

cbs_segments

March 13, 2022

```
[1]: import sys
     sys.path.append('../')

     from helpers.utils import read_docx_to_dict
```

```
[2]: filename = "Analysis_Fire_12_30_21_CBS.docx"

     miami_dallas = "Analysis_12_30_21_Colorado_Fire_CBS_w_DallasMiami.docx"
```

```
[3]: data = read_docx_to_dict(filename)
```

```
[4]: import pandas as pd
     pd.options.display.max_rows = 500

     # create dataframe
     df = pd.DataFrame.from_dict(data)
```

```
[5]: miami_dallas_data = read_docx_to_dict(miami_dallas)
```

```
[6]: miami_dallas_df = pd.DataFrame.from_dict(miami_dallas_data)
```

```
[7]: len(df)
```

```
[7]: 75
```

```
[8]: len(miami_dallas_df)
```

```
[8]: 10
```

```
[9]: df = pd.concat([df, miami_dallas_df], ignore_index=True)
     len(df)
```

```
[9]: 85
```

```
[10]: df.tail(10)
```

```
[10]:
```

	time	location	station \
75	2021-12-30 6:37 PM	Miami-Ft. Lauderdale	WFOR

```

76 2021-12-30 6:07 PM Miami-Ft. Lauderdale WFOR
77 2021-12-30 6:06 PM Dallas-Ft. Worth KTVT
78 2021-12-30 5:37 PM Dallas-Ft. Worth KTVT
79 2021-12-30 5:33 PM Dallas-Ft. Worth KTVT
80 2021-12-30 5:33 PM Dallas-Ft. Worth KTVT
81 2021-12-30 5:19 PM Miami-Ft. Lauderdale WFOR
82 2021-12-30 5:19 PM Miami-Ft. Lauderdale WFOR
83 2021-12-30 5:18 PM Miami-Ft. Lauderdale WFOR
84 2021-12-30 5:18 PM Miami-Ft. Lauderdale WFOR

```

text

```

75 THE EVENING ENDING OVERNIGHT. SEVERAL COMMUNIT...
76 WOULDN'T BE THAT LONG IN DURATION. AND THEN TH...
77 THE OMICRON VARIANT. >>> THE CITY OF PARIS SOU...
78 THE EVENING ENDING OVERNIGHT. SEVERAL COMMUNIT...
79 THE EVENING ENDING OVERNIGHT. SEVERAL COMMUNIT...
80 SHORTAGE. WITH FINAL PREPS UNDERWAY FOR THE NE...
81 >> Reporter: NEWS OF THE BOY BEING BADLY BURNE...
82 >> Reporter: NEWS OF THE BOY BEING BADLY BURNE...
83 AND WE SUPPORT C.C. S. O. AND THE OFFICER EARL...

```

[w:drawing]

```
[11]: df['text'] = df['text'].str.lower()
```

```
[12]: df.head(5)
```

```
[12]:
```

	time	location	station	\
0	2021-12-30 6:49 PM	Phoenix (Prescott)	KPHO	
1	2021-12-30 6:48 PM	Phoenix (Prescott)	KPHO	
2	2021-12-30 6:37 PM	Ft. Myers-Naples	WINK	
3	2021-12-30 6:37 PM	Greensboro-H.Point-W.Salem	WFMY	
4	21-12-30 6:37 PM	Atlanta	WGCL	

text

```

0 have been destroyed. ap, the associated press ...
1 kusa, the nbc affiliate in denver, this is in ...
2 several communities were evacuated because of ...
3 the evening ending overnight. several communit...
4 the evening ending overnight. several communit...

```

```
[13]: from helpers.utils import check_text_likeness
```

```
[14]: df['matches'] = df.apply(lambda row: check_text_likeness(df, row['text']),
                                axis=1)
```

```
[15]: df.head(5)
```

```
[15]:
```

	time	location	station	\
0	2021-12-30 6:49 PM	Phoenix (Prescott)	KPHO	
1	2021-12-30 6:48 PM	Phoenix (Prescott)	KPHO	
2	2021-12-30 6:37 PM	Ft. Myers-Naples	WINK	
3	2021-12-30 6:37 PM	Greensboro-H.Point-W.Salem	WFMY	
4	21-12-30 6:37 PM	Atlanta	WGCL	

	text	\
0	have been destroyed. ap, the associated press ...	
1	kusa, the nbc affiliate in denver, this is in ...	
2	several communities were evacuated because of ...	
3	the evening ending overnight. several communit...	
4	the evening ending overnight. several communit...	

	matches
0	[0]
1	[1]
2	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...
3	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...
4	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...

```
[28]: len(df['matches'])
```

```
[28]: 95
```

```
[ ]:
```

```
[16]: from helpers.utils import fetch_biggest_text

df['row_to_use'] = df.apply(lambda row: fetch_biggest_text(df, row['matches']),
                             axis=1)
```

```
[ ]:
```

```
[ ]:
```

```
[17]: df
```

```
[17]:
```

	time	location	station	\
0	2021-12-30 6:49 PM	Phoenix (Prescott)	KPHO	
1	2021-12-30 6:48 PM	Phoenix (Prescott)	KPHO	
2	2021-12-30 6:37 PM	Ft. Myers-Naples	WINK	
3	2021-12-30 6:37 PM	Greensboro-H.Point-W.Salem	WFMY	
4	21-12-30 6:37 PM	Atlanta	WGCL	
5	2021-12-30 6:37 PM	Orlando-Daytona Bch-Melbrn	WKMG	
6	2021-12-30 6:37 PM	New York	WCBS	
7	2021-12-30 6:37 PM	Detroit	WWJ	
8	2021-12-30 6:37 PM	Boston (Manchester)	WBZ	

9	021-12-30 6:37 PM	Sacramnto-Stkton-Modesto	KOVR
10	2021-12-30 6:37 PM	Hartford & New Haven	WFSB
11	2021-12-30 6:37 PM	Baltimore	WJZ
12	021-12-30 6:37 PM	Raleigh-Durham (Fayetteville)	WNCN
13	2021-12-30 6:37 PM	Pittsburgh	KDKA
14	2021-12-30 6:37 PM	Portland, OR	KOIN
15	2021-12-30 6:37 PM	Seattle-Tacoma	KIRO
16	2021-12-30 6:37 PM	Los Angeles	KCBS
17	021-12-30 6:37 PM	San Francisco-Oak-San Jose	KPIX
18	21-12-30 6:37 PM	Washington, DC (Hagrstwn)	WUSA
19	2021-12-30 6:37 PM	Louisville	WLKY
20	2021-12-30 6:37 PM	Providence-New Bedford	WPRI
21	2021-12-30 6:37 PM	Knoxville	WVLT
22	2021-12-30 6:37 PM	Indianapolis	WTTV
23	2021-12-30 6:37 PM	Indianapolis	WTTV
24	2021-12-30 6:37 PM	Indianapolis	WTTV
25	2021-12-30 6:37 PM	Norfolk-Portsmouth-Newport News	WTKR
26	2021-12-30 6:37 PM	West Palm Beach-Ft. Pierce	WPEC
27	2021-12-30 6:37 PM	West Palm Beach-Ft. Pierce	WPEC
28	2021-12-30 6:37 PM	Tampa-St. Pete (Sarasota)	WTSP
29	2021-12-30 6:37 PM	Buffalo	WIVB
30	2021-12-30 6:37 PM	Greenville-Spartanburg-Asheville-Anderson	WSPA
31	2021-12-30 6:37 PM	Jacksonville	WJAX
32	2021-12-30 6:37 PM	Charlotte	WBTV
33	2021-12-30 6:37 PM	Cincinnati	WKRC
34	2021-12-30 6:37 PM	Columbus, OH	WBNS
35	2021-12-30 6:37 PM	Columbus, OH	WBNS
36	2021-12-30 6:37 PM	Columbus, OH	WBNS
37	2021-12-30 6:37 PM	Philadelphia	KYW
38	2021-12-30 6:12 PM	Sacramnto-Stkton-Modesto	KOVR
39	2021-12-30 6:20 PM	Sacramnto-Stkton-Modesto	KOVR
40	021-12-30 6:07 PM	Phoenix (Prescott)	KPHO
41	2021-12-30 6:00 PM	Sacramnto-Stkton-Modesto	KOVR
42	2021-12-30 5:59 PM	Sacramnto-Stkton-Modesto	KOV
43	2021-12-30 5:55 PM	Ft. Myers-Naples	WINK
44	2021-12-30 5:44 PM	Baltimore	WJZ
45	2021-12-30 5:44 PM	Baltimore	WJZ
46	2021-12-30 5:44 PM	Baltimore	WJZ
47	2021-12-30 5:43 PM	Seattle-Tacoma	KIRO
48	2021-12-30 5:37 PM	Houston	KHOU
49	2021-12-30 5:37 PM	Dallas-Ft. Worth	KTVT
50	2021-12-30 5:37 PM	Tulsa	KOTV
51	2021-12-30 5:37 PM	San Antonio	KENS
52	2021-12-30 5:37 PM	Milwaukee	WDJT
53	2021-12-30 5:37 PM	Austin	KEYE
54	2021-12-30 5:37 PM	Chicago	WBBM
55	2021-12-30 5:37 PM	Memphis	WREG

56	2021-12-30 5:37 PM	San Diego	KFMB
57	2021-12-30 5:37 PM	Minneapolis-St. Paul	WCCO
58	2021-12-30 5:37 PM	Kansas City	KCTV
59	2021-12-30 5:37 PM	Las Vegas	KLAS
60	2021-12-30 5:37 PM	New Orleans	WWL
61	2021-12-30 5:37 PM	Las Vegas	KLAS
62	2021-12-30 5:37 PM	New Orleans	WWL
63	2021-12-30 5:37 PM	St. Louis	KMOV
64	2021-12-30 5:37 PM	Oklahoma City	KWTW
65	2021-12-30 5:34 PM	Pittsburgh	KDKA
66	2021-12-30 5:31 PM	Sacramnto-Stkton-Modesto	KOVR
67	2021-12-30 5:17 PM	Tulsa	KOTV
68	2021-12-30 5:14 PM	Portland, OR	KOIN
69	2021-12-30 5:12 PM	San Diego	KFMB
70	2021-12-30 5:07 PM	Albuquerque-Santa Fe	KRQE
71	2021-12-30 5:06 PM	Phoenix (Prescott)	KPHO
72	2021-12-30 5:04 PM	Sacramnto-Stkton-Modesto	KOVR
73	2021-12-30 5:02 PM	Sacramnto-Stkton-Modesto	KOVR
74	2021-12-30 5:01 PM	Sacramnto-Stkton-Modesto	KOVR
75	2021-12-30 6:37 PM	Miami-Ft. Lauderdale	WFOR
76	2021-12-30 6:07 PM	Miami-Ft. Lauderdale	WFOR
77	2021-12-30 6:06 PM	Dallas-Ft. Worth	KTVT
78	2021-12-30 5:37 PM	Dallas-Ft. Worth	KTVT
79	2021-12-30 5:33 PM	Dallas-Ft. Worth	KTVT
80	2021-12-30 5:33 PM	Dallas-Ft. Worth	KTVT
81	2021-12-30 5:19 PM	Miami-Ft. Lauderdale	WFOR
82	2021-12-30 5:19 PM	Miami-Ft. Lauderdale	WFOR
83	2021-12-30 5:18 PM	Miami-Ft. Lauderdale	WFOR
84	2021-12-30 5:18 PM	Miami-Ft. Lauderdale	WFOR

text \

```

0  have been destroyed. ap, the associated press ...
1  kusa, the nbc affiliate in denver, this is in ...
2  several communities were evacuated because of ...
3  the evening ending overnight. several communit...
4  the evening ending overnight. several communit...
5  the evening ending overnight. several communit...
6  several communities were evacuated because of ...
7  the evening ending overnight. several communit...
8  the evening ending overnight. several communit...
9  mudslides, major. >> garrett: jonathan vigliot...
10 and here in southern california, the rain will...
11 the evening ending overnight. several communit...
12 the evening ending overnight. several communit...
13 the evening ending overnight. several communit...
14 the rain will continue through the evening end...
15 the rain will continue through the evening end...

```

16 like this, and the risk of mudslides, major. >...
17 evacuated because of flooding like this, and t...
18 several communities were evacuated because of ...
19 the evening ending overnight. several communit...
20 and here in southern california, the rain will...
21 the evening ending overnight. several communit...
22 the evening ending overnight. several communit...
23 the evening ending overnight. several communit...
24 the evening ending overnight. several communit...
25 the evening ending overnight. several communit...
26 the evening ending overnight. several communit...
27 the evening ending overnight. several communit...
28 the evening ending overnight. several communit...
29 and here in southern california, the rain will...
30 the evening ending overnight. several communit...
31 and here in southern california, the rain will...
32 the evening ending overnight. several communit...
33 the evening ending overnight. several communit...
34 the evening ending overnight. several communit...
35 the evening ending overnight. several communit...
36 the evening ending overnight. several communit...
37 the evening ending overnight. several communit...
38 tracking the devastating fires in colorado and...
39 here we have evacuations. i will tell you a li...
40 hoping to ski into the new year. and here in s...
41 about. we just talked about the hospital where...
42 together. this house, burning to the ground. 3...
43 rubenstein law, injured in a car accident or s...
44 e2021-12-30 5:44 pm
45 wildfire-that engulfed parts of the area. in a...
46 wildfire-that engulfed parts of the area. in a...
47 outside of denver have been told to evacuate. ...
48 the evening ending overnight. several communit...
49 the evening ending overnight. several communit...
50 several communities were evacuated because of ...
51 the evening ending overnight. several communit...
52 the evening ending overnight. several communit...
53 the evening ending overnight. several communit...
54 the evening ending overnight. several communit...
55 the evening ending overnight. several communit...
56 mudslides, major. >> garrett: jonathan vigliot...
57 the evening ending overnight. several communit...
58 and here in southern california, the rain will...
59 mudslides, major. >> garrett: jonathan vigliot...
60 the evening ending overnight. several communit...
61 mudslides, major. >> garrett: jonathan vigliot...
62 the evening ending overnight. several communit...

63 and here in southern california, the rain will...
 64 several communities were evacuated because of ...
 65 they are calling this now the marshall-fire-an...
 66 signatures, here. this is where we have had fl...
 67 sun. morning satellite imagery right now we ha...
 68 >> to breaking news now, this is a live look i...
 69 thousands of people have been ordered to evacu...
 70 the evening ending overnight. several communit...
 71 into the central rockies, and it is a dangerou...
 72 we couple that with relative humidity anywhere...
 73 much damage will be assessed as well. >> we st...
 74 >> the-fire-crews now left the scene knowing t...
 75 the evening ending overnight. several communit...
 76 wouldn't be that long in duration. and then th...
 77 the omicron variant. >>> the city of paris sou...
 78 the evening ending overnight. several communit...
 79 the evening ending overnight. several communit...
 80 shortage. with final preps underway for the ne...
 81 >> reporter: news of the boy being badly burne...
 82 >> reporter: news of the boy being badly burne...
 83 and we support c.c. s. o. and the officer earl...
 84 [w:drawing]

	matches	row_to_use
0	[0]	0
1	[1]	1
2	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
3	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
4	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
5	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
6	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
7	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
8	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
9	[2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 1...	14
10	[2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, ...	14
11	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
12	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
13	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
14	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
15	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
16	[2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 1...	14
17	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
18	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
19	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
20	[2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, ...	14
21	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
22	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14

23	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
24	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
25	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
26	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
27	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
28	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
29	[2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, ...	14
30	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
31	[2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, ...	14
32	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
33	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
34	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
35	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
36	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
37	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
38	[38]	38
39	[39]	39
40	[2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, ...	14
41	[41]	41
42	[42]	42
43	[43]	43
44	[44]	44
45	[45, 46]	45
46	[45, 46]	45
47	[47]	47
48	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
49	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
50	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
51	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
52	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
53	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
54	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
55	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
56	[2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 1...	14
57	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
58	[2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, ...	14
59	[2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 1...	14
60	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
61	[2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 1...	14
62	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
63	[2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, ...	14
64	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14
65	[65]	65
66	[66]	66
67	[67]	67
68	[68]	68
69	[69]	69


```

70 [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...      14
71                                     [71]          71
72                                     [72]          72
73                                     [73]          73
74                                     [74]          74
75 [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...      14
76                                     [76]          76
77                                     [77]          77
78 [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...      14
79 [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...      14
80                                     [80]          80
81                                     [81, 82]        81
82                                     [81, 82]        81
83                                     [83]          83
84                                     [84]          84

```

```
[ ]:
```

```
[ ]:
```

```
[18]: from helpers.utils import mark_use_row
mark_use_row(df)
```

```
[18]: 'done'
```

```
[19]: df.head(20)
```

```
[19]:
```

	time	location	station \
0	2021-12-30 6:49 PM	Phoenix (Prescott)	KPHO
1	2021-12-30 6:48 PM	Phoenix (Prescott)	KPHO
2	2021-12-30 6:37 PM	Ft. Myers-Naples	WINK
3	2021-12-30 6:37 PM	Greensboro-H.Point-W.Salem	WFMY
4	21-12-30 6:37 PM	Atlanta	WGCL
5	2021-12-30 6:37 PM	Orlando-Daytona Bch-Melbrn	WKMG
6	2021-12-30 6:37 PM	New York	WCBS
7	2021-12-30 6:37 PM	Detroit	WWJ
8	2021-12-30 6:37 PM	Boston (Manchester)	WBZ
9	021-12-30 6:37 PM	Sacramnto-Stkton-Modesto	KOVR
10	2021-12-30 6:37 PM	Hartford & New Haven	WFSB
11	2021-12-30 6:37 PM	Baltimore	WJZ
12	021-12-30 6:37 PM	Raleigh-Durham (Fayetvllle)	WNCN
13	2021-12-30 6:37 PM	Pittsburgh	KDKA
14	2021-12-30 6:37 PM	Portland, OR	KOIN
15	2021-12-30 6:37 PM	Seattle-Tacoma	KIRO
16	2021-12-30 6:37 PM	Los Angeles	KCBS
17	021-12-30 6:37 PM	San Francisco-Oak-San Jose	KPIX
18	21-12-30 6:37 PM	Washington, DC (Hagrstwn)	WUSA

text \

0 have been destroyed. ap, the associated press ...
 1 kusa, the nbc affiliate in denver, this is in ...
 2 several communities were evacuated because of ...
 3 the evening ending overnight. several communit...
 4 the evening ending overnight. several communit...
 5 the evening ending overnight. several communit...
 6 several communities were evacuated because of ...
 7 the evening ending overnight. several communit...
 8 the evening ending overnight. several communit...
 9 mudslides, major. >> garrett: jonathan vigliot...
 10 and here in southern california, the rain will...
 11 the evening ending overnight. several communit...
 12 the evening ending overnight. several communit...
 13 the evening ending overnight. several communit...
 14 the rain will continue through the evening end...
 15 the rain will continue through the evening end...
 16 like this, and the risk of mudslides, major. >...
 17 evacuated because of flooding like this, and t...
 18 several communities were evacuated because of ...
 19 the evening ending overnight. several communit...

	matches	row_to_use	use_row
0	[0]	0	True
1	[1]	1	True
2	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
3	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
4	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
5	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
6	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
7	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
8	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
9	[2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 1...	14	False
10	[2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, ...	14	False
11	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
12	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
13	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
14	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	True
15	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
16	[2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 1...	14	False
17	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
18	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
19	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False

```
[20]: df['words'] = df['text'].str.lower().str.replace(',', '').str.replace('>', '').
      ↪str.replace('.', '').str.replace('\n', '').str.replace("'", '').str.replace(
      ↪'!', '').str.replace('?', '').str.replace('%', '').str.replace(')', '').str.
      ↪replace('(', '').str.replace('_', '').str.replace(':', '').str.strip().str.
      ↪split(' ')
```

/Users/loren/.pyenv/versions/3.7.4/lib/python3.7/site-packages/ipykernel_launcher.py:1: FutureWarning: The default value of regex will change from True to False in a future version. In addition, single character regular expressions will *not* be treated as literal strings when regex=True.

"""Entry point for launching an IPython kernel.

/Users/loren/.pyenv/versions/3.7.4/lib/python3.7/site-packages/ipykernel_launcher.py:2: FutureWarning: The default value of regex will change from True to False in a future version. In addition, single character regular expressions will *not* be treated as literal strings when regex=True.

```
[21]: df.head(20)
```

```
[21]:
```

	time	location	station \
0	2021-12-30 6:49 PM	Phoenix (Prescott)	KPHO
1	2021-12-30 6:48 PM	Phoenix (Prescott)	KPHO
2	2021-12-30 6:37 PM	Ft. Myers-Naples	WINK
3	2021-12-30 6:37 PM	Greensboro-H.Point-W.Salem	WFMY
4	21-12-30 6:37 PM	Atlanta	WGCL
5	2021-12-30 6:37 PM	Orlando-Daytona Bch-Melbrn	WKMG
6	2021-12-30 6:37 PM	New York	WCBS
7	2021-12-30 6:37 PM	Detroit	WWJ
8	2021-12-30 6:37 PM	Boston (Manchester)	WBZ
9	021-12-30 6:37 PM	Sacramnto-Stkton-Modesto	KOVR
10	2021-12-30 6:37 PM	Hartford & New Haven	WFSB
11	2021-12-30 6:37 PM	Baltimore	WJZ
12	021-12-30 6:37 PM	Raleigh-Durham (Fayetvll)	WNCN
13	2021-12-30 6:37 PM	Pittsburgh	KDKA
14	2021-12-30 6:37 PM	Portland, OR	KOIN
15	2021-12-30 6:37 PM	Seattle-Tacoma	KIRO
16	2021-12-30 6:37 PM	Los Angeles	KCBS
17	021-12-30 6:37 PM	San Francisco-Oak-San Jose	KPIX
18	21-12-30 6:37 PM	Washington, DC (Hagrstwn)	WUSA
19	2021-12-30 6:37 PM	Louisville	WLKY

```

                                text \
0  have been destroyed. ap, the associated press ...
1  kusa, the nbc affiliate in denver, this is in ...
2  several communities were evacuated because of ...
3  the evening ending overnight. several communit...
4  the evening ending overnight. several communit...
5  the evening ending overnight. several communit...
```

6 several communities were evacuated because of ...
7 the evening ending overnight. several communit...
8 the evening ending overnight. several communit...
9 mudslides, major. >> garrett: jonathan vigliot...
10 and here in southern california, the rain will...
11 the evening ending overnight. several communit...
12 the evening ending overnight. several communit...
13 the evening ending overnight. several communit...
14 the rain will continue through the evening end...
15 the rain will continue through the evening end...
16 like this, and the risk of mudslides, major. >...
17 evacuated because of flooding like this, and t...
18 several communities were evacuated because of ...
19 the evening ending overnight. several communit...

	matches	row_to_use	use_row	\
0	[0]	0	True	
1	[1]	1	True	
2	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False	
3	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False	
4	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False	
5	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False	
6	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False	
7	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False	
8	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False	
9	[2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 1...	14	False	
10	[2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, ...	14	False	
11	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False	
12	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False	
13	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False	
14	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	True	
15	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False	
16	[2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 1...	14	False	
17	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False	
18	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False	
19	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False	

words

0 [have, been, destroyed, ap, the, associated, p...
1 [kusa, the, nbc, affiliate, in, denver, this, ...
2 [several, communities, were, evacuated, becaus...
3 [the, evening, ending, overnight, several, com...
4 [the, evening, ending, overnight, several, com...
5 [the, evening, ending, overnight, several, com...
6 [several, communities, were, evacuated, becaus...
7 [the, evening, ending, overnight, several, com...
8 [the, evening, ending, overnight, several, com...

```

9  [mudslides, major, , garrett, jonathan, viglio...
10 [and, here, in, southern, california, the, rai...
11 [the, evening, ending, overnight, several, com...
12 [the, evening, ending, overnight, several, com...
13 [the, evening, ending, overnight, several, com...
14 [the, rain, will, continue, through, the, even...
15 [the, rain, will, continue, through, the, even...
16 [like, this, and, the, risk, of, mudslides, ma...
17 [evacuated, because, of, flooding, like, this,...
18 [several, communities, were, evacuated, becaus...
19 [the, evening, ending, overnight, several, com...

```

```
[ ]:
```

```
[22]: from helpers.utils import parse_words
df['clean_words'] = df.apply(lambda row: parse_words(row['words']), axis=1)
```

```
[23]: df.head(20)
```

```
[23]:
```

	time	location	station \
0	2021-12-30 6:49 PM	Phoenix (Prescott)	KPHO
1	2021-12-30 6:48 PM	Phoenix (Prescott)	KPHO
2	2021-12-30 6:37 PM	Ft. Myers-Naples	WINK
3	2021-12-30 6:37 PM	Greensboro-H.Point-W.Salem	WFMY
4	21-12-30 6:37 PM	Atlanta	WGCL
5	2021-12-30 6:37 PM	Orlando-Daytona Bch-Melbrn	WKMG
6	2021-12-30 6:37 PM	New York	WCBS
7	2021-12-30 6:37 PM	Detroit	WWJ
8	2021-12-30 6:37 PM	Boston (Manchester)	WBZ
9	021-12-30 6:37 PM	Sacramnto-Stkton-Modesto	KOVR
10	2021-12-30 6:37 PM	Hartford & New Haven	WFSB
11	2021-12-30 6:37 PM	Baltimore	WJZ
12	021-12-30 6:37 PM	Raleigh-Durham (Fayetvll)	WNCN
13	2021-12-30 6:37 PM	Pittsburgh	KDKA
14	2021-12-30 6:37 PM	Portland, OR	KOIN
15	2021-12-30 6:37 PM	Seattle-Tacoma	KIRO
16	2021-12-30 6:37 PM	Los Angeles	KCBS
17	021-12-30 6:37 PM	San Francisco-Oak-San Jose	KPIX
18	21-12-30 6:37 PM	Washington, DC (Hagrstwn)	WUSA
19	2021-12-30 6:37 PM	Louisville	WLKY

```

text \
0  have been destroyed. ap, the associated press ...
1  kusa, the nbc affiliate in denver, this is in ...
2  several communities were evacuated because of ...
3  the evening ending overnight. several communit...
4  the evening ending overnight. several communit...

```

5 the evening ending overnight. several communit...
 6 several communities were evacuated because of ...
 7 the evening ending overnight. several communit...
 8 the evening ending overnight. several communit...
 9 mudslides, major. >> garrett: jonathan vigliot...
 10 and here in southern california, the rain will...
 11 the evening ending overnight. several communit...
 12 the evening ending overnight. several communit...
 13 the evening ending overnight. several communit...
 14 the rain will continue through the evening end...
 15 the rain will continue through the evening end...
 16 like this, and the risk of mudslides, major. >...
 17 evacuated because of flooding like this, and t...
 18 several communities were evacuated because of ...
 19 the evening ending overnight. several communit...

		matches	row_to_use	use_row	\
0		[0]	0	True	
1		[1]	1	True	
2	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...		14	False	
3	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...		14	False	
4	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...		14	False	
5	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...		14	False	
6	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...		14	False	
7	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...		14	False	
8	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...		14	False	
9	[2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 1...		14	False	
10	[2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, ...		14	False	
11	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...		14	False	
12	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...		14	False	
13	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...		14	False	
14	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...		14	True	
15	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...		14	False	
16	[2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 1...		14	False	
17	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...		14	False	
18	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...		14	False	
19	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...		14	False	

		words	\
0	[have, been, destroyed, ap, the, associated, p...		
1	[kusa, the, nbc, affiliate, in, denver, this, ...		
2	[several, communities, were, evacuated, becaus...		
3	[the, evening, ending, overnight, several, com...		
4	[the, evening, ending, overnight, several, com...		
5	[the, evening, ending, overnight, several, com...		
6	[several, communities, were, evacuated, becaus...		
7	[the, evening, ending, overnight, several, com...		

```

8 [the, evening, ending, overnight, several, com...
9 [mudslides, major, , garrett, jonathan, viglio...
10 [and, here, in, southern, california, the, rai...
11 [the, evening, ending, overnight, several, com...
12 [the, evening, ending, overnight, several, com...
13 [the, evening, ending, overnight, several, com...
14 [the, rain, will, continue, through, the, even...
15 [the, rain, will, continue, through, the, even...
16 [like, this, and, the, risk, of, mudslides, ma...
17 [evacuated, because, of, flooding, like, this,...
18 [several, communities, were, evacuated, becaus...
19 [the, evening, ending, overnight, several, com...

```

clean_words

```

0 [have, been, destroyed, ap, the, associated, p...
1 [kusa, the, nbc, affiliate, in, denver, this, ...
2 [several, communities, were, evacuated, becaus...
3 [the, evening, ending, overnight, several, com...
4 [the, evening, ending, overnight, several, com...
5 [the, evening, ending, overnight, several, com...
6 [several, communities, were, evacuated, becaus...
7 [the, evening, ending, overnight, several, com...
8 [the, evening, ending, overnight, several, com...
9 [mudslides, major, garrett, jonathan, vigliott...
10 [and, here, in, southern, california, the, rai...
11 [the, evening, ending, overnight, several, com...
12 [the, evening, ending, overnight, several, com...
13 [the, evening, ending, overnight, several, com...
14 [the, rain, will, continue, through, the, even...
15 [the, rain, will, continue, through, the, even...
16 [like, this, and, the, risk, of, mudslides, ma...
17 [evacuated, because, of, flooding, like, this,...
18 [several, communities, were, evacuated, becaus...
19 [the, evening, ending, overnight, several, com...

```

```
[24]: from helpers.utils import fetch_climate_words_in_words,
      ↪ fetch_climate_phrases_in_text
```

```
[25]: fetch_climate_words_in_words(["adapt", "for", "climate", "change"])
      # segment_df['climate_words_found'] = segment_df.apply(lambda row:
      ↪ fetch_climate_words_in_text(row['clean_words']), axis=1)
```

```
[25]: ['adapt', 'climate']
```

```
[26]: fetch_climate_phrases_in_text("adapt for climate change")
```

```
[26]: ['climate change']
```

```
[27]: df['climate_words_found'] = df.apply(lambda row:
    ↪ fetch_climate_words_in_words(row['clean_words']), axis=1)
```

```
[28]: df.head(20)
```

```
[28]:
```

	time	location	station \
0	2021-12-30 6:49 PM	Phoenix (Prescott)	KPHO
1	2021-12-30 6:48 PM	Phoenix (Prescott)	KPHO
2	2021-12-30 6:37 PM	Ft. Myers-Naples	WINK
3	2021-12-30 6:37 PM	Greensboro-H.Point-W.Salem	WFMY
4	21-12-30 6:37 PM	Atlanta	WGCL
5	2021-12-30 6:37 PM	Orlando-Daytona Bch-Melbrn	WKMG
6	2021-12-30 6:37 PM	New York	WCBS
7	2021-12-30 6:37 PM	Detroit	WWJ
8	2021-12-30 6:37 PM	Boston (Manchester)	WBZ
9	021-12-30 6:37 PM	Sacramnto-Stkton-Modesto	KOVR
10	2021-12-30 6:37 PM	Hartford & New Haven	WFSB
11	2021-12-30 6:37 PM	Baltimore	WJZ
12	021-12-30 6:37 PM	Raleigh-Durham (Fayetvll)	WNCN
13	2021-12-30 6:37 PM	Pittsburgh	KDKA
14	2021-12-30 6:37 PM	Portland, OR	KOIN
15	2021-12-30 6:37 PM	Seattle-Tacoma	KIRO
16	2021-12-30 6:37 PM	Los Angeles	KCBS
17	021-12-30 6:37 PM	San Francisco-Oak-San Jose	KPIX
18	21-12-30 6:37 PM	Washington, DC (Hagrstwn)	WUSA
19	2021-12-30 6:37 PM	Louisville	WLKY

```

                                text \
0  have been destroyed. ap, the associated press ...
1  kusa, the nbc affiliate in denver, this is in ...
2  several communities were evacuated because of ...
3  the evening ending overnight. several communit...
4  the evening ending overnight. several communit...
5  the evening ending overnight. several communit...
6  several communities were evacuated because of ...
7  the evening ending overnight. several communit...
8  the evening ending overnight. several communit...
9  mudslides, major. >> garrett: jonathan vigliot...
10 and here in southern california, the rain will...
11 the evening ending overnight. several communit...
12 the evening ending overnight. several communit...
13 the evening ending overnight. several communit...
14 the rain will continue through the evening end...
15 the rain will continue through the evening end...
16 like this, and the risk of mudslides, major. >...
17 evacuated because of flooding like this, and t...
18 several communities were evacuated because of ...

```


19 the evening ending overnight. several communit...

	matches	row_to_use	use_row \
0	[0]	0	True
1	[1]	1	True
2	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
3	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
4	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
5	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
6	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
7	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
8	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
9	[2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 1...	14	False
10	[2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, ...	14	False
11	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
12	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
13	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
14	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	True
15	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
16	[2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 1...	14	False
17	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
18	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
19	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False

	words \
0	[have, been, destroyed, ap, the, associated, p...
1	[kusa, the, nbc, affiliate, in, denver, this, ...
2	[several, communities, were, evacuated, becaus...
3	[the, evening, ending, overnight, several, com...
4	[the, evening, ending, overnight, several, com...
5	[the, evening, ending, overnight, several, com...
6	[several, communities, were, evacuated, becaus...
7	[the, evening, ending, overnight, several, com...
8	[the, evening, ending, overnight, several, com...
9	[mudslides, major, , garrett, jonathan, viglio...
10	[and, here, in, southern, california, the, rai...
11	[the, evening, ending, overnight, several, com...
12	[the, evening, ending, overnight, several, com...
13	[the, evening, ending, overnight, several, com...
14	[the, rain, will, continue, through, the, even...
15	[the, rain, will, continue, through, the, even...
16	[like, this, and, the, risk, of, mudslides, ma...
17	[evacuated, because, of, flooding, like, this,...
18	[several, communities, were, evacuated, becaus...
19	[the, evening, ending, overnight, several, com...

clean_words \

```

0 [have, been, destroyed, ap, the, associated, p...
1 [kusa, the, nbc, affiliate, in, denver, this, ...
2 [several, communities, were, evacuated, becaus...
3 [the, evening, ending, overnight, several, com...
4 [the, evening, ending, overnight, several, com...
5 [the, evening, ending, overnight, several, com...
6 [several, communities, were, evacuated, becaus...
7 [the, evening, ending, overnight, several, com...
8 [the, evening, ending, overnight, several, com...
9 [mudslides, major, garrett, jonathan, vigliott...
10 [and, here, in, southern, california, the, rai...
11 [the, evening, ending, overnight, several, com...
12 [the, evening, ending, overnight, several, com...
13 [the, evening, ending, overnight, several, com...
14 [the, rain, will, continue, through, the, even...
15 [the, rain, will, continue, through, the, even...
16 [like, this, and, the, risk, of, mudslides, ma...
17 [evacuated, because, of, flooding, like, this,...
18 [several, communities, were, evacuated, becaus...
19 [the, evening, ending, overnight, several, com...

```

```

climate_words_found
0 []
1 []
2 [severe]
3 [severe]
4 [severe]
5 [severe]
6 [severe]
7 [severe]
8 [severe]
9 [high, severe, temperature]
10 [severe]
11 [severe]
12 [severe]
13 [severe]
14 [severe]
15 [severe]
16 [severe]
17 [severe, temperature]
18 [severe]
19 [severe]

```

```
[29]: df['climate_phrases_found'] = df.apply(lambda row:
    ↪ fetch_climate_phrases_in_text(row['text']), axis=1)
```

```
[30]: df.head(20)
```

```
[30]:
```

	time	location	station \
0	2021-12-30 6:49 PM	Phoenix (Prescott)	KPHO
1	2021-12-30 6:48 PM	Phoenix (Prescott)	KPHO
2	2021-12-30 6:37 PM	Ft. Myers-Naples	WINK
3	2021-12-30 6:37 PM	Greensboro-H.Point-W.Salem	WFMY
4	21-12-30 6:37 PM	Atlanta	WGCL
5	2021-12-30 6:37 PM	Orlando-Daytona Bch-Melbrn	WKMG
6	2021-12-30 6:37 PM	New York	WCBS
7	2021-12-30 6:37 PM	Detroit	WWJ
8	2021-12-30 6:37 PM	Boston (Manchester)	WBZ
9	021-12-30 6:37 PM	Sacramnto-Stkton-Modesto	KOVR
10	2021-12-30 6:37 PM	Hartford & New Haven	WFSB
11	2021-12-30 6:37 PM	Baltimore	WJZ
12	021-12-30 6:37 PM	Raleigh-Durham (Fayetvll)	WNCN
13	2021-12-30 6:37 PM	Pittsburgh	KDKA
14	2021-12-30 6:37 PM	Portland, OR	KOIN
15	2021-12-30 6:37 PM	Seattle-Tacoma	KIRO
16	2021-12-30 6:37 PM	Los Angeles	KCBS
17	021-12-30 6:37 PM	San Francisco-Oak-San Jose	KPIX
18	21-12-30 6:37 PM	Washington, DC (Hagrstwn)	WUSA
19	2021-12-30 6:37 PM	Louisville	WLKY

	text \
0	have been destroyed. ap, the associated press ...
1	kusa, the nbc affiliate in denver, this is in ...
2	several communities were evacuated because of ...
3	the evening ending overnight. several communit...
4	the evening ending overnight. several communit...
5	the evening ending overnight. several communit...
6	several communities were evacuated because of ...
7	the evening ending overnight. several communit...
8	the evening ending overnight. several communit...
9	mudslides, major. >> garrett: jonathan vigliot...
10	and here in southern california, the rain will...
11	the evening ending overnight. several communit...
12	the evening ending overnight. several communit...
13	the evening ending overnight. several communit...
14	the rain will continue through the evening end...
15	the rain will continue through the evening end...
16	like this, and the risk of mudslides, major. >...
17	evacuated because of flooding like this, and t...
18	several communities were evacuated because of ...
19	the evening ending overnight. several communit...

	matches	row_to_use	use_row \
0	[0]	0	True
1	[1]	1	True

2	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
3	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
4	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
5	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
6	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
7	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
8	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
9	[2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 1...	14	False
10	[2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, ...	14	False
11	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
12	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
13	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
14	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	True
15	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
16	[2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 1...	14	False
17	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
18	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False
19	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	False

words \

0	[have, been, destroyed, ap, the, associated, p...
1	[kusa, the, nbc, affiliate, in, denver, this, ...
2	[several, communities, were, evacuated, becaus...
3	[the, evening, ending, overnight, several, com...
4	[the, evening, ending, overnight, several, com...
5	[the, evening, ending, overnight, several, com...
6	[several, communities, were, evacuated, becaus...
7	[the, evening, ending, overnight, several, com...
8	[the, evening, ending, overnight, several, com...
9	[mudslides, major, , garrett, jonathan, viglio...
10	[and, here, in, southern, california, the, rai...
11	[the, evening, ending, overnight, several, com...
12	[the, evening, ending, overnight, several, com...
13	[the, evening, ending, overnight, several, com...
14	[the, rain, will, continue, through, the, even...
15	[the, rain, will, continue, through, the, even...
16	[like, this, and, the, risk, of, mudslides, ma...
17	[evacuated, because, of, flooding, like, this,...
18	[several, communities, were, evacuated, becaus...
19	[the, evening, ending, overnight, several, com...

clean_words \

0	[have, been, destroyed, ap, the, associated, p...
1	[kusa, the, nbc, affiliate, in, denver, this, ...
2	[several, communities, were, evacuated, becaus...
3	[the, evening, ending, overnight, several, com...
4	[the, evening, ending, overnight, several, com...

```

5 [the, evening, ending, overnight, several, com...
6 [several, communities, were, evacuated, becaus...
7 [the, evening, ending, overnight, several, com...
8 [the, evening, ending, overnight, several, com...
9 [mudslides, major, garrett, jonathan, vigliott...
10 [and, here, in, southern, california, the, rai...
11 [the, evening, ending, overnight, several, com...
12 [the, evening, ending, overnight, several, com...
13 [the, evening, ending, overnight, several, com...
14 [the, rain, will, continue, through, the, even...
15 [the, rain, will, continue, through, the, even...
16 [like, this, and, the, risk, of, mudslides, ma...
17 [evacuated, because, of, flooding, like, this,...
18 [several, communities, were, evacuated, becaus...
19 [the, evening, ending, overnight, several, com...

```

```

          climate_words_found climate_phrases_found
0                                []                    []
1                                []                    []
2                        [severe]                    []
3                        [severe]                    []
4                        [severe]                    []
5                        [severe]                    []
6                        [severe]                    []
7                        [severe]                    []
8                        [severe]                    []
9 [high, severe, temperature]                    []
10                        [severe]                    []
11                        [severe]                    []
12                        [severe]                    []
13                        [severe]                    []
14                        [severe]                    []
15                        [severe]                    []
16                        [severe]                    []
17 [severe, temperature]                    []
18                        [severe]                    []
19                        [severe]                    []

```

```
[31]: # save data to csv
df.to_csv('reports/cbs_all.csv', encoding='utf-8')
```

```
[32]: unique_df = df[df['use_row']]
```

```
[33]: unique_df
```

```
[33]:
          time                location station \
0  2021-12-30 6:49 PM      Phoenix (Prescott)  KPHO
```

1	2021-12-30 6:48 PM	Phoenix (Prescott)	KPHO
14	2021-12-30 6:37 PM	Portland, OR	KOIN
38	2021-12-30 6:12 PM	Sacramnto-Stkton-Modesto	KOVR
39	2021-12-30 6:20 PM	Sacramnto-Stkton-Modesto	KOVR
41	2021-12-30 6:00 PM	Sacramnto-Stkton-Modesto	KOVR
42	2021-12-30 5:59 PM	Sacramnto-Stkton-Modesto	KOV
43	2021-12-30 5:55 PM	Ft. Myers-Naples	WINK
44	2021-12-30 5:44 PM	Baltimore	WJZ
45	2021-12-30 5:44 PM	Baltimore	WJZ
47	2021-12-30 5:43 PM	Seattle-Tacoma	KIRO
65	2021-12-30 5:34 PM	Pittsburgh	KDKA
66	2021-12-30 5:31 PM	Sacramnto-Stkton-Modesto	KOVR
67	2021-12-30 5:17 PM	Tulsa	KOTV
68	2021-12-30 5:14 PM	Portland, OR	KOIN
69	2021-12-30 5:12 PM	San Diego	KFMB
71	2021-12-30 5:06 PM	Phoenix (Prescott)	KPHO
72	2021-12-30 5:04 PM	Sacramnto-Stkton-Modesto	KOVR
73	2021-12-30 5:02 PM	Sacramnto-Stkton-Modesto	KOVR
74	2021-12-30 5:01 PM	Sacramnto-Stkton-Modesto	KOVR
76	2021-12-30 6:07 PM	Miami-Ft. Lauderdale	WFOR
77	2021-12-30 6:06 PM	Dallas-Ft. Worth	KTVT
80	2021-12-30 5:33 PM	Dallas-Ft. Worth	KTVT
81	2021-12-30 5:19 PM	Miami-Ft. Lauderdale	WFOR
83	2021-12-30 5:18 PM	Miami-Ft. Lauderdale	WFOR
84	2021-12-30 5:18 PM	Miami-Ft. Lauderdale	WFOR

text \

0 have been destroyed. ap, the associated press ...
1 kusa, the nbc affiliate in denver, this is in ...
14 the rain will continue through the evening end...
38 tracking the devastating fires in colorado and...
39 here we have evacuations. i will tell you a li...
41 about. we just talked about the hospital where...
42 together. this house, burning to the ground. 3...
43 rubenstein law, injured in a car accident or s...
44 e2021-12-30 5:44 pm
45 wildfire-that engulfed parts of the area. in a...
47 outside of denver have been told to evacuate. ...
65 they are calling this now the marshall-fire-an...
66 signatures, here. this is where we have had fl...
67 sun. morning satellite imagery right now we ha...
68 >> to breaking news now, this is a live look i...
69 thousands of people have been ordered to evacu...
71 into the central rockies, and it is a dangerou...
72 we couple that with relative humidity anywhere...
73 much damage will be assessed as well. >> we st...
74 >> the-fire-crews now left the scene knowing t...

76 wouldn't be that long in duration. and then th...
 77 the omicron variant. >>> the city of paris sou...
 80 shortage. with final preps underway for the ne...
 81 >> reporter: news of the boy being badly burne...
 83 and we support c.c. s. o. and the officer earl...
 84 [w:drawing]

	matches	row_to_use	use_row	\
0	[0]	0	True	
1	[1]	1	True	
14	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	True	
38	[38]	38	True	
39	[39]	39	True	
41	[41]	41	True	
42	[42]	42	True	
43	[43]	43	True	
44	[44]	44	True	
45	[45, 46]	45	True	
47	[47]	47	True	
65	[65]	65	True	
66	[66]	66	True	
67	[67]	67	True	
68	[68]	68	True	
69	[69]	69	True	
71	[71]	71	True	
72	[72]	72	True	
73	[73]	73	True	
74	[74]	74	True	
76	[76]	76	True	
77	[77]	77	True	
80	[80]	80	True	
81	[81, 82]	81	True	
83	[83]	83	True	
84	[84]	84	True	

	words	\
0	[have, been, destroyed, ap, the, associated, p...	
1	[kusa, the, nbc, affiliate, in, denver, this, ...	
14	[the, rain, will, continue, through, the, even...	
38	[tracking, the, devastating, fires, in, colora...	
39	[here, we, have, evacuations, i, will, tell, y...	
41	[about, we, just, talked, about, the, hospital...	
42	[together, this, house, burning, to, the, grou...	
43	[rubenstein, law, injured, in, a, car, acciden...	
44	[e2021-12-30, 544, pm]	
45	[wildfire-that, engulfed, parts, of, the, area...	
47	[outside, of, denver, have, been, told, to, ev...	

65 [they, are, calling, this, now, the, marshall-...
66 [signatures, here, this, is, where, we, have, ...
67 [sun, morning, satellite, imagery, right, now,...
68 [to, breaking, news, now, this, is, a, live, l...
69 [thousands, of, people, have, been, ordered, t...
71 [into, the, central, rockies, and, it, is, a, ...
72 [we, couple, that, with, relative, humidity, a...
73 [much, damage, will, be, assessed, as, well, ,...
74 [the-fire-crews, now, left, the, scene, knowin...
76 [wouldn't, be, that, long, in, duration, and, ...
77 [the, omicron, variant, , the, city, of, paris...
80 [shortage, with, final, preps, underway, for, ...
81 [reporter, news, of, the, boy, being, badly, b...
83 [and, we, support, cc, s, o, and, the, officer...
84 [[wdrawing]]

clean_words \

0 [have, been, destroyed, ap, the, associated, p...
1 [kusa, the, nbc, affiliate, in, denver, this, ...
14 [the, rain, will, continue, through, the, even...
38 [tracking, the, devastating, fires, in, colora...
39 [here, we, have, evacuations, i, will, tell, y...
41 [about, we, just, talked, about, the, hospital...
42 [together, this, house, burning, to, the, grou...
43 [rubenstein, law, injured, in, a, car, acciden...
44 [e2021, 12, 30, pm]
45 [wildfire, that, engulfed, parts, of, the, are...
47 [outside, of, denver, have, been, told, to, ev...
65 [they, are, calling, this, now, the, marshall,...
66 [signatures, here, this, is, where, we, have, ...
67 [sun, morning, satellite, imagery, right, now,...
68 [to, breaking, news, now, this, is, a, live, l...
69 [thousands, of, people, have, been, ordered, t...
71 [into, the, central, rockies, and, it, is, a, ...
72 [we, couple, that, with, relative, humidity, a...
73 [much, damage, will, be, assessed, as, well, w...
74 [the, fire, crews, now, left, the, scene, know...
76 [wouldn't, be, that, long, in, duration, and, ...
77 [the, omicron, variant, the, city, of, paris, ...
80 [shortage, with, final, preps, underway, for, ...
81 [reporter, news, of, the, boy, being, badly, b...
83 [and, we, support, cc, s, o, and, the, officer...
84 [[wdrawing]]

climate_words_found climate_phrases_found

0 [] []
1 [] []


```

14         [severe]
38         []
39         []
41         [effect]
42         []
43         []
44         []
45         [high]
47         [ice]
65         []
66         [record]
67         [energy, high, severe]
68         [high, record]
69         [record]         [record breaking]
71         [effect]
72         []
73         [drought, extreme]
74         [heat]
76         [heat, warm]
77         []
80         [history, record, rising]
81         []
83         []
84         []

```

```

[34]: total_words = unique_df['clean_words'].str.len().sum()
      total_words

```

```

[34]: 3558

```

```

[35]: def words_found_master_list(df_clean_words):
      """Given a column of words, aggregate master list"""
      words_found = list()
      for chunk in df_clean_words:
          words_found += chunk

      return words_found

```

```

[37]: words_found = words_found_master_list(unique_df['clean_words'])
      len(words_found)

```

```

[37]: 3558

```

```

[39]: from nltk.corpus import stopwords
      from nltk import pos_tag
      from nltk.stem import WordNetLemmatizer
      from nltk.tokenize import word_tokenize

```

```

from wordcloud import STOPWORDS

def master_stopwords_list():
    """Creates a master list of stopwords from pre-existing stopwords found in
    ↪ nltk and wordcloud"""
    stop_words = set(stopwords.words("english"))
    final_stopwords = list(STOPWORDS) + list(stop_words)
    return [i.lower() for i in set(final_stopwords)]

def lemmatize_words(words):
    """Given a list of words, distill to root words"""
    lem = WordNetLemmatizer()

    lemma_list = []
    for word, tag in pos_tag(words):
        wntag = tag[0].lower()
        wntag = wntag if wntag in ['a', 'r', 'n', 'v'] else None
        if not wntag:
            lemma = word
        else:
            lemma = lem.lemmatize(word, pos=wntag)
        lemma_list.append(lemma)
    return lemma_list

def clean_lemmatized_words(lemma_words):
    """Removes stop words from the lemma list"""
    nonstop_lemma_words = []
    final_stopwords = master_stopwords_list()

    for word in lemma_words:
        if word not in final_stopwords:
            nonstop_lemma_words.append(word)

    return list(filter(None, nonstop_lemma_words))

```

```
[40]: clean_lemma_words = clean_lemmatized_words(lemmatize_words(words_found))
```

```
[41]: from nltk.probability import FreqDist
```

```

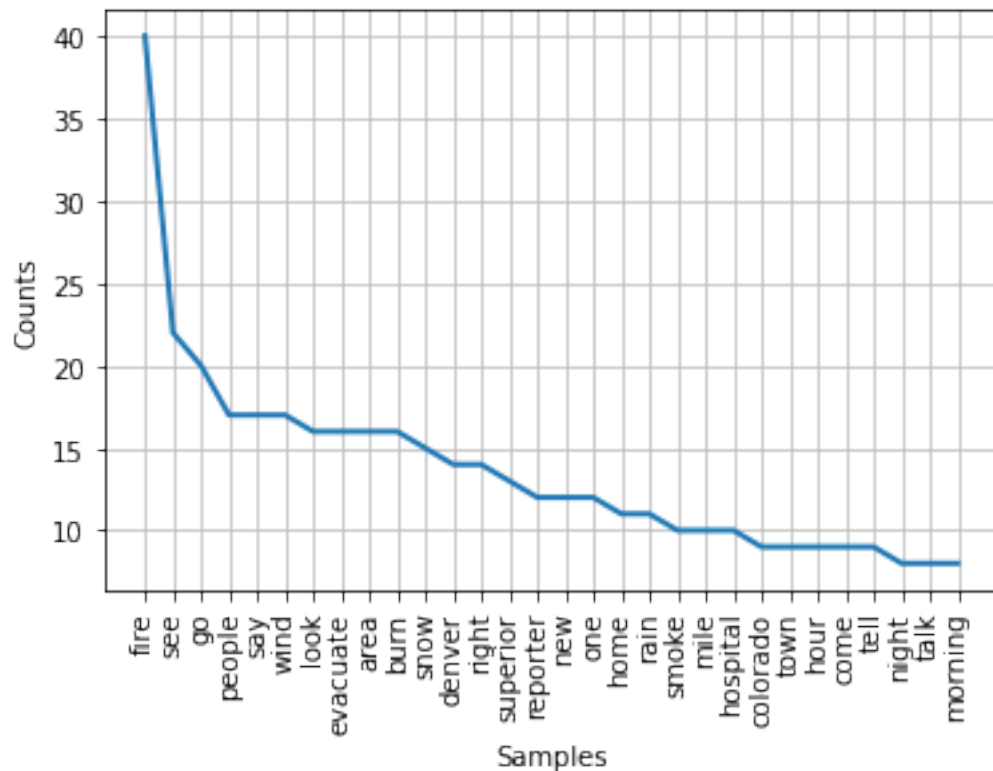
lfdist = FreqDist(clean_lemma_words)
lfdist

```

```
[41]: FreqDist({'fire': 40, 'see': 22, 'go': 20, 'people': 17, 'say': 17, 'wind': 17,
'look': 16, 'evacuate': 16, 'area': 16, 'burn': 16, ...})
```

```
[42]: import matplotlib.pyplot as plt
lfdist.plot(30, cumulative=False)
```

```
plt.show()
```



```
[43]: from wordcloud import WordCloud
from wordcloud import ImageColorGenerator
from wordcloud import STOPWORDS
import matplotlib.pyplot as plt

wordcloud = WordCloud(width = 3000, height = 2000, random_state=1,
    ↳background_color='black', colormap='Set2', collocations=False, stopwords =
    ↳master_stopwords_list()).generate_from_frequencies(lfdist)

# Plot
plt.figure( figsize=(15,10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()

#plt.savefig('word_cloud.png')
```



```

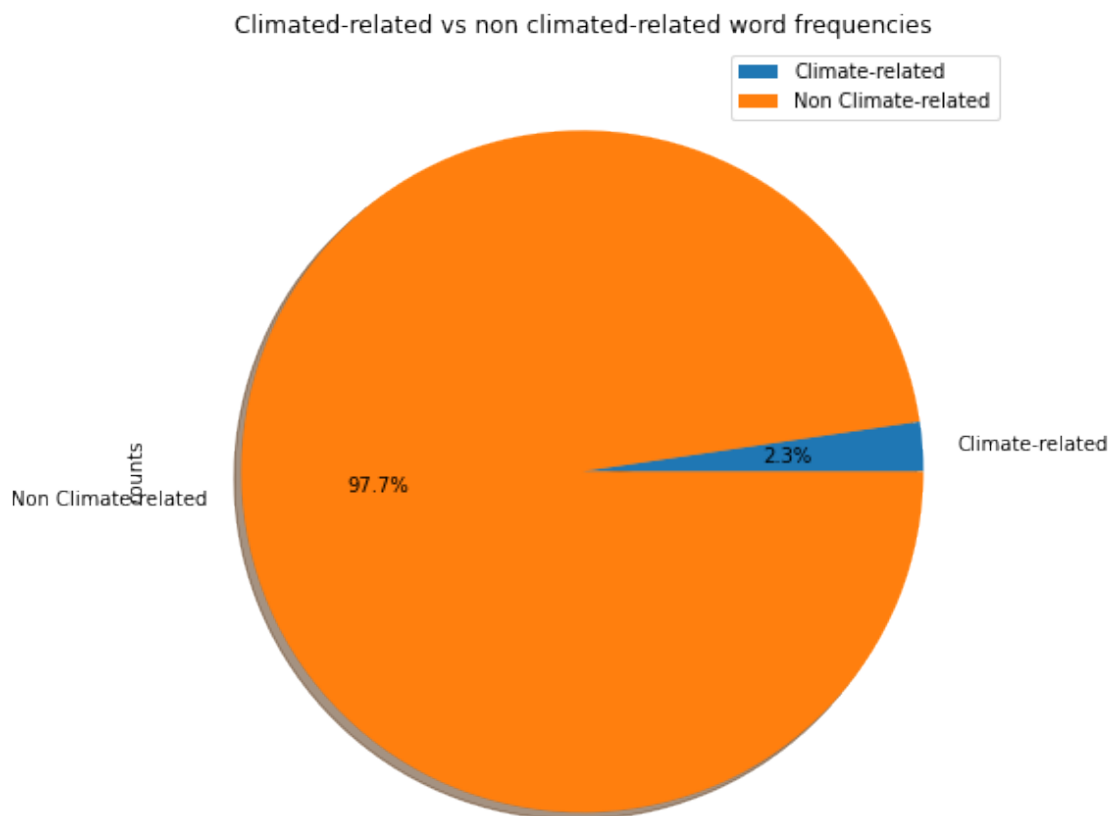
comparison_df = pd.DataFrame({'Words': ['Climate-related', 'Non-Climate-related'],
                               'counts': [climate_words_count, non_climate_words_count]})
comparison_df.set_index('Words', inplace=True)
print(comparison_df)

plot = comparison_df.plot.pie(y='counts', title="Climate-related vs non-Climate-related word frequencies",
                              legend=True, autopct='%1.1f%%', shadow=True, figsize=(8, 8))

fig = plot.get_figure()
fig.savefig("comparison.png")

```

Words	counts
Climate-related	41
Non Climate-related	1713

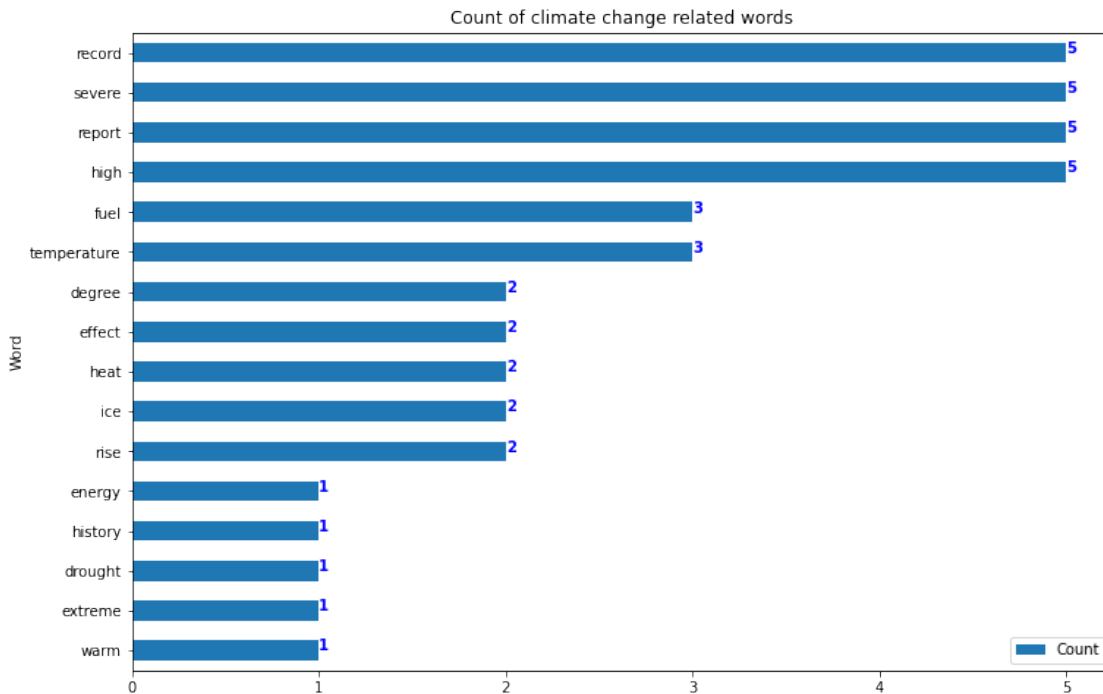


```
[46]: # find climate related word frequencies

# set figure size
fig, ax = plt.subplots(figsize=(12, 8))
# plot horizontal bar plot
climate_change_words_df.sort_values(by='Count').plot.barh(x="Word", y="Count",
    ↪ax=ax)
# set the title
plt.title("Count of climate change related words")

for i, v in enumerate(climate_change_words_df['Count'].sort_values()):
    ax.text(v, i, str(v),
            color = 'blue', fontweight = 'bold')

plt.show()
# plt.savefig('climate-related-words-breakdown.png', transparent=False)
```



```
[47]: # find segments
climate_change_words_found = list(climate_change_words_df['Word'].unique())
climate_change_words_found
```

```
[47]: ['high',
       'report',
       'severe',
```

```

'record',
'temperature',
'fuel',
'rise',
'ice',
'heat',
'effect',
'degree',
'warm',
'extreme',
'drought',
'history',
'energy']

```

```
[49]: unique_df[unique_df["climate_words_found"].str.len() != 0].to_csv('reports/
↳ cbs_final.csv', encoding='utf-8')
```

```
[50]: unique_df[unique_df["climate_words_found"].str.len() != 0]
```

```
[50]:
```

	time	location	station	\
14	2021-12-30 6:37 PM	Portland, OR	KOIN	
41	2021-12-30 6:00 PM	Sacramnto-Stkton-Modesto	KOVR	
45	2021-12-30 5:44 PM	Baltimore	WJZ	
47	2021-12-30 5:43 PM	Seattle-Tacoma	KIRO	
66	2021-12-30 5:31 PM	Sacramnto-Stkton-Modesto	KOVR	
67	2021-12-30 5:17 PM	Tulsa	KOTV	
68	2021-12-30 5:14 PM	Portland, OR	KOIN	
69	2021-12-30 5:12 PM	San Diego	KFMB	
71	2021-12-30 5:06 PM	Phoenix (Prescott)	KPHO	
73	2021-12-30 5:02 PM	Sacramnto-Stkton-Modesto	KOVR	
74	2021-12-30 5:01 PM	Sacramnto-Stkton-Modesto	KOVR	
76	2021-12-30 6:07 PM	Miami-Ft. Lauderdale	WFOR	
80	2021-12-30 5:33 PM	Dallas-Ft. Worth	KTVT	

```

text \
14 the rain will continue through the evening end...
41 about. we just talked about the hospital where...
45 wildfire-that engulfed parts of the area. in a...
47 outside of denver have been told to evacuate. ...
66 signatures, here. this is where we have had fl...
67 sun. morning satellite imagery right now we ha...
68 >> to breaking news now, this is a live look i...
69 thousands of people have been ordered to evacu...
71 into the central rockies, and it is a dangerou...
73 much damage will be assessed as well. >> we st...
74 >> the-fire-crews now left the scene knowing t...
76 wouldn't be that long in duration. and then th...

```

80 shortage. with final preps underway for the ne...

	matches	row_to_use	use_row \
14	[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1...	14	True
41	[41]	41	True
45	[45, 46]	45	True
47	[47]	47	True
66	[66]	66	True
67	[67]	67	True
68	[68]	68	True
69	[69]	69	True
71	[71]	71	True
73	[73]	73	True
74	[74]	74	True
76	[76]	76	True
80	[80]	80	True

	words \
14	[the, rain, will, continue, through, the, even...
41	[about, we, just, talked, about, the, hospital...
45	[wildfire-that, engulfed, parts, of, the, area...
47	[outside, of, denver, have, been, told, to, ev...
66	[signatures, here, this, is, where, we, have, ...
67	[sun, morning, satellite, imagery, right, now,...
68	[to, breaking, news, now, this, is, a, live, l...
69	[thousands, of, people, have, been, ordered, t...
71	[into, the, central, rockies, and, it, is, a, ...
73	[much, damage, will, be, assessed, as, well, ,...
74	[the-fire-crews, now, left, the, scene, knowin...
76	[wouldn't, be, that, long, in, duration, and, ...
80	[shortage, with, final, preps, underway, for, ...

	clean_words \
14	[the, rain, will, continue, through, the, even...
41	[about, we, just, talked, about, the, hospital...
45	[wildfire, that, engulfed, parts, of, the, are...
47	[outside, of, denver, have, been, told, to, ev...
66	[signatures, here, this, is, where, we, have, ...
67	[sun, morning, satellite, imagery, right, now,...
68	[to, breaking, news, now, this, is, a, live, l...
69	[thousands, of, people, have, been, ordered, t...
71	[into, the, central, rockies, and, it, is, a, ...
73	[much, damage, will, be, assessed, as, well, w...
74	[the, fire, crews, now, left, the, scene, know...
76	[wouldn't, be, that, long, in, duration, and, ...
80	[shortage, with, final, preps, underway, for, ...

	climate_words_found	climate_phrases_found
14	[severe]	[]
41	[effect]	[]
45	[high]	[]
47	[ice]	[]
66	[record]	[]
67	[energy, high, severe]	[]
68	[high, record]	[]
69	[record]	[record breaking]
71	[effect]	[]
73	[drought, extreme]	[]
74	[heat]	[]
76	[heat, warm]	[]
80	[history, record, rising]	[]

[]:

[]:

[]: