In [1]:
```
############ CMSC320 Project1: Fly Me To The Moon ############
############ Group Students: Ang Zhou; Han Yang Shen; Jixiang Fan
##########

################################################ PART 1 ############
############################
#### step1 ####
# description: Using requests to get the URL; Use BeautifulSoup to
change it to html; Since there are
############ only 1 table tag in this html, we use find("table") t
o get the table HTML we need;
############ Use pandas to make this table html to a dataframe; Re
name the columns and index;
############ Print out the dataframe: df
import requests as rq
from bs4 import BeautifulSoup as bs
import pandas as pd
import numpy as np
import re
# part1 step 1
r = rq.get("https://www.spaceweatherlive.com/en/solar-activity/top-
50-solar-flares")
soup = bs(r.content, "lxml")
df = pd.read_html(soup.find("table").prettify())[0]
df.columns = ["rank","x_class","date","region","start_time","max_ti
me","end_time","movie"]
df.index = range(1, 51)
print(df)
```

```
    rank x_class        date region start_time max_time end_time
\
1      1   X28.0  2003/11/04   0486      19:29    19:53    20:06
2      2     X20  2001/04/02   9393      21:32    21:51    22:03
3      3   X17.2  2003/10/28   0486      09:51    11:10    11:24
4      4   X17.0  2005/09/07   0808      17:17    17:40    18:03
5      5   X14.4  2001/04/15   9415      13:19    13:50    13:55
6      6   X10.0  2003/10/29   0486      20:37    20:49    21:01
7      7    X9.4  1997/11/06      -      11:49    11:55    12:01
8      8    X9.0  2006/12/05   0930      10:18    10:35    10:45
9      9    X8.3  2003/11/02   0486      17:03    17:25    17:39
10    10    X7.1  2005/01/20   0720      06:36    07:01    07:26
11    11    X6.9  2011/08/09   1263      07:48    08:05    08:08
12    12    X6.5  2006/12/06   0930      18:29    18:47    19:00
13    13    X6.2  2005/09/09   0808      19:13    20:04    20:36
14    14    X6.2  2001/12/13   9733      14:20    14:30    14:35
15    15    X5.7  2000/07/14   9077      10:03    10:24    10:43
16    16    X5.6  2001/04/06   9415      19:10    19:21    19:31
17    17    X5.4  2003/10/23   0486      08:19    08:35    08:49
18    18    X5.4  2005/09/08   0808      20:52    21:06    21:17
19    19    X5.4  2012/03/07   1429      00:02    00:24    00:40
20    20    X5.3  2001/08/25   9591      16:23    16:45    17:04
```

21	21	X4.9	1998/08/18	8307	22:10	22:19	22:28
22	22	X4.9	2014/02/25	1990	00:39	00:49	01:03
23	23	X4.8	2002/07/23	0039	00:18	00:35	00:47
24	24	X4.0	2000/11/26	9236	16:34	16:48	16:56
25	25	X3.9	1998/08/19	-	21:35	21:45	21:50
26	26	X3.9	2003/11/03	0488	09:43	09:55	10:19
27	27	X3.8	2005/01/17	0720	06:59	09:52	10:07
28	28	X3.7	1998/11/22	-	06:30	06:42	06:49
29	29	X3.6	2003/05/28	0365	00:17	00:27	00:39
30	30	X3.6	2004/07/16	0649	13:49	13:55	14:01
31	31	X3.6	2005/09/09	0808	09:42	09:59	10:08
32	32	X3.4	2006/12/13	0930	02:14	02:40	02:57
33	33	X3.4	2001/12/28	-	20:02	20:45	21:32
34	34	X3.3	1998/11/28	-	04:54	05:52	06:13
35	35	X3.3	2002/07/20	-	21:04	21:30	21:54
36	36	X3.3	2013/11/05	1890	22:07	22:12	22:15
37	37	X3.2	2013/05/14	1748	00:00	01:11	01:20
38	38	X3.1	2014/10/24	2192	21:07	21:41	22:13
39	39	X3.1	2002/08/24	0069	00:49	01:12	01:31
40	40	X3.0	2002/07/15	0030	19:59	20:08	20:14
41	41	X2.8	1998/08/18	8307	08:14	08:24	08:32
42	42	X2.8	2001/12/11	9733	07:58	08:08	08:14
43	43	X2.8	2013/05/13	1748	15:48	16:05	16:16
44	44	X2.7	2015/05/05	2339	22:05	22:11	22:15
45	45	X2.7	2003/11/03	0488	01:09	01:30	01:45
46	46	X2.7	1998/05/06	-	07:58	08:09	08:20
47	47	X2.6	1997/11/27	8113	12:59	13:17	13:20
48	48	X2.6	2001/09/24	9632	09:32	10:38	11:09
49	49	X2.6	2005/01/15	0720	22:25	23:02	23:31
50	50	X2.5	2004/11/10	0696	01:59	02:13	02:20

	movie
1	Movie View archive
2	Movie View archive
3	Movie View archive
4	Movie View archive
5	Movie View archive
6	Movie View archive
7	Movie View archive
8	Movie View archive
9	Movie View archive
10	Movie View archive
11	Movie View archive
12	Movie View archive
13	Movie View archive
14	Movie View archive
15	Movie View archive
16	Movie View archive
17	Movie View archive
18	Movie View archive
19	Movie View archive
20	Movie View archive
21	View archive
22	Movie View archive

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

In [2]:
```
#### step2 ####
# description: Remove the "movie" column by using drop. Iterate eac
h rows to change the time strings
############## into datatime objects; Replace empty cell by "NaN"; R
eorder the columns.
############## Print out the dataframe: df2
df2=df.drop("movie",1)
for i, row in df2.iterrows():
    new_st=row["date"]+" "+row["start_time"]
    new_mt=row["date"]+" "+row["max_time"]
    new_et=row["date"]+" "+row["end_time"]
    df2.set_value(i,"start_time",pd.to_datetime(new_st))
    df2.set_value(i,"max_time",pd.to_datetime(new_mt))
    df2.set_value(i,"end_time",pd.to_datetime(new_et))

df2=df2.drop("date",1)
df2=df2.replace("-","NaN")

cols = list(df2)
# move the column to head of list using index, pop and insert
cols.insert(5, cols.pop(cols.index('region')))
df2 = df2.ix[:, cols]
print(df2)
```

```
    rank x class          start time            max time
```

```
          _                    _                    _
end_time  \
1       1   X28.0 2003-11-04 19:29:00 2003-11-04 19:53:00 2003-11-0
4 20:06:00
2       2     X20 2001-04-02 21:32:00 2001-04-02 21:51:00 2001-04-0
2 22:03:00
3       3   X17.2 2003-10-28 09:51:00 2003-10-28 11:10:00 2003-10-2
8 11:24:00
4       4   X17.0 2005-09-07 17:17:00 2005-09-07 17:40:00 2005-09-0
7 18:03:00
5       5   X14.4 2001-04-15 13:19:00 2001-04-15 13:50:00 2001-04-1
5 13:55:00
6       6   X10.0 2003-10-29 20:37:00 2003-10-29 20:49:00 2003-10-2
9 21:01:00
7       7    X9.4 1997-11-06 11:49:00 1997-11-06 11:55:00 1997-11-0
6 12:01:00
8       8    X9.0 2006-12-05 10:18:00 2006-12-05 10:35:00 2006-12-0
5 10:45:00
9       9    X8.3 2003-11-02 17:03:00 2003-11-02 17:25:00 2003-11-0
2 17:39:00
10     10    X7.1 2005-01-20 06:36:00 2005-01-20 07:01:00 2005-01-2
0 07:26:00
11     11    X6.9 2011-08-09 07:48:00 2011-08-09 08:05:00 2011-08-0
9 08:08:00
12     12    X6.5 2006-12-06 18:29:00 2006-12-06 18:47:00 2006-12-0
6 19:00:00
13     13    X6.2 2005-09-09 19:13:00 2005-09-09 20:04:00 2005-09-0
9 20:36:00
14     14    X6.2 2001-12-13 14:20:00 2001-12-13 14:30:00 2001-12-1
3 14:35:00
15     15    X5.7 2000-07-14 10:03:00 2000-07-14 10:24:00 2000-07-1
4 10:43:00
16     16    X5.6 2001-04-06 19:10:00 2001-04-06 19:21:00 2001-04-0
6 19:31:00
17     17    X5.4 2003-10-23 08:19:00 2003-10-23 08:35:00 2003-10-2
3 08:49:00
18     18    X5.4 2005-09-08 20:52:00 2005-09-08 21:06:00 2005-09-0
8 21:17:00
19     19    X5.4 2012-03-07 00:02:00 2012-03-07 00:24:00 2012-03-0
7 00:40:00
20     20    X5.3 2001-08-25 16:23:00 2001-08-25 16:45:00 2001-08-2
5 17:04:00
21     21    X4.9 1998-08-18 22:10:00 1998-08-18 22:19:00 1998-08-1
8 22:28:00
22     22    X4.9 2014-02-25 00:39:00 2014-02-25 00:49:00 2014-02-2
5 01:03:00
23     23    X4.8 2002-07-23 00:18:00 2002-07-23 00:35:00 2002-07-2
3 00:47:00
24     24    X4.0 2000-11-26 16:34:00 2000-11-26 16:48:00 2000-11-2
6 16:56:00
25     25    X3.9 1998-08-19 21:35:00 1998-08-19 21:45:00 1998-08-1
9 21:50:00
26     26    X3.9 2003-11-03 09:43:00 2003-11-03 09:55:00 2003-11-0
3 10:19:00
```

```
27    27    X3.8 2005-01-17 06:59:00 2005-01-17 09:52:00 2005-01-1
7 10:07:00
28    28    X3.7 1998-11-22 06:30:00 1998-11-22 06:42:00 1998-11-2
2 06:49:00
29    29    X3.6 2003-05-28 00:17:00 2003-05-28 00:27:00 2003-05-2
8 00:39:00
30    30    X3.6 2004-07-16 13:49:00 2004-07-16 13:55:00 2004-07-1
6 14:01:00
31    31    X3.6 2005-09-09 09:42:00 2005-09-09 09:59:00 2005-09-0
9 10:08:00
32    32    X3.4 2006-12-13 02:14:00 2006-12-13 02:40:00 2006-12-1
3 02:57:00
33    33    X3.4 2001-12-28 20:02:00 2001-12-28 20:45:00 2001-12-2
8 21:32:00
34    34    X3.3 1998-11-28 04:54:00 1998-11-28 05:52:00 1998-11-2
8 06:13:00
35    35    X3.3 2002-07-20 21:04:00 2002-07-20 21:30:00 2002-07-2
0 21:54:00
36    36    X3.3 2013-11-05 22:07:00 2013-11-05 22:12:00 2013-11-0
5 22:15:00
37    37    X3.2 2013-05-14 00:00:00 2013-05-14 01:11:00 2013-05-1
4 01:20:00
38    38    X3.1 2014-10-24 21:07:00 2014-10-24 21:41:00 2014-10-2
4 22:13:00
39    39    X3.1 2002-08-24 00:49:00 2002-08-24 01:12:00 2002-08-2
4 01:31:00
40    40    X3.0 2002-07-15 19:59:00 2002-07-15 20:08:00 2002-07-1
5 20:14:00
41    41    X2.8 1998-08-18 08:14:00 1998-08-18 08:24:00 1998-08-1
8 08:32:00
42    42    X2.8 2001-12-11 07:58:00 2001-12-11 08:08:00 2001-12-1
1 08:14:00
43    43    X2.8 2013-05-13 15:48:00 2013-05-13 16:05:00 2013-05-1
3 16:16:00
44    44    X2.7 2015-05-05 22:05:00 2015-05-05 22:11:00 2015-05-0
5 22:15:00
45    45    X2.7 2003-11-03 01:09:00 2003-11-03 01:30:00 2003-11-0
3 01:45:00
46    46    X2.7 1998-05-06 07:58:00 1998-05-06 08:09:00 1998-05-0
6 08:20:00
47    47    X2.6 1997-11-27 12:59:00 1997-11-27 13:17:00 1997-11-2
7 13:20:00
48    48    X2.6 2001-09-24 09:32:00 2001-09-24 10:38:00 2001-09-2
4 11:09:00
49    49    X2.6 2005-01-15 22:25:00 2005-01-15 23:02:00 2005-01-1
5 23:31:00
50    50    X2.5 2004-11-10 01:59:00 2004-11-10 02:13:00 2004-11-1
0 02:20:00

   region
1    0486
2    9393
3    0486
4    0808
```

```
5     9415
6     0486
7      NaN
8     0930
9     0486
10    0720
11    1263
12    0930
13    0808
14    9733
15    9077
16    9415
17    0486
18    0808
19    1429
20    9591
21    8307
22    1990
23    0039
24    9236
25     NaN
26    0488
27    0720
28     NaN
29    0365
30    0649
31    0808
32    0930
33     NaN
34     NaN
35     NaN
36    1890
37    1748
38    2192
39    0069
40    0030
41    8307
42    9733
43    1748
44    2339
45    0488
46     NaN
47    8113
48    9632
49    0720
50    0696
```

In [3]:
```
#### step3 ####
# description: Using requests to get the URL; Use BeautifulSoup to
change it to html;
############## Since there are only 1 pre tag in this html, we use f
ind('pre') to get the HTML we need;
############## Use get_text() to extract the text; Split the line an
d split by "PHTX" to get the text
############## before "PHTX"; then split by whitespace and form the
dataframe: df3
r3=rq.get("http://www.hcbravo.org/IntroDataSci/misc/waves_type2.htm
l")
soup3 = bs(r3.content,"lxml")
text3=soup3.find('pre').get_text()
text3=text3.split("\n")

df3 = pd.DataFrame(columns=\
                   ['start_date','start_time','end_date','end_time'
,'start_frequency',\
                    'end_frequency','flare_location','flare_region'
,'flare_classification',\
                    'cme_date','cme_time','cme_angle','cme_width','
cme_speed'], index=range(1,483))
Lines=[]
count=12
while (count<=493):
    str_list=(text3[count].split('PHTX'))[0]
    str_list=str_list.split()
    Lines.append(str_list)
    df3.loc[count-11]=Lines[count-12]
    count = count+1

print(df3)
```

```
     start_date start_time end_date end_time start_frequency end_f
requency  \
1    1997/04/01      14:00    04/01    14:15            8000
4000
2    1997/04/07      14:30    04/07    17:30           11000
1000
3    1997/05/12      05:15    05/14    16:00           12000
80
4    1997/05/21      20:20    05/21    22:00            5000
500
5    1997/09/23      21:53    09/23    22:16            6000
2000
6    1997/11/03      05:15    11/03    12:00           14000
250
7    1997/11/03      10:30    11/03    11:30           14000
5000
8    1997/11/04      06:00    11/05    04:30           14000
100
9    1997/11/06      12:20    11/07    08:30           14000
100
```

```
10   1997/11/27      13:30    11/27    14:00           14000
7000
11   1997/12/12      22:45    12/12    23:20           14000
8000
12   1998/01/25      15:03    01/25    15:18           14000
10000
13   1998/03/29      03:40    03/29    03:52           14000
7000
14   1998/04/20      10:25    04/22    06:00           10000
35
15   1998/04/23      06:00    04/23    15:30           14000
200
16   1998/04/24      09:17    04/24    09:25            4700
2600
17   1998/04/27      09:20    04/27    10:00           10000
1000
18   1998/04/29      16:30    04/29    17:00           10000
2000
19   1998/05/02      14:25    05/02    14:50            5000
3000
20   1998/05/06      08:25    05/06    08:35           14000
5000
21   1998/05/09      03:35    05/09    10:00            9000
400
22   1998/05/11      21:40    05/11    22:00           10000
1000
23   1998/05/19      10:00    05/19    11:30           14000
3000
24   1998/05/27      13:30    05/27    14:20            4000
1000
25   1998/06/11      10:15    06/11    10:20            8000
4000
26   1998/06/16      18:20    06/17    21:00           12000
50
27   1998/06/20      19:39    06/20    20:00            2600
1800
28   1998/06/22      07:15    06/22    09:20            6000
2000
29   1998/11/02      14:00    11/02    14:40           14000
4000
30   1998/11/05      22:00    11/07    08:00            5000
50
..          ...       ...      ...      ...             ...
...
453  2014/04/04      14:02    04/04    14:07           14000
11000
454  2014/04/18      13:05    04/18    22:50           14000
150
455  2014/05/07      16:24    05/07    23:18           16000
200
456  2014/05/08      03:21    05/08    05:26           16000
1100
457  2014/05/09      02:40    05/09    04:30           14000
500
```

```
458  2014/05/10      04:32    05/10    08:37           16000
400
459  2014/06/10      12:58    06/10    15:00           16000
1000
460  2014/06/12      22:14    06/12    22:35           14000
6000
461  2014/07/30      07:44    07/30    08:00            6300
4500
462  2014/08/01      18:58    08/02    05:00            1000
150
463  2014/08/22      10:37    08/22    11:18           14000
6000
464  2014/08/25      15:20    08/25    16:02           14000
4000
465  2014/08/25      20:43    08/25    21:00           14000
7200
466  2014/08/28      17:05    08/28    22:08           16000
600
467  2014/09/01      11:12    09/01    22:35           16000
600
468  2014/09/09      00:05    09/09    13:00           11000
100
469  2014/09/10      17:45    09/11    12:00           14000
100
470  2014/09/20      05:10    09/20    05:30           14000
9700
471  2014/09/22      06:13    09/22    06:50           16000
4900
472  2014/09/23      23:41    09/23    23:47           14000
12000
473  2014/09/24      20:54    09/24    23:48           16000
500
474  2014/10/02      21:34    10/02    21:56            3300
1900
475  2014/10/10      18:11    10/10    18:33            2500
1500
476  2014/10/21      12:33    10/21    13:01           14000
8900
477  2014/11/08      16:57    11/08    17:18           14000
7800
478  2014/12/13      14:27    12/13    14:51           14000
3900
479  2014/12/17      04:09    12/17    04:19            2900
2100
480  2014/12/17      05:00    12/17    05:09           14000
11500
481  2014/12/18      22:31    12/18    22:54            5100
1300
482  2014/12/21      12:05    12/21    12:28           14000
7400

    flare_location flare_region flare_classification cme_date cme_
time  \
1           S25E16         8026                 M1.3    04/01    1
```

5:18
2 S28E19 8027 C6.8 04/07 1
4:27
3 N21W08 8038 C1.3 05/12 0
5:30
4 N05W12 8040 M1.3 05/21 2
1:00
5 S29E25 8088 C1.4 09/23 2
2:02
6 S20W13 8100 C8.6 11/03 0
5:28
7 S16W21 8100 M4.2 11/03 1
1:11
8 S14W33 8100 X2.1 11/04 0
6:10
9 S18W63 8100 X9.4 11/06 1
2:10
10 N17E63 8113 X2.6 11/27 1
3:56
11 N25W52 8116 B9.4 12/13 0
0:26
12 N21E25 8141 C1.1 01/25 1
5:26
13 SW90 ----- ---- 03/29 0
3:48
14 S22W90 8194 M1.4 04/20 1
0:07
15 S17E90 8210 X1.2 04/23 0
5:55
16 S10E90 8210 C8.9 04/24 0
8:55
17 S16E50 8210 X1.0 04/27 0
8:56
18 S18E20 8210 M6.8 04/29 1
6:58
19 S15W15 8210 X1.1 05/02 1
4:06
20 S11W65 8210 X2.7 05/06 0
8:29
21 S14W89 8210 M7.7 05/09 0
3:35
22 N32W90 8214 B6.6 05/11 2
1:55
23 N29W46 8222 B7.9 05/19 1
0:27
24 N19W62 8226 C7.5 05/27 1
3:45
25 N16E86 8243 M1.4 06/11 1
0:28
26 S22W90 8232 M1.0 06/16 1
8:27
27 Back ----- ---- 06/20 1
8:20
28 N16W46 8243 C2.9 06/22 0

```
5:01
29          S25E47          8373                  C4.4    11/02    1
4:18
30          N22W18          8375                  M8.4    11/05    2
0:44
..             ...           ...                   ...      ...
...
453         N13E26         12027                  C8.3    04/04    1
4:12
454         S20W34         12036                  M7.3    04/18    1
3:25
455         N11E53         12027                  ----    05/07    1
6:24
456        S09W108         12051                  ----    05/08    0
3:24
457        S11W122         12051                  ----    05/09    0
2:48
458        S11W136         12051                  ----    05/10    0
4:36
459         S17E82         12087                  X1.5    06/10    1
3:30
460         S20W55         12085                  M3.1    06/12    2
2:12
461         N10E30           EP?                  C1.5    07/30    0
7:00
462         S10E11         12127                  M1.5    08/01    1
8:36
463         N12E01         12146                  C2.2    08/22    1
1:12
464         N05W36         12146                  M2.0    08/25    1
5:36
465         N07W43         12146                  M3.9    08/25    2
0:48
466        S19E162         12157                  ----    08/28    1
7:24
467        N14E127         12158                  ----    09/01    1
1:12
468         N12E29         12158                  M4.5    09/09    0
0:06
469         N14E02         12158                  X1.6    09/10    1
8:00
470         S11W89         12164                  ----    09/20    0
5:24
471        N14W142         12158                  ----    09/22    0
6:12
472         S13E33         12172                  M2.3    09/23    2
3:36
473        N13E179           EP                  ----    09/24    2
1:30
474         S17W82         12173                  M7.3    10/02    1
9:12
475         S20W51           EP                   C3.0    10/10    1
6:12
476         S18E36         12192                  C4.4    10/21    1
```

```
2:48
477           W90b        12203                 ----    11/08    1
6:36
478           W90b        -----                 ----    12/13    1
4:24
479         S11E33        12241                 M1.1    12/17    0
2:00
480         S20E09        12242                 M8.7    12/17    0
5:00
481         S11E15        12241                 M6.9    12/19    0
1:04
482         S14W25        12241                 M1.0    12/21    1
2:12
```

	cme_angle	cme_width	cme_speed
1	74	79	312
2	Halo	360	878
3	Halo	360	464
4	263	165	296
5	133	155	712
6	240	109	227
7	233	122	352
8	Halo	360	785
9	Halo	360	1556
10	98	91	441
11	278	73	191
12	Halo	360	693
13	Halo	360	1397
14	284	165	1863
15	Halo	360	1691
16	100	84	1184
17	Halo	360	1385
18	Halo	360	1374
19	Halo	360	938
20	309	190	1099
21	262	178	2331
22	208	>301	830
23	268	139	801
24	175	268	878
25	123	177	1223
26	341	281	1484
27	Halo	360	964
28	265	59	206
29	116	169	661
30	Halo	360	1118
..	...	...	...
453	54	96	467
454	Halo	360	1203
455	Halo	360	923
456	Halo	360	847
457	Halo	360	1099
458	Halo	360	1086
459	Halo	360	1469
460	233	186	684

```
461        13       254       700
462      Halo      360h       789
463      Halo       360       600
464      Halo       360       555
465       273       177       711
466      Halo       360       766
467      Halo       360      1901
468      Halo       360       920
469      Halo       360      1267
470       292        87       426
471       342       252       618
472       109       134       331
473      Halo       360      1350
474       264       159       513
475       309      >210       782
476       152       142       260
477       305       141       426
478      Halo       360      2222
479       107       108       869
480      Halo       360       587
481      Halo       360      1195
482      Halo       360       669

[482 rows x 14 columns]
```

In [4]:

```
#### step4 ####
# description: Replace all the missing information by "NaN"; Make i
s_halo and width_lower_bound columns;
############## change those columns with time string to datetime obj
ects; remove un-need columns and
############## re-order the columns of df4
from datetime import datetime, timedelta
df4=df3.copy()
df4=df4.replace('----',"NaN")
df4=df4.replace("-----","NaN")
df4=df4.replace("------","NaN")
df4=df4.replace("--/--","NaN")
df4=df4.replace("--:--","NaN")
df4=df4.replace("Back","NaN")
df4=df4.replace("BACK","NaN")
df4=df4.replace("????","NaN")

is_halo=[]
for i in range(1,483):
    if df4["cme_angle"][i]=="Halo":
        is_halo.append("True")
    else:
        is_halo.append("False")
df4["is_halo"]=is_halo
df4=df4.replace("Halo","NaN")

width_lower_bound=[]
for i in range(1,483):
```

```
    if re.search(r">",df4["cme_width"][i])==None:
        width_lower_bound.append("False")
    else:
        width_lower_bound.append("True")
df4["width_lower_bound"]=width_lower_bound

df4["start_datetime"]=df4["start_date"].map(str)+" "+df4["start_tim
e"].map(str)

ed4 = list(map(lambda x,y,z: x.split("/")[0]+"/"+y+" "+z,\
               df4["start_datetime"],df4["end_date"],df4["end_time"
]))

for i in range (0,482):
    if(ed4[i].find('--')!= -1):
        ed4[i] = ('NaN')
    else:
        if (ed4[i].find('24:')!= -1):
            ed4[i] = ed4[i].replace('24:','00:')
            date = datetime.strptime(ed4[i],"%Y/%m/%d %H:%M")
            modified_date = date + timedelta(days = 1)
            ed4[i] = datetime.strftime(modified_date,"%Y/%m/%d
%H:%M:%S")
            ed4[i] = pd.to_datetime(ed4[i]);
        else:
            ed4[i] = pd.to_datetime(ed4[i]);

cme_datetime=list(map(lambda x,y,z: "NaN" if x=="NaN" or y=="NaN" o
r z=="NaN" else \
                      pd.to_datetime(x.split("/")[0]+"/"+y+" "+z),d
f4["start_datetime"],\
                      df4["cme_date"],df4["cme_time"]))
df4["cme_datetime"]=cme_datetime

df4["start_datetime"] = pd.to_datetime(df4["start_datetime"])
df4["end_datetime"] = ed4
df4["cme_datetime"] = pd.to_datetime(df4["cme_datetime"])

cols4 = list(df4)
cols4.insert(1, cols4.pop(cols4.index('end_datetime')))
cols4.insert(1, cols4.pop(cols4.index('start_datetime')))
cols4.pop(cols4.index('start_date'))
cols4.pop(cols4.index('start_time'))
cols4.pop(cols4.index('end_date'))
cols4.pop(cols4.index('end_time'))
cols4.insert(8, cols4.pop(cols4.index('cme_datetime')))
cols4.pop(cols4.index('cme_date'))
cols4.pop(cols4.index('cme_time'))
df4 = df4.ix[:, cols4]
print(df4)
################################################ PART 1 (End) ######
#################################
```

```
         start_datetime        end_datetime start_frequency end_fr
```

```
equency  \
1   1997-04-01 14:00:00 1997-04-01 14:15:00            8000
4000
2   1997-04-07 14:30:00 1997-04-07 17:30:00           11000
1000
3   1997-05-12 05:15:00 1997-05-14 16:00:00           12000
80
4   1997-05-21 20:20:00 1997-05-21 22:00:00            5000
500
5   1997-09-23 21:53:00 1997-09-23 22:16:00            6000
2000
6   1997-11-03 05:15:00 1997-11-03 12:00:00           14000
250
7   1997-11-03 10:30:00 1997-11-03 11:30:00           14000
5000
8   1997-11-04 06:00:00 1997-11-05 04:30:00           14000
100
9   1997-11-06 12:20:00 1997-11-07 08:30:00           14000
100
10  1997-11-27 13:30:00 1997-11-27 14:00:00           14000
7000
11  1997-12-12 22:45:00 1997-12-12 23:20:00           14000
8000
12  1998-01-25 15:03:00 1998-01-25 15:18:00           14000
10000
13  1998-03-29 03:40:00 1998-03-29 03:52:00           14000
7000
14  1998-04-20 10:25:00 1998-04-22 06:00:00           10000
35
15  1998-04-23 06:00:00 1998-04-23 15:30:00           14000
200
16  1998-04-24 09:17:00 1998-04-24 09:25:00            4700
2600
17  1998-04-27 09:20:00 1998-04-27 10:00:00           10000
1000
18  1998-04-29 16:30:00 1998-04-29 17:00:00           10000
2000
19  1998-05-02 14:25:00 1998-05-02 14:50:00            5000
3000
20  1998-05-06 08:25:00 1998-05-06 08:35:00           14000
5000
21  1998-05-09 03:35:00 1998-05-09 10:00:00            9000
400
22  1998-05-11 21:40:00 1998-05-11 22:00:00           10000
1000
23  1998-05-19 10:00:00 1998-05-19 11:30:00           14000
3000
24  1998-05-27 13:30:00 1998-05-27 14:20:00            4000
1000
25  1998-06-11 10:15:00 1998-06-11 10:20:00            8000
4000
26  1998-06-16 18:20:00 1998-06-17 21:00:00           12000
50
27  1998-06-20 19:39:00 1998-06-20 20:00:00            2600
```

```
1800
28  1998-06-22 07:15:00 1998-06-22 09:20:00            6000
2000
29  1998-11-02 14:00:00 1998-11-02 14:40:00           14000
4000
30  1998-11-05 22:00:00 1998-11-07 08:00:00            5000
50
..                  ...                 ...             ...
...
453 2014-04-04 14:02:00 2014-04-04 14:07:00           14000
11000
454 2014-04-18 13:05:00 2014-04-18 22:50:00           14000
150
455 2014-05-07 16:24:00 2014-05-07 23:18:00           16000
200
456 2014-05-08 03:21:00 2014-05-08 05:26:00           16000
1100
457 2014-05-09 02:40:00 2014-05-09 04:30:00           14000
500
458 2014-05-10 04:32:00 2014-05-10 08:37:00           16000
400
459 2014-06-10 12:58:00 2014-06-10 15:00:00           16000
1000
460 2014-06-12 22:14:00 2014-06-12 22:35:00           14000
6000
461 2014-07-30 07:44:00 2014-07-30 08:00:00            6300
4500
462 2014-08-01 18:58:00 2014-08-02 05:00:00            1000
150
463 2014-08-22 10:37:00 2014-08-22 11:18:00           14000
6000
464 2014-08-25 15:20:00 2014-08-25 16:02:00           14000
4000
465 2014-08-25 20:43:00 2014-08-25 21:00:00           14000
7200
466 2014-08-28 17:05:00 2014-08-28 22:08:00           16000
600
467 2014-09-01 11:12:00 2014-09-01 22:35:00           16000
600
468 2014-09-09 00:05:00 2014-09-09 13:00:00           11000
100
469 2014-09-10 17:45:00 2014-09-11 12:00:00           14000
100
470 2014-09-20 05:10:00 2014-09-20 05:30:00           14000
9700
471 2014-09-22 06:13:00 2014-09-22 06:50:00           16000
4900
472 2014-09-23 23:41:00 2014-09-23 23:47:00           14000
12000
473 2014-09-24 20:54:00 2014-09-24 23:48:00           16000
500
474 2014-10-02 21:34:00 2014-10-02 21:56:00            3300
1900
475 2014-10-10 18:11:00 2014-10-10 18:33:00            2500
```

```
1500
476 2014-10-21 12:33:00 2014-10-21 13:01:00           14000
8900
477 2014-11-08 16:57:00 2014-11-08 17:18:00           14000
7800
478 2014-12-13 14:27:00 2014-12-13 14:51:00           14000
3900
479 2014-12-17 04:09:00 2014-12-17 04:19:00            2900
2100
480 2014-12-17 05:00:00 2014-12-17 05:09:00           14000
11500
481 2014-12-18 22:31:00 2014-12-18 22:54:00            5100
1300
482 2014-12-21 12:05:00 2014-12-21 12:28:00           14000
7400

    flare_location flare_region flare_classification        cme_da
tetime  \
1           S25E16         8026                  M1.3 1997-04-01 15
:18:00
2           S28E19         8027                  C6.8 1997-04-07 14
:27:00
3           N21W08         8038                  C1.3 1997-05-12 05
:30:00
4           N05W12         8040                  M1.3 1997-05-21 21
:00:00
5           S29E25         8088                  C1.4 1997-09-23 22
:02:00
6           S20W13         8100                  C8.6 1997-11-03 05
:28:00
7           S16W21         8100                  M4.2 1997-11-03 11
:11:00
8           S14W33         8100                  X2.1 1997-11-04 06
:10:00
9           S18W63         8100                  X9.4 1997-11-06 12
:10:00
10          N17E63         8113                  X2.6 1997-11-27 13
:56:00
11          N25W52         8116                  B9.4 1997-12-13 00
:26:00
12          N21E25         8141                  C1.1 1998-01-25 15
:26:00
13            SW90          NaN                   NaN 1998-03-29 03
:48:00
14          S22W90         8194                  M1.4 1998-04-20 10
:07:00
15          S17E90         8210                  X1.2 1998-04-23 05
:55:00
16          S10E90         8210                  C8.9 1998-04-24 08
:55:00
17          S16E50         8210                  X1.0 1998-04-27 08
:56:00
18          S18E20         8210                  M6.8 1998-04-29 16
:58:00
```

```
19          S15W15         8210                X1.1 1998-05-02 14
:06:00
20          S11W65         8210                X2.7 1998-05-06 08
:29:00
21          S14W89         8210                M7.7 1998-05-09 03
:35:00
22          N32W90         8214                B6.6 1998-05-11 21
:55:00
23          N29W46         8222                B7.9 1998-05-19 10
:27:00
24          N19W62         8226                C7.5 1998-05-27 13
:45:00
25          N16E86         8243                M1.4 1998-06-11 10
:28:00
26          S22W90         8232                M1.0 1998-06-16 18
:27:00
27             NaN          NaN                 NaN 1998-06-20 18
:20:00
28          N16W46         8243                C2.9 1998-06-22 05
:01:00
29          S25E47         8373                C4.4 1998-11-02 14
:18:00
30          N22W18         8375                M8.4 1998-11-05 20
:44:00
..             ...          ...                 ...
...
453         N13E26        12027                C8.3 2014-04-04 14
:12:00
454         S20W34        12036                M7.3 2014-04-18 13
:25:00
455         N11E53        12027                 NaN 2014-05-07 16
:24:00
456        S09W108        12051                 NaN 2014-05-08 03
:24:00
457        S11W122        12051                 NaN 2014-05-09 02
:48:00
458        S11W136        12051                 NaN 2014-05-10 04
:36:00
459         S17E82        12087                X1.5 2014-06-10 13
:30:00
460         S20W55        12085                M3.1 2014-06-12 22
:12:00
461         N10E30          EP?                C1.5 2014-07-30 07
:00:00
462         S10E11        12127                M1.5 2014-08-01 18
:36:00
463         N12E01        12146                C2.2 2014-08-22 11
:12:00
464         N05W36        12146                M2.0 2014-08-25 15
:36:00
465         N07W43        12146                M3.9 2014-08-25 20
:48:00
466        S19E162        12157                 NaN 2014-08-28 17
:24:00
```

```
467        N14E127        12158                  NaN 2014-09-01 11
:12:00
468         N12E29        12158                M4.5 2014-09-09 00
:06:00
469         N14E02        12158                X1.6 2014-09-10 18
:00:00
470         S11W89        12164                  NaN 2014-09-20 05
:24:00
471        N14W142        12158                  NaN 2014-09-22 06
:12:00
472         S13E33        12172                M2.3 2014-09-23 23
:36:00
473        N13E179           EP                  NaN 2014-09-24 21
:30:00
474         S17W82        12173                M7.3 2014-10-02 19
:12:00
475         S20W51           EP                C3.0 2014-10-10 16
:12:00
476         S18E36        12192                C4.4 2014-10-21 12
:48:00
477           W90b        12203                  NaN 2014-11-08 16
:36:00
478           W90b          NaN                  NaN 2014-12-13 14
:24:00
479         S11E33        12241                M1.1 2014-12-17 02
:00:00
480         S20E09        12242                M8.7 2014-12-17 05
:00:00
481         S11E15        12241                M6.9 2014-12-19 01
:04:00
482         S14W25        12241                M1.0 2014-12-21 12
:12:00
```

	cme_angle	cme_width	cme_speed	is_halo	width_lower_bound
1	74	79	312	False	False
2	NaN	360	878	True	False
3	NaN	360	464	True	False
4	263	165	296	False	False
5	133	155	712	False	False
6	240	109	227	False	False
7	233	122	352	False	False
8	NaN	360	785	True	False
9	NaN	360	1556	True	False
10	98	91	441	False	False
11	278	73	191	False	False
12	NaN	360	693	True	False
13	NaN	360	1397	True	False
14	284	165	1863	False	False
15	NaN	360	1691	True	False
16	100	84	1184	False	False
17	NaN	360	1385	True	False
18	NaN	360	1374	True	False
19	NaN	360	938	True	False
20	309	190	1099	False	False

```
21        262       178      2331   False             False
22        208      >301       830   False              True
23        268       139       801   False             False
24        175       268       878   False             False
25        123       177      1223   False             False
26        341       281      1484   False             False
27        NaN       360       964    True             False
28        265        59       206   False             False
29        116       169       661   False             False
30        NaN       360      1118    True             False
..        ...       ...       ...     ...               ...
453        54        96       467   False             False
454       NaN       360      1203    True             False
455       NaN       360       923    True             False
456       NaN       360       847    True             False
457       NaN       360      1099    True             False
458       NaN       360      1086    True             False
459       NaN       360      1469    True             False
460       233       186       684   False             False
461        13       254       700   False             False
462       NaN      360h       789    True             False
463       NaN       360       600    True             False
464       NaN       360       555    True             False
465       273       177       711   False             False
466       NaN       360       766    True             False
467       NaN       360      1901    True             False
468       NaN       360       920    True             False
469       NaN       360      1267    True             False
470       292        87       426   False             False
471       342       252       618   False             False
472       109       134       331   False             False
473       NaN       360      1350    True             False
474       264       159       513   False             False
475       309      >210       782   False              True
476       152       142       260   False             False
477       305       141       426   False             False
478       NaN       360      2222    True             False
479       107       108       869   False             False
480       NaN       360       587    True             False
481       NaN       360      1195    True             False
482       NaN       360       669    True             False

[482 rows x 13 columns]
```

In [5]:
```
################################################ PART 2 ############
#####################################
############################## Question 1 ##############################
# description: change data like X28. to X28.0; construct a nature s
orting function to sort the
############## classification column and update the dataframe df5; r
eorder the columns and remove
############## the un-need columns; present the last 4 digit of regi
ons
```

```
df5=df4.copy()
df5['flare_classification']=list(map(lambda x: x+"0" if x[-1]=="."
else x, df5['flare_classification']))

def sort_df(df, column_idx, key):
    '''Takes dataframe, column index and custom function for sortin
g,
    returns dataframe sorted by this column using this function'''

    col = df.ix[:,column_idx]
    temp = pd.DataFrame([])
    temp[0] = col
    temp[1] = df.index
    temp = temp.values.tolist()
    list=sorted(temp, key=key)
    list.reverse()
    df = df.ix[[i[1] for i in list]]

    return df

def atoi(text):
    return float(text) if text.isdigit() else text

def natural_keys(text):
    return [ atoi(c) for c in re.split('(\d+)', text[0]) ]

df5=sort_df(df5,'flare_classification',natural_keys)
df5=df5.drop(df5.index[range(50,482)])
df5.index=range(1,51)

df5['rank']=df5.index
cols5 = list(df5)
# move the column to head of list using index, pop and insert
cols5.insert(0,cols5.pop(cols5.index('rank')))
df5 = df5.ix[:, cols5]

df5['x_class']=df5['flare_classification']
cols5 = list(df5)
# move the column to head of list using index, pop and insert
cols5.insert(1,cols5.pop(cols5.index('x_class')))
df5 = df5.ix[:, cols5]

df5['max_datetime']=df5['cme_datetime']
cols5 = list(df5)
# move the column to head of list using index, pop and insert
cols5.insert(3,cols5.pop(cols5.index('max_datetime')))
df5 = df5.ix[:, cols5]

df5=df5.drop("start_frequency",1)
df5=df5.drop("end_frequency",1)
df5=df5.drop("flare_location",1)
df5=df5.drop("flare_classification",1)
df5=df5.drop("cme_angle",1)
df5=df5.drop("cme_width",1)
```

```
df5=df5.drop("cme_speed",1)
df5=df5.drop("is_halo",1)
df5=df5.drop("width_lower_bound",1)
df5=df5.drop("cme_datetime",1)
df5['flare_region']=list(map(lambda x: x[-4:], df5['flare_region'])
)
print(df5)
```

```
    rank x_class      start_datetime        max_datetime        en
d_datetime  \
1      1   X28.0 2003-11-04 20:00:00 2003-11-04 19:54:00 2003-11-0
5 00:00:00
2      2   X20.0 2001-04-02 22:05:00 2001-04-02 22:06:00 2001-04-0
3 02:30:00
3      3   X17.0 2003-10-28 11:10:00 2003-10-28 11:30:00 2003-10-3
0 00:00:00
4      4   X14.0 2001-04-15 14:05:00 2001-04-15 14:06:00 2001-04-1
6 13:00:00
5      5   X10.0 2003-10-29 20:55:00 2003-10-29 20:54:00 2003-10-3
0 00:00:00
6      6    X9.4 1997-11-06 12:20:00 1997-11-06 12:10:00 1997-11-0
7 08:30:00
7      7    X9.0 2006-12-05 10:50:00                NaT 2006-12-0
5 20:00:00
8      8    X8.3 2003-11-02 17:30:00 2003-11-02 17:30:00 2003-11-0
3 01:00:00
9      9    X7.1 2005-01-20 07:15:00 2005-01-20 06:54:00 2005-01-2
0 16:30:00
10    10    X6.9 2011-08-09 08:20:00 2011-08-09 08:12:00 2011-08-0
9 08:35:00
11    11    X6.5 2006-12-06 19:00:00                NaT 2006-12-0
9 00:00:00
12    12    X6.2 2005-09-09 19:45:00 2005-09-09 19:48:00 2005-09-0
9 22:00:00
13    13    X5.7 2000-07-14 10:30:00 2000-07-14 10:54:00 2000-07-1
5 14:30:00
14    14    X5.6 2001-04-06 19:35:00 2001-04-06 19:30:00 2001-04-0
7 01:50:00
15    15    X5.4 2012-03-07 01:00:00 2012-03-07 00:24:00 2012-03-0
8 19:00:00
16    16    X5.3 2001-08-25 16:50:00 2001-08-25 16:50:00 2001-08-2
5 23:00:00
17    17    X4.9 2014-02-25 00:56:00 2014-02-25 01:25:00 2014-02-2
5 11:28:00
18    18    X4.8 2002-07-23 00:50:00 2002-07-23 00:42:00 2002-07-2
3 04:00:00
19    19    X4.0 2000-11-26 17:00:00 2000-11-26 17:06:00 2000-11-2
6 17:15:00
20    20    X3.9 2003-11-03 10:00:00 2003-11-03 10:06:00 2003-11-0
3 12:30:00
21    21    X3.8 2005-01-17 10:00:00 2005-01-17 09:54:00 2005-01-1
7 10:35:00
22    22    X3.6 2003-05-28 01:00:00 2003-05-28 00:50:00 2003-05-2
9 00:30:00
```

```
23    23    X3.4 2006-12-13 02:45:00 2006-12-13 02:54:00 2006-12-1
3 10:40:00
24    24    X3.4 2001-12-28 20:35:00 2001-12-28 20:30:00 2001-12-2
9 03:00:00
25    25    X3.3 2002-07-20 21:30:00 2002-07-20 22:06:00 2002-07-2
0 22:20:00
26    26    X3.2 2013-05-14 01:16:00 2013-05-14 01:25:00 2013-05-1
4 02:35:00
27    27    X3.1 2002-08-24 01:45:00 2002-08-24 01:27:00 2002-08-2
4 03:25:00
28    28    X2.8 2013-05-13 16:15:00 2013-05-13 16:07:00 2013-05-1
3 19:10:00
29    29    X2.7 2003-11-03 01:15:00 2003-11-03 01:59:00 2003-11-0
3 01:25:00
30    30    X2.7 1998-05-06 08:25:00 1998-05-06 08:29:00 1998-05-0
6 08:35:00
31    31    X2.6 2005-01-15 23:00:00 2005-01-15 23:06:00 2005-01-1
5 00:00:00
32    32    X2.6 2001-09-24 10:45:00 2001-09-24 10:30:00 2001-09-2
5 20:00:00
33    33    X2.6 1997-11-27 13:30:00 1997-11-27 13:56:00 1997-11-2
7 14:00:00
34    34    X2.5 2004-11-10 02:25:00 2004-11-10 02:26:00 2004-11-1
0 03:40:00
35    35    X2.3 2001-04-10 05:24:00 2001-04-10 05:30:00 2001-04-1
1 00:00:00
36    36    X2.3 2000-11-24 15:25:00 2000-11-24 15:30:00 2000-11-2
4 22:00:00
37    37    X2.3 2000-06-06 15:20:00 2000-06-06 15:54:00 2000-06-0
8 09:00:00
38    38    X2.2 2011-02-15 02:10:00 2011-02-15 02:24:00 2011-02-1
5 07:00:00
39    39    X2.1 2013-10-25 15:08:00 2013-10-25 15:12:00 2013-10-2
5 22:32:00
40    40    X2.1 2011-09-06 22:30:00 2011-09-06 23:05:00 2011-09-0
7 15:40:00
41    41    X2.1 2005-09-10 21:45:00 2005-09-10 21:52:00 2005-09-1
0 01:00:00
42    42    X2.1 1997-11-04 06:00:00 1997-11-04 06:10:00 1997-11-0
5 04:30:00
43    43    X2.0 2005-01-17 09:25:00 2005-01-17 09:30:00 2005-01-1
7 16:00:00
44    44    X2.0 2004-11-07 16:25:00 2004-11-07 16:54:00 2004-11-0
8 20:00:00
45    45    X2.0 2001-04-12 10:20:00 2001-04-12 10:31:00 2001-04-1
2 10:40:00
46    46    X2.0 2000-11-24 05:10:00 2000-11-24 05:30:00 2000-11-2
4 15:00:00
47    47    X1.9 2000-11-25 19:00:00 2000-11-25 19:31:00 2000-11-2
5 19:35:00
48    48    X1.8 2002-07-18 07:55:00 2002-07-18 08:06:00 2002-07-1
8 08:45:00
49    49    X1.8 2000-11-24 22:24:00 2000-11-24 22:06:00 2000-11-2
4 22:36:00
```

```
50    50    X1.8 1999-10-14 09:10:00 1999-10-14 09:26:00 1999-10-1
4 10:00:00
```

	flare_region
1	0486
2	9393
3	0486
4	9415
5	0486
6	8100
7	0930
8	0486
9	0720
10	1263
11	0930
12	0808
13	9077
14	9415
15	1429
16	9591
17	1990
18	0039
19	9236
20	0488
21	0720
22	0365
23	0930
24	9756
25	0039
26	1748
27	0069
28	1748
29	0488
30	8210
31	0720
32	9632
33	8113
34	0696
35	9415
36	9236
37	9026
38	1158
39	1882
40	1283
41	0808
42	8100
43	0720
44	0696
45	9415
46	9236
47	9236
48	0030
49	9236
50	8731

In [6]:
```
############################################### PART 2 ############
###################################
############################## Question 2 ##############################
# description: copy the NASA data and extract the information of x-
class,start date, flare region,
############## start time ,max-time, and end time; reindex the new
dataset, for the top 50 solar flares,
############## firstly sort each datetime(%Y%M%D) in top 50 solar f
lares dataset, if we dont find
############## any date matched , we use x_class for sort; if match
ed datetimes or matched x_class are
############## bigger than two selections, we use the the differenc
es between the selected dates and the
############## SpaceWeather date or match the x_class is order to f
ind the best choice. If there is still
############## no selected row, then we print 'No match'.
############## Finally, we add the found SpaceWeatherLive rank into
the NASA dataframe.
dfSF = df[['x_class', 'date', 'region', 'start_time', 'max_time', '
end_time']]
dfNS = df3[['flare_classification', 'start_date', 'flare_region', '
start_time', 'cme_time', 'end_time']].copy()
dfNS['row'] = range(1, 483)
dfNS = dfNS[['row', 'flare_classification', 'start_date', 'flare_re
gion', 'start_time', 'cme_time', 'end_time']]

dfNSSF = dfNS.copy();
dfNSSF['Rank in SpaceWeatherLive'] = ('');

print(['row', 'x', 'start_date', 'region', 'start_time', 'max_time'
, 'end_time'])

for i in range (1,51):
    sf = dfSF.loc[[i], :].values
    x = sf[0][0];
    date = sf[0][1];
    ns = dfNS.loc[dfNS['start_date'] == date].values

    if(ns.size == 0):
        ns = dfNS.loc[dfNS['flare_classification'] == x].values

    if(ns.size == 6):
        ns = ns[0];

    if (ns.size > 6):
        temp = ns[0];
        for j in range(0,int((ns.size/7))):
            if(abs(pd.to_datetime(ns[j][2]) - pd.to_datetime(date))
< \
               abs(pd.to_datetime(temp[2]) - pd.to_datetime(date)))
:
                temp = ns[j];
```

```
            elif(abs(pd.to_datetime(ns[j][2]) - pd.to_datetime(date
)) == \
                 abs(pd.to_datetime(temp[2]) - pd.to_datetime(date)
)):
                if(ns[j][1][0] == x[0] and float(ns[j][1][1:]) == f
loat(x[1:])):
                    temp = ns[j];

            else:
                temp = ns[j];

        ns = temp;

    if(ns.size == 0):
        print(i);
        print('No match');
        print('---');
    else:
        print(i);
        print(ns);
        print('----------------------------------------------------
--------');
        r_i = (ns[0]);
        dfNSSF.loc[r_i, 'Rank in SpaceWeatherLive'] = ('rank: '+ st
r(i));
print(dfNSSF)
```

```
['row', 'x', 'start_date', 'region', 'start_time', 'max_time', 'en
d_time']
1
[243 'X28.' '2003/11/04' '10486' '20:00' '19:54' '24:00']
------------------------------------------------------------
2
[120 'X20.' '2001/04/02' '9393' '22:05' '22:06' '02:30']
------------------------------------------------------------
3
[235 'X17.' '2003/10/28' '10486' '11:10' '11:30' '24:00']
------------------------------------------------------------
4
[319 'X1.7' '2005/09/07' '10808' '18:05' '--:--' '00:00']
------------------------------------------------------------
5
[129 'X14.' '2001/04/15' '9415' '14:05' '14:06' '13:00']
------------------------------------------------------------
6
[236 'X10.' '2003/10/29' '10486' '20:55' '20:54' '24:00']
------------------------------------------------------------
7
[9 'X9.4' '1997/11/06' '8100' '12:20' '12:10' '08:30']
------------------------------------------------------------
8
[331 'X9.0' '2006/12/05' '10930' '10:50' '--:--' '20:00']
------------------------------------------------------------
```

```
9
[239 'X8.3' '2003/11/02' '10486' '17:30' '17:30' '01:00']
------------------------------------------------------------
10
[291 'X7.1' '2005/01/20' '10720' '07:15' '06:54' '16:30']
------------------------------------------------------------
11
[361 'X6.9' '2011/08/09' '11263' '08:20' '08:12' '08:35']
------------------------------------------------------------
12
[334 'X6.5' '2006/12/06' '10930' '19:00' '--:--' '24:00']
------------------------------------------------------------
13
[320 'X6.2' '2005/09/09' '10808' '19:45' '19:48' '22:00']
------------------------------------------------------------
14
[320 'X6.2' '2005/09/09' '10808' '19:45' '19:48' '22:00']
------------------------------------------------------------
15
[84 'X5.7' '2000/07/14' '9077' '10:30' '10:54' '14:30']
------------------------------------------------------------
16
[124 'X5.6' '2001/04/06' '9415' '19:35' '19:30' '01:50']
------------------------------------------------------------
17
[377 'X5.4' '2012/03/07' '11429' '01:00' '00:24' '19:00']
------------------------------------------------------------
18
[377 'X5.4' '2012/03/07' '11429' '01:00' '00:24' '19:00']
------------------------------------------------------------
19
[377 'X5.4' '2012/03/07' '11429' '01:00' '00:24' '19:00']
------------------------------------------------------------
20
[138 'X5.3' '2001/08/25' '9591' '16:50' '16:50' '23:00']
------------------------------------------------------------
21
[445 'X4.9' '2014/02/25' '11990' '00:56' '01:25' '11:28']
------------------------------------------------------------
22
[445 'X4.9' '2014/02/25' '11990' '00:56' '01:25' '11:28']
------------------------------------------------------------
23
[196 'X4.8' '2002/07/23' '10039' '00:50' '00:42' '04:00']
------------------------------------------------------------
24
[107 'X4.0' '2000/11/26' '9236' '17:00' '17:06' '17:15']
------------------------------------------------------------
25
[241 'X3.9' '2003/11/03' '10488' '10:00' '10:06' '12:30']
------------------------------------------------------------
26
[241 'X3.9' '2003/11/03' '10488' '10:00' '10:06' '12:30']
------------------------------------------------------------
```

```
27
[290 'X3.8' '2005/01/17' '10720' '10:00' '09:54' '10:35']
------------------------------------------------------------
28
No match
---
29
[224 'X3.6' '2003/05/28' '10365' '01:00' '00:50' '00:30']
------------------------------------------------------------
30
[224 'X3.6' '2003/05/28' '10365' '01:00' '00:50' '00:30']
------------------------------------------------------------
31
[320 'X6.2' '2005/09/09' '10808' '19:45' '19:48' '22:00']
------------------------------------------------------------
32
[335 'X3.4' '2006/12/13' '10930' '02:45' '02:54' '10:40']
------------------------------------------------------------
33
[163 'X3.4' '2001/12/28' '9756' '20:35' '20:30' '03:00']
------------------------------------------------------------
34
[195 'X3.3' '2002/07/20' '10039' '21:30' '22:06' '22:20']
------------------------------------------------------------
35
[195 'X3.3' '2002/07/20' '10039' '21:30' '22:06' '22:20']
------------------------------------------------------------
36
[195 'X3.3' '2002/07/20' '10039' '21:30' '22:06' '22:20']
------------------------------------------------------------
37
[406 'X3.2' '2013/05/14' '11748' '01:16' '01:25' '02:35']
------------------------------------------------------------
38
[203 'X3.1' '2002/08/24' '10069' '01:45' '01:27' '03:25']
------------------------------------------------------------
39
[203 'X3.1' '2002/08/24' '10069' '01:45' '01:27' '03:25']
------------------------------------------------------------
40
[190 'M1.8' '2002/07/15' '10030' '21:15' '21:30' '05:00']
------------------------------------------------------------
41
[405 'X2.8' '2013/05/13' '11748' '16:15' '16:07' '19:10']
------------------------------------------------------------
42
[160 '----' '2001/12/11' '-----' '12:45' '09:54' '17:00']
------------------------------------------------------------
43
[405 'X2.8' '2013/05/13' '11748' '16:15' '16:07' '19:10']
------------------------------------------------------------
44
[240 'X2.7' '2003/11/03' '10488' '01:15' '01:59' '01:25']
------------------------------------------------------------
```

```
45
[240 'X2.7' '2003/11/03' '10488' '01:15' '01:59' '01:25']
------------------------------------------------------------
46
[20 'X2.7' '1998/05/06' '8210' '08:25' '08:29' '08:35']
------------------------------------------------------------
47
[10 'X2.6' '1997/11/27' '8113' '13:30' '13:56' '14:00']
------------------------------------------------------------
48
[145 'X2.6' '2001/09/24' '9632' '10:45' '10:30' '20:00']
------------------------------------------------------------
49
[287 'X2.6' '2005/01/15' '10720' '23:00' '23:06' '00:00']
------------------------------------------------------------
50
[279 'X2.5' '2004/11/10' '10696' '02:25' '02:26' '03:40']
------------------------------------------------------------
     row flare_classification  start_date flare_region start_time
cme_time  \
1      1                M1.3  1997/04/01         8026      14:00
15:18
2      2                C6.8  1997/04/07         8027      14:30
14:27
3      3                C1.3  1997/05/12         8038      05:15
05:30
4      4                M1.3  1997/05/21         8040      20:20
21:00
5      5                C1.4  1997/09/23         8088      21:53
22:02
6      6                C8.6  1997/11/03         8100      05:15
05:28
7      7                M4.2  1997/11/03         8100      10:30
11:11
8      8                X2.1  1997/11/04         8100      06:00
06:10
9      9                X9.4  1997/11/06         8100      12:20
12:10
10    10                X2.6  1997/11/27         8113      13:30
13:56
11    11                B9.4  1997/12/12         8116      22:45
00:26
12    12                C1.1  1998/01/25         8141      15:03
15:26
13    13                ----  1998/03/29        -----      03:40
03:48
14    14                M1.4  1998/04/20         8194      10:25
10:07
15    15                X1.2  1998/04/23         8210      06:00
05:55
16    16                C8.9  1998/04/24         8210      09:17
08:55
17    17                X1.0  1998/04/27         8210      09:20
08:56
```

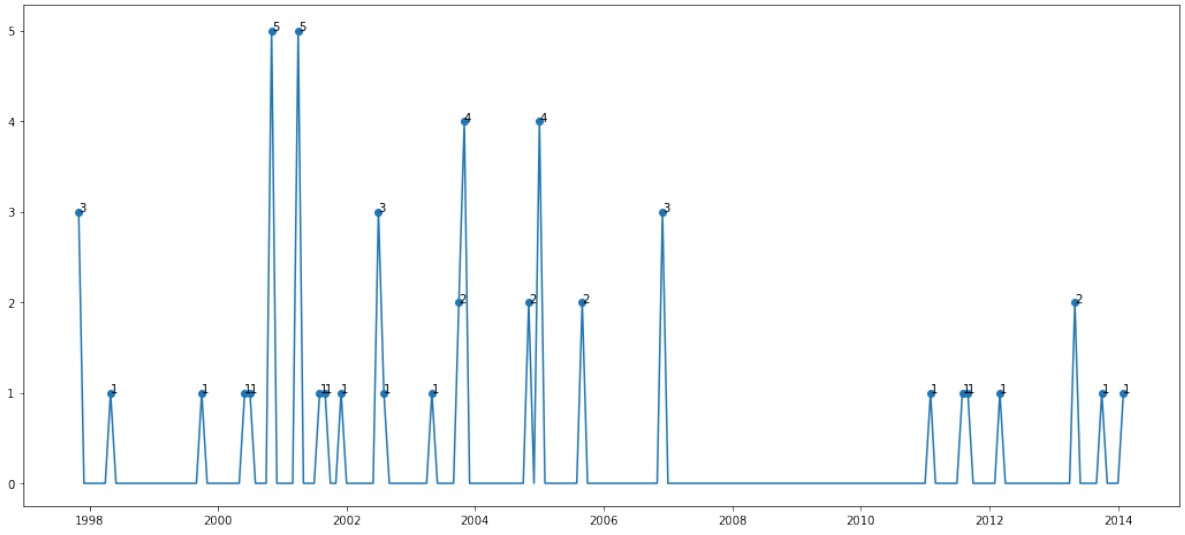
```
18    18                  M6.8  1998/04/29         8210       16:30
16:58
19    19                  X1.1  1998/05/02         8210       14:25
14:06
20    20                  X2.7  1998/05/06         8210       08:25
08:29
21    21                  M7.7  1998/05/09         8210       03:35
03:35
22    22                  B6.6  1998/05/11         8214       21:40
21:55
23    23                  B7.9  1998/05/19         8222       10:00
10:27
24    24                  C7.5  1998/05/27         8226       13:30
13:45
25    25                  M1.4  1998/06/11         8243       10:15
10:28
26    26                  M1.0  1998/06/16         8232       18:20
18:27
27    27                  ----  1998/06/20        -----       19:39
18:20
28    28                  C2.9  1998/06/22         8243       07:15
05:01
29    29                  C4.4  1998/11/02         8373       14:00
14:18
30    30                  M8.4  1998/11/05         8375       22:00
20:44
..   ...                   ...         ...          ...        ...
...
453  453                  C8.3  2014/04/04        12027       14:02
14:12
454  454                  M7.3  2014/04/18        12036       13:05
13:25
455  455                  ----  2014/05/07        12027       16:24
16:24
456  456                  ----  2014/05/08        12051       03:21
03:24
457  457                  ----  2014/05/09        12051       02:40
02:48
458  458                  ----  2014/05/10        12051       04:32
04:36
459  459                  X1.5  2014/06/10        12087       12:58
13:30
460  460                  M3.1  2014/06/12        12085       22:14
22:12
461  461                  C1.5  2014/07/30          EP?       07:44
07:00
462  462                  M1.5  2014/08/01        12127       18:58
18:36
463  463                  C2.2  2014/08/22        12146       10:37
11:12
464  464                  M2.0  2014/08/25        12146       15:20
15:36
465  465                  M3.9  2014/08/25        12146       20:43
20:48
```

```
466  466                 ----  2014/08/28        12157      17:05
17:24
467  467                 ----  2014/09/01        12158      11:12
11:12
468  468                 M4.5  2014/09/09        12158      00:05
00:06
469  469                 X1.6  2014/09/10        12158      17:45
18:00
470  470                 ----  2014/09/20        12164      05:10
05:24
471  471                 ----  2014/09/22        12158      06:13
06:12
472  472                 M2.3  2014/09/23        12172      23:41
23:36
473  473                 ----  2014/09/24           EP      20:54
21:30
474  474                 M7.3  2014/10/02        12173      21:34
19:12
475  475                 C3.0  2014/10/10           EP      18:11
16:12
476  476                 C4.4  2014/10/21        12192      12:33
12:48
477  477                 ----  2014/11/08        12203      16:57
16:36
478  478                 ----  2014/12/13        -----      14:27
14:24
479  479                 M1.1  2014/12/17        12241      04:09
02:00
480  480                 M8.7  2014/12/17        12242      05:00
05:00
481  481                 M6.9  2014/12/18        12241      22:31
01:04
482  482                 M1.0  2014/12/21        12241      12:05
12:12

    end_time Rank in SpaceWeatherLive
1      14:15
2      17:30
3      16:00
4      22:00
5      22:16
6      12:00
7      11:30
8      04:30
9      08:30                rank: 7
10     14:00               rank: 47
11     23:20
12     15:18
13     03:52
14     06:00
15     15:30
16     09:25
17     10:00
18     17:00
```

```
19     14:50
20     08:35                  rank: 46
21     10:00
22     22:00
23     11:30
24     14:20
25     10:20
26     21:00
27     20:00
28     09:20
29     14:40
30     08:00
..       ...                     ...
453    14:07
454    22:50
455    23:18
456    05:26
457    04:30
458    08:37
459    15:00
460    22:35
461    08:00
462    05:00
463    11:18
464    16:02
465    21:00
466    22:08
467    22:35
468    13:00
469    12:00
470    05:30
471    06:50
472    23:47
473    23:48
474    21:56
475    18:33
476    13:01
477    17:18
478    14:51
479    04:19
480    05:09
481    22:54
482    12:28

[482 rows x 8 columns]
```

In [7]:
```
############################## Question 3 ##############################
# description: We plot the number of flares per month over time, ad
d a graphical element to indicate
############## the number of strong flares (in the top 50 from NASA
) to see if they cluster. The months,
############## which don't have a strong flare would be plot the va
lue 0. If the curves of the plot fluctuate
```

```
############## in a large scale, then it implies that the strong fl
ares cluster in some specific months.
############## The x-coordinate of our plot is the months from 1997
-11-01 to 2014-02-01, the y-coordinate
############## is the number of strong flares (in the top 50 from N
ASA) occurred in that month. The points
############## with a numbers on the courves are the graphical elem
ents to indicate the number of the strong
############## flares. According to the graph we plot, we can see s
everal strong flares cluster in several
############## months between 2000 and 2005, especially in 2000-11-
01, 2001-04-01, 2003-11-01, and 2005-01-01.
import matplotlib.pyplot as plt
from dateutil.relativedelta import relativedelta
month_list=list(map(lambda x: x.strftime("%Y-%m"),df5['start_dateti
me']))
month_index=[]
for e in month_list:
       if e not in month_index:
           month_index.append(e)
month_index.sort()
month_index=pd.to_datetime(month_index)

result = []
current = month_index[0]
last_month = month_index[-1]

while current <= last_month:
    result.append(current)
    current += relativedelta(months=1)
result=pd.to_datetime(result)

df6 = pd.DataFrame(0,columns=['month_value'], index=result)
df7 = pd.DataFrame(0,columns=['month_value'], index=month_index)
for e in month_list:
    if e in month_index:
        df6.xs(e)['month_value']=df6.xs(e)['month_value']+1;
        df7.xs(e)['month_value']=df7.xs(e)['month_value']+1;

plt.figure(figsize=(18,8))
plt.plot(result,df6['month_value'])

plt.scatter(month_index,df7['month_value'])
month_index=list(map(lambda x: (x.strftime("%Y-%m")),month_index))
for i, txt in enumerate(df7['month_value']):
    plt.annotate(txt, (month_index[i],df7['month_value'][i]))

plt.show()
############################################## PART 2 (End) ######
###################################
```


5
4
3
2
1
0
1998
2000
2002
2004
2006
2008
2010
2012
2014