Job Reports 2022

Table of Contents

```python
import pandas as pd
import os, json
import pandas as pd
import numpy as np
import glob
import re
import sys

pd.set_option('display.max_colwidth', None)
pd.set_option('display.max_rows', None)
# pd.set_option('display.max_columns', None)
# pd.set_option('display.width', None)
np.set_printoptions(threshold=sys.maxsize)
```

```python
companies = {
    '1': 'Safaricom',
    '2': 'EABL',
    '3': 'UN, United Nations',
    '4': 'KRA',
    '5': 'KenGen',
    '6': 'KPC, Kenya Pipeline Company',
    '7': 'Coca cola',
    '8':'Deloitte',
    '9': 'Unilever',
    '10': 'Kplc, Kenya Power',
    '11': 'Google',
    '12': 'KPA, Kenya Ports Authority',
    '13': 'KCB, Kenya Commercial Bank',
    '14': 'Microsoft',
    '15': 'Bidco',
    '16': 'RMS, Royal Media Services',
    '17': 'KEMRI, Kenya Medical Research Institute',
    '18': 'Amref'
}
```

Export tweets (json format)

```python
# f = open('./dayta/1-json.json')
arr = []
path_to_json = './dayta/'
json_pattern = os.path.join(path_to_json, '*.json')
file_list = glob.glob(json_pattern)

# print(file_list)

for file in file_list:
    with open(file) as f:
        temp = json.load(f)
        for line in temp['data']:
            arr.append(line)



df = pd.DataFrame(data=arr)
# save to csv
df.to_csv('./out/collated.csv')

df = df[['text', 'created_at']]

# df.head(100)
```

Text Cleaning and Extraction

Pre-processing

```python
txt_lower = df['text'].str.lower()
txt_l_cln = txt_lower.str.replace("\@wanjikureports", "") #replace wanjiku
mentions
txt_l_cln2 = txt_l_cln.str.replace("(\n*\d\.)", '') #replace ordered
listings e.g. 1., 2. etc
narr = txt_l_cln2.to_numpy()
```

Pseudocode

Each listitem in list

- Remove all stop words
- Split all words/numbers
- Check whether word in company list, if in list add to output var -> use eigenvalues, check for closeness of word
- check each # with company list key, add value to output list

```python
from difflib import SequenceMatcher
import jellyfish


def similar(a, b):
    return jellyfish.jaro_distance(a, b)  # prob threshhold set at 0.75
    # return SequenceMatcher(None, a, b).ratio()


def returnFromNumeric(numstr):
    if 1 <= int(numstr) <= 18:
        splt = companies[numstr].split(',')
        return splt[0]

# https://stackoverflow.com/questions/51217909/removing-all-emojis-from-
text
def replaceEmoji(text):
    emoji_pattern = re.compile("["
                                u"\U0001F600-\U0001F64F"  # emoticons
                                u"\U0001F300-\U0001F5FF"  # symbols &
pictographs
                                u"\U0001F680-\U0001F6FF"  # transport & map
symbols
                                u"\U0001F1E0-\U0001F1FF"  # flags (iOS)
                                u"\U0001F1F2-\U0001F1F4"  # Macau flag
                                u"\U0001F1E6-\U0001F1FF"  # flags
                                u"\U0001F600-\U0001F64F"
                                u"\U00002702-\U000027B0"
                                u"\U000024C2-\U0001F251"
                                u"\U0001f926-\U0001f937"
                                u"\U0001F1F2"
                                u"\U0001F1F4"
                                u"\U0001F620"
                                u"\u200d"
                                u"\u2640-\u2642"
                                "]+", flags=re.UNICODE)


    return emoji_pattern.sub(r'', text)


def returnFromStr(thestr):
    output = None

    for val in companies.values():
        splt = val.split(',')

        selected = splt[0]
        prob = similar(thestr, splt[0].lower())

        if prob < 0.75 and len(splt) > 1:
            # selected = splt[1]
            prob = similar(thestr, splt[1].lower())
```

```python
        if prob >= .75:
            output = selected

    return output
```

```python
similar('kenya_ports', 'Kenya Ports Authority'.lower())

# thevar = 'KPA, Kenya pipes'.split(',')
# print(thevar[0])
# if (len(thevar) > 1):
#     print(thevar[1])
```

```
0.7950937950937952
```

```python
output_arr=[]
test = narr[:5]

for i,txt in enumerate(narr):
    if i not in [908, 1073, 1096]:
        tokens = word_tokenize(txt)
        print(i)
        for token in tokens:
            if token.isnumeric():
                output_arr.append(returnFromNumeric(token))
            else:
                output_arr.append(returnFromStr(token))
```

```
0
1
2
3
4
5
6
7
8
9
10
11
12
13
```

14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64

```
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
```

```
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
```

167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217

```
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
```

269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319

```
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
```

```
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
```

```
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
```

```
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
```

```
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
```

575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625

```
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
```

677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727

```
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
```

779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829

830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880

```
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
```

933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983

984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034

```
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
```

```
1087
1088
1089
1090
1091
1092
1093
1094
1095
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
```

```
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
```

1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240

```
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
```

```
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
```

```
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
```

1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444

```
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
```

```
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
```

```
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
```

```
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
```

```python
# tt= narr[908]
# nt = replaceEmoji(tt)
# print(nt)

# word_tokenize(nt)
# # narr[908]
```

```python
finalarr = [txt for txt in output_arr if txt is not None]

finalarr
```

```
['KPA',
 'KPC',
 'Coca cola',
 'Amref',
 'Amref',
 'KPA',
 'EABL',
 'Google',
 'Microsoft',
 'Deloitte',
 'Unilever',
 'KPA',
 'KCB',
 'KCB',
 'Amref',
 'Google',
 'UN',
```

```
'Google',
'UN',
'Kplc',
'Google',
'Safaricom',
'EABL',
'KPA',
'UN',
'KPC',
'Unilever',
'Safaricom',
'KPC',
'UN',
'EABL',
'Safaricom',
'EABL',
'Kplc',
'Bidco',
'Coca cola',
'KRA',
'Google',
'Microsoft',
'Amref',
'EABL',
'Amref',
'KPA',
'KenGen',
'RMS',
'UN',
'Unilever',
'Kplc',
'EABL',
'EABL',
'EABL',
'Unilever',
'KCB',
'Amref',
'Safaricom',
'Amref',
'Safaricom',
'KPA',
'KenGen',
'EABL',
'Safaricom',
'Coca cola',
'KCB',
'Google',
'Microsoft',
'Safaricom',
'UN',
'KenGen',
```

```
'Unilever',
'Unilever',
'UN',
'Deloitte',
'Google',
'Deloitte',
'KRA',
'Deloitte',
'EABL',
'Coca cola',
'Safaricom',
'KPA',
'Unilever',
'Safaricom',
'KPA',
'Google',
'EABL',
'UN',
'Amref',
'Unilever',
'UN',
'Google',
'UN',
'Microsoft',
'Unilever',
'Amref',
'Microsoft',
'Microsoft',
'UN',
'KenGen',
'UN',
'Google',
'KPA',
'KenGen',
'Google',
'KCB',
'EABL',
'Bidco',
'KPA',
'EABL',
'UN',
'Amref',
'Google',
'Amref',
'Microsoft',
'UN',
'Bidco',
'UN',
'Safaricom',
'KenGen',
'UN',
```

```
    'UN',
    'KCB',
    'KRA',
    'KRA',
    'KPC',
    'Unilever',
    'Safaricom',
    'Google',
    'Microsoft',
    'Coca cola',
    'KCB',
    'Amref',
    'Amref',
    'KenGen',
    'EABL',
    'KenGen',
    'KenGen',
    'UN',
    'EABL',
    'KPA',
    'EABL',
    'KenGen',
    'EABL',
    'Unilever',
    'Unilever',
    'UN',
    'KPA',
    'Safaricom',
    'Safaricom',
    'KenGen',
    'KEMRI',
    'Unilever',
    'Google',
    'Microsoft',
    'KPA',
    'Google',
    'Amref',
    'EABL',
    'Safaricom',
    'EABL',
    'KPA',
    'KCB',
    'Unilever',
    'Coca cola',
    'EABL',
    'Google',
    'KenGen',
    'KenGen',
    'Kplc',
    'KenGen',
    'Unilever',
```

```
'Deloitte',
'KPA',
'Unilever',
'Safaricom',
'KenGen',
'Google',
'Microsoft',
'Amref',
'Unilever',
'UN',
'Google',
'KenGen',
'Microsoft',
'Microsoft',
'KenGen',
'Deloitte',
'UN',
'KRA',
'UN',
'EABL',
'Google',
'Microsoft',
'Safaricom',
'KenGen',
'Bidco',
'EABL',
'EABL',
'Google',
'Kplc',
'Google',
'KCB',
'Microsoft',
'Safaricom',
'KenGen',
'UN',
'Amref',
'Deloitte',
'Amref',
'Safaricom',
'Safaricom',
'EABL',
'Amref',
'KCB',
'KCB',
'KenGen',
'EABL',
'EABL',
'EABL',
'UN',
'UN',
'EABL',
```

```
'KCB',
'Safaricom',
'Safaricom',
'KPA',
'KRA',
'Safaricom',
'Unilever',
'Google',
'Unilever',
'KPA',
'Microsoft',
'Safaricom',
'Unilever',
'KCB',
'Kplc',
'Safaricom',
'EABL',
'Coca cola',
'Coca cola',
'Unilever',
'Google',
'Unilever',
'Unilever',
'UN',
'Google',
'Kplc',
'KPC',
'KenGen',
'EABL',
'KPC',
'EABL',
'EABL',
'EABL',
'Unilever',
'Coca cola',
'Bidco',
'Bidco',
'Microsoft',
'Google',
'Unilever',
'KRA',
'KRA',
'Google',
'Unilever',
'Unilever',
'UN',
'KEMRI',
'Deloitte',
'Google',
'UN',
'KenGen',
```

```
'EABL',
'KPC',
'Unilever',
'Unilever',
'EABL',
'Kplc',
'Bidco',
'Bidco',
'Amref',
'Google',
'KCB',
'Amref',
'RMS',
'Unilever',
'Google',
'Google',
'Amref',
'KenGen',
'Amref',
'Safaricom',
'Coca cola',
'EABL',
'UN',
'Microsoft',
'Amref',
'Coca cola',
'KPA',
'KenGen',
'Kplc',
'UN',
'Google',
'Bidco',
'Safaricom',
'Safaricom',
'EABL',
'Kplc',
'EABL',
'Kplc',
'Microsoft',
'Safaricom',
'EABL',
'Google',
'Safaricom',
'Unilever',
'Google',
'KCB',
'Amref',
'Google',
'Unilever',
'Safaricom',
'Unilever',
```

```
'Deloitte',
'EABL',
'Deloitte',
'EABL',
'Google',
'Deloitte',
'Deloitte',
'Amref',
'Safaricom',
'KRA',
'Unilever',
'Amref',
'Deloitte',
'EABL',
'Google',
'Unilever',
'Amref',
'Safaricom',
'EABL',
'UN',
'Coca cola',
'Deloitte',
'Unilever',
'Google',
'Unilever',
'KEMRI',
'KenGen',
'KenGen',
'Google',
'Unilever',
'KPA',
'UN',
'Google',
'Google',
'KPA',
'KRA',
'UN',
'UN',
'KPA',
'KRA',
'Unilever',
'Amref',
'Amref',
'KRA',
'KCB',
'RMS',
'EABL',
'Kplc',
'KenGen',
'Unilever',
'KEMRI',
```

```
'EABL',
'UN',
'Coca cola',
'Amref',
'EABL',
'EABL',
'Deloitte',
'Safaricom',
'Amref',
'Google',
'KenGen',
'KPC',
'KRA',
'KRA',
'Google',
'UN',
'Google',
'KCB',
'KRA',
'Unilever',
'Deloitte',
'Safaricom',
'KenGen',
'UN',
'UN',
'KenGen',
'KPC',
'Safaricom',
'KRA',
'KenGen',
'UN',
'KenGen',
'KPA',
'Kplc',
'UN',
'RMS',
'EABL',
'KenGen',
'KRA',
'UN',
'Unilever',
'Amref',
'KenGen',
'EABL',
'KenGen',
'Deloitte',
'Microsoft',
'UN',
'Kplc',
'Amref',
'Unilever',
```

```
'KenGen',
'KenGen',
'Unilever',
'Unilever',
'Unilever',
'Amref',
'UN',
'Safaricom',
'Safaricom',
'KEMRI',
'KEMRI',
'EABL',
'Amref',
'Deloitte',
'Kplc',
'KEMRI',
'EABL',
'UN',
'Microsoft',
'Deloitte',
'Google',
'KenGen',
'Google',
'Safaricom',
'Safaricom',
'Safaricom',
'KEMRI',
'KEMRI',
'EABL',
'Unilever',
'KEMRI',
'Coca cola',
'EABL',
'UN',
'Bidco',
'KenGen',
'Google',
'Microsoft',
'KenGen',
'Unilever',
'EABL',
'Safaricom',
'Google',
'Amref',
'UN',
'EABL',
'Google',
'Microsoft',
'Unilever',
'KenGen',
'Amref',
```

```
    'KRA',
    'KenGen',
    'Amref',
    'Amref',
    'KRA',
    'Coca cola',
    'Safaricom',
    'Google',
    'Kplc',
    'Amref',
    'Coca cola',
    'UN',
    'KenGen',
    'EABL',
    'Deloitte',
    'Amref',
    'KCB',
    'Google',
    'EABL',
    'KCB',
    'Amref',
    'KEMRI',
    'Amref',
    'Google',
    'Unilever',
    'UN',
    'KRA',
    'KenGen',
    'EABL',
    'EABL',
    'EABL',
    'Safaricom',
    'EABL',
    'Unilever',
    'EABL',
    'Bidco',
    'EABL',
    'RMS',
    'Microsoft',
    'Microsoft',
    'KenGen',
    'Amref',
    'EABL',
    'KenGen',
    'Microsoft',
    'KPC',
    'EABL',
    'Microsoft',
    'Unilever',
    'KPA',
    'Unilever',
```

```
'EABL',
'UN',
'Kplc',
'Amref',
'Google',
'Coca cola',
'EABL',
'Amref',
'Microsoft',
'Bidco',
'Unilever',
'Unilever',
'Kplc',
'KPA',
'KenGen',
'EABL',
'Amref',
'UN',
'Google',
'EABL',
'Google',
'Google',
'Bidco',
'Amref',
'KPA',
'Unilever',
'KRA',
'KRA',
'Kplc',
'UN',
'Coca cola',
'UN',
'KPA',
'Kplc',
'KenGen',
'Kplc',
'Amref',
'KPA',
'UN',
'Deloitte',
'Unilever',
'Safaricom',
'KCB',
'Safaricom',
'Safaricom',
'Google',
'Bidco',
'UN',
'Unilever',
'Kplc',
'Unilever',
```

```
        'Amref',
        'EABL',
        'Amref',
        'Deloitte',
        'Amref',
        'Amref',
        'Kplc',
        'Unilever',
        'Unilever',
        'Safaricom',
        'EABL',
        'Amref',
        'Amref',
        'Amref',
        'Safaricom',
        'Google',
        'KCB',
        'KenGen',
        'EABL',
        'Amref',
        'EABL',
        'EABL',
        'EABL',
        'Amref',
        'Google',
        'Microsoft',
        'UN',
        'Kplc',
        'EABL',
        'UN',
        'Unilever',
        'Amref',
        'Unilever',
        'UN',
        'EABL',
        'KEMRI',
        'EABL',
        'Safaricom',
        'Bidco',
        'EABL',
        'Amref',
        'KPA',
        'UN',
        'Deloitte',
        'Coca cola',
        'Google',
        'Safaricom',
        'KRA',
        'Amref',
        'UN',
        'Amref',
```

```
'Safaricom',
'UN',
'UN',
'Deloitte',
'Google',
'Microsoft',
'UN',
'Amref',
'EABL',
'Amref',
'EABL',
'Bidco',
'Microsoft',
'EABL',
'Amref',
'KPC',
'Kplc',
'Bidco',
'Safaricom',
'EABL',
'Unilever',
'EABL',
'EABL',
'EABL',
'Coca cola',
'Kplc',
'Unilever',
'Safaricom',
'Safaricom',
'EABL',
'Unilever',
'Google',
'EABL',
'Safaricom',
'Safaricom',
'Unilever',
'Unilever',
'KPA',
'EABL',
'Unilever',
'Amref',
'KPA',
'Unilever',
'Kplc',
'Unilever',
'KPA',
'Unilever',
'Amref',
'EABL',
'Kplc',
'Amref',
```

```
'Safaricom',
'KPA',
'KPA',
'KPA',
'Google',
'Microsoft',
'EABL',
'Bidco',
'EABL',
'EABL',
'KenGen',
'Google',
'KenGen',
'KPC',
'EABL',
'Google',
'Deloitte',
'Deloitte',
'EABL',
'Deloitte',
'Microsoft',
'KenGen',
'Google',
'Amref',
'UN',
'Unilever',
'Safaricom',
'Deloitte',
'KCB',
'UN',
'Unilever',
'Unilever',
'KCB',
'Kplc',
'EABL',
'Unilever',
'Bidco',
'KPA',
'Amref',
'EABL',
'Safaricom',
'UN',
'Unilever',
'Kplc',
'KenGen',
'EABL',
'Microsoft',
'Google',
'Google',
'EABL',
'EABL',
```

```
'EABL',
'Amref',
'Unilever',
'KPA',
'Safaricom',
'Safaricom',
'EABL',
'UN',
'KCB',
'Unilever',
'Google',
'Microsoft',
'Google',
'KCB',
'Google',
'KenGen',
'KRA',
'Bidco',
'EABL',
'KPA',
'Google',
'Microsoft',
'Google',
'Microsoft',
'Bidco',
'Bidco',
'Unilever',
'Google',
'Bidco',
'EABL',
'KPA',
'Unilever',
'Microsoft',
'Amref',
'Amref',
'EABL',
'UN',
'Safaricom',
'UN',
'Amref',
'Google',
'Google',
'Safaricom',
'Safaricom',
'Unilever',
'EABL',
'EABL',
'Deloitte',
'Deloitte',
'KenGen',
'EABL',
```

```
    'Deloitte',
    'Microsoft',
    'Microsoft',
    'UN',
    'KPA',
    'KPA',
    'Microsoft',
    'UN',
    'Deloitte',
    'Kplc',
    'UN',
    'KenGen',
    'Safaricom',
    'Bidco',
    'Unilever',
    'Unilever',
    'UN',
    'KenGen',
    'Safaricom',
    'EABL',
    'KenGen',
    'Unilever',
    'EABL',
    'Kplc',
    'Coca cola',
    'EABL',
    'Safaricom',
    'EABL',
    'Deloitte',
    'Unilever',
    'Kplc',
    'Unilever',
    'Amref',
    'EABL',
    'KRA',
    'Safaricom',
    'Unilever',
    'Amref',
    'KenGen',
    'Microsoft',
    'Microsoft',
    'Deloitte',
    'Amref',
    'EABL',
    'Google',
    'Kplc',
    'Deloitte',
    'Deloitte',
    'EABL',
    'Coca cola',
    'KenGen',
```

```
'Kplc',
'KenGen',
'UN',
'Bidco',
'Amref',
'Kplc',
'Microsoft',
'Safaricom',
'KPA',
'Safaricom',
'Bidco',
'KRA',
'EABL',
'Bidco',
'KPA',
'Deloitte',
'KRA',
'Amref',
'Google',
'Amref',
'Microsoft',
'Kplc',
'Unilever',
'KEMRI',
'Safaricom',
'Coca cola',
'Google',
'UN',
'Safaricom',
'UN',
'UN',
'KRA',
'KRA',
'Microsoft',
'Google',
'KRA',
'UN',
'EABL',
'UN',
'Google',
'KEMRI',
'Unilever',
'Amref',
'KRA',
'Amref',
'Deloitte',
'Microsoft',
'KenGen',
'EABL',
'Deloitte',
'Google',
```

```
'Amref',
'Unilever',
'EABL',
'Google',
'KenGen',
'KenGen',
'Safaricom',
'Safaricom',
'Safaricom',
'KenGen',
'Amref',
'UN',
'Safaricom',
'Microsoft',
'EABL',
'UN',
'Kplc',
'Unilever',
'KEMRI',
'Amref',
'KPA',
'Unilever',
'EABL',
'EABL',
'Amref',
'UN',
'Google',
'Kplc',
'Google',
'Google',
'Bidco',
'Kplc',
'Kplc',
'Google',
'KRA',
'KPA',
'KenGen',
'EABL',
'Deloitte',
'Deloitte',
'Unilever',
'Coca cola',
'EABL',
'UN',
'Coca cola',
'Unilever',
'EABL',
'Amref',
'KEMRI',
'EABL',
'Kplc',
```

```
    'KenGen',
    'Safaricom',
    'Safaricom',
    'Safaricom',
    'EABL',
    'Safaricom',
    'Safaricom',
    'KenGen',
    'Amref',
    'Unilever',
    'UN',
    'EABL',
    'Deloitte',
    'Microsoft',
    'Safaricom',
    'Google',
    'Bidco',
    'KPA',
    'KenGen',
    'EABL',
    'KCB',
    'UN',
    'KRA',
    'Microsoft',
    'KenGen',
    'EABL',
    'Amref',
    'UN',
    'Amref',
    'Bidco',
    'EABL',
    'KRA',
    'Google',
    'Unilever',
    'Deloitte',
    'KenGen',
    'Safaricom',
    'Unilever',
    'KenGen',
    'KRA',
    'Amref',
    'KCB',
    'KenGen',
    'Kplc',
    'Amref',
    'Kplc',
    'Unilever',
    'Unilever',
    'Microsoft',
    'Unilever',
    'Amref',
```

```
 'Google',
 'Google',
 'Bidco',
 'Safaricom',
 'Deloitte',
 'Unilever',
 'Coca cola',
 'Google',
 'KPC',
 'Amref',
 'Amref',
 'EABL',
 'Google',
 'KenGen',
 ...]
```

```
corpus = "; ".join(finalarr)
corpus
```

```
'KPA; KPC; Coca cola; Amref; Amref; KPA; EABL; Google; Microsoft;
Deloitte; Unilever; KPA; KCB; KCB; Amref; Google; UN; Google; UN; Kplc;
Google; Safaricom; EABL; KPA; UN; KPC; Unilever; Safaricom; KPC; UN; EABL;
Safaricom; EABL; Kplc; Bidco; Coca cola; KRA; Google; Microsoft; Amref;
EABL; Amref; KPA; KenGen; RMS; UN; Unilever; Kplc; EABL; EABL; EABL;
Unilever; KCB; Amref; Safaricom; Amref; Safaricom; KPA; KenGen; EABL;
Safaricom; Coca cola; KCB; Google; Microsoft; Safaricom; UN; KenGen;
Unilever; Unilever; UN; Deloitte; Google; Deloitte; KRA; Deloitte; EABL;
Coca cola; Safaricom; KPA; Unilever; Safaricom; KPA; Google; EABL; UN;
Amref; Unilever; UN; Google; UN; Microsoft; Unilever; Amref; Microsoft;
Microsoft; UN; KenGen; UN; Google; KPA; KenGen; Google; KCB; EABL; Bidco;
KPA; EABL; UN; Amref; Google; Amref; Microsoft; UN; Bidco; UN; Safaricom;
KenGen; UN; UN; KCB; KRA; KRA; KPC; Unilever; Safaricom; Google;
Microsoft; Coca cola; KCB; Amref; Amref; KenGen; EABL; KenGen; KenGen; UN;
EABL; KPA; EABL; KenGen; EABL; Unilever; Unilever; UN; KPA; Safaricom;
Safaricom; KenGen; KEMRI; Unilever; Google; Microsoft; KPA; Google; Amref;
EABL; Safaricom; EABL; KPA; KCB; Unilever; Coca cola; EABL; Google;
KenGen; KenGen; Kplc; KenGen; Unilever; Deloitte; KPA; Unilever;
Safaricom; KenGen; Google; Microsoft; Amref; Unilever; UN; Google; KenGen;
Microsoft; Microsoft; KenGen; Deloitte; UN; KRA; UN; EABL; Google;
Microsoft; Safaricom; KenGen; Bidco; EABL; EABL; Google; Kplc; Google;
KCB; Microsoft; Safaricom; KenGen; UN; Amref; Deloitte; Amref; Safaricom;
Safaricom; EABL; Amref; KCB; KCB; KenGen; EABL; EABL; EABL; UN; UN; EABL;
KCB; Safaricom; Safaricom; KPA; KRA; Safaricom; Unilever; Google;
Unilever; KPA; Microsoft; Safaricom; Unilever; KCB; Kplc; Safaricom; EABL;
Coca cola; Coca cola; Unilever; Google; Unilever; Unilever; UN; Google;
Kplc; KPC; KenGen; EABL; KPC; EABL; EABL; EABL; Unilever; Coca cola;
Bidco; Bidco; Microsoft; Google; Unilever; KRA; KRA; Google; Unilever;
```

Unilever; UN; KEMRI; Deloitte; Google; UN; KenGen; EABL; KPC; Unilever;
Unilever; EABL; Kplc; Bidco; Bidco; Amref; Google; KCB; Amref; RMS;
Unilever; Google; Google; Amref; KenGen; Amref; Safaricom; Coca cola;
EABL; UN; Microsoft; Amref; Coca cola; KPA; KenGen; Kplc; UN; Google;
Bidco; Safaricom; Safaricom; EABL; Kplc; EABL; Kplc; Microsoft; Safaricom;
EABL; Google; Safaricom; Unilever; Google; KCB; Amref; Google; Unilever;
Safaricom; Unilever; Deloitte; EABL; Deloitte; EABL; Google; Deloitte;
Deloitte; Amref; Safaricom; KRA; Unilever; Amref; Deloitte; EABL; Google;
Unilever; Amref; Safaricom; EABL; UN; Coca cola; Deloitte; Unilever;
Google; Unilever; KEMRI; KenGen; KenGen; Google; Unilever; KPA; UN;
Google; Google; KPA; KRA; UN; UN; KPA; KRA; Unilever; Amref; Amref; KRA;
KCB; RMS; EABL; Kplc; KenGen; Unilever; KEMRI; EABL; UN; Coca cola; Amref;
EABL; EABL; Deloitte; Safaricom; Amref; Google; KenGen; KPC; KRA; KRA;
Google; UN; Google; KCB; KRA; Unilever; Deloitte; Safaricom; KenGen; UN;
UN; KenGen; KPC; Safaricom; KRA; KenGen; UN; KenGen; KPA; Kplc; UN; RMS;
EABL; KenGen; KRA; UN; Unilever; Amref; KenGen; EABL; KenGen; Deloitte;
Microsoft; UN; Kplc; Amref; Unilever; KenGen; KenGen; Unilever; Unilever;
Unilever; Amref; UN; Safaricom; Safaricom; KEMRI; KEMRI; EABL; Amref;
Deloitte; Kplc; KEMRI; EABL; UN; Microsoft; Deloitte; Google; KenGen;
Google; Safaricom; Safaricom; Safaricom; KEMRI; KEMRI; EABL; Unilever;
KEMRI; Coca cola; EABL; UN; Bidco; KenGen; Google; Microsoft; KenGen;
Unilever; EABL; Safaricom; Google; Amref; UN; EABL; Google; Microsoft;
Unilever; KenGen; Amref; KRA; KenGen; Amref; Amref; KRA; Coca cola;
Safaricom; Google; Kplc; Amref; Coca cola; UN; KenGen; EABL; Deloitte;
Amref; KCB; Google; EABL; KCB; Amref; KEMRI; Amref; Google; Unilever; UN;
KRA; KenGen; EABL; EABL; EABL; Safaricom; EABL; Unilever; EABL; Bidco;
EABL; RMS; Microsoft; Microsoft; KenGen; Amref; EABL; KenGen; Microsoft;
KPC; EABL; Microsoft; Unilever; KPA; Unilever; EABL; UN; Kplc; Amref;
Google; Coca cola; EABL; Amref; Microsoft; Bidco; Unilever; Unilever;
Kplc; KPA; KenGen; EABL; Amref; UN; Google; EABL; Google; Google; Bidco;
Amref; KPA; Unilever; KRA; KRA; Kplc; UN; Coca cola; UN; KPA; Kplc;
KenGen; Kplc; Amref; KPA; UN; Deloitte; Unilever; Safaricom; KCB;
Safaricom; Safaricom; Google; Bidco; UN; Unilever; Kplc; Unilever; Amref;
EABL; Amref; Deloitte; Amref; Amref; Kplc; Unilever; Unilever; Safaricom;
EABL; Amref; Amref; Amref; Safaricom; Google; KCB; KenGen; EABL; Amref;
EABL; EABL; EABL; Amref; Google; Microsoft; UN; Kplc; EABL; UN; Unilever;
Amref; Unilever; UN; EABL; KEMRI; EABL; Safaricom; Bidco; EABL; Amref;
KPA; UN; Deloitte; Coca cola; Google; Safaricom; KRA; Amref; UN; Amref;
Safaricom; UN; UN; Deloitte; Google; Microsoft; UN; Amref; EABL; Amref;
EABL; Bidco; Microsoft; EABL; Amref; KPC; Kplc; Bidco; Safaricom; EABL;
Unilever; EABL; EABL; EABL; Coca cola; Kplc; Unilever; Safaricom;
Safaricom; EABL; Unilever; Google; EABL; Safaricom; Safaricom; Unilever;
Unilever; KPA; EABL; Unilever; Amref; KPA; Unilever; Kplc; Unilever; KPA;
Unilever; Amref; EABL; Kplc; Amref; Safaricom; KPA; KPA; KPA; Google;
Microsoft; EABL; Bidco; EABL; EABL; KenGen; Google; KenGen; KPC; EABL;
Google; Deloitte; Deloitte; EABL; Deloitte; Microsoft; KenGen; Google;
Amref; UN; Unilever; Safaricom; Deloitte; KCB; UN; Unilever; Unilever;
KCB; Kplc; EABL; Unilever; Bidco; KPA; Amref; EABL; Safaricom; UN;
Unilever; Kplc; KenGen; EABL; Microsoft; Google; Google; EABL; EABL; EABL;
Amref; Unilever; KPA; Safaricom; Safaricom; EABL; UN; KCB; Unilever;
Google; Microsoft; Google; KCB; Google; KenGen; KRA; Bidco; EABL; KPA;

Google; Microsoft; Google; Microsoft; Bidco; Bidco; Unilever; Google;
Bidco; EABL; KPA; Unilever; Microsoft; Amref; Amref; EABL; UN; Safaricom;
UN; Amref; Google; Google; Safaricom; Safaricom; Unilever; EABL; EABL;
Deloitte; Deloitte; KenGen; EABL; Deloitte; Microsoft; Microsoft; UN; KPA;
KPA; Microsoft; UN; Deloitte; Kplc; UN; KenGen; Safaricom; Bidco;
Unilever; Unilever; UN; KenGen; Safaricom; EABL; KenGen; Unilever; EABL;
Kplc; Coca cola; EABL; Safaricom; EABL; Deloitte; Unilever; Kplc;
Unilever; Amref; EABL; KRA; Safaricom; Unilever; Amref; KenGen; Microsoft;
Microsoft; Deloitte; Amref; EABL; Google; Kplc; Deloitte; Deloitte; EABL;
Coca cola; KenGen; Kplc; KenGen; UN; Bidco; Amref; Kplc; Microsoft;
Safaricom; KPA; Safaricom; Bidco; KRA; EABL; Bidco; KPA; Deloitte; KRA;
Amref; Google; Amref; Microsoft; Kplc; Unilever; KEMRI; Safaricom; Coca
cola; Google; UN; Safaricom; UN; UN; KRA; KRA; Microsoft; Google; KRA; UN;
EABL; UN; Google; KEMRI; Unilever; Amref; KRA; Amref; Deloitte; Microsoft;
KenGen; EABL; Deloitte; Google; Amref; Unilever; EABL; Google; KenGen;
KenGen; Safaricom; Safaricom; Safaricom; KenGen; Amref; UN; Safaricom;
Microsoft; EABL; UN; Kplc; Unilever; KEMRI; Amref; KPA; Unilever; EABL;
EABL; Amref; UN; Google; Kplc; Google; Google; Bidco; Kplc; Kplc; Google;
KRA; KPA; KenGen; EABL; Deloitte; Deloitte; Unilever; Coca cola; EABL; UN;
Coca cola; Unilever; EABL; Amref; KEMRI; EABL; Kplc; KenGen; Safaricom;
Safaricom; Safaricom; EABL; Safaricom; Safaricom; KenGen; Amref; Unilever;
UN; EABL; Deloitte; Microsoft; Safaricom; Google; Bidco; KPA; KenGen;
EABL; KCB; UN; KRA; Microsoft; KenGen; EABL; Amref; UN; Amref; Bidco;
EABL; KRA; Google; Unilever; Deloitte; KenGen; Safaricom; Unilever;
KenGen; KRA; Amref; KCB; KenGen; Kplc; Amref; Kplc; Unilever; Unilever;
Microsoft; Unilever; Amref; Google; Google; Bidco; Safaricom; Deloitte;
Unilever; Coca cola; Google; KPC; Amref; Amref; EABL; Google; KenGen; UN;
Amref; KPA; KEMRI; EABL; KPA; EABL; Kplc; KenGen; Google; Google;
Unilever; KPA; Amref; EABL; Amref; Google; Deloitte; Unilever; Google;
Microsoft; Google; Deloitte; EABL; Kplc; Google; Google; Amref; Google;
Google; Google; EABL; Kplc; Deloitte; EABL; Deloitte; Deloitte; Safaricom;
Unilever; Amref; KRA; EABL; Kplc; Kplc; UN; UN; Amref; Safaricom; KenGen;
KPA; Coca cola; Safaricom; KCB; KenGen; EABL; Google; EABL; UN; Kplc;
KenGen; KenGen; Amref; Unilever; Kplc; Safaricom; Unilever; Amref; EABL;
EABL; EABL; Amref; KPA; Safaricom; Unilever; Kplc; UN; KRA; KenGen; EABL;
Unilever; EABL; Amref; Amref; EABL; Google; Coca cola; UN; Microsoft;
KenGen; Microsoft; KenGen; Amref; KRA; EABL; UN; Amref; EABL; Safaricom;
Amref; KEMRI; Google; Amref; Microsoft; Amref; Coca cola; Amref; Amref;
Amref; Deloitte; KPA; Coca cola; Bidco; Bidco; Bidco; Bidco; Unilever;
KenGen; Google; Google; Kplc; EABL; UN; Amref; KPA; EABL; Safaricom;
Amref; Deloitte; KPC; Bidco; KCB; Deloitte; Unilever; EABL; Google; Amref;
Deloitte; KenGen; Deloitte; Kplc; UN; Unilever; Amref; Amref; EABL; KPA;
Google; Google; KenGen; KenGen; EABL; Safaricom; Amref; Safaricom;
Safaricom; EABL; UN; UN; KenGen; KCB; Deloitte; Coca cola; KPA; UN;
Safaricom; Kplc; EABL; Deloitte; EABL; Unilever; KenGen; Kplc; UN; Amref;
Microsoft; KRA; KenGen; Amref; KPA; Kplc; UN; KPA; Deloitte; Deloitte;
KEMRI; KEMRI; Unilever; KPA; Amref; KPA; Unilever; Unilever; Amref;
Unilever; Microsoft; KRA; Kplc; UN; Coca cola; Unilever; Safaricom;
Deloitte; EABL; Coca cola; Google; Safaricom; Unilever; EABL; KenGen; Coca
cola; Deloitte; Unilever; Deloitte; EABL; KEMRI; EABL; EABL; UN; KPA;
Kplc; Google; KenGen; Amref; Unilever; Coca cola; UN; Google; Microsoft;

Safaricom; Safaricom; KPA; KenGen; UN; KRA; EABL; KPA; Unilever; Kplc;
Microsoft; Google; UN; KPA; UN; UN; KenGen; EABL; KRA; KPA; KPA;
Safaricom; Coca cola; KenGen; UN; Google; KenGen; UN; Amref; KRA;
Unilever; Unilever; Safaricom; Unilever; Amref; EABL; Safaricom;
Safaricom; KPA; Unilever; Kplc; Amref; KEMRI; KEMRI; KEMRI; Amref; Amref;
Unilever; Kplc; Google; Amref; Kplc; Google; Google; Unilever; Google;
KPA; Unilever; Unilever; UN; Amref; Amref; Unilever; UN; Google;
Microsoft; Unilever; EABL; UN; EABL; Unilever; EABL; Deloitte; Amref;
Unilever; Amref; EABL; Safaricom; Safaricom; Google; Google; EABL; KCB;
EABL; Bidco; KenGen; Coca cola; Unilever; KPA; KPA; KenGen; Microsoft;
Microsoft; Deloitte; Google; Safaricom; Microsoft; Safaricom; Unilever;
Google; Unilever; KRA; KRA; EABL; Safaricom; KEMRI; Unilever; Unilever;
Unilever; Safaricom; UN; Coca cola; RMS; KenGen; Unilever; Amref; RMS;
Google; KRA; Kplc; EABL; Kplc; Amref; RMS; KEMRI; Unilever; UN; EABL; KPC;
KCB; KRA; KPA; Amref; Google; KenGen; KRA; Google; Bidco; Unilever;
Unilever; RMS; Safaricom; Microsoft; UN; Amref; Unilever; Coca cola;
Amref; Safaricom; Amref; Amref; EABL; EABL; EABL; KenGen; Deloitte; Amref;
KPA; Google; Amref; Unilever; Amref; EABL; Safaricom; UN; Google; KPC;
KPA; KPA; Coca cola; KEMRI; Unilever; KPC; Amref; Microsoft; Microsoft;
Unilever; EABL; Safaricom; Google; KenGen; Amref; KPA; Unilever; Bidco;
Amref; Safaricom; Google; KPA; Google; Microsoft; KRA; EABL; Coca cola;
Deloitte; UN; Google; Microsoft; KenGen; UN; Unilever; Google; Google;
Kplc; UN; Safaricom; Google; Microsoft; EABL; Unilever; Google; Unilever;
Safaricom; Unilever; Amref; Microsoft; KCB; Safaricom; KPA; Google;
KenGen; Safaricom; Kplc; Kplc; UN; Microsoft; Google; Safaricom; Unilever;
KPA; KPC; Google; Deloitte; Microsoft; KenGen; KPC; Kplc; EABL; UN;
Unilever; Safaricom; KPA; EABL; Deloitte; KPC; EABL; EABL; EABL; UN;
Unilever; KRA; KenGen; Kplc; EABL; Kplc; Microsoft; Kplc; Unilever; EABL;
KEMRI; Unilever; Microsoft; Amref; EABL; KRA; Coca cola; KPA; Deloitte;
Amref; Amref; KenGen; KPC; Safaricom; UN; Safaricom; EABL; EABL; Kplc;
Kplc; Microsoft; KEMRI; Amref; UN; UN; UN; UN; KCB; Unilever; Safaricom;
KRA; KEMRI; EABL; Unilever; UN; KPA; KenGen; EABL; EABL; Unilever; KPA;
UN; Amref; Safaricom; Safaricom; Coca cola; Coca cola; KPC; KEMRI; Amref;
Amref; EABL; Amref; EABL; Amref; UN; Google; KPA; KenGen; Google; Kplc;
KenGen; KenGen; KPC; KPC; Deloitte; Deloitte; Amref; Safaricom; UN; EABL;
KPA; Safaricom; UN; KenGen; EABL; Bidco; Kplc; Kplc; KPA; KEMRI; EABL;
EABL; Unilever; EABL; UN; KPA; Safaricom; UN; Unilever; Google; Microsoft;
Kplc; Google; KenGen; Amref; KenGen; Kplc; Microsoft; KenGen; Amref; EABL;
KPA; Deloitte; KPC; Microsoft; KenGen; UN; UN; Unilever; KCB; Safaricom;
KCB; Coca cola; Google; UN; Google; KPA; EABL; EABL; Unilever; Amref;
EABL; Safaricom; EABL; Google; Safaricom; Bidco; Unilever; Bidco;
Safaricom; Bidco; EABL; Google; KenGen; KPA; UN; Deloitte; KenGen;
Unilever; Safaricom; Safaricom; Unilever; Unilever; KEMRI; KRA; Unilever;
EABL; Kplc; KenGen; Unilever; Kplc; Safaricom; Google; Amref; Google;
KenGen; Google; Deloitte; Deloitte; Deloitte; Microsoft; Unilever; KenGen;
Unilever; EABL; Kplc; Google; KenGen; Google; EABL; Bidco; KRA; Google;
Amref; Amref; KEMRI; Google; Kplc; EABL; Deloitte; UN; Coca cola; EABL;
KenGen; Deloitte; Kplc; Safaricom; EABL; UN; Kplc; Google; EABL; KenGen;
KPC; Safaricom; Safaricom; EABL; Google; Deloitte; KRA; Safaricom; Coca
cola; Unilever; KRA; Amref; Amref; Amref; EABL; Safaricom; UN; UN; KEMRI;
Microsoft; Google; Microsoft; Amref; Amref; UN; Coca cola; Amref; EABL;

```
Safaricom; KEMRI; Unilever; UN; KenGen; Unilever; Unilever; Kplc; Amref;
Microsoft; Google; Deloitte; KenGen; Kplc; Amref; UN; Kplc; UN; KPC;
Unilever; Google; Google; Microsoft; Deloitte; Google; Microsoft; Coca
cola; Coca cola; EABL; EABL; KPA; Unilever; Coca cola; EABL; KEMRI; KPA;
UN; Deloitte; RMS; EABL; Microsoft; KenGen; EABL; Safaricom; UN; KRA;
Google; UN; Google; Microsoft; Bidco; Kplc; UN; EABL; Amref; Unilever; UN;
EABL; Kplc; KRA; KRA; Kplc; EABL; KenGen; KPA; Bidco; Unilever; EABL;
EABL; Unilever; Deloitte; Deloitte; UN; Amref; Safaricom; KRA; Amref;
Deloitte; Microsoft; KRA; KenGen; KPC; Kplc; KPA; KRA; Amref; EABL;
Safaricom; KRA; UN; Unilever; Safaricom; Microsoft; Safaricom; Coca cola;
KPA; KenGen; Deloitte; Amref; UN; Kplc; UN; Google; Kplc; Google; Google;
EABL; EABL; EABL; Google; Amref; Safaricom; UN; Amref; EABL; EABL; Amref;
KPA; Amref; Safaricom; Deloitte; EABL; EABL; Microsoft; Google; UN;
Microsoft; Safaricom; RMS; UN; Google; KenGen; KenGen; UN; Coca cola;
Unilever; KenGen; Google; Microsoft; Amref; UN; KPA; UN; KRA; Unilever;
Deloitte; KenGen; KCB; EABL; KPA; EABL; EABL; Unilever; Kplc; Kplc; KPA;
KenGen; Kplc; UN; Amref; Microsoft; KCB; Google; KenGen; EABL; Bidco;
EABL; KenGen; EABL; Safaricom; Coca cola; KenGen; EABL; EABL; Kplc; Kplc;
Microsoft; KenGen; Safaricom; Microsoft; Google; Amref; Amref; UN;
Safaricom; Google; Microsoft; Safaricom; Coca cola; Kplc; KPA; Microsoft;
Microsoft; KPA; Deloitte; KRA; Google; KRA; Deloitte; Deloitte; Unilever;
Amref; Amref; KRA; Bidco; EABL; Safaricom; KenGen; Kplc; KenGen; UN;
KenGen; KEMRI; KEMRI; Safaricom; Deloitte; Amref; UN; EABL; Unilever; KRA;
KPA; KPA; KRA; KPC; KenGen; UN; Unilever; EABL; Unilever; Google;
Safaricom; KenGen; EABL; Microsoft; Safaricom; KenGen; Kplc; Microsoft;
Amref; KPA; KPA; UN; Unilever; Unilever; KPA; Safaricom; KPA; KenGen; KPC;
KEMRI; Amref; EABL; RMS; RMS; Amref; KCB; UN; Deloitte; Coca cola; UN;
KRA; Google; KRA; KenGen; KRA; Safaricom; EABL; UN; UN; KenGen; Coca cola;
KPA; UN; Safaricom; KRA; Coca cola; Deloitte; UN; Amref; Unilever; KPA;
Kplc; KRA; Kplc; Safaricom; Unilever; Google; Microsoft; KCB; KPA'
```

```python
cols = ['Companies']
df_out = pd.DataFrame(finalarr, columns = cols )
df_grp = df_out.groupby(cols).size().reset_index(name='Count')
```

```python
df_grp.sort_values(by='Count', ascending=False)
```

```css
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	Companies	Count
4	EABL	249
17	Unilever	194
0	Amref	193
5	Google	178
16	UN	172
15	Safaricom	161
11	KenGen	147
8	KPA	107
12	Kplc	104
13	Microsoft	99
3	Deloitte	93
10	KRA	74
2	Coca cola	58
1	Bidco	48
6	KCB	40
7	KEMRI	40
9	KPC	30
14	RMS	13

```python
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
from collections import Counter
import matplotlib.pyplot as plt

word_could_dict=Counter(finalarr)

wordcloud =
WordCloud(background_color="white").generate_from_frequencies(word_could_d
ict) #.generate(corpus) #.generate_from_frequencies(word_could_dict)
#Post processing
#ii. plot word cloud model
plt.figure(figsize = (22,22))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```

Notes

Its important to note changes to metrics the last 3 days - Amref moved from 2nd to 3rd etc

Note these interpretations are time-bound and decay over time hence it's important to keep updating this data

Save to csv

```
df_grp.sort_values(by='Count',
ascending=False).to_csv('./out/sortedout.csv')
```