amp', 'to', 'allows'], 10], [['trump', 'didnt', 'create', 'amp', 'allows'], 10], [['didnt', 'create', 'amp', 'allows', 'us'], 10], [['didnt', 'create', 'amp', 'allows', 'what'], 10], [['didnt', 'youranoncentral', 'create', 'amp', 'allows'], 10], [['create', 'amp', 'go\r', 'allows', 'he'], 10], [['impunity', 'create', 'amp', 'go\r', 'allows'], 10], [['is', 'create', 'amp', 'go\r', 'allows'], 10], [['police', 'create', 'amp', 'go\r', 'allows'], 10], [['racists', 'create', 'amp', 'go\r', 'allows'], 10], [['create', 'system', 'amp', 'go\r', 'allows'], 10], [['the', 'create', 'amp', 'go\r', 'allows'], 10], [['create', 'amp', 'go\r', 'to', 'allows'], 10], [['trump', 'create', 'amp', 'go\r', 'allows'], 10], [['create', 'amp',

'go\r', 'allows', 'us'], 10], [['create', 'amp', 'go\r', 'allows', 'what'], 10],  
[['youranoncentral', 'create', 'amp', 'go\r', 'allows'], 10], [['impunity', 'create',  
'amp', 'allows', 'he'], 10], [['is', 'create', 'amp', 'allows', 'he'], 10], [['police',  
'create', 'amp', 'allows', 'he'], 10], [['racists', 'create', 'amp', 'allows', 'he'], 10],  
[['create', 'system', 'amp', 'allows', 'he'], 10], [['the', 'create', 'amp', 'allows', 'he'],  
10], [['create', 'amp', 'to', 'allows', 'he'], 10], [['trump', 'create', 'amp', 'allows',  
'he'], 10], [['create', 'amp', 'allows', 'us', 'he'], 10], [['create', 'amp', 'he', 'allows',  
'what'], 10], [['youranoncentral', 'create', 'amp', 'allows', 'he'], 10], [['impunity',  
'is', 'create', 'amp', 'allows'], 10], [['impunity', 'police', 'create', 'amp', 'allows'],  
10], [['impunity', 'racists', 'create', 'amp', 'allows'], 10], [['impunity', 'create',  
'system', 'amp', 'allows'], 10], [['impunity', 'the', 'create', 'amp', 'allows'], 10],  
[['impunity', 'create', 'amp', 'to', 'allows'], 10], [['trump', 'impunity', 'create', 'amp',  
'allows'], 10], [['impunity', 'create', 'amp', 'allows', 'us'], 10], [['impunity', 'create',  
'amp', 'allows', 'what'], 10], [['impunity', 'youranoncentral', 'create', 'amp',  
'allows'], 10], [['is', 'police', 'create', 'amp', 'allows'], 10], [['racists', 'is', 'create',  
'amp', 'allows'], 10], [['is', 'create', 'system', 'amp', 'allows'], 10], [['is', 'the',  
'create', 'amp', 'allows'], 10], [['is', 'create', 'amp', 'to', 'allows'], 10], [['trump', 'is',  
'create', 'amp', 'allows'], 10], [['is', 'create', 'amp', 'allows', 'us'], 10], [['is',  
'create', 'amp', 'allows', 'what'], 10], [['is', 'youranoncentral', 'create', 'amp',  
'allows'], 10], [['racists', 'police', 'create', 'amp', 'allows'], 10], [['police', 'create',  
'system', 'amp', 'allows'], 10], [['police', 'the', 'create', 'amp', 'allows'], 10],  
[['police', 'create', 'amp', 'to', 'allows'], 10], [['trump', 'police', 'create', 'amp',  
'allows'], 10], [['police', 'create', 'amp', 'allows', 'us'], 10], [['police', 'create',  
'amp', 'allows', 'what'], 10], [['police', 'youranoncentral', 'create', 'amp', 'allows'],  
10], [['racists', 'create', 'system', 'amp', 'allows'], 10], [['racists', 'the', 'create',  
'amp', 'allows'], 10], [['racists', 'create', 'amp', 'to', 'allows'], 10], [['trump',  
'racists', 'create', 'amp', 'allows'], 10], [['racists', 'create', 'amp', 'allows', 'us'],  
10]]



### The top 100 length-1 patterns for D2:

[[['covid19'], 729], [['to'], 474], [['in'], 450], [['is'], 377], [['a'], 327], [['the'], 318], [['this'], 304], [['pandemic'], 260], [['an'], 248], [['testing'], 246], [['protesters'], 242], [['la'], 240], [['centers'], 239], [['closing'], 238], [['punish'], 238], [['of'], 237], [['theyre'], 236], [['attempt'], 233], [['kissychalamets'], 233], [['sil\r'], 233], [['weaponizing'], 233], [['r'], 198], [['and'], 171], [['you'], 170], [['i'], 145], [['for'], 138], [['if'], 131], [['not'], 125], [['healthcare'], 111], [['at'], 104], [['providers'], 95], [['were'], 94], [['can'], 93], [['do'], 87], [['many'], 87], [['care'], 84], [['protest'], 84], [['c\r'], 82], [['need'], 82], [['we'], 82], [['insurance'], 78], [['plans'], 78], [['telling'], 78], [['with'], 78], [['deny'], 77], [['recommend'], 77], [['tipptipphorray'], 77], [['who'], 75], [['are'], 69], [['get'], 68], [['has'], 60], [['from'], 55], [['like'], 53], [['my'], 52], [['on'], 51], [['two'], 51], [['our'], 48], [['what'], 48], [['that'], 47], [['as'], 46], [['just'], 45], [['country'], 44], [['all'], 40], [['be'], 40], [['new'], 40], [['will'], 40], [['have'], 39], [['love'], 38], [['so'], 36], [['blood'], 35], [['here'], 35], [['cases'], 34], [['charge'], 34], [['day'], 34], [['it'], 34], [['one'], 34], [['still'], 34], [['more'], 33], [['thought'], 33], [['would'], 33], [['after'], 31], [['anyone'], 31], [['away'], 31], [['2'], 30], [['gone'], 30], [['now'], 30], [['tested'], 30], [['amp'], 29], [['hasnt'], 29], [['no'], 29], [['police'], 29], [['protested'], 29], [['bryanna711'], 28], [['downtown'], 28], [['elemental'], 28], [['folks'], 28], [['mom'], 28], [['over'], 28], [['testin\r'], 28], [['bill'], 27]]

### The top 100 length-2 patterns for D2:

[[['covid19', 'to'], 407], [['covid19', 'in'], 396], [['covid19', 'is'], 344], [['to', 'in'], 335], [['to', 'is'], 302], [['in', 'is'], 290], [['covid19', 'this'], 282], [['this', 'to'], 262], [['this', 'is'], 258], [['this', 'in'], 254], [['covid19', 'pandemic'], 252], [['to', 'pandemic'], 247], [['covid19', 'an'], 246], [['pandemic', 'in'], 246], [['covid19', 'testing'], 244], [['protesters', 'covid19'], 242], [['testing', 'to'], 242], [['pandemic', 'is'], 240], [['this', 'pandemic'], 240], [['protesters', 'to'], 240], [['covid19', 'centers'], 239], [['centers', 'closing'], 238], [['protesters', 'centers'], 238], [['centers', 'punish'], 238], [['centers', 'testing'], 238], [['centers', 'to'], 238], [['covid19', 'closing'], 238], [['protesters', 'closing'], 238], [['closing', 'punish'], 238], [['testing', 'closing'], 238], [['to', 'closing'], 238], [['covid19', 'la'], 238], [['covid19', 'punish'], 238], [['protesters', 'punish'], 238], [['protesters', 'testing'], 238], [['testing', 'punish'], 238], [['to', 'punish'], 238], [['an', 'in'], 236], [['testing', 'in'], 236], [['theyre', 'is'], 236], [['an', 'is'], 235], [['covid19', 'theyre'], 235], [['la', 'testing'], 235], [['this', 'la'], 235], [['this', 'testing'], 235], [['an', 'pandemic'], 234], [['an', 'to'], 234], [['la', 'in'], 234], [['la', 'is'], 234], [['testing', 'is'], 234], [['la', 'to'], 234], [['this', 'theyre'], 234], [['theyre', 'to'], 234], [['an', 'attempt'], 233], [['an', 'centers'], 233], [['an', 'closing'], 233], [['kissychalamets', 'an'], 233], [['an', 'la'],

233], [['protesters', 'an'], 233], [['an', 'punish'], 233], [['an', 'sil\r'], 233], [['an', 'testing'], 233], [['an', 'theyre'], 233], [['an', 'this'], 233], [['an', 'weaponizing'], 233], [['centers', 'attempt'], 233], [['attempt', 'closing'], 233], [['covid19', 'attempt'], 233], [['attempt', 'in'], 233], [['attempt', 'is'], 233], [['kissychalamets', 'attempt'], 233], [['la', 'attempt'], 233], [['attempt', 'pandemic'], 233], [['protesters', 'attempt'], 233], [['attempt', 'punish'], 233], [['sil\r', 'attempt'], 233], [['attempt', 'testing'], 233], [['theyre', 'attempt'], 233], [['this', 'attempt'], 233], [['attempt', 'to'], 233], [['attempt', 'weaponizing'], 233], [['centers', 'in'], 233], [['centers', 'is'], 233], [['kissychalamets', 'centers'], 233], [['centers', 'la'], 233], [['centers', 'pandemic'], 233], [['sil\r', 'centers'], 233], [['centers', 'theyre'], 233], [['this', 'centers'], 233], [['centers', 'weaponizing'], 233], [['closing', 'in'], 233], [['closing', 'is'], 233], [['kissychalamets', 'closing'], 233], [['la', 'closing'], 233], [['pandemic', 'closing'], 233], [['sil\r', 'closing'], 233], [['theyre', 'closing'], 233], [['this', 'closing'], 233], [['weaponizing', 'closing'], 233], [['covid19', 'kissychalamets'], 233]]

### The top 100 length-3 patterns for D2:

[[['covid19', 'to', 'in'], 306], [['covid19', 'to', 'is'], 287], [['covid19', 'in', 'is'], 282], [['to', 'in', 'is'], 271], [['covid19', 'this', 'to'], 258], [['covid19', 'this', 'is'], 250], [['covid19', 'this', 'in'], 248], [['this', 'to', 'in'], 244], [['this', 'in', 'is'], 243], [['covid19', 'to', 'pandemic'], 242], [['this', 'to', 'is'], 242], [['covid19', 'testing', 'to'], 241], [['protesters', 'covid19', 'to'], 240], [['to', 'pandemic', 'in'], 240], [['covid19', 'pandemic', 'in'], 239], [['covid19', 'pandemic', 'is'], 239], [['covid19', 'centers', 'closing'], 238], [['protesters', 'centers', 'closing'], 238], [['centers', 'closing', 'punish'], 238], [['centers', 'testing', 'closing'], 238], [['to', 'centers', 'closing'], 238], [['protesters', 'centers', 'covid19'], 238], [['covid19', 'centers', 'punish'], 238], [['covid19', 'centers', 'testing'], 238], [['covid19', 'centers', 'to'], 238], [['protesters', 'centers', 'punish'], 238], [['protesters', 'centers', 'testing'], 238], [['protesters', 'centers', 'to'], 238], [['centers', 'testing', 'punish'], 238], [['centers', 'to', 'punish'], 238], [['centers', 'testing', 'to'], 238], [['protesters', 'covid19', 'closing'], 238], [['covid19', 'closing', 'punish'], 238], [['covid19', 'testing', 'closing'], 238], [['covid19', 'to', 'closing'], 238], [['protesters', 'closing', 'punish'], 238], [['protesters', 'testing', 'closing'], 238], [['protesters', 'to', 'closing'], 238], [['testing', 'closing', 'punish'], 238], [['to', 'closing', 'punish'], 238], [['to', 'testing', 'closing'], 238], [['covid19', 'this', 'pandemic'], 238], [['protesters', 'covid19', 'punish'], 238], [['protesters', 'covid19', 'testing'], 238], [['covid19', 'testing', 'punish'], 238], [['covid19', 'to', 'punish'], 238], [['to', 'pandemic', 'is'], 238], [['protesters', 'testing', 'punish'], 238], [['protesters', 'to', 'punish'], 238], [['protesters', 'testing', 'to'], 238], [['testing', 'to', 'punish'], 238], [['this',

'pandemic', 'is'], 237], [['to', 'this', 'pandemic'], 237], [['covid19', 'an', 'in'], 235], [['covid19', 'theyre', 'is'], 235], [['covid19', 'la', 'testing'], 235], [['covid19', 'this', 'la'], 235], [['covid19', 'this', 'testing'], 235], [['pandemic', 'in', 'is'], 235], [['this', 'pandemic', 'in'], 235], [['testing', 'to', 'in'], 235], [['this', 'la', 'testing'], 235], [['covid19', 'an', 'is'], 234], [['covid19', 'an', 'pandemic'], 234], [['covid19', 'an', 'to'], 234], [['an', 'is', 'in'], 234], [['an', 'pandemic', 'in'], 234], [['covid19', 'testing', 'in'], 234], [['la', 'in', 'is'], 234], [['testing', 'in', 'is'], 234], [['this', 'theyre', 'is'], 234], [['theyre', 'to', 'is'], 234], [['this', 'theyre', 'to'], 234], [['an', 'centers', 'attempt'], 233], [['an', 'attempt', 'closing'], 233], [['covid19', 'an', 'attempt'], 233], [['an', 'attempt', 'in'], 233], [['an', 'attempt', 'is'], 233], [['kissychalamets', 'an', 'attempt'], 233], [['an', 'la', 'attempt'], 233], [['an', 'attempt', 'pandemic'], 233], [['protesters', 'an', 'attempt'], 233], [['an', 'attempt', 'punish'], 233], [['an', 'attempt', 'sil\r'], 233], [['an', 'attempt', 'testing'], 233], [['an', 'theyre', 'attempt'], 233], [['an', 'this', 'attempt'], 233], [['an', 'attempt', 'to'], 233], [['an', 'attempt', 'weaponizing'], 233], [['an', 'centers', 'closing'], 233], [['covid19', 'an', 'centers'], 233], [['an', 'centers', 'in'], 233], [['an', 'centers', 'is'], 233], [['kissychalamets', 'an', 'centers'], 233], [['an', 'centers', 'la'], 233], [['an', 'centers', 'pandemic'], 233], [['protesters', 'an', 'centers'], 233], [['an', 'centers', 'punish'], 233], [['an', 'centers', 'sil\r'], 233], [['an', 'centers', 'testing'], 233]]

### The top 100 length-4 patterns for D2:

[[['covid19', 'to', 'in', 'is'], 268], [['covid19', 'this', 'to', 'in'], 243], [['covid19', 'this', 'in', 'is'], 240], [['covid19', 'this', 'to', 'is'], 239], [['protesters', 'centers', 'covid19', 'closing'], 238], [['covid19', 'centers', 'closing', 'punish'], 238], [['covid19', 'centers', 'testing', 'closing'], 238], [['covid19', 'to', 'centers', 'closing'], 238], [['protesters', 'centers', 'closing', 'punish'], 238], [['protesters', 'centers', 'testing', 'closing'], 238], [['protesters', 'to', 'centers', 'closing'], 238], [['centers', 'testing', 'closing', 'punish'], 238], [['to', 'centers', 'closing', 'punish'], 238], [['to', 'centers', 'testing', 'closing'], 238], [['protesters', 'centers', 'covid19', 'punish'], 238], [['protesters', 'covid19', 'centers', 'testing'], 238], [['protesters', 'centers', 'covid19', 'to'], 238], [['covid19', 'centers', 'testing', 'punish'], 238], [['covid19', 'to', 'centers', 'punish'], 238], [['covid19', 'centers', 'testing', 'to'], 238], [['protesters', 'centers', 'testing', 'punish'], 238], [['protesters', 'centers', 'to', 'punish'], 238], [['protesters', 'centers', 'testing', 'to'], 238], [['centers', 'testing', 'to', 'punish'], 238], [['protesters', 'covid19', 'closing', 'punish'], 238], [['protesters', 'covid19', 'testing', 'closing'], 238], [['protesters', 'to', 'covid19', 'closing'], 238], [['covid19', 'testing', 'closing', 'punish'], 238], [['covid19', 'to', 'closing', 'punish'], 238], [['covid19', 'to', 'testing', 'closing'], 238], [['protesters', 'testing', 'closing', 'punish'], 238], [['protesters', 'to', 'closing', 'punish'], 238],

[['protesters', 'to', 'testing', 'closing'], 238], [['to', 'testing', 'closing', 'punish'], 238], [['covid19', 'to', 'pandemic', 'is'], 238], [['protesters', 'covid19', 'testing', 'punish'], 238], [['protesters', 'covid19', 'to', 'punish'], 238], [['protesters', 'covid19', 'testing', 'to'], 238], [['covid19', 'testing', 'to', 'punish'], 238], [['protesters', 'testing', 'to', 'punish'], 238], [['covid19', 'this', 'pandemic', 'to'], 237], [['this', 'to', 'in', 'is'], 237], [['covid19', 'this', 'pandemic', 'is'], 236], [['to', 'this', 'pandemic', 'is'], 236], [['covid19', 'to', 'pandemic', 'in'], 235], [['covid19', 'this', 'la', 'testing'], 235], [['covid19', 'an', 'is', 'in'], 234], [['covid19', 'an', 'pandemic', 'in'], 234], [['covid19', 'pandemic', 'in', 'is'], 234], [['covid19', 'this', 'pandemic', 'in'], 234], [['covid19', 'testing', 'to', 'in'], 234], [['this', 'pandemic', 'in', 'is'], 234], [['to', 'pandemic', 'in', 'is'], 234], [['this', 'theyre', 'to', 'is'], 234], [['an', 'centers', 'attempt', 'closing'], 233], [['covid19', 'an', 'centers', 'attempt'], 233], [['an', 'centers', 'attempt', 'in'], 233], [['an', 'centers', 'attempt', 'is'], 233], [['kissychalamets', 'an', 'centers', 'attempt'], 233], [['centers', 'an', 'la', 'attempt'], 233], [['an', 'centers', 'attempt', 'pandemic'], 233], [['protesters', 'an', 'centers', 'attempt'], 233], [['an', 'centers', 'attempt', 'punish'], 233], [['attempt', 'an', 'centers', 'sil\r'], 233], [['attempt', 'an', 'centers', 'testing'], 233], [['centers', 'an', 'theyre', 'attempt'], 233], [['centers', 'an', 'this', 'attempt'], 233], [['an', 'centers', 'attempt', 'to'], 233], [['an', 'centers', 'attempt', 'weaponizing'], 233], [['covid19', 'an', 'attempt', 'closing'], 233], [['an', 'attempt', 'closing', 'in'], 233], [['an', 'attempt', 'closing', 'is'], 233], [['kissychalamets', 'an', 'attempt', 'closing'], 233], [['an', 'la', 'attempt', 'closing'], 233], [['pandemic', 'an', 'attempt', 'closing'], 233], [['protesters', 'an', 'attempt', 'closing'], 233], [['an', 'attempt', 'closing', 'punish'], 233], [['an', 'attempt', 'sil\r', 'closing'], 233], [['an', 'attempt', 'testing', 'closing'], 233], [['an', 'theyre', 'attempt', 'closing'], 233], [['an', 'this', 'attempt', 'closing'], 233], [['to', 'an', 'attempt', 'closing'], 233], [['an', 'attempt', 'closing', 'weaponizing'], 233], [['covid19', 'an', 'attempt', 'in'], 233], [['covid19', 'an', 'attempt', 'is'], 233], [['kissychalamets', 'an', 'attempt', 'covid19'], 233], [['covid19', 'an', 'la', 'attempt'], 233], [['covid19', 'an', 'attempt', 'pandemic'], 233], [['protesters', 'an', 'attempt', 'covid19'], 233], [['covid19', 'an', 'attempt', 'punish'], 233], [['covid19', 'an', 'attempt', 'sil\r'], 233], [['covid19', 'an', 'attempt', 'testing'], 233], [['covid19', 'an', 'theyre', 'attempt'], 233], [['covid19', 'an', 'this', 'attempt'], 233], [['covid19', 'an', 'attempt', 'to'], 233], [['covid19', 'an', 'attempt', 'weaponizing'], 233], [['an', 'attempt', 'in', 'is'], 233], [['kissychalamets', 'an', 'attempt', 'in'], 233], [['an', 'la', 'attempt', 'in'], 233], [['an', 'attempt', 'pandemic', 'in'], 233]]

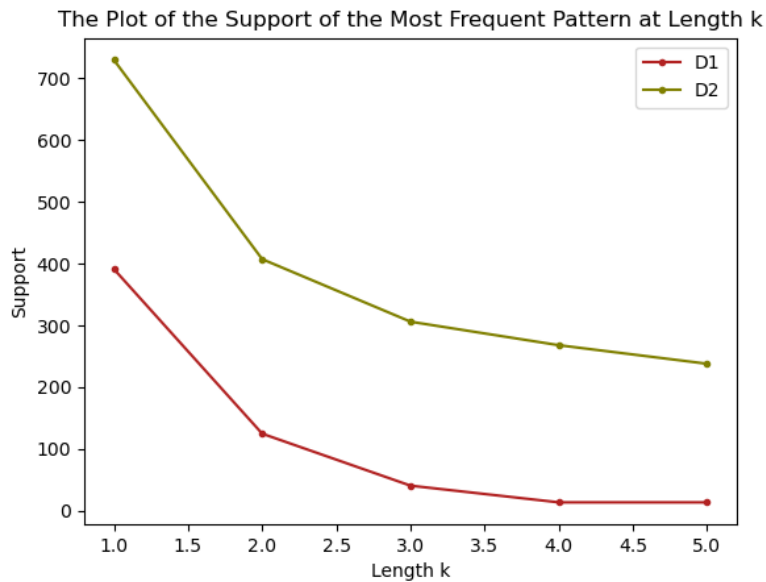
### The top 100 length-5 patterns for D2:

[[['closing', 'punish', 'protesters', 'covid19', 'centers'], 238], [['closing',

'protesters', 'testing', 'covid19', 'centers'], 238], [['closing', 'protesters', 'to', 'covid19', 'centers'], 238], [['closing', 'punish', 'testing', 'covid19', 'centers'], 238], [['closing', 'punish', 'to', 'covid19', 'centers'], 238], [['closing', 'testing', 'to', 'covid19', 'centers'], 238], [['closing', 'punish', 'protesters', 'testing', 'centers'], 238], [['closing', 'punish', 'protesters', 'to', 'centers'], 238], [['closing', 'protesters', 'testing', 'to', 'centers'], 238], [['closing', 'punish', 'testing', 'to', 'centers'], 238], [['punish', 'protesters', 'testing', 'covid19', 'centers'], 238], [['punish', 'protesters', 'to', 'covid19', 'centers'], 238], [['protesters', 'testing', 'to', 'covid19', 'centers'], 238], [['punish', 'testing', 'to', 'covid19', 'centers'], 238], [['punish', 'protesters', 'testing', 'to', 'centers'], 238], [['closing', 'punish', 'protesters', 'testing', 'covid19'], 238], [['closing', 'punish', 'protesters', 'to', 'covid19'], 238], [['closing', 'protesters', 'testing', 'to', 'covid19'], 238], [['closing', 'punish', 'testing', 'to', 'covid19'], 238], [['closing', 'punish', 'protesters', 'testing', 'to'], 238], [['punish', 'protesters', 'testing', 'to', 'covid19'], 238], [['this', 'is', 'to', 'covid19', 'in'], 236], [['this', 'is', 'pandemic', 'to', 'covid19'], 236], [['is', 'pandemic', 'to', 'covid19', 'in'], 234], [['attempt', 'closing', 'an', 'covid19', 'centers'], 233], [['attempt', 'closing', 'an', 'centers', 'in'], 233], [['is', 'attempt', 'closing', 'an', 'centers'], 233], [['attempt', 'closing', 'an', 'kissychalamets', 'centers'], 233], [['la', 'attempt', 'closing', 'an', 'centers'], 233], [['attempt', 'pandemic', 'closing', 'an', 'centers'], 233], [['attempt', 'closing', 'protesters', 'an', 'centers'], 233], [['attempt', 'closing', 'punish', 'an', 'centers'], 233], [['attempt', 'closing', 'an', 'silr', 'centers'], 233], [['attempt', 'closing', 'an', 'testing', 'centers'], 233], [['attempt', 'closing', 'an', 'theyre', 'centers'], 233], [['this', 'attempt', 'closing', 'an', 'centers'], 233], [['attempt', 'closing', 'an', 'to', 'centers'], 233], [['weaponizing', 'attempt', 'closing', 'an', 'centers'], 233], [['attempt', 'an', 'covid19', 'centers', 'in'], 233], [['is', 'attempt', 'an', 'covid19', 'centers'], 233], [['covid19', 'attempt', 'an', 'kissychalamets', 'centers'], 233], [['la', 'attempt', 'an', 'covid19', 'centers'], 233], [['attempt', 'pandemic', 'an', 'covid19', 'centers'], 233], [['attempt', 'protesters', 'an', 'covid19', 'centers'], 233], [['attempt', 'punish', 'an', 'covid19', 'centers'], 233], [['attempt', 'an', 'covid19', 'silr', 'centers'], 233], [['attempt', 'an', 'testing', 'covid19', 'centers'], 233], [['attempt', 'an', 'theyre', 'covid19', 'centers'], 233], [['this', 'attempt', 'an', 'covid19', 'centers'], 233], [['attempt', 'an', 'to', 'covid19', 'centers'], 233], [['weaponizing', 'attempt', 'an', 'covid19', 'centers'], 233], [['is', 'attempt', 'an', 'centers', 'in'], 233], [['attempt', 'an', 'kissychalamets', 'centers', 'in'], 233], [['la', 'attempt', 'an', 'centers', 'in'], 233], [['attempt', 'pandemic', 'an', 'centers', 'in'], 233], [['attempt', 'protesters', 'an', 'centers', 'in'], 233], [['attempt', 'punish', 'an', 'centers', 'in'], 233], [['attempt', 'an', 'silr', 'centers', 'in'], 233], [['attempt', 'an', 'testing', 'centers', 'in'], 233], [['attempt', 'an', 'theyre', 'centers', 'in'], 233], [['this', 'attempt', 'an',

'centers', 'in'], 233], [['attempt', 'an', 'to', 'centers', 'in'], 233], [['weaponizing', 'attempt', 'an', 'centers', 'in'], 233], [['is', 'attempt', 'an', 'kissychalamets', 'centers'], 233], [['la', 'is', 'attempt', 'an', 'centers'], 233], [['is', 'attempt', 'pandemic', 'an', 'centers'], 233], [['is', 'attempt', 'protesters', 'an', 'centers'], 233], [['is', 'attempt', 'punish', 'an', 'centers'], 233], [['is', 'attempt', 'an', 'silr', 'centers'], 233], [['is', 'attempt', 'an', 'testing', 'centers'], 233], [['is', 'attempt', 'an', 'theyre', 'centers'], 233], [['this', 'is', 'attempt', 'an', 'centers'], 233], [['is', 'attempt', 'an', 'to', 'centers'], 233], [['weaponizing', 'is', 'attempt', 'an', 'centers'], 233], [['la', 'attempt', 'an', 'kissychalamets', 'centers'], 233], [['attempt', 'pandemic', 'an', 'kissychalamets', 'centers'], 233], [['attempt', 'protesters', 'an', 'kissychalamets', 'centers'], 233], [['attempt', 'punish', 'an', 'kissychalamets', 'centers'], 233], [['attempt', 'an', 'kissychalamets', 'silr', 'centers'], 233], [['attempt', 'an', 'testing', 'kissychalamets', 'centers'], 233], [['attempt', 'an', 'theyre', 'kissychalamets', 'centers'], 233], [['this', 'attempt', 'an', 'kissychalamets', 'centers'], 233], [['attempt', 'an', 'to', 'kissychalamets', 'centers'], 233], [['weaponizing', 'attempt', 'an', 'kissychalamets', 'centers'], 233], [['la', 'attempt', 'pandemic', 'an', 'centers'], 233], [['la', 'attempt', 'protesters', 'an', 'centers'], 233], [['la', 'attempt', 'punish', 'an', 'centers'], 233], [['la', 'attempt', 'an', 'silr', 'centers'], 233], [['la', 'attempt', 'an', 'testing', 'centers'], 233], [['la', 'attempt', 'an', 'theyre', 'centers'], 233], [['this', 'la', 'attempt', 'an', 'centers'], 233], [['la', 'attempt', 'an', 'to', 'centers'], 233], [['la', 'weaponizing', 'attempt', 'an', 'centers'], 233], [['attempt', 'pandemic', 'protesters', 'an', 'centers'], 233], [['attempt', 'pandemic', 'punish', 'an', 'centers'], 233], [['attempt', 'pandemic', 'an', 'silr', 'centers'], 233], [['attempt', 'pandemic', 'an', 'testing', 'centers'], 233], [['attempt', 'pandemic', 'an', 'theyre', 'centers'], 233], [['this', 'attempt', 'pandemic', 'an', 'centers'], 233], [['attempt', 'pandemic', 'an', 'to', 'centers'], 233]]

3.



The power law distribution is  $f(x) = ax^{-c}$ . To find the parameter  $a$  and  $c$ , I use the function `curve_fit()` find to the parameters of the distribution.

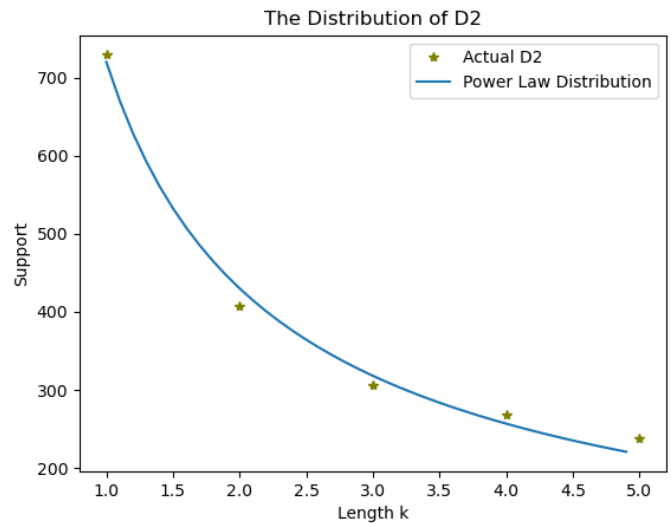
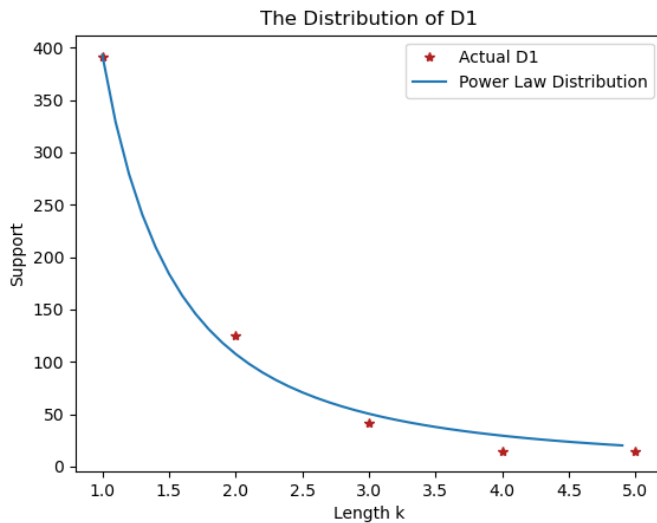
The curves of D1 and D2 fit the power law distribution.

The parameters for the power law distribution of D1:

$a = 393.19221869516963$  and  $c = 1.871479994340617$

The parameters for the power law distribution of D2:

$a = 720.0108535015611$  and  $c = 0.7435016639252436$





## Question 4

1. For this question, we can find all patterns of length up to 4 for D1 with a minimum support 5 and all patterns of length up to 4 for D2 using the Apriori algorithm. We can add another column “odd”. For each frequent pattern in D1, calculate and store its *odd* value by finding its support in D2. Then sort those frequent patterns by their *odd* in descending order. We can take the first 100 patterns.

Since  $\text{sup}_{D_1}(X) \geq 5$ ,  $\text{odd}(X) = \frac{\text{sup}_{D_2}(X)}{\text{sup}_{D_1}(X)}$ , if  $\text{sup}_{D_2}(X) < s$ , then there must be  $\text{odd}(X) < \frac{s}{\text{sup}_{D_1}(X)} \leq \frac{s}{5}$ . If we can make sure that odd values of top 100 patterns are all higher than  $\frac{s}{5}$ , there will be no effect for ignoring the support value less than  $s$ . To reduce the search space, when finding all patterns of length up to 4 for D2, I let the value of the minimum support for D2 be 20. The 100<sup>th</sup> highest odd of pattern is 4.083333333333333, which is higher than  $\frac{20}{5} = 4$ . Therefore, setting the minimum support as 20 is no effect on the result.

The code is in the file “code.py”. I modified the code from assignment 1.

2. `[[['covid19'], 121.5], [['covid19', 'to'], 81.4], [['an', 'is', 'this'], 46.6], [['protesters', 'to'], 40.0], [['in', 'an', 'is'], 39.0], [['la'], 34.285714285714285], [['in', 'an'], 33.714285714285715], [['theyre', 'to'], 29.25], [['an', 'this'], 29.125], [['in', 'is', 'to'], 27.1], [['theyre'], 23.6], [['in', 'this', 'to'], 22.181818181818183], [['protesters'], 20.166666666666668], [['in', 'is', 'this'], 18.692307692307693], [['an', 'to'], 18.0], [['not', 'a', 'i'], 16.8], [['an', 'is'], 16.785714285714285], [['a', 'were'], 16.2], [['can', 'i'], 15.6], [['do', 'a', 'you'], 15.6], [['a', 'if', 'protest'], 15.6], [['were', 'you'], 13.5], [['is', 'to', 'this'], 13.444444444444445], [['not', 'if'],`



13.166666666666666], [['a', 'need'], 13.0], [['if', 'i', 'you'],  
11.285714285714286], [['telling'], 11.142857142857142], [['if', 'protest'],  
11.142857142857142], [['do', 'a'], 9.875], [['care'], 9.333333333333334], [['a',  
'if', 'you'], 9.272727272727273], [['at', 'i'], 9.222222222222221], [['not', 'do'],  
8.777777777777779], [['a', 'can'], 8.7], [['a', 'protest'], 8.4], [['do', 'you'], 8.1],  
[['in', 'this'], 7.696969696969697], [['in', 'is'], 7.631578947368421], [['country'],  
7.333333333333333], [['in', 'to'], 7.127659574468085], [['here', 'the'], 7.0],  
[['an'], 6.702702702702703], [['can', 'you'], 6.583333333333333], [['like', 'to'],  
6.428571428571429], [['in', 'get'], 6.4], [['in', 'and', 'who'], 6.4], [['many'],  
6.214285714285714], [['what', 'you', 'the'], 6.2], [['not', 'i'], 6.0], [['get', 'for'],  
6.0], [['from', 'who'], 6.0], [['this', 'to'], 5.954545454545454], [['a', 'as', 'of'], 5.8],  
[['and', 'get', 'to'], 5.8], [['in', 'is', 'my'], 5.8], [['and', 'of', 'who'], 5.6], [['not', 'you',  
'for'], 5.6], [['not', 'you'], 5.578947368421052], [['here', 'for'], 5.4], [['if', 'i'], 5.25],  
[['what', 'for'], 5.2], [['anyone'], 5.166666666666667], [['for', 'has'],  
5.166666666666667], [['two'], 5.1], [['we', 'a', 'of'], 4.857142857142857], [['in',  
'of', 'and'], 4.833333333333333], [['in', 'of', 'to'], 4.8], [['from', 'you'],  
4.666666666666667], [['of', 'like'], 4.625], [['we', 'in', 'the'], 4.6], [['of', 'get'],  
4.571428571428571], [['the', 'get'], 4.571428571428571], [['in', 'and', 'is', 'to'],  
4.571428571428571], [['need'], 4.555555555555555], [['in', 'is', 'of'],  
4.555555555555555], [['police', 'for'], 4.5], [['not', 'who'], 4.5], [['from', 'police',  
'the'], 4.5], [['new'], 4.444444444444445], [['away'], 4.428571428571429],  
[['and', 'two'], 4.428571428571429], [['is', 'get'], 4.428571428571429], [['a', 'at'],  
4.421052631578948], [['in', 'as', 'a'], 4.4], [['in', 'like'], 4.333333333333333],  
[['one', 'is'], 4.333333333333333], [['not', 'police'], 4.333333333333333], [['the',  
'you', 'who'], 4.333333333333333], [['my', 'of', 'and'], 4.285714285714286],  
[['is', 'who'], 4.25], [['the', 'still'], 4.2], [['in', 'all', 'to'], 4.2], [['we', 'a', 'and', 'the'],  
4.2], [['is', 'to'], 4.194444444444445], [['not', 'a'], 4.190476190476191], [['i',  
'you'], 4.15], [['as', 'of'], 4.125], [['a', 'if'], 4.115384615384615], [['were'],  
4.086956521739131], [['a', 'has'], 4.083333333333333]]