```python
#Part 1
#Step 1
import json
import requests
import pandas as pd
from bs4 import BeautifulSoup
r = requests.get("https://cmsc320.github.io/files/top-50-solar-flares.html")

rSoup = BeautifulSoup(r.content, 'html.parser')
#rSoup.prettify()

rTable = rSoup.find('table')


from IPython.display import display
# df = pd.DataFrame(columns=['rank', 'x_class', 'date', 'region', 'start', 'max_time', 'end', 'movie'])
tables = pd.read_html(r.content)
df = tables[0]
display(df)
```

```
    Unnamed: 0 Unnamed: 1  Unnamed: 2  Region  Start Maximum    End  \
0            1       X28+  2003/11/04     486  19:29   19:53  20:06   
1            2       X20+  2001/04/02    9393  21:32   21:51  22:03   
2            3     X17.2+  2003/10/28     486  09:51   11:10  11:24   
3            4       X17+  2005/09/07     808  17:17   17:40  18:03   
4            5      X14.4  2001/04/15    9415  13:19   13:50  13:55   
5            6        X10  2003/10/29     486  20:37   20:49  21:01   
6            7       X9.4  1997/11/06    8100  11:49   11:55  12:01   
7            8       X9.3  2017/09/06    2673  11:53   12:02  12:10   
8            9         X9  2006/12/05     930  10:18   10:35  10:45   
9           10       X8.3  2003/11/02     486  17:03   17:25  17:39   
10          11       X8.2  2017/09/10    2673  15:35   16:06  16:31   
11          12       X7.1  2005/01/20     720  06:36   07:01  07:26   
12          13       X6.9  2011/08/09    1263  07:48   08:05  08:08   
13          14       X6.5  2006/12/06     930  18:29   18:47  19:00   
14          15       X6.2  2005/09/09     808  19:13   20:04  20:36   
15          16       X6.2  2001/12/13    9733  14:20   14:30  14:35   
16          17       X5.7  2000/07/14    9077  10:03   10:24  10:43   
17          18       X5.6  2001/04/06    9415  19:10   19:21  19:31   
18          19       X5.4  2012/03/07    1429  00:02   00:24  00:40   
19          20       X5.4  2005/09/08     808  20:52   21:06  21:17   
20          21       X5.4  2003/10/23     486  08:19   08:35  08:49   
21          22       X5.3  2001/08/25    9591  16:23   16:45  17:04   
22          23       X4.9  2014/02/25    1990  00:39   00:49  01:03   
23          24       X4.9  1998/08/18    8307  22:10   22:19  22:28   
24          25       X4.8  2002/07/23      39  00:18   00:35  00:47   
25          26         X4  2000/11/26    9236  16:34   16:48  16:56   
26          27       X3.9  2003/11/03     488  09:43   09:55  10:19   
27          28       X3.9  1998/08/19    8307  21:35   21:45  21:50   
```

```
28          29       X3.8  2005/01/17     720  06:59   09:52  10:07
29          30       X3.7  1998/11/22    8384  06:30   06:42  06:49
30          31       X3.6  2005/09/09     808  09:42   09:59  10:08
31          32       X3.6  2004/07/16     649  13:49   13:55  14:01
32          33       X3.6  2003/05/28     365  00:17   00:27  00:39
33          34       X3.4  2006/12/13     930  02:14   02:40  02:57
34          35       X3.4  2001/12/28    9767  20:02   20:45  21:32
35          36       X3.3  2013/11/05    1890  22:07   22:12  22:15
36          37       X3.3  2002/07/20      39  21:04   21:30  21:54
37          38       X3.3  1998/11/28    8395  04:54   05:52  06:13
38          39       X3.2  2013/05/14    1748  00:00   01:11  01:20
39          40       X3.1  2014/10/24    2192  21:07   21:41  22:13
40          41       X3.1  2002/08/24      69  00:49   01:12  01:31
41          42         X3  2002/07/15      30  19:59   20:08  20:14
42          43       X2.8  2013/05/13    1748  15:48   16:05  16:16
43          44       X2.8  2001/12/11    9733  07:58   08:08  08:14
44          45       X2.8  1998/08/18    8307  08:14   08:24  08:32
45          46       X2.7  2015/05/05    2339  22:05   22:11  22:15
46          47       X2.7  2003/11/03     488  01:09   01:30  01:45
47          48       X2.7  1998/05/06    8210  07:58   08:09  08:20
48          49       X2.6  2005/01/15     720  22:25   23:02  23:31
49          50       X2.6  2001/09/24    9632  09:32   10:38  11:09

           Unnamed: 7
0   MovieView archive
1   MovieView archive
2   MovieView archive
3   MovieView archive
4   MovieView archive
5   MovieView archive
6   MovieView archive
7   MovieView archive
8   MovieView archive
9   MovieView archive
10  MovieView archive
11  MovieView archive
12  MovieView archive
13  MovieView archive
14  MovieView archive
15  MovieView archive
16  MovieView archive
17  MovieView archive
18  MovieView archive
19  MovieView archive
20  MovieView archive
21  MovieView archive
22  MovieView archive
23       View archive
24  MovieView archive
25  MovieView archive
```

```
26  MovieView archive
27       View archive
28  MovieView archive
29  MovieView archive
30  MovieView archive
31  MovieView archive
32  MovieView archive
33  MovieView archive
34  MovieView archive
35  MovieView archive
36  MovieView archive
37  MovieView archive
38  MovieView archive
39  MovieView archive
40  MovieView archive
41  MovieView archive
42  MovieView archive
43  MovieView archive
44       View archive
45  MovieView archive
46  MovieView archive
47  MovieView archive
48  MovieView archive
49  MovieView archive
```

```
#Step 2
#df = df.drop('Unnamed: 7', 1)
tidydf = df.rename({'Unnamed: 0': 'Rank', 'Unnamed: 1': 'X Class',
'Unnamed: 2': 'Date'}, axis=1)
display(tidydf)
```

```
    Rank X Class        Date  Region  Start Maximum    End
0      1    X28+  2003/11/04     486  19:29   19:53  20:06
1      2    X20+  2001/04/02    9393  21:32   21:51  22:03
2      3  X17.2+  2003/10/28     486  09:51   11:10  11:24
3      4    X17+  2005/09/07     808  17:17   17:40  18:03
4      5   X14.4  2001/04/15    9415  13:19   13:50  13:55
5      6     X10  2003/10/29     486  20:37   20:49  21:01
6      7    X9.4  1997/11/06    8100  11:49   11:55  12:01
7      8    X9.3  2017/09/06    2673  11:53   12:02  12:10
8      9      X9  2006/12/05     930  10:18   10:35  10:45
9     10    X8.3  2003/11/02     486  17:03   17:25  17:39
10    11    X8.2  2017/09/10    2673  15:35   16:06  16:31
11    12    X7.1  2005/01/20     720  06:36   07:01  07:26
12    13    X6.9  2011/08/09    1263  07:48   08:05  08:08
13    14    X6.5  2006/12/06     930  18:29   18:47  19:00
14    15    X6.2  2005/09/09     808  19:13   20:04  20:36
15    16    X6.2  2001/12/13    9733  14:20   14:30  14:35
16    17    X5.7  2000/07/14    9077  10:03   10:24  10:43
17    18    X5.6  2001/04/06    9415  19:10   19:21  19:31
18    19    X5.4  2012/03/07    1429  00:02   00:24  00:40
```

```
19    20    X5.4  2005/09/08     808  20:52   21:06  21:17
20    21    X5.4  2003/10/23     486  08:19   08:35  08:49
21    22    X5.3  2001/08/25    9591  16:23   16:45  17:04
22    23    X4.9  2014/02/25    1990  00:39   00:49  01:03
23    24    X4.9  1998/08/18    8307  22:10   22:19  22:28
24    25    X4.8  2002/07/23      39  00:18   00:35  00:47
25    26      X4  2000/11/26    9236  16:34   16:48  16:56
26    27    X3.9  2003/11/03     488  09:43   09:55  10:19
27    28    X3.9  1998/08/19    8307  21:35   21:45  21:50
28    29    X3.8  2005/01/17     720  06:59   09:52  10:07
29    30    X3.7  1998/11/22    8384  06:30   06:42  06:49
30    31    X3.6  2005/09/09     808  09:42   09:59  10:08
31    32    X3.6  2004/07/16     649  13:49   13:55  14:01
32    33    X3.6  2003/05/28     365  00:17   00:27  00:39
33    34    X3.4  2006/12/13     930  02:14   02:40  02:57
34    35    X3.4  2001/12/28    9767  20:02   20:45  21:32
35    36    X3.3  2013/11/05    1890  22:07   22:12  22:15
36    37    X3.3  2002/07/20      39  21:04   21:30  21:54
37    38    X3.3  1998/11/28    8395  04:54   05:52  06:13
38    39    X3.2  2013/05/14    1748  00:00   01:11  01:20
39    40    X3.1  2014/10/24    2192  21:07   21:41  22:13
40    41    X3.1  2002/08/24      69  00:49   01:12  01:31
41    42      X3  2002/07/15      30  19:59   20:08  20:14
42    43    X2.8  2013/05/13    1748  15:48   16:05  16:16
43    44    X2.8  2001/12/11    9733  07:58   08:08  08:14
44    45    X2.8  1998/08/18    8307  08:14   08:24  08:32
45    46    X2.7  2015/05/05    2339  22:05   22:11  22:15
46    47    X2.7  2003/11/03     488  01:09   01:30  01:45
47    48    X2.7  1998/05/06    8210  07:58   08:09  08:20
48    49    X2.6  2005/01/15     720  22:25   23:02  23:31
49    50    X2.6  2001/09/24    9632  09:32   10:38  11:09
```

```python
st_value = pd.to_datetime(tidydf['Date'] + ' ' + tidydf['Start'])
mt_value = pd.to_datetime(tidydf['Date'] + ' ' + tidydf['Maximum'])
et_value = pd.to_datetime(tidydf['Date'] + ' ' + tidydf['End'])

tidydf['Start DateTime'] = st_value
tidydf['Max DateTime'] = mt_value
tidydf['End DateTime'] = et_value

tidydf = tidydf.drop('Date', 1)
tidydf = tidydf.drop('Start', 1)
tidydf = tidydf.drop('Maximum', 1)
tidydf = tidydf.drop('End', 1)
```

```
/var/folders/yl/0_18jst15nb9gbl2n_j_z5tw0000gn/T/
ipykernel_3763/3178710839.py:9: FutureWarning: In a future version of
pandas all arguments of DataFrame.drop except for the argument
'labels' will be keyword-only.
  tidydf = tidydf.drop('Date', 1)
```

```
/var/folders/yl/0_18jst15nb9gbl2n_j_z5tw0000gn/T/ipykernel_3763/317871
0839.py:10: FutureWarning: In a future version of pandas all arguments
of DataFrame.drop except for the argument 'labels' will be keyword-
only.
  tidydf = tidydf.drop('Start', 1)
/var/folders/yl/0_18jst15nb9gbl2n_j_z5tw0000gn/T/ipykernel_3763/317871
0839.py:11: FutureWarning: In a future version of pandas all arguments
of DataFrame.drop except for the argument 'labels' will be keyword-
only.
  tidydf = tidydf.drop('Maximum', 1)
/var/folders/yl/0_18jst15nb9gbl2n_j_z5tw0000gn/T/ipykernel_3763/317871
0839.py:12: FutureWarning: In a future version of pandas all arguments
of DataFrame.drop except for the argument 'labels' will be keyword-
only.
  tidydf = tidydf.drop('End', 1)
```

```
tidydf = tidydf[['Rank', 'X Class', 'Start DateTime', 'Max DateTime',
'End DateTime', 'Region']]
display(tidydf)
```

```
    Rank X Class      Start DateTime        Max DateTime        End
DateTime  \
0      1    X28+ 2003-11-04 19:29:00 2003-11-04 19:53:00 2003-11-04
20:06:00
1      2    X20+ 2001-04-02 21:32:00 2001-04-02 21:51:00 2001-04-02
22:03:00
2      3  X17.2+ 2003-10-28 09:51:00 2003-10-28 11:10:00 2003-10-28
11:24:00
3      4    X17+ 2005-09-07 17:17:00 2005-09-07 17:40:00 2005-09-07
18:03:00
4      5   X14.4 2001-04-15 13:19:00 2001-04-15 13:50:00 2001-04-15
13:55:00
5      6     X10 2003-10-29 20:37:00 2003-10-29 20:49:00 2003-10-29
21:01:00
6      7    X9.4 1997-11-06 11:49:00 1997-11-06 11:55:00 1997-11-06
12:01:00
7      8    X9.3 2017-09-06 11:53:00 2017-09-06 12:02:00 2017-09-06
12:10:00
8      9      X9 2006-12-05 10:18:00 2006-12-05 10:35:00 2006-12-05
10:45:00
9     10    X8.3 2003-11-02 17:03:00 2003-11-02 17:25:00 2003-11-02
17:39:00
10    11    X8.2 2017-09-10 15:35:00 2017-09-10 16:06:00 2017-09-10
16:31:00
11    12    X7.1 2005-01-20 06:36:00 2005-01-20 07:01:00 2005-01-20
07:26:00
12    13    X6.9 2011-08-09 07:48:00 2011-08-09 08:05:00 2011-08-09
08:08:00
13    14    X6.5 2006-12-06 18:29:00 2006-12-06 18:47:00 2006-12-06
19:00:00
14    15    X6.2 2005-09-09 19:13:00 2005-09-09 20:04:00 2005-09-09
```

```
20:36:00
15    16    X6.2 2001-12-13 14:20:00 2001-12-13 14:30:00 2001-12-13
14:35:00
16    17    X5.7 2000-07-14 10:03:00 2000-07-14 10:24:00 2000-07-14
10:43:00
17    18    X5.6 2001-04-06 19:10:00 2001-04-06 19:21:00 2001-04-06
19:31:00
18    19    X5.4 2012-03-07 00:02:00 2012-03-07 00:24:00 2012-03-07
00:40:00
19    20    X5.4 2005-09-08 20:52:00 2005-09-08 21:06:00 2005-09-08
21:17:00
20    21    X5.4 2003-10-23 08:19:00 2003-10-23 08:35:00 2003-10-23
08:49:00
21    22    X5.3 2001-08-25 16:23:00 2001-08-25 16:45:00 2001-08-25
17:04:00
22    23    X4.9 2014-02-25 00:39:00 2014-02-25 00:49:00 2014-02-25
01:03:00
23    24    X4.9 1998-08-18 22:10:00 1998-08-18 22:19:00 1998-08-18
22:28:00
24    25    X4.8 2002-07-23 00:18:00 2002-07-23 00:35:00 2002-07-23
00:47:00
25    26      X4 2000-11-26 16:34:00 2000-11-26 16:48:00 2000-11-26
16:56:00
26    27    X3.9 2003-11-03 09:43:00 2003-11-03 09:55:00 2003-11-03
10:19:00
27    28    X3.9 1998-08-19 21:35:00 1998-08-19 21:45:00 1998-08-19
21:50:00
28    29    X3.8 2005-01-17 06:59:00 2005-01-17 09:52:00 2005-01-17
10:07:00
29    30    X3.7 1998-11-22 06:30:00 1998-11-22 06:42:00 1998-11-22
06:49:00
30    31    X3.6 2005-09-09 09:42:00 2005-09-09 09:59:00 2005-09-09
10:08:00
31    32    X3.6 2004-07-16 13:49:00 2004-07-16 13:55:00 2004-07-16
14:01:00
32    33    X3.6 2003-05-28 00:17:00 2003-05-28 00:27:00 2003-05-28
00:39:00
33    34    X3.4 2006-12-13 02:14:00 2006-12-13 02:40:00 2006-12-13
02:57:00
34    35    X3.4 2001-12-28 20:02:00 2001-12-28 20:45:00 2001-12-28
21:32:00
35    36    X3.3 2013-11-05 22:07:00 2013-11-05 22:12:00 2013-11-05
22:15:00
36    37    X3.3 2002-07-20 21:04:00 2002-07-20 21:30:00 2002-07-20
21:54:00
37    38    X3.3 1998-11-28 04:54:00 1998-11-28 05:52:00 1998-11-28
06:13:00
38    39    X3.2 2013-05-14 00:00:00 2013-05-14 01:11:00 2013-05-14
01:20:00
39    40    X3.1 2014-10-24 21:07:00 2014-10-24 21:41:00 2014-10-24
```

```
22:13:00
40    41    X3.1 2002-08-24 00:49:00 2002-08-24 01:12:00 2002-08-24
01:31:00
41    42      X3 2002-07-15 19:59:00 2002-07-15 20:08:00 2002-07-15
20:14:00
42    43    X2.8 2013-05-13 15:48:00 2013-05-13 16:05:00 2013-05-13
16:16:00
43    44    X2.8 2001-12-11 07:58:00 2001-12-11 08:08:00 2001-12-11
08:14:00
44    45    X2.8 1998-08-18 08:14:00 1998-08-18 08:24:00 1998-08-18
08:32:00
45    46    X2.7 2015-05-05 22:05:00 2015-05-05 22:11:00 2015-05-05
22:15:00
46    47    X2.7 2003-11-03 01:09:00 2003-11-03 01:30:00 2003-11-03
01:45:00
47    48    X2.7 1998-05-06 07:58:00 1998-05-06 08:09:00 1998-05-06
08:20:00
48    49    X2.6 2005-01-15 22:25:00 2005-01-15 23:02:00 2005-01-15
23:31:00
49    50    X2.6 2001-09-24 09:32:00 2001-09-24 10:38:00 2001-09-24
11:09:00

    Region
0      486
1     9393
2      486
3      808
4     9415
5      486
6     8100
7     2673
8      930
9      486
10    2673
11     720
12    1263
13     930
14     808
15    9733
16    9077
17    9415
18    1429
19     808
20     486
21    9591
22    1990
23    8307
24      39
25    9236
26     488
```

```
27    8307
28     720
29    8384
30     808
31     649
32     365
33     930
34    9767
35    1890
36      39
37    8395
38    1748
39    2192
40      69
41      30
42    1748
43    9733
44    8307
45    2339
46     488
47    8210
48     720
49    9632
```

```
#Part 3
from bs4 import BeautifulSoup
import urllib3
import requests
#http = urllib3.PoolManager()
r = requests.get("https://cmsc320.github.io/files/waves_type2.html")
rSoup2 = BeautifulSoup(r.content, 'html.parser')
rTable2 = rSoup2.find('pre')
rText = rTable2.get_text()
entries = rText.split('\n')
row_num = 0
for i in range(0,12):
    entries.pop(0)
entries.pop(len(entries)-1)
entries.pop(len(entries)-1)
entries.pop(len(entries)-1)
df2 = pd.DataFrame(index=range(0, len(entries)), columns=['Start Date', 'Start Time', 'End Date', 'End Time', 'Start Freq', 'End Freq', 'Flare Location', 'Flare Region', 'Flare Class', 'CME Date', 'CME Time', 'CME Angle', 'CME Width', 'CME Speed'])
for entry in entries:
    cols = entry.split(' ')
    while '' in cols:
        cols.remove('')
    col_num = 0
    while col_num < 14:
        df2.iat[row_num, col_num] = cols[col_num]
        col_num += 1
    row_num += 1

df2
```

```
     Start Date Start Time End Date End Time Start Freq End Freq  \
0    1997/04/01      14:00    04/01    14:15       8000     4000   
1    1997/04/07      14:30    04/07    17:30      11000     1000   
2    1997/05/12      05:15    05/14    16:00      12000       80   
3    1997/05/21      20:20    05/21    22:00       5000      500   
4    1997/09/23      21:53    09/23    22:16       6000     2000   
..          ...        ...      ...      ...        ...      ...   
512  2017/07/23      05:27    07/23    06:12       4400      900   
513  2017/09/04      20:27    09/05    04:54      14000      210   
514  2017/09/06      12:05    09/07    08:00      16000       70   
515  2017/09/10      16:02    09/11    06:50      16000      150   
516  2017/09/12      07:38    09/12    07:43      16000    13000   

    Flare Location Flare Region Flare Class CME Date CME Time CME
Angle  \
0           S25E16         8026        M1.3    04/01    15:18
74   
1           S28E19         8027        C6.8    04/07    14:27
Halo   
2           N21W08         8038        C1.3    05/12    05:30
Halo   
3           N05W12         8040        M1.3    05/21    21:00
263   
4           S29E25         8088        C1.4    09/23    22:02
133   
..             ...          ...         ...      ...      ...       ..
.   
512           BACK        -----        ----    07/23    04:48
Halo   
513         S10W12        12673        M5.5    09/04    20:12
Halo   
514         S08W33        12673        X9.3    09/06    12:24
Halo   
515         S09W92        -----        X8.3    09/10    16:00
Halo   
516         N08E48        12680        C3.0    09/12    08:03
124   

    CME Width CME Speed  
0          79       312  
1         360       878  
2         360       464  
3         165       296  
4         155       712  
```

```
..        ...       ...
512       360      1848
513       360      1418
514       360      1571
515       360      3163
516        96       252

[517 rows x 14 columns]
```

```
#Step 4
tidydf2 = df2.replace('----', 'NaN').replace('-----', 'NaN')
tidydf2['Is Halo?'] = tidydf2['CME Angle'].map(lambda i: i == 'Halo')
tidydf2 = tidydf2.replace('Halo', 'NA')
#if first char of string is >, it is a lower bound
tidydf2['Width Lower Bound?'] = tidydf2['CME Width'].map(lambda i:
(str(i))[0] == '>')
tidydf2['CME Width'] = tidydf2['CME Width'].map(lambda i:
"".join(filter(str.isdigit, i)))
tidydf2['Start DateTime'] = pd.to_datetime(tidydf2['Start Date'] + ' '
+ tidydf2['Start Time'])
tidydf2['End DateTime'] = ""
tidydf2['CME DateTime'] = ""

#to get the date right so we can to pd.to_datetime
for row in range(0,len(entries)):
    dateArray = tidydf2.values[row][0].split('/')
    year = dateArray[0]
    end_daymonth = tidydf2.values[row][2]
    end_dt = year + '/' + end_daymonth
    tidydf2.iat[row, 2] = end_dt
    cme_daymonth = tidydf2.values[row][9]
    cme_dt = year + '/' + cme_daymonth
    tidydf2.iat[row, 9] = cme_dt

display(tidydf2)
```

```
     Start Date Start Time    End Date End Time Start Freq End Freq  \
0    1997/04/01      14:00  1997/04/01    14:15       8000     4000
1    1997/04/07      14:30  1997/04/07    17:30      11000     1000
2    1997/05/12      05:15  1997/05/14    16:00      12000       80
3    1997/05/21      20:20  1997/05/21    22:00       5000      500
4    1997/09/23      21:53  1997/09/23    22:16       6000     2000
..          ...        ...         ...      ...        ...      ...
512  2017/07/23      05:27  2017/07/23    06:12       4400      900
513  2017/09/04      20:27  2017/09/05    04:54      14000      210
514  2017/09/06      12:05  2017/09/07    08:00      16000       70
515  2017/09/10      16:02  2017/09/11    06:50      16000      150
516  2017/09/12      07:38  2017/09/12    07:43      16000    13000

    Flare Location Flare Region Flare Class    CME Date CME Time CME
```

```
Angle  \
0           S25E16         8026        M1.3  1997/04/01    15:18
74
1           S28E19         8027        C6.8  1997/04/07    14:27
NA
2           N21W08         8038        C1.3  1997/05/12    05:30
NA
3           N05W12         8040        M1.3  1997/05/21    21:00
263
4           S29E25         8088        C1.4  1997/09/23    22:02
133
..             ...          ...         ...         ...      ...
...
512           BACK          NaN         NaN  2017/07/23    04:48
NA
513         S10W12        12673        M5.5  2017/09/04    20:12
NA
514         S08W33        12673        X9.3  2017/09/06    12:24
NA
515         S09W92          NaN        X8.3  2017/09/10    16:00
NA
516         N08E48        12680        C3.0  2017/09/12    08:03
124

    CME Width CME Speed  Is Halo?  Width Lower Bound?      Start
DateTime  \
0          79       312     False               False 1997-04-01
14:00:00
1         360       878      True               False 1997-04-07
14:30:00
2         360       464      True               False 1997-05-12
05:15:00
3         165       296     False               False 1997-05-21
20:20:00
4         155       712     False               False 1997-09-23
21:53:00
..        ...       ...       ...                 ...
...
512       360      1848      True               False 2017-07-23
05:27:00
513       360      1418      True               False 2017-09-04
20:27:00
514       360      1571      True               False 2017-09-06
12:05:00
515       360      3163      True               False 2017-09-10
16:02:00
516        96       252     False               False 2017-09-12
07:38:00

    End DateTime CME DateTime
```

```
0
1
2
3
4
..           ...          ...
512
513
514
515
516

[517 rows x 19 columns]
```

```
from datetime import datetime
tidydf2 = tidydf2[['End Date', 'End Time', 'CME Date', 'CME Time',
'Start DateTime', 'End DateTime', 'Start Freq', 'End Freq', 'Flare
Location', 'Flare Region', 'Flare Class', 'CME DateTime', 'CME Angle',
'CME Width', 'CME Speed', 'Is Halo?', 'Width Lower Bound?']]
for row in range(0, len(entries)):
    timeCheckEnd = tidydf2.values[row][1].split(':')
    timeCheckCME = tidydf2.values[row][3].split(':')
    #if the time is 24, leave blank
    if (int(timeCheckEnd[0]) < 24):
        tidydf2.iat[row, 5] = tidydf2.values[row][0].replace('/','-')
+ ' ' + tidydf2.values[row][1] + ':00'
    if (str(timeCheckCME[0]) == "--"):
        tidydf2.iat[row, 11] = ""
    else: tidydf2.iat[row, 11] = tidydf2.values[row]
[2].replace('/','-') + ' ' + tidydf2.values[row][3] + ':00'
tidydf2 = tidydf2.drop('End Date', 1)
tidydf2 = tidydf2.drop('End Time', 1)
tidydf2 = tidydf2.drop('CME Date', 1)
tidydf2 = tidydf2.drop('CME Time', 1)
display(tidydf2)
```

```
/var/folders/yl/0_18jst15nb9gbl2n_j_z5tw0000gn/T/
ipykernel_3763/1317046877.py:12: FutureWarning: In a future version of
pandas all arguments of DataFrame.drop except for the argument
'labels' will be keyword-only.
  tidydf2 = tidydf2.drop('End Date', 1)
/var/folders/yl/0_18jst15nb9gbl2n_j_z5tw0000gn/T/ipykernel_3763/131704
6877.py:13: FutureWarning: In a future version of pandas all arguments
of DataFrame.drop except for the argument 'labels' will be keyword-
only.
  tidydf2 = tidydf2.drop('End Time', 1)
/var/folders/yl/0_18jst15nb9gbl2n_j_z5tw0000gn/T/ipykernel_3763/131704
6877.py:14: FutureWarning: In a future version of pandas all arguments
of DataFrame.drop except for the argument 'labels' will be keyword-
only.
```

```
  tidydf2 = tidydf2.drop('CME Date', 1)
/var/folders/yl/0_18jst15nb9gbl2n_j_z5tw0000gn/T/ipykernel_3763/131704
6877.py:15: FutureWarning: In a future version of pandas all arguments
of DataFrame.drop except for the argument 'labels' will be keyword-
only.
  tidydf2 = tidydf2.drop('CME Time', 1)

         Start DateTime         End DateTime Start Freq End Freq  \
0   1997-04-01 14:00:00  1997-04-01 14:15:00       8000     4000   
1   1997-04-07 14:30:00  1997-04-07 17:30:00      11000     1000   
2   1997-05-12 05:15:00  1997-05-14 16:00:00      12000       80   
3   1997-05-21 20:20:00  1997-05-21 22:00:00       5000      500   
4   1997-09-23 21:53:00  1997-09-23 22:16:00       6000     2000   
..                  ...                  ...        ...      ...   
512 2017-07-23 05:27:00  2017-07-23 06:12:00       4400      900   
513 2017-09-04 20:27:00  2017-09-05 04:54:00      14000      210   
514 2017-09-06 12:05:00  2017-09-07 08:00:00      16000       70   
515 2017-09-10 16:02:00  2017-09-11 06:50:00      16000      150   
516 2017-09-12 07:38:00  2017-09-12 07:43:00      16000    13000   

    Flare Location Flare Region Flare Class         CME DateTime CME
Angle  \
0           S25E16         8026        M1.3  1997-04-01 15:18:00
74   
1           S28E19         8027        C6.8  1997-04-07 14:27:00
NA   
2           N21W08         8038        C1.3  1997-05-12 05:30:00
NA   
3           N05W12         8040        M1.3  1997-05-21 21:00:00
263   
4           S29E25         8088        C1.4  1997-09-23 22:02:00
133   
..             ...          ...         ...                  ...
...   
512           BACK          NaN         NaN  2017-07-23 04:48:00
NA   
513         S10W12        12673        M5.5  2017-09-04 20:12:00
NA   
514         S08W33        12673        X9.3  2017-09-06 12:24:00
NA   
515         S09W92          NaN        X8.3  2017-09-10 16:00:00
NA   
516         N08E48        12680        C3.0  2017-09-12 08:03:00
124   

    CME Width CME Speed  Is Halo?  Width Lower Bound?  
0          79       312     False               False  
1         360       878      True               False  
2         360       464      True               False  
3         165       296     False               False  
```

```
4         155       712     False               False
..        ...       ...       ...                 ...
512       360      1848      True               False
513       360      1418      True               False
514       360      1571      True               False
515       360      3163      True               False
516        96       252     False               False

[517 rows x 13 columns]
```

```
#Part 2
#Question 1
top50SF = tidydf2.loc[tidydf2['Flare Class'].str.contains('X')]
top50SF['Flare Class'] = top50SF['Flare Class'].str.lstrip('X')
#float to compare
top50SF = top50SF.astype({'Flare Class': float})
top50SF = top50SF.sort_values('Flare Class', ascending = False)
top50SF = top50SF.head(50)
top50SF = top50SF.astype({'Flare Class': str})
top50SF['Flare Class'] = "X" + top50SF['Flare Class']
display(top50SF)
```

```
/var/folders/yl/0_18jst15nb9gbl2n_j_z5tw0000gn/T/
ipykernel_3763/1127006976.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  top50SF['Flare Class'] = top50SF['Flare Class'].str.lstrip('X')
```

```
         Start DateTime         End DateTime Start Freq End Freq  \
240 2003-11-04 20:00:00                           10000      200
117 2001-04-02 22:05:00  2001-04-03 02:30:00      14000      250
233 2003-10-28 11:10:00                           14000       40
126 2001-04-15 14:05:00  2001-04-16 13:00:00      14000       40
234 2003-10-29 20:55:00                           11000      500
8   1997-11-06 12:20:00  1997-11-07 08:30:00      14000      100
514 2017-09-06 12:05:00  2017-09-07 08:00:00      16000       70
328 2006-12-05 10:50:00  2006-12-05 20:00:00      14000      250
237 2003-11-02 17:30:00  2003-11-03 01:00:00      12000      250
515 2017-09-10 16:02:00  2017-09-11 06:50:00      16000      150
288 2005-01-20 07:15:00  2005-01-20 16:30:00      14000       25
359 2011-08-09 08:20:00  2011-08-09 08:35:00      16000     4000
331 2006-12-06 19:00:00                           16000       30
317 2005-09-09 19:45:00  2005-09-09 22:00:00      10000       50
82  2000-07-14 10:30:00  2000-07-15 14:30:00      14000       80
121 2001-04-06 19:35:00  2001-04-07 01:50:00      14000      230
375 2012-03-07 01:00:00  2012-03-08 19:00:00      16000       30
135 2001-08-25 16:50:00  2001-08-25 23:00:00       8000      170
```

```
443 2014-02-25 00:56:00  2014-02-25 11:28:00      14000      100
193 2002-07-23 00:50:00  2002-07-23 04:00:00      11000      400
104 2000-11-26 17:00:00  2000-11-26 17:15:00      14000     7000
239 2003-11-03 10:00:00  2003-11-03 12:30:00       6000      400
286 2005-01-17 10:00:00  2005-01-17 10:35:00       6100     1500
222 2003-05-28 01:00:00  2003-05-29 00:30:00       1000      200
332 2006-12-13 02:45:00  2006-12-13 10:40:00      12000      150
160 2001-12-28 20:35:00  2001-12-29 03:00:00      14000      350
192 2002-07-20 21:30:00  2002-07-20 22:20:00      10000     2000
404 2013-05-14 01:16:00  2013-05-14 08:20:00      16000      240
201 2002-08-24 01:45:00  2002-08-24 03:25:00       5000      400
403 2013-05-13 16:15:00  2013-05-13 19:10:00      16000      300
487 2015-05-05 22:24:00  2015-05-05 23:14:00      14000      500
19  1998-05-06 08:25:00  1998-05-06 08:35:00      14000     5000
238 2003-11-03 01:15:00  2003-11-03 01:25:00       3000     1500
284 2005-01-15 23:00:00  2005-01-17 00:00:00       3000       40
142 2001-09-24 10:45:00  2001-09-25 20:00:00       7000       30
9   1997-11-27 13:30:00  1997-11-27 14:00:00      14000     7000
276 2004-11-10 02:25:00  2004-11-10 03:40:00      14000     1000
123 2001-04-10 05:24:00                           14000      100
99  2000-11-24 15:25:00  2000-11-24 22:00:00      14000      200
73  2000-06-06 15:20:00  2000-06-08 09:00:00      14000       40
345 2011-02-15 02:10:00  2011-02-15 07:00:00      16000      400
318 2005-09-10 21:45:00  2005-09-11 01:00:00      14000      200
361 2011-09-06 22:30:00  2011-09-07 15:40:00      16000      150
420 2013-10-25 15:08:00  2013-10-25 22:32:00      16000      200
7   1997-11-04 06:00:00  1997-11-05 04:30:00      14000      100
98  2000-11-24 05:10:00  2000-11-24 15:00:00      14000      100
125 2001-04-12 10:20:00  2001-04-12 10:40:00      14000     7000
274 2004-11-07 16:25:00  2004-11-08 20:00:00      14000       60
285 2005-01-17 09:25:00  2005-01-17 16:00:00      14000       30
102 2000-11-25 19:00:00  2000-11-25 19:35:00       6000     2000

    Flare Location Flare Region Flare Class         CME DateTime CME
Angle  \
240         S19W83        10486       X28.0  2003-11-04 19:54:00
NA
117         N19W72         9393       X20.0  2001-04-02 22:06:00
261
233         S16E08        10486       X17.0  2003-10-28 11:30:00
NA
126         S20W85         9415       X14.0  2001-04-15 14:06:00
245
234         S15W02        10486       X10.0  2003-10-29 20:54:00
NA
8           S18W63         8100        X9.4  1997-11-06 12:10:00
NA
514         S08W33        12673        X9.3  2017-09-06 12:24:00
NA
328         S07E68        10930        X9.0
```

```
NaN
237         S14W56        10486        X8.3  2003-11-02 17:30:00
NA
515         S09W92          NaN        X8.3  2017-09-10 16:00:00
NA
288         N14W61        10720        X7.1  2005-01-20 06:54:00
NA
359         N17W69        11263        X6.9  2011-08-09 08:12:00
NA
331         S05E64        10930        X6.5
NaN
317         S12E67        10808        X6.2  2005-09-09 19:48:00
NA
82          N22W07         9077        X5.7  2000-07-14 10:54:00
NA
121         S21E31         9415        X5.6  2001-04-06 19:30:00
NA
375         N17E27        11429        X5.4  2012-03-07 00:24:00
NA
135         S17E34         9591        X5.3  2001-08-25 16:50:00
NA
443         S12E82        11990        X4.9  2014-02-25 01:25:00
NA
193         S13E72        10039        X4.8  2002-07-23 00:42:00
NA
104         N18W38         9236        X4.0  2000-11-26 17:06:00
NA
239         N08W77        10488        X3.9  2003-11-03 10:06:00
293
286         N15W25        10720        X3.8  2005-01-17 09:54:00
NA
222         S07W20        10365        X3.6  2003-05-28 00:50:00
NA
332         S06W23        10930        X3.4  2006-12-13 02:54:00
NA
160         S26E90         9756        X3.4  2001-12-28 20:30:00
NA
192         S13E90        10039        X3.3  2002-07-20 22:06:00
NA
404         N08E77        11748        X3.2  2013-05-14 01:25:00
NA
201         S02W81        10069        X3.1  2002-08-24 01:27:00
NA
403         N11E85        11748        X2.8  2013-05-13 16:07:00
NA
487         N15E79        12339        X2.7  2015-05-05 22:24:00
NA
19          S11W65         8210        X2.7  1998-05-06 08:29:00
309
238         N10W83        10488        X2.7  2003-11-03 01:59:00
```

```
304
284         N15W05        10720        X2.6  2005-01-15 23:06:00
NA
142         S16E23         9632        X2.6  2001-09-24 10:30:00
NA
9           N17E63         8113        X2.6  1997-11-27 13:56:00
98
276         N09W49        10696        X2.5  2004-11-10 02:26:00
NA
123         S23W09         9415        X2.3  2001-04-10 05:30:00
NA
99          N22W07         9236        X2.3  2000-11-24 15:30:00
NA
73          N20E18         9026        X2.3  2000-06-06 15:54:00
NA
345         S20W12        11158        X2.2  2011-02-15 02:24:00
NA
318         S13E47        10808        X2.1  2005-09-10 21:52:00
NA
361         N14W18        11283        X2.1  2011-09-06 23:05:00
NA
420         S06E69        11882        X2.1  2013-10-25 15:12:00
NA
7           S14W33         8100        X2.1  1997-11-04 06:10:00
NA
98          N20W05         9236        X2.0  2000-11-24 05:30:00
NA
125         S19W43         9415        X2.0  2001-04-12 10:31:00
NA
274         N09W17        10696        X2.0  2004-11-07 16:54:00
NA
285         N15W25        10720        X2.0  2005-01-17 09:30:00
NA
102         N20W23         9236        X1.9  2000-11-25 19:31:00
NA

    CME Width CME Speed  Is Halo?  Width Lower Bound?
240       360      2657      True               False
117       244      2505     False               False
233       360      2459      True               False
126       167      1199     False               False
234       360      2029      True               False
8         360      1556      True               False
514       360      1571      True               False
328                 NaN     False               False
237       360      2598      True               False
515       360      3163      True               False
288       360       882      True               False
359       360      1610      True               False
331                 NaN     False               False
```

```
317       360      2257      True               False
82        360      1674      True               False
121       360      1270      True               False
375       360      2684      True               False
135       360      1433      True               False
443       360      2147      True               False
193       360      2285      True               False
104       360       980      True               False
239       103      1420     False               False
286       360      2547      True               False
222       360      1366      True               False
332       360      1774      True               False
160       360      2216      True               False
192       360      1941      True               False
404       360      2625      True               False
201       360      1913      True               False
403       360      1850      True               False
487       360       715      True               False
19        190      1099     False               False
238        65       827     False               False
284       360      2861      True               False
142       360      2402      True               False
9          91       441     False               False
276       360      3387      True               False
123       360      2411      True               False
99        360      1245      True               False
73        360      1119      True               False
345       360       669      True               False
318       360      1893      True               False
361       360       575      True               False
420       360      1081      True               False
7         360       785      True               False
98        360      1289      True               False
125       360      1184      True               False
274       360      1759      True               False
285       360      2094      True               False
102       360       671      True               False
```

```python
compareNASA = top50SF[['Flare Class', 'Start DateTime', 'CME
DateTime', 'End DateTime', 'Flare Region']]
compareNASA = compareNASA.rename(columns={'Flare Class': 'X Class',
'CME DateTime': 'Max DateTime', 'Flare Region': 'Region'})
compareSWL = tidydf[['X Class', 'Start DateTime', 'Max DateTime', 'End
DateTime', 'Region']]
compareSWL.iat[0,0] = 'X28.0'
compareSWL.iat[1,0] = 'X20.0'
compareSWL.iat[2,0] = 'X17.2'
compareSWL.iat[3,0] = 'X17.0'
for row in range(0,50):
    if (compareNASA.iat[row,4] != 'NaN'):
        tempInt = int(compareNASA.iat[row,4])
        if (tempInt >= 10000):
            tempInt -= 10000
            compareNASA.iat[row,4] = tempInt
display(compareSWL)
display(compareNASA)
#Most of the top 50 solar flares on spaceweatherlive.com are the same
ones that I acquired from NASA.
#While some of the information is missing (end datetimes, cme
datetimes), most of the relevant information is there.
#Also, while most of the regions are different, those that have the
same region in the SWL top 50
#also contain the same region in NASA's data.
```

```
   X Class      Start DateTime        Max DateTime        End DateTime
Region
0    X28.0 2003-11-04 19:29:00 2003-11-04 19:53:00 2003-11-04 20:06:00
486
1    X20.0 2001-04-02 21:32:00 2001-04-02 21:51:00 2001-04-02 22:03:00
9393
2    X17.2 2003-10-28 09:51:00 2003-10-28 11:10:00 2003-10-28 11:24:00
486
3    X17.0 2005-09-07 17:17:00 2005-09-07 17:40:00 2005-09-07 18:03:00
808
4    X14.4 2001-04-15 13:19:00 2001-04-15 13:50:00 2001-04-15 13:55:00
9415
5      X10 2003-10-29 20:37:00 2003-10-29 20:49:00 2003-10-29 21:01:00
486
6     X9.4 1997-11-06 11:49:00 1997-11-06 11:55:00 1997-11-06 12:01:00
8100
7     X9.3 2017-09-06 11:53:00 2017-09-06 12:02:00 2017-09-06 12:10:00
2673
8       X9 2006-12-05 10:18:00 2006-12-05 10:35:00 2006-12-05 10:45:00
930
9     X8.3 2003-11-02 17:03:00 2003-11-02 17:25:00 2003-11-02 17:39:00
486
10    X8.2 2017-09-10 15:35:00 2017-09-10 16:06:00 2017-09-10 16:31:00
2673
11    X7.1 2005-01-20 06:36:00 2005-01-20 07:01:00 2005-01-20 07:26:00
720
12    X6.9 2011-08-09 07:48:00 2011-08-09 08:05:00 2011-08-09 08:08:00
1263
13    X6.5 2006-12-06 18:29:00 2006-12-06 18:47:00 2006-12-06 19:00:00
930
14    X6.2 2005-09-09 19:13:00 2005-09-09 20:04:00 2005-09-09 20:36:00
808
15    X6.2 2001-12-13 14:20:00 2001-12-13 14:30:00 2001-12-13 14:35:00
9733
16    X5.7 2000-07-14 10:03:00 2000-07-14 10:24:00 2000-07-14 10:43:00
9077
```

```
17     X5.6 2001-04-06 19:10:00 2001-04-06 19:21:00 2001-04-06 19:31:00
9415
18     X5.4 2012-03-07 00:02:00 2012-03-07 00:24:00 2012-03-07 00:40:00
1429
19     X5.4 2005-09-08 20:52:00 2005-09-08 21:06:00 2005-09-08 21:17:00
808
20     X5.4 2003-10-23 08:19:00 2003-10-23 08:35:00 2003-10-23 08:49:00
486
21     X5.3 2001-08-25 16:23:00 2001-08-25 16:45:00 2001-08-25 17:04:00
9591
22     X4.9 2014-02-25 00:39:00 2014-02-25 00:49:00 2014-02-25 01:03:00
1990
23     X4.9 1998-08-18 22:10:00 1998-08-18 22:19:00 1998-08-18 22:28:00
8307
24     X4.8 2002-07-23 00:18:00 2002-07-23 00:35:00 2002-07-23 00:47:00
39
25       X4 2000-11-26 16:34:00 2000-11-26 16:48:00 2000-11-26 16:56:00
9236
26     X3.9 2003-11-03 09:43:00 2003-11-03 09:55:00 2003-11-03 10:19:00
488
27     X3.9 1998-08-19 21:35:00 1998-08-19 21:45:00 1998-08-19 21:50:00
8307
28     X3.8 2005-01-17 06:59:00 2005-01-17 09:52:00 2005-01-17 10:07:00
720
29     X3.7 1998-11-22 06:30:00 1998-11-22 06:42:00 1998-11-22 06:49:00
8384
30     X3.6 2005-09-09 09:42:00 2005-09-09 09:59:00 2005-09-09 10:08:00
808
31     X3.6 2004-07-16 13:49:00 2004-07-16 13:55:00 2004-07-16 14:01:00
649
32     X3.6 2003-05-28 00:17:00 2003-05-28 00:27:00 2003-05-28 00:39:00
365
33     X3.4 2006-12-13 02:14:00 2006-12-13 02:40:00 2006-12-13 02:57:00
930
34     X3.4 2001-12-28 20:02:00 2001-12-28 20:45:00 2001-12-28 21:32:00
9767
35     X3.3 2013-11-05 22:07:00 2013-11-05 22:12:00 2013-11-05 22:15:00
1890
36     X3.3 2002-07-20 21:04:00 2002-07-20 21:30:00 2002-07-20 21:54:00
39
37     X3.3 1998-11-28 04:54:00 1998-11-28 05:52:00 1998-11-28 06:13:00
8395
38     X3.2 2013-05-14 00:00:00 2013-05-14 01:11:00 2013-05-14 01:20:00
1748
39     X3.1 2014-10-24 21:07:00 2014-10-24 21:41:00 2014-10-24 22:13:00
2192
40     X3.1 2002-08-24 00:49:00 2002-08-24 01:12:00 2002-08-24 01:31:00
69
41       X3 2002-07-15 19:59:00 2002-07-15 20:08:00 2002-07-15 20:14:00
30
```

```
42    X2.8 2013-05-13 15:48:00 2013-05-13 16:05:00 2013-05-13 16:16:00
1748
43    X2.8 2001-12-11 07:58:00 2001-12-11 08:08:00 2001-12-11 08:14:00
9733
44    X2.8 1998-08-18 08:14:00 1998-08-18 08:24:00 1998-08-18 08:32:00
8307
45    X2.7 2015-05-05 22:05:00 2015-05-05 22:11:00 2015-05-05 22:15:00
2339
46    X2.7 2003-11-03 01:09:00 2003-11-03 01:30:00 2003-11-03 01:45:00
488
47    X2.7 1998-05-06 07:58:00 1998-05-06 08:09:00 1998-05-06 08:20:00
8210
48    X2.6 2005-01-15 22:25:00 2005-01-15 23:02:00 2005-01-15 23:31:00
720
49    X2.6 2001-09-24 09:32:00 2001-09-24 10:38:00 2001-09-24 11:09:00
9632

    X Class      Start DateTime         Max DateTime         End
DateTime  \
240   X28.0 2003-11-04 20:00:00  2003-11-04 19:54:00

117   X20.0 2001-04-02 22:05:00  2001-04-02 22:06:00  2001-04-03
02:30:00
233   X17.0 2003-10-28 11:10:00  2003-10-28 11:30:00

126   X14.0 2001-04-15 14:05:00  2001-04-15 14:06:00  2001-04-16
13:00:00
234   X10.0 2003-10-29 20:55:00  2003-10-29 20:54:00

8      X9.4 1997-11-06 12:20:00  1997-11-06 12:10:00  1997-11-07
08:30:00
514    X9.3 2017-09-06 12:05:00  2017-09-06 12:24:00  2017-09-07
08:00:00
328    X9.0 2006-12-05 10:50:00                       2006-12-05
20:00:00
237    X8.3 2003-11-02 17:30:00  2003-11-02 17:30:00  2003-11-03
01:00:00
515    X8.3 2017-09-10 16:02:00  2017-09-10 16:00:00  2017-09-11
06:50:00
288    X7.1 2005-01-20 07:15:00  2005-01-20 06:54:00  2005-01-20
16:30:00
359    X6.9 2011-08-09 08:20:00  2011-08-09 08:12:00  2011-08-09
08:35:00
331    X6.5 2006-12-06 19:00:00

317    X6.2 2005-09-09 19:45:00  2005-09-09 19:48:00  2005-09-09
22:00:00
82     X5.7 2000-07-14 10:30:00  2000-07-14 10:54:00  2000-07-15
14:30:00
121    X5.6 2001-04-06 19:35:00  2001-04-06 19:30:00  2001-04-07
```

```
01:50:00
375    X5.4 2012-03-07 01:00:00  2012-03-07 00:24:00  2012-03-08
19:00:00
135    X5.3 2001-08-25 16:50:00  2001-08-25 16:50:00  2001-08-25
23:00:00
443    X4.9 2014-02-25 00:56:00  2014-02-25 01:25:00  2014-02-25
11:28:00
193    X4.8 2002-07-23 00:50:00  2002-07-23 00:42:00  2002-07-23
04:00:00
104    X4.0 2000-11-26 17:00:00  2000-11-26 17:06:00  2000-11-26
17:15:00
239    X3.9 2003-11-03 10:00:00  2003-11-03 10:06:00  2003-11-03
12:30:00
286    X3.8 2005-01-17 10:00:00  2005-01-17 09:54:00  2005-01-17
10:35:00
222    X3.6 2003-05-28 01:00:00  2003-05-28 00:50:00  2003-05-29
00:30:00
332    X3.4 2006-12-13 02:45:00  2006-12-13 02:54:00  2006-12-13
10:40:00
160    X3.4 2001-12-28 20:35:00  2001-12-28 20:30:00  2001-12-29
03:00:00
192    X3.3 2002-07-20 21:30:00  2002-07-20 22:06:00  2002-07-20
22:20:00
404    X3.2 2013-05-14 01:16:00  2013-05-14 01:25:00  2013-05-14
08:20:00
201    X3.1 2002-08-24 01:45:00  2002-08-24 01:27:00  2002-08-24
03:25:00
403    X2.8 2013-05-13 16:15:00  2013-05-13 16:07:00  2013-05-13
19:10:00
487    X2.7 2015-05-05 22:24:00  2015-05-05 22:24:00  2015-05-05
23:14:00
19     X2.7 1998-05-06 08:25:00  1998-05-06 08:29:00  1998-05-06
08:35:00
238    X2.7 2003-11-03 01:15:00  2003-11-03 01:59:00  2003-11-03
01:25:00
284    X2.6 2005-01-15 23:00:00  2005-01-15 23:06:00  2005-01-17
00:00:00
142    X2.6 2001-09-24 10:45:00  2001-09-24 10:30:00  2001-09-25
20:00:00
9      X2.6 1997-11-27 13:30:00  1997-11-27 13:56:00  1997-11-27
14:00:00
276    X2.5 2004-11-10 02:25:00  2004-11-10 02:26:00  2004-11-10
03:40:00
123    X2.3 2001-04-10 05:24:00  2001-04-10 05:30:00

99     X2.3 2000-11-24 15:25:00  2000-11-24 15:30:00  2000-11-24
22:00:00
73     X2.3 2000-06-06 15:20:00  2000-06-06 15:54:00  2000-06-08
09:00:00
345    X2.2 2011-02-15 02:10:00  2011-02-15 02:24:00  2011-02-15
```

```
07:00:00
318    X2.1 2005-09-10 21:45:00  2005-09-10 21:52:00  2005-09-11
01:00:00
361    X2.1 2011-09-06 22:30:00  2011-09-06 23:05:00  2011-09-07
15:40:00
420    X2.1 2013-10-25 15:08:00  2013-10-25 15:12:00  2013-10-25
22:32:00
7      X2.1 1997-11-04 06:00:00  1997-11-04 06:10:00  1997-11-05
04:30:00
98     X2.0 2000-11-24 05:10:00  2000-11-24 05:30:00  2000-11-24
15:00:00
125    X2.0 2001-04-12 10:20:00  2001-04-12 10:31:00  2001-04-12
10:40:00
274    X2.0 2004-11-07 16:25:00  2004-11-07 16:54:00  2004-11-08
20:00:00
285    X2.0 2005-01-17 09:25:00  2005-01-17 09:30:00  2005-01-17
16:00:00
102    X1.9 2000-11-25 19:00:00  2000-11-25 19:31:00  2000-11-25
19:35:00

    Region
240    486
117   9393
233    486
126   9415
234    486
8     8100
514   2673
328    930
237    486
515    NaN
288    720
359   1263
331    930
317    808
82    9077
121   9415
375   1429
135   9591
443   1990
193     39
104   9236
239    488
286    720
222    365
332    930
160   9756
192     39
404   1748
201     69
```

```
403   1748
487   2339
19    8210
238    488
284    720
142   9632
9     8113
276    696
123   9415
99    9236
73    9026
345   1158
318    808
361   1283
420   1882
7     8100
98    9236
125   9415
274    696
285    720
102   9236
```

```
#Question 2
compareNASA['Best Match?'] = pd.Series('NaN', index=compareNASA.index)
#display(compareNASA)
#Function that finds which flare it best matches with (if any)
def bestMatch(index):
    for row in range(0,50):
        #temp = row[5][1:]
        if (str(compareNASA.iat[index,0]) ==
str(compareSWL.iat[row,0]) and int(compareNASA.iat[index,4]) ==
int(compareSWL.iat[row,4])):
            #return best match rank if found
            return row
    # if no best match is found
    return 'NaN'
for i in range(0,50):
    if (compareNASA.iat[i, 4] != 'NaN'):
        temp = bestMatch(i)
        compareNASA.iat[i, 5] = temp
display(compareNASA)
```

```
    X Class      Start DateTime         Max DateTime         End
DateTime  \
240   X28.0 2003-11-04 20:00:00  2003-11-04 19:54:00

117   X20.0 2001-04-02 22:05:00  2001-04-02 22:06:00  2001-04-03
02:30:00
233   X17.0 2003-10-28 11:10:00  2003-10-28 11:30:00
```

```
126   X14.0 2001-04-15 14:05:00  2001-04-15 14:06:00  2001-04-16
13:00:00
234   X10.0 2003-10-29 20:55:00  2003-10-29 20:54:00

8      X9.4 1997-11-06 12:20:00  1997-11-06 12:10:00  1997-11-07
08:30:00
514    X9.3 2017-09-06 12:05:00  2017-09-06 12:24:00  2017-09-07
08:00:00
328    X9.0 2006-12-05 10:50:00                       2006-12-05
20:00:00
237    X8.3 2003-11-02 17:30:00  2003-11-02 17:30:00  2003-11-03
01:00:00
515    X8.3 2017-09-10 16:02:00  2017-09-10 16:00:00  2017-09-11
06:50:00
288    X7.1 2005-01-20 07:15:00  2005-01-20 06:54:00  2005-01-20
16:30:00
359    X6.9 2011-08-09 08:20:00  2011-08-09 08:12:00  2011-08-09
08:35:00
331    X6.5 2006-12-06 19:00:00

317    X6.2 2005-09-09 19:45:00  2005-09-09 19:48:00  2005-09-09
22:00:00
82     X5.7 2000-07-14 10:30:00  2000-07-14 10:54:00  2000-07-15
14:30:00
121    X5.6 2001-04-06 19:35:00  2001-04-06 19:30:00  2001-04-07
01:50:00
375    X5.4 2012-03-07 01:00:00  2012-03-07 00:24:00  2012-03-08
19:00:00
135    X5.3 2001-08-25 16:50:00  2001-08-25 16:50:00  2001-08-25
23:00:00
443    X4.9 2014-02-25 00:56:00  2014-02-25 01:25:00  2014-02-25
11:28:00
193    X4.8 2002-07-23 00:50:00  2002-07-23 00:42:00  2002-07-23
04:00:00
104    X4.0 2000-11-26 17:00:00  2000-11-26 17:06:00  2000-11-26
17:15:00
239    X3.9 2003-11-03 10:00:00  2003-11-03 10:06:00  2003-11-03
12:30:00
286    X3.8 2005-01-17 10:00:00  2005-01-17 09:54:00  2005-01-17
10:35:00
222    X3.6 2003-05-28 01:00:00  2003-05-28 00:50:00  2003-05-29
00:30:00
332    X3.4 2006-12-13 02:45:00  2006-12-13 02:54:00  2006-12-13
10:40:00
160    X3.4 2001-12-28 20:35:00  2001-12-28 20:30:00  2001-12-29
03:00:00
192    X3.3 2002-07-20 21:30:00  2002-07-20 22:06:00  2002-07-20
22:20:00
404    X3.2 2013-05-14 01:16:00  2013-05-14 01:25:00  2013-05-14
08:20:00
```

```
201    X3.1 2002-08-24 01:45:00  2002-08-24 01:27:00  2002-08-24
03:25:00
403    X2.8 2013-05-13 16:15:00  2013-05-13 16:07:00  2013-05-13
19:10:00
487    X2.7 2015-05-05 22:24:00  2015-05-05 22:24:00  2015-05-05
23:14:00
19     X2.7 1998-05-06 08:25:00  1998-05-06 08:29:00  1998-05-06
08:35:00
238    X2.7 2003-11-03 01:15:00  2003-11-03 01:59:00  2003-11-03
01:25:00
284    X2.6 2005-01-15 23:00:00  2005-01-15 23:06:00  2005-01-17
00:00:00
142    X2.6 2001-09-24 10:45:00  2001-09-24 10:30:00  2001-09-25
20:00:00
9      X2.6 1997-11-27 13:30:00  1997-11-27 13:56:00  1997-11-27
14:00:00
276    X2.5 2004-11-10 02:25:00  2004-11-10 02:26:00  2004-11-10
03:40:00
123    X2.3 2001-04-10 05:24:00  2001-04-10 05:30:00

99     X2.3 2000-11-24 15:25:00  2000-11-24 15:30:00  2000-11-24
22:00:00
73     X2.3 2000-06-06 15:20:00  2000-06-06 15:54:00  2000-06-08
09:00:00
345    X2.2 2011-02-15 02:10:00  2011-02-15 02:24:00  2011-02-15
07:00:00
318    X2.1 2005-09-10 21:45:00  2005-09-10 21:52:00  2005-09-11
01:00:00
361    X2.1 2011-09-06 22:30:00  2011-09-06 23:05:00  2011-09-07
15:40:00
420    X2.1 2013-10-25 15:08:00  2013-10-25 15:12:00  2013-10-25
22:32:00
7      X2.1 1997-11-04 06:00:00  1997-11-04 06:10:00  1997-11-05
04:30:00
98     X2.0 2000-11-24 05:10:00  2000-11-24 05:30:00  2000-11-24
15:00:00
125    X2.0 2001-04-12 10:20:00  2001-04-12 10:31:00  2001-04-12
10:40:00
274    X2.0 2004-11-07 16:25:00  2004-11-07 16:54:00  2004-11-08
20:00:00
285    X2.0 2005-01-17 09:25:00  2005-01-17 09:30:00  2005-01-17
16:00:00
102    X1.9 2000-11-25 19:00:00  2000-11-25 19:31:00  2000-11-25
19:35:00

    Region Best Match?
240    486           0
117   9393           1
233    486         NaN
126   9415         NaN
```

```
234    486          NaN
8     8100            6
514   2673            7
328    930          NaN
237    486            9
515    NaN          NaN
288    720           11
359   1263           12
331    930           13
317    808           14
82    9077           16
121   9415           17
375   1429           18
135   9591           21
443   1990           22
193     39           24
104   9236          NaN
239    488           26
286    720           28
222    365           32
332    930           33
160   9756          NaN
192     39           36
404   1748           38
201     69           40
403   1748           42
487   2339           45
19    8210           47
238    488           46
284    720           48
142   9632           49
9     8113          NaN
276    696          NaN
123   9415          NaN
99    9236          NaN
73    9026          NaN
345   1158          NaN
318    808          NaN
361   1283          NaN
420   1882          NaN
7     8100          NaN
98    9236          NaN
125   9415          NaN
274    696          NaN
285    720          NaN
102   9236          NaN
```

```
#I did a couple things here to determine if the data I got from NASA best matched with
#any ranks in the SWL top 50. To do this, I changed those with classes
```

```
of X__+ to X__.0,
#if they had not already had a decimal, as well as decreasing any
region with a number
#more than 10000 by 10000, since I had realized the numbers were the
same otherwise.
#I assume this was likely due to some number of digits maximum, since
it only applied to
#those with regions that were 5 digits. Since many of the DateTime
values were also either
#missing or close to the actual times, I decided to first only focus
on those which had
#the same X Class and Region values, classifying those as flares with
a "Best Match".

#Basically, I went through every row in my top 50 solar flares I got
from NASA. For each
#row, I iterated through every row in the SWL data. If the X Class and
Region values
#are the same, I would put that ranking under the 'Best Match?' column
in my new dataframe.
#Therefore, there are many "best matches" in my NASA dataframe. Since
there is more than one
#SWL entry that best matches, I then looked towards the start and end
datetimes to determine
#which one is the "best match". I then compared all of the start/end
datetimes of best matches
#to see which one had the least difference. This, using the code
below, was determined to be
#solar flares 23 on the NASA data and 46 on SWL's list.
import datetime
bestMatches = pd.DataFrame(index=range(0,26), columns=['X Class',
'Region', 'Matching Rank', 'StartDT Diff', 'EndDT Diff', 'Total
DateTime Difference'])
bestMatches = bestMatches.astype({'StartDT Diff': str, 'EndDT Diff':
str, 'Total DateTime Difference': float})
i = 0
for row in range(0,50):
    if (compareNASA.iat[row,5] != 'NaN' and compareNASA.iat[row,3] !=
''):
        bestMatches.iat[i,0] = compareNASA.iat[row,0]
        bestMatches.iat[i,1] = compareNASA.iat[row,4]
        bestMatches.iat[i,2] = compareNASA.iat[row,5]
        for rank in range(0,50):
            if (str(compareNASA.iat[row,0]) ==
str(compareSWL.iat[rank,0]) and int(compareNASA.iat[row,4]) ==
int(compareSWL.iat[rank,4])):
                swlStartDT = compareSWL.iat[rank, 1]
                nasaStartDT = compareNASA.iat[row, 1]
                a = datetime.datetime.strptime(str(swlStartDT), "%Y-
%m-%d %H:%M:%S")
```

```
                b = datetime.datetime.strptime(str(nasaStartDT), "%Y-
%m-%d %H:%M:%S")
                c = a-b
                swlEndDT = compareSWL.iat[rank, 3]
                nasaEndDT = compareNASA.iat[row, 3]
                d= datetime.datetime.strptime(str(swlEndDT), "%Y-%m-%d
%H:%M:%S")
                e = datetime.datetime.strptime(str(nasaEndDT), "%Y-%m-
%d %H:%M:%S")
                f = d-e
                bestMatches.iat[i,3] = c.total_seconds()
                bestMatches.iat[i,4] = f.total_seconds()
        i += 1
bestMatches = bestMatches.astype({'Total DateTime Difference': float})
for j in range(0,26):
    bestMatches.iat[j,5] = float(abs(bestMatches.iat[j,3]) +
abs(bestMatches.iat[j,4]))
bestMatches = bestMatches.sort_values('Total DateTime Difference',
ascending = True)
display(bestMatches)
```

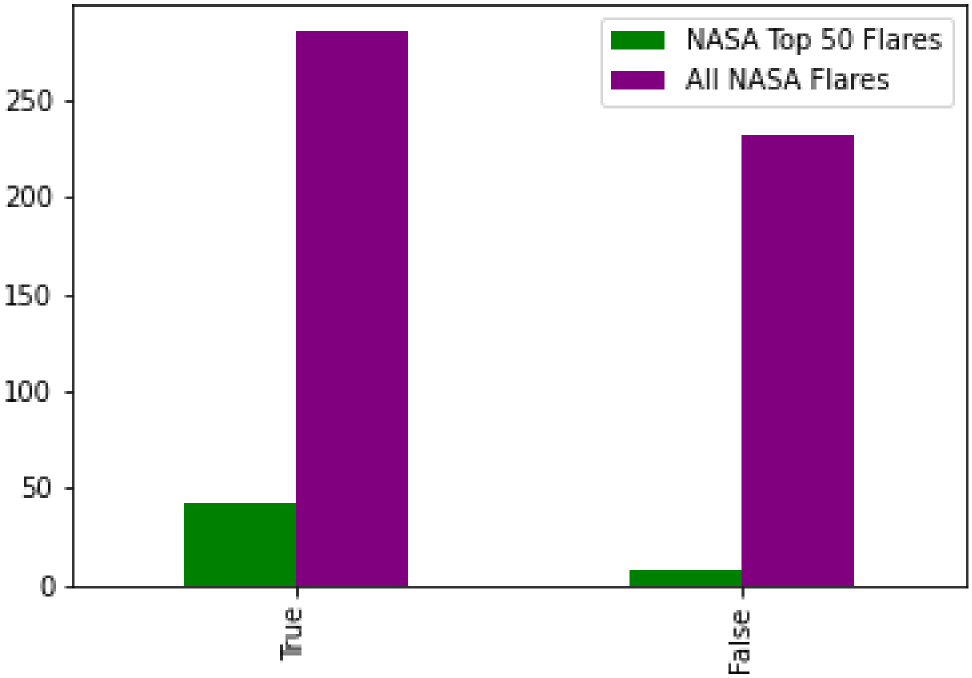
```
   X Class  Region  Matching Rank  StartDT Diff  EndDT Diff  \
23    X2.7     488             46        -360.0      1200.0
22    X2.7    8210             47       -1620.0      -900.0
17    X3.3      39             36       -1560.0     -1560.0
5     X6.9    1263             12       -1920.0     -1620.0
21    X2.7    2339             45       -1140.0     -3540.0
6     X6.2     808             14       -1920.0     -5040.0
13    X3.9     488             26       -1020.0     -7860.0
19    X3.1      69             40       -3360.0     -6840.0
20    X2.8    1748             42       -1620.0    -10440.0
14    X3.8     720             28      -10860.0     -1680.0
12    X4.8      39             24       -1920.0    -11580.0
0    X20.0    9393              1       -1980.0    -16020.0
10    X5.3    9591             21       -1620.0    -21360.0
8     X5.6    9415             17       -1500.0    -22740.0
3     X8.3     486              9       -1620.0    -26460.0
16    X3.4     930             33       -1860.0    -27780.0
18    X3.2    1748             38       -4560.0    -25200.0
4     X7.1     720             11       -2340.0    -32640.0
11    X4.9    1990             22       -1020.0    -37500.0
2     X9.3    2673              7        -720.0    -71400.0
1     X9.4    8100              6       -1860.0    -73740.0
15    X3.6     365             32       -2580.0    -85860.0
24    X2.6     720             48       -2100.0    -88140.0
7     X5.7    9077             16       -1620.0   -100020.0
25    X2.6    9632             49       -4380.0   -118260.0
9     X5.4    1429             18       -3480.0   -152400.0
```

```
    Total DateTime Difference
23                     1560.0
22                     2520.0
17                     3120.0
5                      3540.0
21                     4680.0
6                      6960.0
13                     8880.0
19                    10200.0
20                    12060.0
14                    12540.0
12                    13500.0
0                     18000.0
10                    22980.0
8                     24240.0
3                     28080.0
16                    29640.0
18                    29760.0
4                     34980.0
11                    38520.0
2                     72120.0
1                     75600.0
15                    88440.0
24                    90240.0
7                    101640.0
25                   122640.0
9                    155880.0
```

```
#Question 3
import numpy as np
import matplotlib_inline
import matplotlib
trueForSome = 0
falseforSome = trueForAll = falseForAll = trueForSome
for index, row in top50SF.iterrows():
    if row['Is Halo?']: trueForSome += 1
    else: falseforSome += 1
for index, row in tidydf2.iterrows():
    if row['Is Halo?']: trueForAll += 1
    else: falseForAll += 1
halo_data = np.array([['','NASA Top 50 Flares','All NASA Flares'],
                [True,trueForSome,trueForAll],
                [False,falseforSome,falseForAll]])
halo_df = pd.DataFrame(data=halo_data[1:,1:],
                  index=halo_data[1:,0],
                  columns=halo_data[0,1:])
halo_df['NASA Top 50 Flares'] = halo_df['NASA Top 50
Flares'].astype(str).astype(int)
halo_df['All NASA Flares'] = halo_df['All NASA
Flares'].astype(str).astype(int)
```

```
halo_df.plot.bar(color=['green', 'purple'])
# This graph signifies that there is a correlation between being a top
50 solar flare
# and having a Halo CME.
```

```
<AxesSubplot:>
```


NASA Top 50 Flares
All NASA Flares
True
False