In [1]:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import datetime
import re
from collections import Counter
```

```
D:\Working\Anaconda\lib\site-packages\scipy\__init__.py:138: UserWarning: A NumP
y version >=1.16.5 and <1.23.0 is required for this version of SciPy (detected v
ersion 1.24.3)
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion} is requ
ired for this version of "
```

In [2]:

```
data =  pd.read_csv("C:/Users/86166/Desktop/v2tone/full.csv")
data['DATE'] = pd.to_datetime(data['DATE'])
```

In [3]:

```
1 data
```

Out[3]:

	DATE	DocumentIdentifier	V2Tone	
0	2020-01-03	https://www.independent.co.uk/life-style/gadge...	-0.437318	SCIENCE;MANM
1	2020-01-06	https://cleantechnica.com/2020/01/05/energy-st...	-0.150038	WB_135_TRANSPO
2	2020-01-07	https://calgaryherald.com/business/local-busin...	0.846561	UNGP_FORESTS_I
3	2020-01-07	https://calgarysun.com/business/local-business...	0.847458	UNGP_FORESTS_I
4	2020-01-07	https://nationalpost.com/business/local-busine...	0.842993	UNGP_FORESTS_I
...	...	...	...	
7280	2016-12-23	http://www.einnews.com/pr_news/359300579/avalo...	0.942655	WB_855_LABOR_M
7281	2016-12-25	http://www.dealstreetasia.com/stories/india-li...	0.619638	WB_135_TRANSI
7282	2016-12-27	https://knowridge.com/2016/12/lost-lithium-des...	2.404526	DRUG_TRADE;WB
7283	2016-12-28	http://www.marketwatch.com/story/lithium-explo...	-0.975081	TAX_DISEASE;T
7284	2016-12-28	http://www.prnewswire.com/news-releases/lithiu...	-1.025057	TAX_DISEASE;T

7285 rows × 6 columns

In [4]:

```
1 Y=pd.read_csv("C:/Users/86166/Desktop/v2tone/lithium_v2tone_2016-2023_rate.csv")
```

In [5]:

```
1 Y = Y[['DATE2','Li2CO3 99%']]
2 Y = Y.rename(columns={'DATE2':'DATE'})
3 Y['DATE'] = pd.to_datetime(Y['DATE'])
```

In [6]:

```
mergedt = pd.merge(data,Y,on='DATE')
```

In [7]:

```
mergedt['Y'] = abs(mergedt['Li2CO3 99%'])
```

In [8]:

```
sort = mergedt.sort_values('Y',ascending=False)
```

In [9]:

```
top = sort[sort['Y']!=1][['DATE','Y','Li2CO3 99%']].drop_duplicates().nlargest(12,keep='al
```

In [10]:

```
top = top.sort_values('DATE')
```

In [11]:

```
top
```

Out[11]:

	DATE	Y	Li2CO3 99%
5384	2016-01-12	0.085837	0.085837
5387	2016-01-19	0.041152	0.041152
5469	2016-03-02	0.068259	0.068259
2966	2023-03-09	0.043290	-0.043290
3116	2023-03-31	0.050420	-0.050420
3132	2023-04-03	0.055432	-0.055432
3407	2023-05-08	0.050000	0.050000
3455	2023-05-10	0.062350	0.062350
3488	2023-05-11	0.073333	0.073333
3497	2023-05-12	0.070248	0.070248
3509	2023-05-15	0.062016	0.062016
3533	2023-05-18	0.055172	0.055172

In [12]:

```
neg = top[top['Li2CO3 99%']<0]
neg
```

Out[12]:

	DATE	Y	Li2CO3 99%
2966	2023-03-09	0.043290	-0.043290
3116	2023-03-31	0.050420	-0.050420
3132	2023-04-03	0.055432	-0.055432

In [13]:

```
pos = top.sort_values('Y').tail(4)
pos
```

Out[13]:

	DATE	Y	Li2CO3 99%
5469	2016-03-02	0.068259	0.068259
3497	2023-05-12	0.070248	0.070248
3488	2023-05-11	0.073333	0.073333
5384	2016-01-12	0.085837	0.085837

Negative

In [14]:

```
neg
```

Out[14]:

	DATE	Y	Li2CO3 99%
2966	2023-03-09	0.043290	-0.043290
3116	2023-03-31	0.050420	-0.050420
3132	2023-04-03	0.055432	-0.055432

In [15]:

```
### t7
t7 = set()
for i in range(neg.shape[0]):
    dt = neg.iloc[i]['DATE']
    j=7
    t7.add(dt+datetime.timedelta(days=-j))
    #day.add(dt+datetime.timedelta(days=1))
print(t7)
d1 = mergedt.loc[mergedt['DATE'].isin(t7)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_t7 = res.most_common(15)
res_t7
```

```
{Timestamp('2023-03-24 00:00:00'), Timestamp('2023-03-02 00:00:00'), Timestamp
('2023-03-27 00:00:00')}

[V2Tone--std]: 1.7486565762185693
[Y--std]: 0.0022471920728855932
```

Out[15]:

```
[('TAX_ETHNICITY', 13),
 ('ENV_MINING', 10),
 ('ECON_STOCKMARKET', 7),
 ('LEGISLATION', 5),
 ('MANMADE_DISASTER_IMPLIED', 4),
 ('GENERAL_GOVERNMENT', 4),
 ('ENV_SOLAR', 4),
 ('TAX_DISEASE', 4),
 ('EDUCATION', 3),
 ('ENV_GREEN', 3),
 ('NATURAL_DISASTER', 3),
 ('SLFID_MINERAL_RESOURCES', 2),
 ('KILL', 2),
 ('DELAY', 2),
 ('GENERAL_HEALTH', 2)]
```

In [16]:

```
top3 = pd.DataFrame(dict(res_t7),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

A

```
TAX_ETHNICITY
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      1.187074     1.750585 -0.017365   0.002115
```

B

```
ENV_MINING
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0       1.18267     1.484589 -0.017019   0.002321
```

C

```
ECON_STOCKMARKET
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      0.899151     1.306862 -0.016538   0.002653
```

In [17]:

```
### t6
t6 = set()
for i in range(neg.shape[0]):
    dt = neg.iloc[i]['DATE']
    j=6
    t6.add(dt+datetime.timedelta(days=-j))
    #day.add(dt+datetime.timedelta(days=1))
print(t6)
d1 = mergedt.loc[mergedt['DATE'].isin(t6)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_t6 = res.most_common(15)
res_t6
```

```
{Timestamp('2023-03-25 00:00:00'), Timestamp('2023-03-28 00:00:00'), Timestamp
('2023-03-03 00:00:00')}

[V2Tone--std]: 1.2787033209061482
[Y--std]: 0.005754809183030011
```

Out[17]:

```
[('ENV_MINING', 20),
 ('TAX_ETHNICITY', 19),
 ('ECON_STOCKMARKET', 13),
 ('MANMADE_DISASTER_IMPLIED', 12),
 ('LEGISLATION', 12),
 ('GENERAL_GOVERNMENT', 12),
 ('KILL', 8),
 ('ENV_GREEN', 8),
 ('GENERAL_HEALTH', 7),
 ('SLFID_MINERAL_RESOURCES', 6),
 ('NATURAL_DISASTER', 5),
 ('ECON_CURRENCY_EXCHANGE_RATE', 4),
 ('TRIAL', 4),
 ('TAX_DISEASE', 4),
 ('ECON_TAXATION', 3)]
```

In [18]:

```
top3 = pd.DataFrame(dict(res_t6),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

C

```
ENV_MINING
    mean_V2Tone  std_V2Tone    mean_Y     std_Y
0     0.609352    1.285618 -0.023455  0.006486
```

B

```
TAX_ETHNICITY
    mean_V2Tone  std_V2Tone   mean_Y     std_Y
0     0.641985    0.884184 -0.02412  0.005921
```

A

```
ECON_STOCKMARKET
    mean_V2Tone  std_V2Tone    mean_Y     std_Y
0     0.713154     1.35368 -0.024184  0.005935
```

In [19]:

```
### t5
t5 = set()
for i in range(neg.shape[0]):
    dt = neg.iloc[i]['DATE']
    j=5
    t5.add(dt+datetime.timedelta(days=-j))
    #day.add(dt+datetime.timedelta(days=1))
print(t5)
d1 = mergedt.loc[mergedt['DATE'].isin(t5)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_t5 = res.most_common(15)
res_t5
```

```
{Timestamp('2023-03-29 00:00:00'), Timestamp('2023-03-26 00:00:00'), Timestamp
('2023-03-04 00:00:00')}

[V2Tone--std]: 1.6336928848964023
[Y--std]: 7.152448122690996e-18
```

Out[19]:

```
[('TAX_ETHNICITY', 13),
 ('ECON_STOCKMARKET', 9),
 ('LEGISLATION', 6),
 ('ENV_MINING', 6),
 ('NATURAL_DISASTER', 5),
 ('SLFID_MINERAL_RESOURCES', 4),
 ('MANMADE_DISASTER_IMPLIED', 4),
 ('GENERAL_GOVERNMENT', 4),
 ('URBAN', 4),
 ('ELECTION', 4),
 ('RATIFY', 4),
 ('TAX_DISEASE', 3),
 ('ECON_ENTREPRENEURSHIP', 2),
 ('ENV_GREEN', 2),
 ('KILL', 2)]
```

In [20]:

```
top3 = pd.DataFrame(dict(res_t5),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

```
TAX_ETHNICITY
    mean_V2Tone  std_V2Tone    mean_Y         std_Y
0     1.031674    1.313241 -0.023346  7.222230e-18

ECON_STOCKMARKET
    mean_V2Tone  std_V2Tone    mean_Y  std_Y
0     1.225434    1.526008 -0.023346    0.0

LEGISLATION
    mean_V2Tone  std_V2Tone    mean_Y  std_Y
0     0.459598    1.427765 -0.023346    0.0
```

In [21]:

```
### t4
t4 = set()
for i in range(neg.shape[0]):
    dt = neg.iloc[i]['DATE']
    j=4
    t4.add(dt+datetime.timedelta(days=-j))
    #day.add(dt+datetime.timedelta(days=1))
print(t4)
d1 = mergedt.loc[mergedt['DATE'].isin(t4)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_t4 = res.most_common(15)
res_t4
```

```
{Timestamp('2023-03-30 00:00:00'), Timestamp('2023-03-05 00:00:00'), Timestamp
('2023-03-27 00:00:00')}

[V2Tone--std]: 1.579902713696243
[Y--std]: 0.004644958091178718
```

Out[21]:

```
[('TAX_ETHNICITY', 21),
 ('ENV_MINING', 18),
 ('ECON_STOCKMARKET', 13),
 ('GENERAL_GOVERNMENT', 12),
 ('LEGISLATION', 11),
 ('MANMADE_DISASTER_IMPLIED', 9),
 ('TAX_DISEASE', 8),
 ('KILL', 6),
 ('ENV_GREEN', 6),
 ('GENERAL_HEALTH', 4),
 ('DELAY', 4),
 ('PROTEST', 4),
 ('SLFID_MINERAL_RESOURCES', 4),
 ('NEGOTIATIONS', 4),
 ('NATURAL_DISASTER', 3)]
```

In [22]:

```
top3 = pd.DataFrame(dict(res_t4),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

C

```
TAX_ETHNICITY
    mean_V2Tone  std_V2Tone    mean_Y     std_Y
0      0.809604    1.699252 -0.024388  0.004718
```

B

```
ENV_MINING
    mean_V2Tone  std_V2Tone    mean_Y    std_Y
0      0.685218    1.656523 -0.025366  0.00437
```

A

```
ECON_STOCKMARKET
    mean_V2Tone  std_V2Tone    mean_Y     std_Y
0      0.863653    1.088693 -0.025812  0.004158
```

In [23]:

```
### t3
t3 = set()
for i in range(neg.shape[0]):
    dt = neg.iloc[i]['DATE']
    j=3
    t3.add(dt+datetime.timedelta(days=-j))
    #day.add(dt+datetime.timedelta(days=1))
print(t3)
d1 = mergedt.loc[mergedt['DATE'].isin(t3)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_t3 = res.most_common(15)
res_t3
```

```
{Timestamp('2023-03-31 00:00:00'), Timestamp('2023-03-28 00:00:00'), Timestamp
('2023-03-06 00:00:00')}

[V2Tone--std]: 1.8608267110233963
[Y--std]: 0.015233517257276398
```

Out[23]:

```
[('TAX_ETHNICITY', 33),
 ('ENV_MINING', 31),
 ('GENERAL_GOVERNMENT', 21),
 ('ECON_STOCKMARKET', 21),
 ('MANMADE_DISASTER_IMPLIED', 18),
 ('LEGISLATION', 12),
 ('SLFID_MINERAL_RESOURCES', 11),
 ('TAX_DISEASE', 10),
 ('GENERAL_HEALTH', 10),
 ('NATURAL_DISASTER', 8),
 ('KILL', 8),
 ('ENV_GREEN', 7),
 ('ECON_TAXATION', 7),
 ('ENV_OIL', 5),
 ('ECON_ENTREPRENEURSHIP', 5)]
```

In [24]:

```
top3 = pd.DataFrame(dict(res_t3),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

A

```
TAX_ETHNICITY
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0       0.29501     1.383319 -0.033592   0.015136
```

C

```
ENV_MINING
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      0.617851     1.290584 -0.032506   0.014975
```

B

```
GENERAL_GOVERNMENT
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      0.461631     1.416169 -0.032854   0.015642
```

In [25]:

```
### t2
t2 = set()
for i in range(neg.shape[0]):
    dt = neg.iloc[i]['DATE']
    j=2
    t2.add(dt+datetime.timedelta(days=-j))
    #day.add(dt+datetime.timedelta(days=1))
print(t2)
d1 = mergedt.loc[mergedt['DATE'].isin(t2)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_t2 = res.most_common(15)
res_t2
```

```
{Timestamp('2023-03-29 00:00:00'), Timestamp('2023-04-01 00:00:00'), Timestamp
('2023-03-07 00:00:00')}

[V2Tone--std]: 1.914955841621569
[Y--std]: 0.009724598284415749
```

Out[25]:

```
[('TAX_ETHNICITY', 31),
 ('ENV_MINING', 21),
 ('MANMADE_DISASTER_IMPLIED', 13),
 ('ECON_STOCKMARKET', 10),
 ('GENERAL_GOVERNMENT', 9),
 ('LEGISLATION', 7),
 ('NATURAL_DISASTER', 6),
 ('EDUCATION', 5),
 ('TAX_DISEASE', 5),
 ('GENERAL_HEALTH', 4),
 ('MEDICAL', 4),
 ('SLFID_MINERAL_RESOURCES', 4),
 ('URBAN', 4),
 ('ELECTION', 4),
 ('RATIFY', 4)]
```

In [26]:

```
top3 = pd.DataFrame(dict(res_t2),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

A

```
TAX_ETHNICITY
    mean_V2Tone   std_V2Tone    mean_Y      std_Y
0      0.131633     1.468065 -0.01217   0.009655
```

C

```
ENV_MINING
    mean_V2Tone   std_V2Tone     mean_Y     std_Y
0     -0.140201     1.324484 -0.009598   0.00891
```

B

```
MANMADE_DISASTER_IMPLIED
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0     -0.908658       2.0961 -0.010021   0.009246
```

In [27]:

```
### t1
t1 = set()
for i in range(neg.shape[0]):
    dt = neg.iloc[i]['DATE']
    j=1
    t1.add(dt+datetime.timedelta(days=-j))
    #day.add(dt+datetime.timedelta(days=1))
print(t1)
d1 = mergedt.loc[mergedt['DATE'].isin(t1)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_t1 = res.most_common(15)
res_t1
```

```
{Timestamp('2023-04-02 00:00:00'), Timestamp('2023-03-08 00:00:00'), Timestamp
('2023-03-30 00:00:00')}

[V2Tone--std]: 1.4626233761691043
[Y--std]: 0.007089198913374481
```

Out[27]:

```
[('ENV_MINING', 16),
 ('TAX_ETHNICITY', 14),
 ('ECON_STOCKMARKET', 14),
 ('GENERAL_GOVERNMENT', 10),
 ('LEGISLATION', 8),
 ('ENV_GREEN', 7),
 ('MANMADE_DISASTER_IMPLIED', 6),
 ('SLFID_MINERAL_RESOURCES', 5),
 ('TAX_DISEASE', 5),
 ('KILL', 5),
 ('DELAY', 4),
 ('NEGOTIATIONS', 4),
 ('CORRUPTION', 3),
 ('ARMEDCONFLICT', 3),
 ('GENERAL_HEALTH', 3)]
```

In [28]:

```
top3 = pd.DataFrame(dict(res_t1),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

B

```
ENV_MINING
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      0.307117     1.712552 -0.025084   0.006269
```

A

```
TAX_ETHNICITY
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      0.300646     1.653841 -0.026889   0.004156
```

C

```
ECON_STOCKMARKET
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0        0.7984     0.799391 -0.023557   0.007291
```

In [29]:

```
### t0
t0 = set()
for i in range(neg.shape[0]):
    dt = neg.iloc[i]['DATE']
    j=0
    t0.add(dt+datetime.timedelta(days=j))
    #day.add(dt+datetime.timedelta(days=1))
print(t0)
d1 = mergedt.loc[mergedt['DATE'].isin(t0)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_t0 = res.most_common(15)
res_t0
```

```
{Timestamp('2023-03-09 00:00:00'), Timestamp('2023-03-31 00:00:00'), Timestamp
('2023-04-03 00:00:00')}

[V2Tone--std]: 2.325190058649111
[Y--std]: 0.004568839967703527
```

Out[29]:

```
[('TAX_ETHNICITY', 29),
 ('ENV_MINING', 20),
 ('TAX_DISEASE', 17),
 ('MANMADE_DISASTER_IMPLIED', 16),
 ('GENERAL_HEALTH', 15),
 ('ECON_STOCKMARKET', 15),
 ('KILL', 12),
 ('GENERAL_GOVERNMENT', 12),
 ('SCIENCE', 11),
 ('EDUCATION', 11),
 ('HEALTH_PANDEMIC', 10),
 ('BAN', 10),
 ('SLFID_MINERAL_RESOURCES', 10),
 ('LEGISLATION', 9),
 ('ARMEDCONFLICT', 8)]
```

In [30]:

```
top3 = pd.DataFrame(dict(res_t0),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

B

TAX_ETHNICITY

	mean_V2Tone	std_V2Tone	mean_Y	std_Y
0	-0.83113	1.853501	-0.049417	0.004645

A

ENV_MINING

	mean_V2Tone	std_V2Tone	mean_Y	std_Y
0	0.488366	1.666214	-0.050247	0.004157

C

TAX_DISEASE

	mean_V2Tone	std_V2Tone	mean_Y	std_Y
0	-1.661974	2.104821	-0.048664	0.005024

In [31]:

```
### tp1
tp1 = set()
for i in range(neg.shape[0]):
    dt = neg.iloc[i]['DATE']
    j=1
    tp1.add(dt+datetime.timedelta(days=j))
    #day.add(dt+datetime.timedelta(days=1))
print(tp1)
d1=mergedt.loc[mergedt['DATE'].isin(tp1)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_tp1 = res.most_common(15)
res_tp1
```

```
{Timestamp('2023-03-10 00:00:00'), Timestamp('2023-04-01 00:00:00'), Timestamp
('2023-04-04 00:00:00')}

[V2Tone--std]: 1.997702467160171
[Y--std]: 0.007543756077995352
```

Out[31]:

```
[('TAX_ETHNICITY', 26),
 ('ENV_MINING', 14),
 ('MANMADE_DISASTER_IMPLIED', 13),
 ('GENERAL_GOVERNMENT', 13),
 ('ECON_STOCKMARKET', 11),
 ('LEGISLATION', 11),
 ('GENERAL_HEALTH', 9),
 ('SCIENCE', 9),
 ('TAX_DISEASE', 8),
 ('ENV_OIL', 7),
 ('WOUND', 6),
 ('KILL', 5),
 ('WATER_SECURITY', 5),
 ('ECON_ENTREPRENEURSHIP', 5),
 ('MEDICAL', 5)]
```

In [32]:

```
top3 = pd.DataFrame(dict(res_tp1),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

A

```
TAX_ETHNICITY
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0     -0.002185     1.863549 -0.021582   0.007646
```

B

```
ENV_MINING
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0     -0.239197     1.861236 -0.021087   0.007723
```

A

```
MANMADE_DISASTER_IMPLIED
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      -1.00774     1.891283 -0.021582   0.007803
```

In [33]:

```
### sum(t-7 --- t-1)
all7 = set()
for i in range(neg.shape[0]):
    dt = neg.iloc[i]['DATE']
    for j in range(1,8):
        all7.add(dt+datetime.timedelta(days=-j))
    #day.add(dt+datetime.timedelta(days=1))
print(all7)
d1 = mergedt.loc[mergedt['DATE'].isin(all7)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_all7 = res.most_common(15)
res_all7
```

```
{Timestamp('2023-03-08 00:00:00'), Timestamp('2023-04-02 00:00:00'), Timestamp
('2023-03-07 00:00:00'), Timestamp('2023-03-29 00:00:00'), Timestamp('2023-03-27
00:00:00'), Timestamp('2023-03-26 00:00:00'), Timestamp('2023-04-01 00:00:00'),
Timestamp('2023-03-06 00:00:00'), Timestamp('2023-03-30 00:00:00'), Timestamp('2
023-03-31 00:00:00'), Timestamp('2023-03-03 00:00:00'), Timestamp('2023-03-04 0
0:00:00'), Timestamp('2023-03-02 00:00:00'), Timestamp('2023-03-24 00:00:00'), T
imestamp('2023-03-28 00:00:00'), Timestamp('2023-03-25 00:00:00'), Timestamp('20
23-03-05 00:00:00')}

[V2Tone--std]: 1.7833960452128648
[Y--std]: 0.014303938443581764
```

Out[33]:

```
[('TAX_ETHNICITY', 94),
 ('ENV_MINING', 82),
 ('ECON_STOCKMARKET', 54),
 ('GENERAL_GOVERNMENT', 46),
 ('MANMADE_DISASTER_IMPLIED', 44),
 ('LEGISLATION', 35),
 ('SLFID_MINERAL_RESOURCES', 25),
 ('TAX_DISEASE', 25),
 ('ENV_GREEN', 22),
 ('KILL', 20),
 ('GENERAL_HEALTH', 19),
 ('NATURAL_DISASTER', 18),
 ('DELAY', 13),
 ('EDUCATION', 12),
 ('ECON_TAXATION', 12)]
```

In [34]:

```
top5 = pd.DataFrame(dict(res_all7),index=['cnt']).columns[0:5]
for i in top5:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

```
TAX_ETHNICITY
    mean_V2Tone  std_V2Tone    mean_Y     std_Y
0     0.355204    1.500698 -0.022558  0.014216

ENV_MINING
    mean_V2Tone  std_V2Tone    mean_Y     std_Y
0     0.356517     1.45844 -0.022244  0.014273

ECON_STOCKMARKET
    mean_V2Tone  std_V2Tone   mean_Y     std_Y
0        0.6603    1.333849 -0.02657  0.012416

GENERAL_GOVERNMENT
    mean_V2Tone  std_V2Tone    mean_Y     std_Y
0     0.065559    1.420073 -0.025491  0.014266

MANMADE_DISASTER_IMPLIED
    mean_V2Tone  std_V2Tone    mean_Y     std_Y
0    -0.383188    1.998733 -0.019974  0.013587
```

Positive

In [35]:

```
pos
```

Out[35]:

	DATE	Y	Li2CO3 99%
5469	2016-03-02	0.068259	0.068259
3497	2023-05-12	0.070248	0.070248
3488	2023-05-11	0.073333	0.073333
5384	2016-01-12	0.085837	0.085837

In [36]:

```
### t7
t7 = set()
for i in range(pos.shape[0]):
    dt = pos.iloc[i]['DATE']
    j=7
    t7.add(dt+datetime.timedelta(days=-j))
    #day.add(dt+datetime.timedelta(days=1))
print(t7)
d1 = mergedt.loc[mergedt['DATE'].isin(t7)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_t7 = res.most_common(15)
res_t7
```

```
{Timestamp('2016-02-24 00:00:00'), Timestamp('2023-05-04 00:00:00'), Timestamp
('2023-05-05 00:00:00'), Timestamp('2016-01-05 00:00:00')}

[V2Tone--std]: 2.132678325877105
[Y--std]: 0.001881917041893799
```

Out[36]:

```
[('ENV_MINING', 17),
 ('TAX_ETHNICITY', 14),
 ('GENERAL_GOVERNMENT', 8),
 ('SLFID_MINERAL_RESOURCES', 8),
 ('ECON_STOCKMARKET', 6),
 ('LEGISLATION', 6),
 ('BAN', 5),
 ('MANMADE_DISASTER_IMPLIED', 5),
 ('WATER_SECURITY', 4),
 ('ENV_OIL', 2),
 ('SLFID_NATURAL_RESOURCES', 2),
 ('PUBLIC_TRANSPORT', 2),
 ('ECON_NATIONALIZE', 2),
 ('SHORTAGE', 2),
 ('GENERAL_HEALTH', 2)]
```

In [37]:

```
top3 = pd.DataFrame(dict(res_t7),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

A

```
ENV_MINING
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      0.809867      1.34515   0.003596   0.001294
```

C

```
TAX_ETHNICITY
    mean_V2Tone   std_V2Tone     mean_Y    std_Y
0      0.348388     1.969314   0.003374   0.0016
```

B

```
GENERAL_GOVERNMENT
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      0.328402     1.364235   0.003474   0.001275
```

In [38]:

```
### t6
t6 = set()
for i in range(pos.shape[0]):
    dt = pos.iloc[i]['DATE']
    j=6
    t6.add(dt+datetime.timedelta(days=-j))
    #day.add(dt+datetime.timedelta(days=1))
print(t6)
d1 = mergedt.loc[mergedt['DATE'].isin(t6)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_t6 = res.most_common(15)
res_t6
```

```
{Timestamp('2016-02-25 00:00:00'), Timestamp('2023-05-05 00:00:00'), Timestamp
('2023-05-06 00:00:00'), Timestamp('2016-01-06 00:00:00')}

[V2Tone--std]: 1.9215851241337818
[Y--std]: 0.0028451666039426573
```

Out[38]:

```
[('ENV_MINING', 10),
 ('TAX_ETHNICITY', 9),
 ('MANMADE_DISASTER_IMPLIED', 8),
 ('SLFID_MINERAL_RESOURCES', 6),
 ('GENERAL_GOVERNMENT', 6),
 ('LEGISLATION', 5),
 ('DELAY', 4),
 ('ECON_STOCKMARKET', 3),
 ('SHORTAGE', 2),
 ('ENV_OIL', 2),
 ('ECON_NATIONALIZE', 1),
 ('GENERAL_HEALTH', 1),
 ('MEDICAL', 1),
 ('EDUCATION', 1),
 ('DRUG_TRADE', 1)]
```

In [39]:

```
top3 = pd.DataFrame(dict(res_t6),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

A

```
ENV_MINING
    mean_V2Tone   std_V2Tone    mean_Y     std_Y
0      0.613783     0.861678   0.00277   0.00292
```

B

```
TAX_ETHNICITY
    mean_V2Tone   std_V2Tone     mean_Y     std_Y
0      0.520421     1.317213   0.002462   0.00292
```

C

```
MANMADE_DISASTER_IMPLIED
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0     -0.456028     2.268122   0.001385   0.002565
```

In [40]:

```
### t5
t5 = set()
for i in range(pos.shape[0]):
    dt = pos.iloc[i]['DATE']
    j=5
    t5.add(dt+datetime.timedelta(days=-j))
    #day.add(dt+datetime.timedelta(days=1))
print(t5)
d1 = mergedt.loc[mergedt['DATE'].isin(t5)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_t5 = res.most_common(15)
res_t5
```

```
{Timestamp('2016-01-07 00:00:00'), Timestamp('2016-02-26 00:00:00'), Timestamp
('2023-05-07 00:00:00'), Timestamp('2023-05-06 00:00:00')}

[V2Tone--std]: 1.725107557477848
[Y--std]: 0.0
```

Out[40]:

```
[('ENV_MINING', 3),
 ('TAX_ETHNICITY', 3),
 ('ECON_STOCKMARKET', 2),
 ('MANMADE_DISASTER_IMPLIED', 2),
 ('GENERAL_GOVERNMENT', 2),
 ('ENV_OIL', 1),
 ('NEGOTIATIONS', 1),
 ('ECON_TAXATION', 1),
 ('EDUCATION', 1),
 ('RETIREMENTS', 1),
 ('RETIREMENT', 1),
 ('ENV_COAL', 1),
 ('BAN', 1),
 ('MARITIME_INCIDENT', 1),
 ('MARITIME', 1)]
```

In [41]:

```
top3 = pd.DataFrame(dict(res_t5),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

```
ENV_MINING
    mean_V2Tone  std_V2Tone  mean_Y  std_Y
0     0.407442    0.369458     0.0    0.0

TAX_ETHNICITY
    mean_V2Tone  std_V2Tone  mean_Y  std_Y
0    -1.090314    2.078333     0.0    0.0

ECON_STOCKMARKET
    mean_V2Tone  std_V2Tone  mean_Y  std_Y
0    -1.743462    2.465628     0.0    0.0
```

In [42]:

```
### t4
t4 = set()
for i in range(pos.shape[0]):
    dt = pos.iloc[i]['DATE']
    j=4
    t4.add(dt+datetime.timedelta(days=-j))
    #day.add(dt+datetime.timedelta(days=1))
print(t4)
d1 = mergedt.loc[mergedt['DATE'].isin(t4)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_t4 = res.most_common(15)
res_t4
```

```
{Timestamp('2016-01-08 00:00:00'), Timestamp('2016-02-27 00:00:00'), Timestamp
('2023-05-07 00:00:00'), Timestamp('2023-05-08 00:00:00')}

[V2Tone--std]: 0.9210569881555004
[Y--std]: 7.152448122690996e-18
```

Out[42]:

```
[('TAX_ETHNICITY', 13),
 ('ECON_STOCKMARKET', 10),
 ('ENV_MINING', 6),
 ('GENERAL_GOVERNMENT', 5),
 ('MANMADE_DISASTER_IMPLIED', 5),
 ('TAX_DISEASE', 4),
 ('ECON_SUBSIDIES', 3),
 ('DELAY', 3),
 ('SLFID_MINERAL_RESOURCES', 3),
 ('GENERAL_HEALTH', 3),
 ('LEGISLATION', 2),
 ('HEALTH_PANDEMIC', 2),
 ('MEDICAL', 2),
 ('STRIKE', 2),
 ('ECON_CURRENCY_EXCHANGE_RATE', 2)]
```

In [43]:

```
top3 = pd.DataFrame(dict(res_t4),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

```
TAX_ETHNICITY
    mean_V2Tone  std_V2Tone  mean_Y  std_Y
0      0.308304     0.95407    0.05    0.0

ECON_STOCKMARKET
    mean_V2Tone  std_V2Tone  mean_Y         std_Y
0      0.237432    0.928135    0.05  7.314236e-18

ENV_MINING
    mean_V2Tone  std_V2Tone  mean_Y         std_Y
0      0.152858    0.975116    0.05  7.601177e-18
```

In [44]:

```
### t3
t3 = set()
for i in range(pos.shape[0]):
    dt = pos.iloc[i]['DATE']
    j=3
    t3.add(dt+datetime.timedelta(days=-j))
    #day.add(dt+datetime.timedelta(days=1))
print(t3)
d1 = mergedt.loc[mergedt['DATE'].isin(t3)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_t3 = res.most_common(15)
res_t3
```

```
{Timestamp('2023-05-09 00:00:00'), Timestamp('2023-05-08 00:00:00'), Timestamp
('2016-01-09 00:00:00'), Timestamp('2016-02-28 00:00:00')}

[V2Tone--std]: 1.1772566393466533
[Y--std]: 0.011071471653217443
```

Out[44]:

```
[('TAX_ETHNICITY', 29),
 ('MANMADE_DISASTER_IMPLIED', 16),
 ('ECON_STOCKMARKET', 13),
 ('TAX_DISEASE', 13),
 ('ENV_MINING', 12),
 ('GENERAL_GOVERNMENT', 11),
 ('ECON_ENTREPRENEURSHIP', 9),
 ('GENERAL_HEALTH', 8),
 ('SLFID_MINERAL_RESOURCES', 6),
 ('LEGISLATION', 6),
 ('DELAY', 5),
 ('ENV_GREEN', 4),
 ('ECON_SUBSIDIES', 3),
 ('HEALTH_PANDEMIC', 3),
 ('ECON_CURRENCY_EXCHANGE_RATE', 3)]
```

In [45]:

```
top3 = pd.DataFrame(dict(res_t3),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

B

```
TAX_ETHNICITY
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      0.200528     1.256901   0.037935   0.011067
```

C

```
MANMADE_DISASTER_IMPLIED
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0     -0.170097     1.282165   0.034966   0.010468
```

A

```
ECON_STOCKMARKET
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      0.317429     1.608971   0.044954   0.009589
```

In [46]:

```
### t2
t2 = set()
for i in range(pos.shape[0]):
    dt = pos.iloc[i]['DATE']
    j=2
    t2.add(dt+datetime.timedelta(days=-j))
    #day.add(dt+datetime.timedelta(days=1))
print(t2)
d1 = mergedt.loc[mergedt['DATE'].isin(t2)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_t2 = res.most_common(15)
res_t2
```

```
{Timestamp('2023-05-09 00:00:00'), Timestamp('2016-02-29 00:00:00'), Timestamp
('2023-05-10 00:00:00'), Timestamp('2016-01-10 00:00:00')}

[V2Tone--std]: 1.8939420204016237
[Y--std]: 0.017768196128823528
```

Out[46]:

```
[('TAX_ETHNICITY', 36),
 ('MANMADE_DISASTER_IMPLIED', 21),
 ('ECON_STOCKMARKET', 19),
 ('ENV_MINING', 17),
 ('TAX_DISEASE', 15),
 ('ECON_ENTREPRENEURSHIP', 11),
 ('GENERAL_GOVERNMENT', 8),
 ('ENV_GREEN', 8),
 ('GENERAL_HEALTH', 7),
 ('LEGISLATION', 6),
 ('AGRICULTURE', 5),
 ('SLFID_MINERAL_RESOURCES', 4),
 ('ECON_TAXATION', 4),
 ('NEW_CONSTRUCTION', 3),
 ('DELAY', 3)]
```

In [47]:

```
top3 = pd.DataFrame(dict(res_t2),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

B

```
TAX_ETHNICITY
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      0.393412     1.284264   0.047143   0.017244
```

C

```
MANMADE_DISASTER_IMPLIED
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0     -0.357598     1.641804   0.044427   0.017511
```

A

```
ECON_STOCKMARKET
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      0.893999     1.492231   0.056947   0.012819
```

In [48]:

```
### t1
t1 = set()
for i in range(pos.shape[0]):
    dt = pos.iloc[i]['DATE']
    j=1
    t1.add(dt+datetime.timedelta(days=-j))
    #day.add(dt+datetime.timedelta(days=1))
print(t1)
d1 = mergedt.loc[mergedt['DATE'].isin(t1)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_t1 = res.most_common(15)
res_t1
```

```
{Timestamp('2023-05-11 00:00:00'), Timestamp('2016-03-01 00:00:00'), Timestamp
('2023-05-10 00:00:00'), Timestamp('2016-01-11 00:00:00')}

[V2Tone--std]: 1.7939166523874135
[Y--std]: 0.01934557761944701
```

Out[48]:

```
[('TAX_ETHNICITY', 31),
 ('ECON_STOCKMARKET', 29),
 ('ENV_MINING', 28),
 ('MANMADE_DISASTER_IMPLIED', 20),
 ('GENERAL_GOVERNMENT', 11),
 ('TAX_DISEASE', 8),
 ('ENV_GREEN', 8),
 ('NEGOTIATIONS', 7),
 ('ECON_TAXATION', 6),
 ('WATER_SECURITY', 6),
 ('ECON_ENTREPRENEURSHIP', 5),
 ('NATURAL_DISASTER', 5),
 ('ETH_INDIGINOUS', 5),
 ('GENERAL_HEALTH', 4),
 ('LEGISLATION', 4)]
```

In [49]:

```
top3 = pd.DataFrame(dict(res_t1),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

B

```
TAX_ETHNICITY
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0       0.68426     1.006618   0.063882   0.012937
```

C

```
ECON_STOCKMARKET
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      1.276498     1.357922   0.059687   0.021266
```

A

```
ENV_MINING
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      0.830999     1.450749   0.066399   0.014088
```

In [50]:

```
### t0
t0 = set()
for i in range(pos.shape[0]):
    dt = pos.iloc[i]['DATE']
    j=0
    t0.add(dt+datetime.timedelta(days=j))
    #day.add(dt+datetime.timedelta(days=1))
print(t0)
d1 = mergedt.loc[mergedt['DATE'].isin(t0)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_t0 = res.most_common(15)
res_t0
```

```
{Timestamp('2023-05-11 00:00:00'), Timestamp('2023-05-12 00:00:00'), Timestamp
('2016-01-12 00:00:00'), Timestamp('2016-03-02 00:00:00')}

[V2Tone--std]: 1.1899149841724204
[Y--std]: 0.002926315864684628
```

Out[50]:

```
[('ENV_MINING', 36),
 ('TAX_ETHNICITY', 28),
 ('ECON_STOCKMARKET', 22),
 ('GENERAL_GOVERNMENT', 20),
 ('MANMADE_DISASTER_IMPLIED', 17),
 ('LEGISLATION', 14),
 ('SLFID_MINERAL_RESOURCES', 14),
 ('SCIENCE', 7),
 ('TAX_DISEASE', 7),
 ('WATER_SECURITY', 7),
 ('PROTEST', 6),
 ('ETH_INDIGINOUS', 6),
 ('ECON_ENTREPRENEURSHIP', 5),
 ('KILL', 5),
 ('NATURAL_DISASTER', 5)]
```

In [51]:

```
top3 = pd.DataFrame(dict(res_t0),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

B

```
ENV_MINING
    mean_V2Tone  std_V2Tone    mean_Y     std_Y
0      1.223447    1.154811  0.071343  0.001919
```

C

```
TAX_ETHNICITY
    mean_V2Tone  std_V2Tone    mean_Y     std_Y
0      1.118152    1.052492  0.070995  0.001918
```

A

```
ECON_STOCKMARKET
    mean_V2Tone  std_V2Tone    mean_Y     std_Y
0      1.373175    1.246953  0.071907  0.003763
```

In [52]:

```
### tp1
tp1 = set()
for i in range(pos.shape[0]):
    dt = pos.iloc[i]['DATE']
    j=1
    tp1.add(dt+datetime.timedelta(days=j))
    #day.add(dt+datetime.timedelta(days=1))
print(tp1)
d1=mergedt.loc[mergedt['DATE'].isin(tp1)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_tp1 = res.most_common(15)
res_tp1
```

```
{Timestamp('2023-05-13 00:00:00'), Timestamp('2016-03-03 00:00:00'), Timestamp
('2023-05-12 00:00:00'), Timestamp('2016-01-13 00:00:00')}

[V2Tone--std]: 1.2125766368956326
[Y--std]: 0.035724420187622584
```

Out[52]:

```
[('ENV_MINING', 25),
 ('TAX_ETHNICITY', 24),
 ('SLFID_MINERAL_RESOURCES', 20),
 ('LEGISLATION', 18),
 ('ECON_STOCKMARKET', 18),
 ('SLFID_NATURAL_RESOURCES', 11),
 ('ELECTION', 10),
 ('SCIENCE', 7),
 ('GENERAL_GOVERNMENT', 7),
 ('TAX_DISEASE', 5),
 ('MANMADE_DISASTER_IMPLIED', 5),
 ('EDUCATION', 5),
 ('PROTEST', 4),
 ('ECON_ENTREPRENEURSHIP', 4),
 ('ECON_TAXATION', 4)]
```

In [53]:

```
top3 = pd.DataFrame(dict(res_tp1),index=['cnt']).columns[0:3]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

A

```
ENV_MINING
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0       1.41703      0.98098   0.042149   0.035124
```

B

```
TAX_ETHNICITY
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      1.190814     1.201358   0.038051   0.035755
```

C

```
SLFID_MINERAL_RESOURCES
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      1.438538     0.666518   0.028099   0.035308
```

In [54]:

```
### sum(t-7 --- t-1)
all7 = set()
for i in range(pos.shape[0]):
    dt = pos.iloc[i]['DATE']
    for j in range(1,8):
        all7.add(dt+datetime.timedelta(days=-j))
    #day.add(dt+datetime.timedelta(days=1))
print(all7)
d1 = mergedt.loc[mergedt['DATE'].isin(all7)].reset_index(drop=True)
print('\n[V2Tone--std]:',d1['V2Tone'].std(),'\n[Y--std]:',d1['Li2CO3 99%'].std())
themes=[]
for i in (d1['FinalThemes'].values):
    for j in i.split(','):
        themes.append(j)
    res = Counter(themes)
res_all7 = res.most_common(15)
res_all7
```

```
{Timestamp('2016-02-29 00:00:00'), Timestamp('2023-05-04 00:00:00'), Timestamp
('2016-01-06 00:00:00'), Timestamp('2016-02-28 00:00:00'), Timestamp('2023-05-05
00:00:00'), Timestamp('2016-01-11 00:00:00'), Timestamp('2016-01-09 00:00:00'),
Timestamp('2023-05-10 00:00:00'), Timestamp('2016-02-26 00:00:00'), Timestamp('2
016-02-24 00:00:00'), Timestamp('2016-01-05 00:00:00'), Timestamp('2016-01-10 0
0:00:00'), Timestamp('2016-03-01 00:00:00'), Timestamp('2016-01-08 00:00:00'), T
imestamp('2016-01-07 00:00:00'), Timestamp('2023-05-06 00:00:00'), Timestamp('20
23-05-09 00:00:00'), Timestamp('2023-05-11 00:00:00'), Timestamp('2016-02-25 00:
00:00'), Timestamp('2016-02-27 00:00:00'), Timestamp('2023-05-07 00:00:00'), Tim
estamp('2023-05-08 00:00:00')}

[V2Tone--std]: 1.8036836053390493
[Y--std]: 0.028589012627860105
```

Out[54]:

```
[('TAX_ETHNICITY', 82),
 ('ENV_MINING', 65),
 ('ECON_STOCKMARKET', 50),
 ('MANMADE_DISASTER_IMPLIED', 49),
 ('GENERAL_GOVERNMENT', 36),
 ('TAX_DISEASE', 22),
 ('SLFID_MINERAL_RESOURCES', 21),
 ('LEGISLATION', 21),
 ('GENERAL_HEALTH', 16),
 ('ECON_ENTREPRENEURSHIP', 15),
 ('ENV_GREEN', 13),
 ('DELAY', 11),
 ('WATER_SECURITY', 10),
 ('NEGOTIATIONS', 9),
 ('ECON_TAXATION', 9)]
```

In [56]:

```
top3 = pd.DataFrame(dict(res_all7),index=['cnt']).columns[0:5]
for i in top3:
    mean_V2Tone = (d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].mean()
    std_V2Tone=(d1.loc[d1['FinalThemes'].str.contains(i)])['V2Tone'].std()
    mean_Y=(d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].mean()
    std_Y = (d1.loc[d1['FinalThemes'].str.contains(i)])['Li2CO3 99%'].std()
    dic = {'mean_V2Tone':mean_V2Tone,'std_V2Tone':std_V2Tone,'mean_Y':mean_Y,'std_Y':std_Y}
    print('%s'%(i),'\n',pd.DataFrame(dic,index=range(4)).head(1),'\n')
```

```
TAX_ETHNICITY
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      0.373404     1.384379   0.038143   0.026663

ENV_MINING
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      0.743897     1.257733   0.036756    0.03087

ECON_STOCKMARKET
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0      0.900576     1.516554   0.046695   0.026509

MANMADE_DISASTER_IMPLIED
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0     -0.378359     1.792738   0.037951    0.02813

GENERAL_GOVERNMENT
    mean_V2Tone   std_V2Tone     mean_Y      std_Y
0       0.26007     1.378118   0.034203   0.029961
```

	Sentiment Avg	Sentiment Volatility	Y Avg	Y Volatility
Mining	-1	L	M	H
SM	HH	M	HH	LL
MD	LL	HH	L	M
GG	L	L	LL	M
E	M	M	M	L