

In [1]:

```
from IPython.display import HTML
HTML('''<script>
code_show_err=false;
function code_toggle_err() {
  if (code_show_err){
    $('div.output_stderr').hide();
  } else {
    $('div.output_stderr').show();
  }
  code_show_err = !code_show_err
}
$( document ).ready(code_toggle_err);
</script>
To toggle on/off output_stderr, click <a href="javascript:code_toggle_err()">here</a>.''' )
```

Out[1]:

To toggle on/off output_stderr, click [here](#).

In [1]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from scipy.stats.stats import pearsonr
from scipy.stats import linregress
%matplotlib inline

df_tweets = pd.read_excel(r'C:\Users\timod\Desktop\GLD_tweets_all.xlsx')
df_price = pd.read_csv(r'C:\Users\timod\Desktop\GLD_historical_data_all.csv')
```

In [2]:

```
df_tweets.head()
```

Out[2]:

	date	text	follower_count	polarity	subjectivity
0	2020-03-27	b'Stock Indexes closed lower today with Dow do...	140	-0.141667	0.241667
1	2020-03-27	b"These Index Charts Will Calm You Down....Tod...	601	0.350000	0.775000
2	2020-03-27	b'\$GOLD \$GLD \$TGCDf https://t.co/wCuIlMnJt1'	77	0.000000	0.000000
3	2020-03-27	b'Stock Market Records Best 3-Day Gains Since ...	35974	0.500000	0.150000
4	2020-03-27	b'HUGE #inflation right around the corner! Cro...	1770	0.369048	0.661905

In [3]:

```
df_price
```

Out[3]:

	Date	Price	Open	High	Low	Vol.	Change %
0	Mar 27, 2020	152.25	153.00	153.55	152.00	9.82M	-0.65%
1	Mar 26, 2020	153.24	153.31	154.81	152.08	12.32M	1.28%
2	Mar 25, 2020	151.30	152.58	152.88	151.01	16.12M	-1.37%
3	Mar 24, 2020	153.40	153.50	155.66	152.05	20.56M	4.85%
4	Mar 23, 2020	146.30	142.68	146.93	142.28	28.12M	4.42%
5	Mar 20, 2020	140.11	140.71	141.17	139.14	20.06M	1.50%

6	Mar 19, 2020	138.04	139.08	140.50	137.85	20.64M	-1.89%
7	Mar 18, 2020	140.70	141.75	142.88	138.50	21.87M	-1.99%
8	Mar 17, 2020	143.56	140.18	146.20	139.93	24.49M	1.36%
9	Mar 16, 2020	141.64	137.56	142.96	136.12	24.50M	-1.14%
10	Mar 13, 2020	143.28	149.42	149.60	141.58	35.18M	-3.05%
11	Mar 12, 2020	147.79	151.53	151.55	146.85	32.89M	-3.99%
12	Mar 11, 2020	153.93	156.25	156.50	153.61	18.35M	-0.36%
13	Mar 10, 2020	154.48	156.18	156.41	154.39	18.81M	-2.11%
14	Mar 09, 2020	157.81	158.60	159.37	156.33	24.87M	0.17%
15	Mar 06, 2020	157.55	158.33	159.25	154.54	26.97M	0.04%
16	Mar 05, 2020	157.49	156.06	157.62	155.72	17.97M	2.16%
17	Mar 04, 2020	154.16	154.40	154.96	153.70	12.32M	0.18%
18	Mar 03, 2020	153.89	150.84	155.24	150.74	28.69M	3.14%
19	Mar 02, 2020	149.20	150.00	150.73	149.04	16.30M	0.55%
20	Feb 28, 2020	148.38	152.71	152.77	147.05	42.71M	-3.65%
21	Feb 27, 2020	154.00	155.58	156.27	153.90	27.88M	0.02%
22	Feb 26, 2020	153.97	153.82	155.28	152.90	17.98M	0.44%
23	Feb 25, 2020	153.30	155.12	156.08	152.91	28.09M	-1.79%
24	Feb 24, 2020	156.09	158.50	158.53	155.36	27.17M	0.90%
25	Feb 21, 2020	154.70	154.43	155.22	154.18	16.35M	1.50%
26	Feb 20, 2020	152.41	152.00	152.83	151.95	8.46M	0.41%
27	Feb 19, 2020	151.79	151.00	151.82	150.81	9.27M	0.58%
28	Feb 18, 2020	150.91	149.88	151.09	149.47	11.53M	1.28%
29	Feb 14, 2020	149.00	148.74	149.13	148.63	5.35M	0.42%
30	Feb 13, 2020	148.38	148.24	148.58	147.95	7.08M	0.57%
31	Feb 12, 2020	147.54	147.42	147.81	147.16	5.04M	-0.08%
32	Feb 11, 2020	147.66	148.03	148.16	147.01	5.18M	-0.34%
33	Feb 10, 2020	148.17	148.21	148.45	147.91	5.79M	0.26%
34	Feb 07, 2020	147.79	147.83	148.18	147.34	6.38M	0.26%
35	Feb 06, 2020	147.40	147.28	147.63	147.00	4.82M	0.54%
36	Feb 05, 2020	146.61	146.38	146.81	146.16	6.16M	0.12%
37	Feb 04, 2020	146.43	146.97	147.03	145.80	10.53M	-1.30%
38	Feb 03, 2020	148.36	148.66	148.78	147.68	9.12M	-0.65%

In [2]:

```
#Gathering the following for each of the first seven days of february:
#number of positive tweets, number of negative tweets, number of neutral tweets, total volume, sentiment polarity

'''
pos0331 = df_tweets.loc[df_tweets.date=='2020-03-31'][df_tweets.polarity > 0].shape[0]
neg0331 = df_tweets.loc[df_tweets.date=='2020-03-31'][df_tweets.polarity < 0].shape[0]
neutral0331 = df_tweets.loc[df_tweets.date=='2020-03-31'][df_tweets.polarity == 0].shape[0]
nonZero0331 = pos0331 + neg0331
px0331 = round(((pos0331-neg0331)/nonZero0331),3)
volume0331 = nonZero0331 + neutral0331

pos0330 = df_tweets.loc[df_tweets.date=='2020-03-30'][df_tweets.polarity > 0].shape[0]
neg0330 = df_tweets.loc[df_tweets.date=='2020-03-30'][df_tweets.polarity < 0].shape[0]
neutral0330 = df_tweets.loc[df_tweets.date=='2020-03-30'][df_tweets.polarity == 0].shape[0]
nonZero0330 = pos0330 + neg0330
px0330 = round(((pos0330-neg0330)/nonZero0330),3)
volume0330 = nonZero0330 + neutral0330

pos0329 = df_tweets.loc[df_tweets.date=='2020-03-29'][df_tweets.polarity > 0].shape[0]
neg0329 = df_tweets.loc[df_tweets.date=='2020-03-29'][df_tweets.polarity < 0].shape[0]
neutral0329 = df_tweets.loc[df_tweets.date=='2020-03-29'][df_tweets.polarity == 0].shape[0]
nonZero0329 = pos0329 + neg0329
```

```

px0329 = round(((pos0329-neg0329)/nonZero0329),3)
volume0329 = nonZero0329 + neutral0329

pos0328 = df_tweets.loc[df_tweets.date=='2020-03-28'][df_tweets.polarity > 0].shape[0]
neg0328 = df_tweets.loc[df_tweets.date=='2020-03-28'][df_tweets.polarity < 0].shape[0]
neutral0328 = df_tweets.loc[df_tweets.date=='2020-03-28'][df_tweets.polarity == 0].shape[0]
nonZero0328 = pos0328 + neg0328
px0328 = round(((pos0328-neg0328)/nonZero0328),3)
volume0328 = nonZero0328 + neutral0328
'''

pos0327 = df_tweets.loc[df_tweets.date=='2020-03-27'][df_tweets.polarity > 0].shape[0]
neg0327 = df_tweets.loc[df_tweets.date=='2020-03-27'][df_tweets.polarity < 0].shape[0]
neutral0327 = df_tweets.loc[df_tweets.date=='2020-03-27'][df_tweets.polarity == 0].shape[0]
nonZero0327 = pos0327 + neg0327
px0327 = round(((pos0327-neg0327)/nonZero0327),3)
volume0327 = nonZero0327 + neutral0327

pos0326 = df_tweets.loc[df_tweets.date=='2020-03-26'][df_tweets.polarity > 0].shape[0]
neg0326 = df_tweets.loc[df_tweets.date=='2020-03-26'][df_tweets.polarity < 0].shape[0]
neutral0326 = df_tweets.loc[df_tweets.date=='2020-03-26'][df_tweets.polarity == 0].shape[0]
nonZero0326 = pos0326 + neg0326
px0326 = round(((pos0326-neg0326)/nonZero0326),3)
volume0326 = nonZero0326 + neutral0326

pos0325 = df_tweets.loc[df_tweets.date=='2020-03-25'][df_tweets.polarity > 0].shape[0]
neg0325 = df_tweets.loc[df_tweets.date=='2020-03-25'][df_tweets.polarity < 0].shape[0]
neutral0325 = df_tweets.loc[df_tweets.date=='2020-03-25'][df_tweets.polarity == 0].shape[0]
nonZero0325 = pos0325 + neg0325
px0325 = round(((pos0325-neg0325)/nonZero0325),3)
volume0325 = nonZero0325 + neutral0325

pos0324 = df_tweets.loc[df_tweets.date=='2020-03-24'][df_tweets.polarity > 0].shape[0]
neg0324 = df_tweets.loc[df_tweets.date=='2020-03-24'][df_tweets.polarity < 0].shape[0]
neutral0324 = df_tweets.loc[df_tweets.date=='2020-03-24'][df_tweets.polarity == 0].shape[0]
nonZero0324 = pos0324 + neg0324
px0324 = round(((pos0324-neg0324)/nonZero0324),3)
volume0324 = nonZero0324 + neutral0324

pos0323 = df_tweets.loc[df_tweets.date=='2020-03-23'][df_tweets.polarity > 0].shape[0]
neg0323 = df_tweets.loc[df_tweets.date=='2020-03-23'][df_tweets.polarity < 0].shape[0]
neutral0323 = df_tweets.loc[df_tweets.date=='2020-03-23'][df_tweets.polarity == 0].shape[0]
nonZero0323 = pos0323 + neg0323
px0323 = round(((pos0323-neg0323)/nonZero0323),3)
volume0323 = nonZero0323 + neutral0323

'''
pos0322 = df_tweets.loc[df_tweets.date=='2020-03-22'][df_tweets.polarity > 0].shape[0]
neg0322 = df_tweets.loc[df_tweets.date=='2020-03-22'][df_tweets.polarity < 0].shape[0]
neutral0322 = df_tweets.loc[df_tweets.date=='2020-03-22'][df_tweets.polarity == 0].shape[0]
nonZero0322 = pos0322 + neg0322
px0322 = round(((pos0322-neg0322)/nonZero0322),3)
volume0322 = nonZero0322 + neutral0322

pos0321 = df_tweets.loc[df_tweets.date=='2020-03-21'][df_tweets.polarity > 0].shape[0]
neg0321 = df_tweets.loc[df_tweets.date=='2020-03-21'][df_tweets.polarity < 0].shape[0]
neutral0321 = df_tweets.loc[df_tweets.date=='2020-03-21'][df_tweets.polarity == 0].shape[0]
nonZero0321 = pos0321 + neg0321
px0321 = round(((pos0321-neg0321)/nonZero0321),3)
volume0321 = nonZero0321 + neutral0321
'''

pos0320 = df_tweets.loc[df_tweets.date=='2020-03-20'][df_tweets.polarity > 0].shape[0]
neg0320 = df_tweets.loc[df_tweets.date=='2020-03-20'][df_tweets.polarity < 0].shape[0]
neutral0320 = df_tweets.loc[df_tweets.date=='2020-03-20'][df_tweets.polarity == 0].shape[0]
nonZero0320 = pos0320 + neg0320
px0320 = round(((pos0320-neg0320)/nonZero0320),3)
volume0320 = nonZero0320 + neutral0320

pos0319 = df_tweets.loc[df_tweets.date=='2020-03-19'][df_tweets.polarity > 0].shape[0]
neg0319 = df_tweets.loc[df_tweets.date=='2020-03-19'][df_tweets.polarity < 0].shape[0]
neutral0319 = df_tweets.loc[df_tweets.date=='2020-03-19'][df_tweets.polarity == 0].shape[0]
nonZero0319 = pos0319 + neg0319
px0319 = round(((pos0319-neg0319)/nonZero0319),3)
volume0319 = nonZero0319 + neutral0319

```

```

pos0318 = df_tweets.loc[df_tweets.date=='2020-03-18'][df_tweets.polarity > 0].shape[0]
neg0318 = df_tweets.loc[df_tweets.date=='2020-03-18'][df_tweets.polarity < 0].shape[0]
neutral0318 = df_tweets.loc[df_tweets.date=='2020-03-18'][df_tweets.polarity == 0].shape[0]
nonZero0318 = pos0318 + neg0318
px0318 = round(((pos0318-neg0318)/nonZero0318),3)
volume0318 = nonZero0318 + neutral0318

pos0317 = df_tweets.loc[df_tweets.date=='2020-03-17'][df_tweets.polarity > 0].shape[0]
neg0317 = df_tweets.loc[df_tweets.date=='2020-03-17'][df_tweets.polarity < 0].shape[0]
neutral0317 = df_tweets.loc[df_tweets.date=='2020-03-17'][df_tweets.polarity == 0].shape[0]
nonZero0317 = pos0317 + neg0317
px0317 = round(((pos0317-neg0317)/nonZero0317),3)
volume0317 = nonZero0317 + neutral0317

pos0316 = df_tweets.loc[df_tweets.date=='2020-03-16'][df_tweets.polarity > 0].shape[0]
neg0316 = df_tweets.loc[df_tweets.date=='2020-03-16'][df_tweets.polarity < 0].shape[0]
neutral0316 = df_tweets.loc[df_tweets.date=='2020-03-16'][df_tweets.polarity == 0].shape[0]
nonZero0316 = pos0316 + neg0316
px0316 = round(((pos0316-neg0316)/nonZero0316),3)
volume0316 = nonZero0316 + neutral0316

'''
pos0315 = df_tweets.loc[df_tweets.date=='2020-03-15'][df_tweets.polarity > 0].shape[0]
neg0315 = df_tweets.loc[df_tweets.date=='2020-03-15'][df_tweets.polarity < 0].shape[0]
neutral0315 = df_tweets.loc[df_tweets.date=='2020-03-15'][df_tweets.polarity == 0].shape[0]
nonZero0315 = pos0315 + neg0315
px0315 = round(((pos0315-neg0315)/nonZero0315),3)
volume0315 = nonZero0315 + neutral0315

pos0314 = df_tweets.loc[df_tweets.date=='2020-03-14'][df_tweets.polarity > 0].shape[0]
neg0314 = df_tweets.loc[df_tweets.date=='2020-03-14'][df_tweets.polarity < 0].shape[0]
neutral0314 = df_tweets.loc[df_tweets.date=='2020-03-14'][df_tweets.polarity == 0].shape[0]
nonZero0314 = pos0314 + neg0314
px0314 = round(((pos0314-neg0314)/nonZero0314),3)
volume0314 = nonZero0314 + neutral0314
'''

pos0313 = df_tweets.loc[df_tweets.date=='2020-03-13'][df_tweets.polarity > 0].shape[0]
neg0313 = df_tweets.loc[df_tweets.date=='2020-03-13'][df_tweets.polarity < 0].shape[0]
neutral0313 = df_tweets.loc[df_tweets.date=='2020-03-13'][df_tweets.polarity == 0].shape[0]
nonZero0313 = pos0313 + neg0313
px0313 = round(((pos0313-neg0313)/nonZero0313),3)
volume0313 = nonZero0313 + neutral0313

pos0312 = df_tweets.loc[df_tweets.date=='2020-03-12'][df_tweets.polarity > 0].shape[0]
neg0312 = df_tweets.loc[df_tweets.date=='2020-03-12'][df_tweets.polarity < 0].shape[0]
neutral0312 = df_tweets.loc[df_tweets.date=='2020-03-12'][df_tweets.polarity == 0].shape[0]
nonZero0312 = pos0312 + neg0312
px0312 = round(((pos0312-neg0312)/nonZero0312),3)
volume0312 = nonZero0312 + neutral0312

pos0311 = df_tweets.loc[df_tweets.date=='2020-03-11'][df_tweets.polarity > 0].shape[0]
neg0311 = df_tweets.loc[df_tweets.date=='2020-03-11'][df_tweets.polarity < 0].shape[0]
neutral0311 = df_tweets.loc[df_tweets.date=='2020-03-11'][df_tweets.polarity == 0].shape[0]
nonZero0311 = pos0311 + neg0311
px0311 = round(((pos0311-neg0311)/nonZero0311),3)
volume0311 = nonZero0311 + neutral0311

pos0310 = df_tweets.loc[df_tweets.date=='2020-03-10'][df_tweets.polarity > 0].shape[0]
neg0310 = df_tweets.loc[df_tweets.date=='2020-03-10'][df_tweets.polarity < 0].shape[0]
neutral0310 = df_tweets.loc[df_tweets.date=='2020-03-10'][df_tweets.polarity == 0].shape[0]
nonZero0310 = pos0310 + neg0310
px0310 = round(((pos0310-neg0310)/nonZero0310),3)
volume0310 = nonZero0310 + neutral0310

pos0309 = df_tweets.loc[df_tweets.date=='2020-03-09'][df_tweets.polarity > 0].shape[0]
neg0309 = df_tweets.loc[df_tweets.date=='2020-03-09'][df_tweets.polarity < 0].shape[0]
neutral0309 = df_tweets.loc[df_tweets.date=='2020-03-09'][df_tweets.polarity == 0].shape[0]
nonZero0309 = pos0309 + neg0309
px0309 = round(((pos0309-neg0309)/nonZero0309),3)
volume0309 = nonZero0309 + neutral0309

'''
pos0308 = df_tweets.loc[df_tweets.date=='2020-03-08'][df_tweets.polarity > 0].shape[0]
neg0308 = df_tweets.loc[df_tweets.date=='2020-03-08'][df_tweets.polarity < 0].shape[0]
neutral0308 = df_tweets.loc[df_tweets.date=='2020-03-08'][df_tweets.polarity == 0].shape[0]
nonZero0308 = pos0308 + neg0308

```

```

px0308 = round(((pos0308-neg0308)/nonZero0308),3)
volume0308 = nonZero0308 + neutral0308

pos0307 = df_tweets.loc[df_tweets.date=='2020-03-07'][df_tweets.polarity > 0].shape[0]
neg0307 = df_tweets.loc[df_tweets.date=='2020-03-07'][df_tweets.polarity < 0].shape[0]
neutral0307 = df_tweets.loc[df_tweets.date=='2020-03-07'][df_tweets.polarity == 0].shape[0]
nonZero0307 = pos0307 + neg0307
px0307 = round(((pos0307-neg0307)/nonZero0307),3)
volume0307 = nonZero0307 + neutral0307
'''

pos0306 = df_tweets.loc[df_tweets.date=='2020-03-06'][df_tweets.polarity > 0].shape[0]
neg0306 = df_tweets.loc[df_tweets.date=='2020-03-06'][df_tweets.polarity < 0].shape[0]
neutral0306 = df_tweets.loc[df_tweets.date=='2020-03-06'][df_tweets.polarity == 0].shape[0]
nonZero0306 = pos0306 + neg0306
px0306 = round(((pos0306-neg0306)/nonZero0306),3)
volume0306 = nonZero0306 + neutral0306

pos0305 = df_tweets.loc[df_tweets.date=='2020-03-05'][df_tweets.polarity > 0].shape[0]
neg0305 = df_tweets.loc[df_tweets.date=='2020-03-05'][df_tweets.polarity < 0].shape[0]
neutral0305 = df_tweets.loc[df_tweets.date=='2020-03-05'][df_tweets.polarity == 0].shape[0]
nonZero0305 = pos0305 + neg0305
px0305 = round(((pos0305-neg0305)/nonZero0305),3)
volume0305 = nonZero0305 + neutral0305

pos0304 = df_tweets.loc[df_tweets.date=='2020-03-04'][df_tweets.polarity > 0].shape[0]
neg0304 = df_tweets.loc[df_tweets.date=='2020-03-04'][df_tweets.polarity < 0].shape[0]
neutral0304 = df_tweets.loc[df_tweets.date=='2020-03-04'][df_tweets.polarity == 0].shape[0]
nonZero0304 = pos0304 + neg0304
px0304 = round(((pos0304-neg0304)/nonZero0304),3)
volume0304 = nonZero0304 + neutral0304

pos0303 = df_tweets.loc[df_tweets.date=='2020-03-03'][df_tweets.polarity > 0].shape[0]
neg0303 = df_tweets.loc[df_tweets.date=='2020-03-03'][df_tweets.polarity < 0].shape[0]
neutral0303 = df_tweets.loc[df_tweets.date=='2020-03-03'][df_tweets.polarity == 0].shape[0]
nonZero0303 = pos0303 + neg0303
px0303 = round(((pos0303-neg0303)/nonZero0303),3)
volume0303 = nonZero0303 + neutral0303

pos0302 = df_tweets.loc[df_tweets.date=='2020-03-02'][df_tweets.polarity > 0].shape[0]
neg0302 = df_tweets.loc[df_tweets.date=='2020-03-02'][df_tweets.polarity < 0].shape[0]
neutral0302 = df_tweets.loc[df_tweets.date=='2020-03-02'][df_tweets.polarity == 0].shape[0]
nonZero0302 = pos0302 + neg0302
px0302 = round(((pos0302-neg0302)/nonZero0302),3)
volume0302 = nonZero0302 + neutral0302

'''

pos0301 = df_tweets.loc[df_tweets.date=='2020-03-01'][df_tweets.polarity > 0].shape[0]
neg0301 = df_tweets.loc[df_tweets.date=='2020-03-01'][df_tweets.polarity < 0].shape[0]
neutral0301 = df_tweets.loc[df_tweets.date=='2020-03-01'][df_tweets.polarity == 0].shape[0]
nonZero0301 = pos0301 + neg0301
px0301 = round(((pos0301-neg0301)/nonZero0301),3)
volume0301 = nonZero0301 + neutral0301

pos0229 = df_tweets.loc[df_tweets.date=='2020-02-29'][df_tweets.polarity > 0].shape[0]
neg0229 = df_tweets.loc[df_tweets.date=='2020-02-29'][df_tweets.polarity < 0].shape[0]
neutral0229 = df_tweets.loc[df_tweets.date=='2020-02-29'][df_tweets.polarity == 0].shape[0]
nonZero0229 = pos0229 + neg0229
px0229 = round(((pos0229-neg0229)/nonZero0229),3)
volume0229 = nonZero0229 + neutral0229
'''

pos0228 = df_tweets.loc[df_tweets.date=='2020-02-28'][df_tweets.polarity > 0].shape[0]
neg0228 = df_tweets.loc[df_tweets.date=='2020-02-28'][df_tweets.polarity < 0].shape[0]
neutral0228 = df_tweets.loc[df_tweets.date=='2020-02-28'][df_tweets.polarity == 0].shape[0]
nonZero0228 = pos0228 + neg0228
px0228 = round(((pos0228-neg0228)/nonZero0228),3)
volume0228 = nonZero0228 + neutral0228

pos0227 = df_tweets.loc[df_tweets.date=='2020-02-27'][df_tweets.polarity > 0].shape[0]
neg0227 = df_tweets.loc[df_tweets.date=='2020-02-27'][df_tweets.polarity < 0].shape[0]
neutral0227 = df_tweets.loc[df_tweets.date=='2020-02-27'][df_tweets.polarity == 0].shape[0]
nonZero0227 = pos0227 + neg0227
px0227 = round(((pos0227-neg0227)/nonZero0227),3)
volume0227 = nonZero0227 + neutral0227

pos0226 = df_tweets.loc[df_tweets.date=='2020-02-26'][df_tweets.polarity > 0].shape[0]

```

```

neg0226 = df_tweets.loc[df_tweets.date=='2020-02-26'][df_tweets.polarity < 0].shape[0]
neutral0226 = df_tweets.loc[df_tweets.date=='2020-02-26'][df_tweets.polarity == 0].shape[0]
nonZero0226 = pos0226 + neg0226
px0226 = round(((pos0226-neg0226)/nonZero0226), 3)
volume0226 = nonZero0226 + neutral0226

pos0225 = df_tweets.loc[df_tweets.date=='2020-02-25'][df_tweets.polarity > 0].shape[0]
neg0225 = df_tweets.loc[df_tweets.date=='2020-02-25'][df_tweets.polarity < 0].shape[0]
neutral0225 = df_tweets.loc[df_tweets.date=='2020-02-25'][df_tweets.polarity == 0].shape[0]
nonZero0225 = pos0225 + neg0225
px0225 = round(((pos0225-neg0225)/nonZero0225), 3)
volume0225 = nonZero0225 + neutral0225

pos0224 = df_tweets.loc[df_tweets.date=='2020-02-24'][df_tweets.polarity > 0].shape[0]
neg0224 = df_tweets.loc[df_tweets.date=='2020-02-24'][df_tweets.polarity < 0].shape[0]
neutral0224 = df_tweets.loc[df_tweets.date=='2020-02-24'][df_tweets.polarity == 0].shape[0]
nonZero0224 = pos0224 + neg0224
px0224 = round(((pos0224-neg0224)/nonZero0224), 3)
volume0224 = nonZero0224 + neutral0224

'''
pos0223 = df_tweets.loc[df_tweets.date=='2020-02-23'][df_tweets.polarity > 0].shape[0]
neg0223 = df_tweets.loc[df_tweets.date=='2020-02-23'][df_tweets.polarity < 0].shape[0]
neutral0223 = df_tweets.loc[df_tweets.date=='2020-02-23'][df_tweets.polarity == 0].shape[0]
nonZero0223 = pos0223 + neg0223
px0223 = round(((pos0223-neg0223)/nonZero0223), 3)
volume0223 = nonZero0223 + neutral0223

pos0222 = df_tweets.loc[df_tweets.date=='2020-02-22'][df_tweets.polarity > 0].shape[0]
neg0222 = df_tweets.loc[df_tweets.date=='2020-02-22'][df_tweets.polarity < 0].shape[0]
neutral0222 = df_tweets.loc[df_tweets.date=='2020-02-22'][df_tweets.polarity == 0].shape[0]
nonZero0222 = pos0222 + neg0222
px0222 = round(((pos0222-neg0222)/nonZero0222), 3)
volume0222 = nonZero0222 + neutral0222
'''

pos0221 = df_tweets.loc[df_tweets.date=='2020-02-21'][df_tweets.polarity > 0].shape[0]
neg0221 = df_tweets.loc[df_tweets.date=='2020-02-21'][df_tweets.polarity < 0].shape[0]
neutral0221 = df_tweets.loc[df_tweets.date=='2020-02-21'][df_tweets.polarity == 0].shape[0]
nonZero0221 = pos0221 + neg0221
px0221 = round(((pos0221-neg0221)/nonZero0221), 3)
volume0221 = nonZero0221 + neutral0221

pos0220 = df_tweets.loc[df_tweets.date=='2020-02-20'][df_tweets.polarity > 0].shape[0]
neg0220 = df_tweets.loc[df_tweets.date=='2020-02-20'][df_tweets.polarity < 0].shape[0]
neutral0220 = df_tweets.loc[df_tweets.date=='2020-02-20'][df_tweets.polarity == 0].shape[0]
nonZero0220 = pos0220 + neg0220
px0220 = round(((pos0220-neg0220)/nonZero0220), 3)
volume0220 = nonZero0220 + neutral0220

pos0219 = df_tweets.loc[df_tweets.date=='2020-02-19'][df_tweets.polarity > 0].shape[0]
neg0219 = df_tweets.loc[df_tweets.date=='2020-02-19'][df_tweets.polarity < 0].shape[0]
neutral0219 = df_tweets.loc[df_tweets.date=='2020-02-19'][df_tweets.polarity == 0].shape[0]
nonZero0219 = pos0219 + neg0219
px0219 = round(((pos0219-neg0219)/nonZero0219), 3)
volume0219 = nonZero0219 + neutral0219

pos0218 = df_tweets.loc[df_tweets.date=='2020-02-18'][df_tweets.polarity > 0].shape[0]
neg0218 = df_tweets.loc[df_tweets.date=='2020-02-18'][df_tweets.polarity < 0].shape[0]
neutral0218 = df_tweets.loc[df_tweets.date=='2020-02-18'][df_tweets.polarity == 0].shape[0]
nonZero0218 = pos0218 + neg0218
px0218 = round(((pos0218-neg0218)/nonZero0218), 3)
volume0218 = nonZero0218 + neutral0218

'''
pos0217 = df_tweets.loc[df_tweets.date=='2020-02-17'][df_tweets.polarity > 0].shape[0]
neg0217 = df_tweets.loc[df_tweets.date=='2020-02-17'][df_tweets.polarity < 0].shape[0]
neutral0217 = df_tweets.loc[df_tweets.date=='2020-02-17'][df_tweets.polarity == 0].shape[0]
nonZero0217 = pos0217 + neg0217
px0217 = round(((pos0217-neg0217)/nonZero0217), 3)
volume0217 = nonZero0217 + neutral0217

pos0216 = df_tweets.loc[df_tweets.date=='2020-02-16'][df_tweets.polarity > 0].shape[0]
neg0216 = df_tweets.loc[df_tweets.date=='2020-02-16'][df_tweets.polarity < 0].shape[0]
neutral0216 = df_tweets.loc[df_tweets.date=='2020-02-16'][df_tweets.polarity == 0].shape[0]
nonZero0216 = pos0216 + neg0216
px0216 = round(((pos0216-neg0216)/nonZero0216), 3)

```

```

volume0216 = nonZero0216 + neutral0216

pos0215 = df_tweets.loc[df_tweets.date=='2020-02-15'][df_tweets.polarity > 0].shape[0]
neg0215 = df_tweets.loc[df_tweets.date=='2020-02-15'][df_tweets.polarity < 0].shape[0]
neutral0215 = df_tweets.loc[df_tweets.date=='2020-02-15'][df_tweets.polarity == 0].shape[0]
nonZero0215 = pos0215 + neg0215
px0215 = round(((pos0215-neg0215)/nonZero0215),3)
volume0215 = nonZero0215 + neutral0215
'''

pos0214 = df_tweets.loc[df_tweets.date=='2020-02-14'][df_tweets.polarity > 0].shape[0]
neg0214 = df_tweets.loc[df_tweets.date=='2020-02-14'][df_tweets.polarity < 0].shape[0]
neutral0214 = df_tweets.loc[df_tweets.date=='2020-02-14'][df_tweets.polarity == 0].shape[0]
nonZero0214 = pos0214 + neg0214
px0214 = round(((pos0214-neg0214)/nonZero0214),3)
volume0214 = nonZero0214 + neutral0214

pos0213 = df_tweets.loc[df_tweets.date=='2020-02-13'][df_tweets.polarity > 0].shape[0]
neg0213 = df_tweets.loc[df_tweets.date=='2020-02-13'][df_tweets.polarity < 0].shape[0]
neutral0213 = df_tweets.loc[df_tweets.date=='2020-02-13'][df_tweets.polarity == 0].shape[0]
nonZero0213 = pos0213 + neg0213
px0213 = round(((pos0213-neg0213)/nonZero0213),3)
volume0213 = nonZero0213 + neutral0213

pos0212 = df_tweets.loc[df_tweets.date=='2020-02-12'][df_tweets.polarity > 0].shape[0]
neg0212 = df_tweets.loc[df_tweets.date=='2020-02-12'][df_tweets.polarity < 0].shape[0]
neutral0212 = df_tweets.loc[df_tweets.date=='2020-02-12'][df_tweets.polarity == 0].shape[0]
nonZero0212 = pos0212 + neg0212
px0212 = round(((pos0212-neg0212)/nonZero0212),3)
volume0212 = nonZero0212 + neutral0212

pos0211 = df_tweets.loc[df_tweets.date=='2020-02-11'][df_tweets.polarity > 0].shape[0]
neg0211 = df_tweets.loc[df_tweets.date=='2020-02-11'][df_tweets.polarity < 0].shape[0]
neutral0211 = df_tweets.loc[df_tweets.date=='2020-02-11'][df_tweets.polarity == 0].shape[0]
nonZero0211 = pos0211 + neg0211
px0211 = round(((pos0211-neg0211)/nonZero0211),3)
volume0211 = nonZero0211 + neutral0211

pos0210 = df_tweets.loc[df_tweets.date=='2020-02-10'][df_tweets.polarity > 0].shape[0]
neg0210 = df_tweets.loc[df_tweets.date=='2020-02-10'][df_tweets.polarity < 0].shape[0]
neutral0210 = df_tweets.loc[df_tweets.date=='2020-02-10'][df_tweets.polarity == 0].shape[0]
nonZero0210 = pos0210 + neg0210
px0210 = round(((pos0210-neg0210)/nonZero0210),3)
volume0210 = nonZero0210 + neutral0210

'''

pos0209 = df_tweets.loc[df_tweets.date=='2020-02-09'][df_tweets.polarity > 0].shape[0]
neg0209 = df_tweets.loc[df_tweets.date=='2020-02-09'][df_tweets.polarity < 0].shape[0]
neutral0209 = df_tweets.loc[df_tweets.date=='2020-02-09'][df_tweets.polarity == 0].shape[0]
nonZero0209 = pos0209 + neg0209
px0209 = round(((pos0209-neg0209)/nonZero0209),3)
volume0209 = nonZero0209 + neutral0209

pos0208 = df_tweets.loc[df_tweets.date=='2020-02-08'][df_tweets.polarity > 0].shape[0]
neg0208 = df_tweets.loc[df_tweets.date=='2020-02-08'][df_tweets.polarity < 0].shape[0]
neutral0208 = df_tweets.loc[df_tweets.date=='2020-02-08'][df_tweets.polarity == 0].shape[0]
nonZero0208 = pos0208 + neg0208
px0208 = round(((pos0208-neg0208)/nonZero0208),3)
volume0208 = nonZero0208 + neutral0208
'''

pos0207 = df_tweets.loc[df_tweets.date=='2020-02-07'][df_tweets.polarity > 0].shape[0]
neg0207 = df_tweets.loc[df_tweets.date=='2020-02-07'][df_tweets.polarity < 0].shape[0]
neutral0207 = df_tweets.loc[df_tweets.date=='2020-02-07'][df_tweets.polarity == 0].shape[0]
nonZero0207 = pos0207 + neg0207
px0207 = round(((pos0207-neg0207)/nonZero0207),3)
volume0207 = nonZero0207 + neutral0207

pos0206 = df_tweets.loc[df_tweets.date=='2020-02-06'][df_tweets.polarity > 0].shape[0]
neg0206 = df_tweets.loc[df_tweets.date=='2020-02-06'][df_tweets.polarity < 0].shape[0]
neutral0206 = df_tweets.loc[df_tweets.date=='2020-02-06'][df_tweets.polarity == 0].shape[0]
nonZero0206 = pos0206 + neg0206
px0206 = round(((pos0206-neg0206)/nonZero0206),3)
volume0206 = nonZero0206 + neutral0206

pos0205 = df_tweets.loc[df_tweets.date=='2020-02-05'][df_tweets.polarity > 0].shape[0]
neg0205 = df_tweets.loc[df_tweets.date=='2020-02-05'][df_tweets.polarity < 0].shape[0]

```



```

neutral0205 = df_tweets.loc[df_tweets.date=='2020-02-05'][df_tweets.polarity == 0].shape[0]
nonZero0205 = pos0205 + neg0205
px0205 = round(((pos0205-neg0205)/nonZero0205),3)
volume0205 = nonZero0205 + neutral0205

pos0204 = df_tweets.loc[df_tweets.date=='2020-02-04'][df_tweets.polarity > 0].shape[0]
neg0204 = df_tweets.loc[df_tweets.date=='2020-02-04'][df_tweets.polarity < 0].shape[0]
neutral0204 = df_tweets.loc[df_tweets.date=='2020-02-04'][df_tweets.polarity == 0].shape[0]
nonZero0204 = pos0204 + neg0204
px0204 = round(((pos0204-neg0204)/nonZero0204),3)
volume0204 = nonZero0204 + neutral0204

pos0203 = df_tweets.loc[df_tweets.date=='2020-02-03'][df_tweets.polarity > 0].shape[0]
neg0203 = df_tweets.loc[df_tweets.date=='2020-02-03'][df_tweets.polarity < 0].shape[0]
neutral0203 = df_tweets.loc[df_tweets.date=='2020-02-03'][df_tweets.polarity == 0].shape[0]
nonZero0203 = pos0203 + neg0203
px0203 = round(((pos0203-neg0203)/nonZero0203),3)
volume0203 = nonZero0203 + neutral0203

'''
pos0202 = df_tweets.loc[df_tweets.date=='2020-02-02'][df_tweets.polarity > 0].shape[0]
neg0202 = df_tweets.loc[df_tweets.date=='2020-02-02'][df_tweets.polarity < 0].shape[0]
neutral0202 = df_tweets.loc[df_tweets.date=='2020-02-02'][df_tweets.polarity == 0].shape[0]
nonZero0202 = pos0202 + neg0202
px0202 = round(((pos0202-neg0202)/nonZero0202),3)
volume0202 = nonZero0202 + neutral0202

pos0201 = df_tweets.loc[df_tweets.date=='2020-02-01'][df_tweets.polarity > 0].shape[0]
neg0201 = df_tweets.loc[df_tweets.date=='2020-02-01'][df_tweets.polarity < 0].shape[0]
neutral0201 = df_tweets.loc[df_tweets.date=='2020-02-01'][df_tweets.polarity == 0].shape[0]
nonZero0201 = pos0201 + neg0201
px0201 = round(((pos0201-neg0201)/nonZero0201),3)
volume0201 = nonZero0201 + neutral0201
'''

```

```

C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:34: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:35: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:36: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:42: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:43: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:44: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:49: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:50: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:51: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:56: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:57: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:58: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:63: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:64: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:65: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:86: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:87: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:88: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:93: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:94: UserWarning: Boolean Series k
ey will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:95: UserWarning: Boolean Series k

```


[illegible]

```

key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:364: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:369: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:370: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:371: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:392: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:393: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:394: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:399: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:400: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:401: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:406: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:407: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:408: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:413: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:414: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:415: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:420: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:421: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.
C:\Users\timod\anaconda3\lib\site-packages\ipykernel_launcher.py:422: UserWarning: Boolean Series
key will be reindexed to match DataFrame index.

```

Out[2]:

```

"npos0202 = df_tweets.loc[df_tweets.date=='2020-02-02'][df_tweets.polarity > 0].shape[0]\nneg0202
= df_tweets.loc[df_tweets.date=='2020-02-02'][df_tweets.polarity < 0].shape[0]\nneutral0202 =
df_tweets.loc[df_tweets.date=='2020-02-02'][df_tweets.polarity == 0].shape[0]\nnonZero0202 =
pos0202 + neg0202\npx0202 = round(((pos0202-neg0202)/nonZero0202),3)\nvolume0202 = nonZero0202 + n
eutral0202\n\npos0201 = df_tweets.loc[df_tweets.date=='2020-02-01'][df_tweets.polarity >
0].shape[0]\nneg0201 = df_tweets.loc[df_tweets.date=='2020-02-01'][df_tweets.polarity <
0].shape[0]\nneutral0201 = df_tweets.loc[df_tweets.date=='2020-02-01'][df_tweets.polarity ==
0].shape[0]\nnonZero0201 = pos0201 + neg0201\npx0201 = round(((pos0201-
neg0201)/nonZero0201),3)\nvolume0201 = nonZero0201 + neutral0201\n"

```

In [3]:

```

#combining daily returns and daily sentiment polarity into one dataframe

returnSeries = df_price['Change %'].apply(lambda x: float(x.strip('%')))
sentimentSeries = [px0327, px0326, px0325, px0324, px0323,
                   px0320, px0319, px0318, px0317, px0316,
                   px0313, px0312, px0311, px0310, px0309,
                   px0306, px0305, px0304, px0303, px0302,
                   px0228, px0227, px0226, px0225, px0224,
                   px0221, px0220, px0219, px0218,
                   px0214, px0213, px0212, px0211, px0210,
                   px0207, px0206, px0205, px0204, px0203]

ts = pd.DataFrame()
ts['Date'], ts['Daily Return in %'], ts['Sentiment Polarity'] = df_price['Date'], returnSeries, sen
timentSeries
ts = ts.iloc[::1].reset_index(drop=True)
ts

```

Out[3]:

Date Daily Return in % Sentiment Polarity

	Date	Daily Return in %	Sentiment Polarity
	Date	Daily Return in %	Sentiment Polarity
0	Feb 03, 2020	-0.65	1.000
1	Feb 04, 2020	-1.30	0.263
2	Feb 05, 2020	0.12	0.425
3	Feb 06, 2020	0.54	0.410
4	Feb 07, 2020	0.26	0.370
5	Feb 10, 2020	0.26	0.370
6	Feb 11, 2020	-0.34	0.319
7	Feb 12, 2020	-0.08	0.451
8	Feb 13, 2020	0.57	0.474
9	Feb 14, 2020	0.42	0.436
10	Feb 18, 2020	1.28	0.558
11	Feb 19, 2020	0.58	0.492
12	Feb 20, 2020	0.41	0.414
13	Feb 21, 2020	1.50	0.535
14	Feb 24, 2020	0.90	0.409
15	Feb 25, 2020	-1.79	0.252
16	Feb 26, 2020	0.44	0.343
17	Feb 27, 2020	0.02	0.209
18	Feb 28, 2020	-3.65	0.316
19	Mar 02, 2020	0.55	0.433
20	Mar 03, 2020	3.14	0.392
21	Mar 04, 2020	0.18	0.457
22	Mar 05, 2020	2.16	0.512
23	Mar 06, 2020	0.04	0.369
24	Mar 09, 2020	0.17	0.426
25	Mar 10, 2020	-2.11	0.536
26	Mar 11, 2020	-0.36	0.218
27	Mar 12, 2020	-3.99	0.280
28	Mar 13, 2020	-3.05	0.268
29	Mar 16, 2020	-1.14	0.329
30	Mar 17, 2020	1.36	0.416
31	Mar 18, 2020	-1.99	0.414
32	Mar 19, 2020	-1.89	0.287
33	Mar 20, 2020	1.50	0.371
34	Mar 23, 2020	4.42	0.446
35	Mar 24, 2020	4.85	0.474
36	Mar 25, 2020	-1.37	0.384
37	Mar 26, 2020	1.28	0.437
38	Mar 27, 2020	-0.65	0.403

In [4]:

```
fig, axs = plt.subplots(2, sharex=True, sharey=False, figsize=(10,7))
#fig.suptitle('Relevance and Stock Price Over Time')

x = ts['Date']
plt.xticks(rotation=90, fontsize=8)

axs[0].yaxis.grid()
axs[1].yaxis.grid()

axs[0].plot(x, ts['Daily Return in %'],color='#5C6BC0')
axs[1].plot(x, ts['Sentiment Polarity'],color='#673AB7')
```

```

axs[0].set_ylabel("Daily Return in %",fontsize=10)
axs[1].set_ylabel("Sentiment Polarity",fontsize=10)

axs[1].set_xlabel("Date",fontsize=10)
plt.show()

```

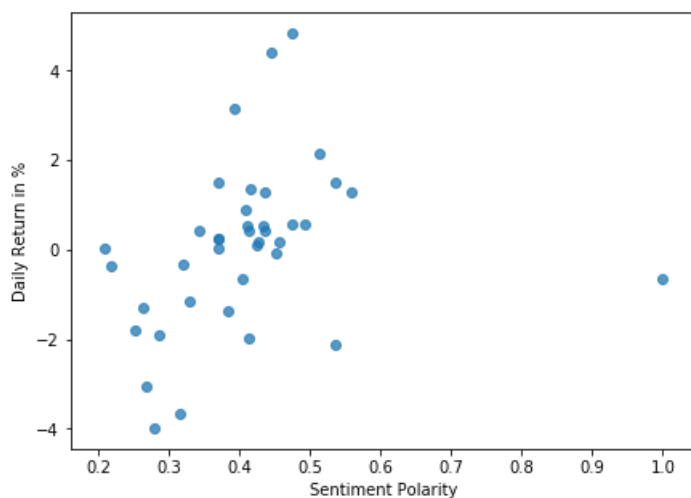


In [5]:

```

plt.figure(figsize=(7,5))
plt.scatter(ts['Sentiment Polarity'],ts['Daily Return in %'], alpha=0.75)
#plt.plot(ts['Sentiment Polarity'][28],ts['Daily Return in %'][34], 'g*')
#plt.plot(ts['Sentiment Polarity'][35],ts['Daily Return in %'][35], 'y*')
#plt.plot(ts['Sentiment Polarity'][29],ts['Daily Return in %'][18], 'r*')
#plt.plot(ts['Sentiment Polarity'][27],ts['Daily Return in %'][27], 'r*')
plt.xlabel("Sentiment Polarity")
plt.ylabel("Daily Return in %")
plt.show()

```



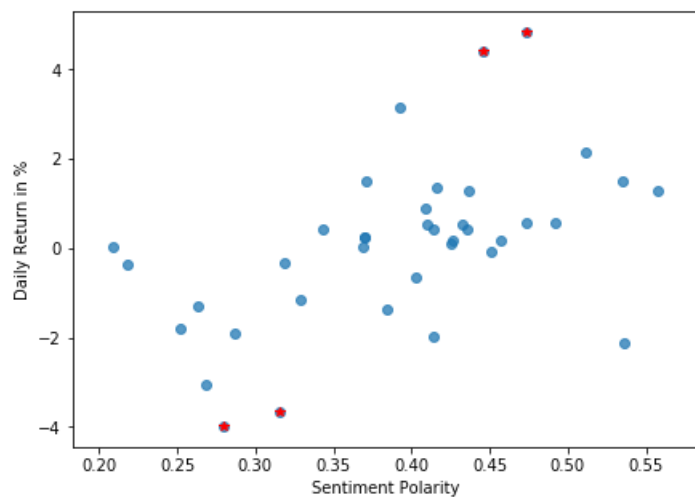
In [8]:

```

plt.figure(figsize=(7,5))
plt.scatter(tsNO['Sentiment Polarity'],tsNO['Daily Return in %'], alpha=0.75)
plt.plot(ts['Sentiment Polarity'][18],ts['Daily Return in %'][18], 'r*')
plt.plot(ts['Sentiment Polarity'][27],ts['Daily Return in %'][27], 'r*')
plt.plot(ts['Sentiment Polarity'][34],ts['Daily Return in %'][34], 'r*')

```

```
plt.plot(ts['Sentiment Polarity'][35],ts['Daily Return in %'][35], 'r*')
plt.xlabel("Sentiment Polarity")
plt.ylabel("Daily Return in %")
plt.show()
```



In [9]:

```
q1 = ts['Daily Return in %'].quantile(0.25)
q3 = ts['Daily Return in %'].quantile(0.75)
iqr = q3 - q1

print(q1-1.5*iqr)
print(q3+1.5*iqr)
```

```
-3.3475
3.1925
```

In [7]:

```
tsNO = pd.DataFrame(ts)
tsNO = tsNO.drop(tsNO.index[[0]])
tsNO
```

Out [7]:

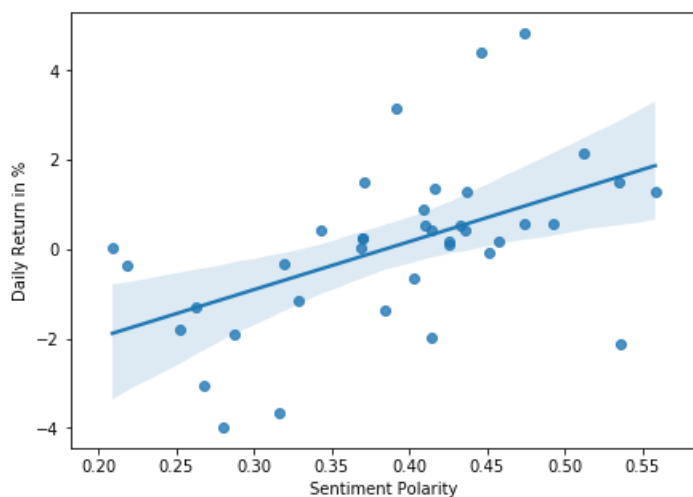
	Date	Daily Return in %	Sentiment Polarity
1	Feb 04, 2020	-1.30	0.263
2	Feb 05, 2020	0.12	0.425
3	Feb 06, 2020	0.54	0.410
4	Feb 07, 2020	0.26	0.370
5	Feb 10, 2020	0.26	0.370
6	Feb 11, 2020	-0.34	0.319
7	Feb 12, 2020	-0.08	0.451
8	Feb 13, 2020	0.57	0.474
9	Feb 14, 2020	0.42	0.436
10	Feb 18, 2020	1.28	0.558
11	Feb 19, 2020	0.58	0.492
12	Feb 20, 2020	0.41	0.414
13	Feb 21, 2020	1.50	0.535
14	Feb 24, 2020	0.90	0.409
15	Feb 25, 2020	-1.79	0.252
16	Feb 26, 2020	0.44	0.343
17	Feb 27, 2020	0.02	0.209
18	Feb 28, 2020	2.65	0.216

10	FEB 20, 2020		-3.00	0.310
	Date	Daily Return in %	Sentiment Polarity	
19	Mar 02, 2020	0.55	0.433	
20	Mar 03, 2020	3.14	0.392	
21	Mar 04, 2020	0.18	0.457	
22	Mar 05, 2020	2.16	0.512	
23	Mar 06, 2020	0.04	0.369	
24	Mar 09, 2020	0.17	0.426	
25	Mar 10, 2020	-2.11	0.536	
26	Mar 11, 2020	-0.36	0.218	
27	Mar 12, 2020	-3.99	0.280	
28	Mar 13, 2020	-3.05	0.268	
29	Mar 16, 2020	-1.14	0.329	
30	Mar 17, 2020	1.36	0.416	
31	Mar 18, 2020	-1.99	0.414	
32	Mar 19, 2020	-1.89	0.287	
33	Mar 20, 2020	1.50	0.371	
34	Mar 23, 2020	4.42	0.446	
35	Mar 24, 2020	4.85	0.474	
36	Mar 25, 2020	-1.37	0.384	
37	Mar 26, 2020	1.28	0.437	
38	Mar 27, 2020	-0.65	0.403	

In [18]:

```
plt.figure(figsize=(7,5))
sns.regplot(data=tsNO, x='Sentiment Polarity', y='Daily Return in %')
print(linregress(tsNO['Sentiment Polarity'],tsNO['Daily Return in %']))
```

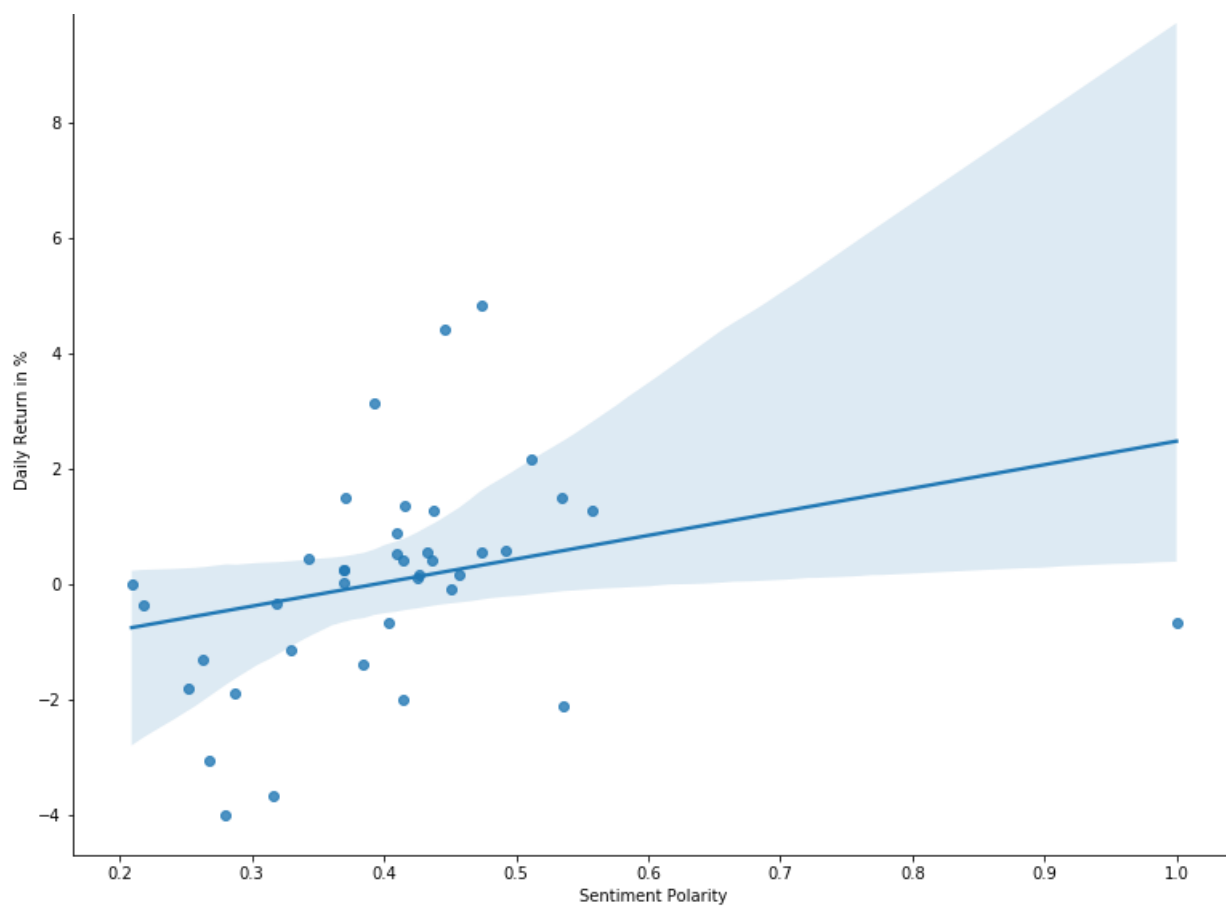
LinregressResult(slope=10.758353998343358, intercept=-4.132577838613667, rvalue=0.5063413475883303, pvalue=0.001180409505985558, stderr=3.0536992485378684)



In [11]:

```
#linear relationship between sentiment polarity and daily returns (%)
plt.figure(figsize=(13,10))
sns.regplot(data=ts, x='Sentiment Polarity', y='Daily Return in %')
print(linregress(ts['Sentiment Polarity'],ts['Daily Return in %']))
```

LinregressResult(slope=4.085972328696994, intercept=-1.5991996944006361, rvalue=0.2885133498240899, pvalue=0.07487355691194837, stderr=2.229238243640697)

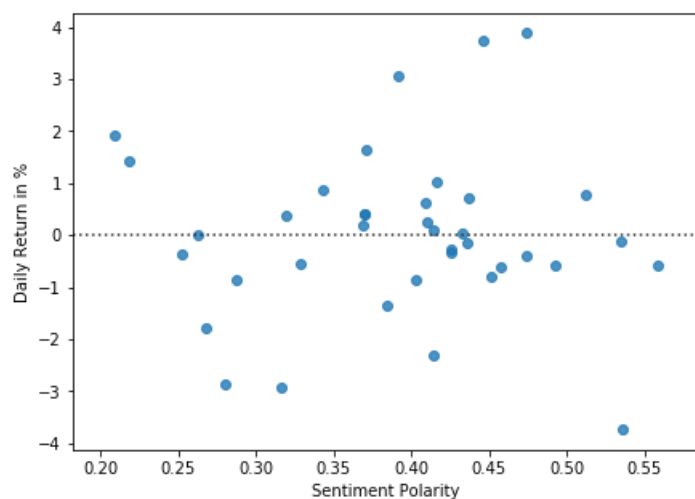


In [17]:

```
plt.figure(figsize=(7,5))
sns.residplot(x='Sentiment Polarity',y='Daily Return in %',data=tsNO)
```

Out[17]:

<matplotlib.axes._subplots.AxesSubplot at 0x23d50462208>



In [15]:

```
tsNO2 = pd.DataFrame(ts)
tsNO2 = tsNO2.drop(tsNO2.index[[18,27,34,35,0]])
tsNO2
```

Out[15]:

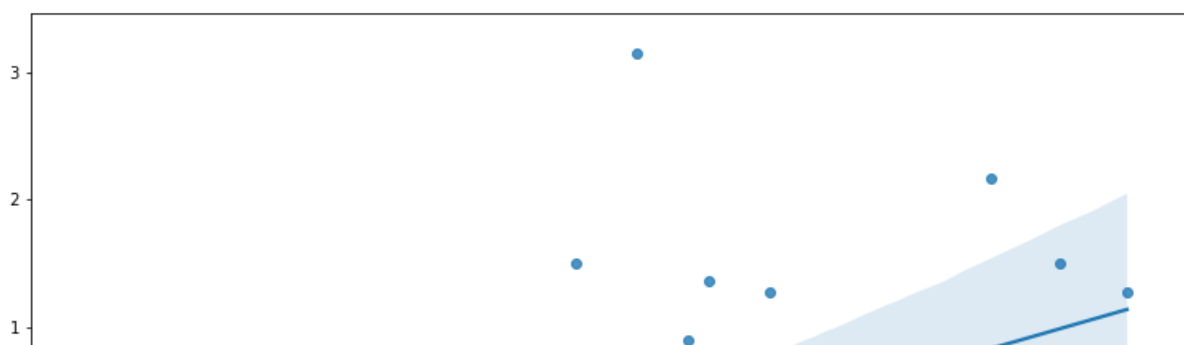
	Date	Daily Return in %	Sentiment Polarity
1	Feb 04. 2020	-1.30	0.263

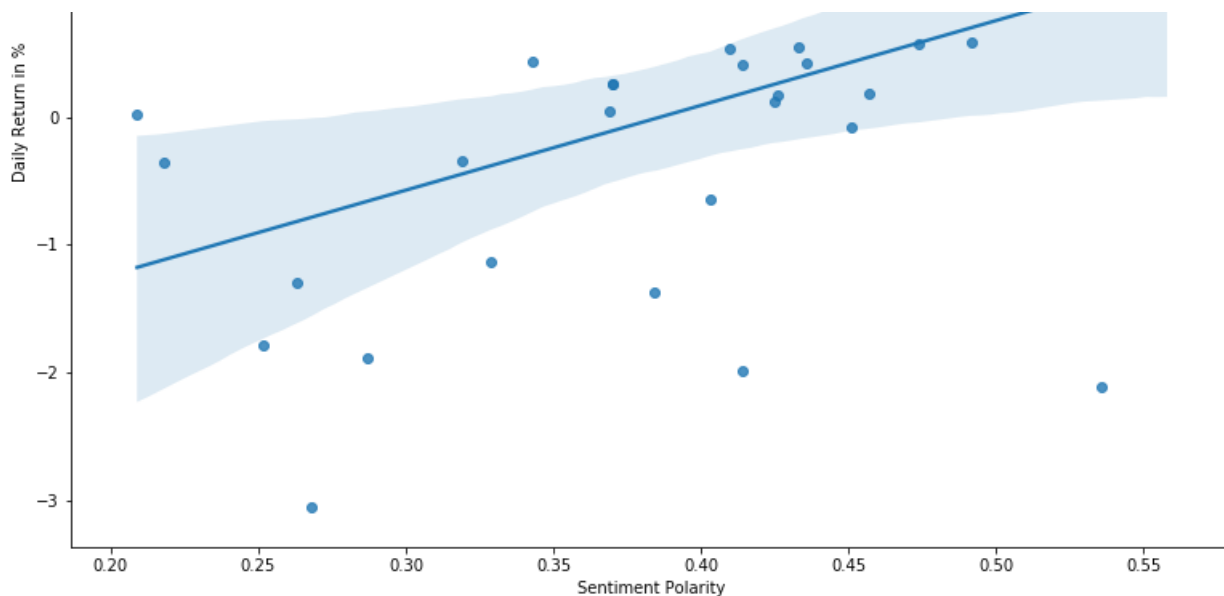
	Date	Daily Return in %	Sentiment Polarity
2	Feb 05, 2020	0.12	0.425
3	Feb 06, 2020	0.54	0.410
4	Feb 07, 2020	0.26	0.370
5	Feb 10, 2020	0.26	0.370
6	Feb 11, 2020	-0.34	0.319
7	Feb 12, 2020	-0.08	0.451
8	Feb 13, 2020	0.57	0.474
9	Feb 14, 2020	0.42	0.436
10	Feb 18, 2020	1.28	0.558
11	Feb 19, 2020	0.58	0.492
12	Feb 20, 2020	0.41	0.414
13	Feb 21, 2020	1.50	0.535
14	Feb 24, 2020	0.90	0.409
15	Feb 25, 2020	-1.79	0.252
16	Feb 26, 2020	0.44	0.343
17	Feb 27, 2020	0.02	0.209
19	Mar 02, 2020	0.55	0.433
20	Mar 03, 2020	3.14	0.392
21	Mar 04, 2020	0.18	0.457
22	Mar 05, 2020	2.16	0.512
23	Mar 06, 2020	0.04	0.369
24	Mar 09, 2020	0.17	0.426
25	Mar 10, 2020	-2.11	0.536
26	Mar 11, 2020	-0.36	0.218
28	Mar 13, 2020	-3.05	0.268
29	Mar 16, 2020	-1.14	0.329
30	Mar 17, 2020	1.36	0.416
31	Mar 18, 2020	-1.99	0.414
32	Mar 19, 2020	-1.89	0.287
33	Mar 20, 2020	1.50	0.371
36	Mar 25, 2020	-1.37	0.384
37	Mar 26, 2020	1.28	0.437
38	Mar 27, 2020	-0.65	0.403

In [16]:

```
#linear relationship between sentiment polarity and daily returns (%)
plt.figure(figsize=(13,10))
sns.regplot(data=tsNO2, x='Sentiment Polarity', y='Daily Return in %')
print(linregress(tsNO2['Sentiment Polarity'],tsNO2['Daily Return in %']))
```

```
LinregressResult(slope=6.640846757905835, intercept=-2.56640621512635, rvalue=0.4468737504287383,
pvalue=0.00806038470507915, stderr=2.350125626456018)
```





In []:

In []:

In [19]:

```
gDelta = pd.read_excel(r'C:\Users\timod\Desktop\GLD_gDelta_all.xlsx')
gDelta['Daily Return in %'] = ts['Daily Return in %']

gDelta
```

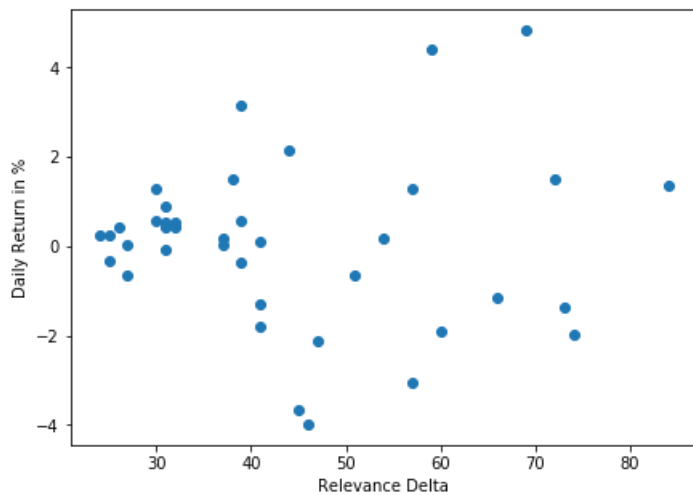
Out[19]:

	Date	Buy Gold	Sell Gold	Delta	Daily Return in %
0	2020-02-03	47	20	27	-0.65
1	2020-02-04	60	19	41	-1.30
2	2020-02-05	56	15	41	0.12
3	2020-02-06	47	16	31	0.54
4	2020-02-07	41	17	24	0.26
5	2020-02-10	46	21	25	0.26
6	2020-02-11	42	17	25	-0.34
7	2020-02-12	47	16	31	-0.08
8	2020-02-13	55	16	39	0.57
9	2020-02-14	50	24	26	0.42
10	2020-02-18	50	20	30	1.28
11	2020-02-19	54	24	30	0.58
12	2020-02-20	55	24	31	0.41
13	2020-02-21	60	22	38	1.50
14	2020-02-24	56	25	31	0.90
15	2020-02-25	66	25	41	-1.79
16	2020-02-26	53	21	32	0.44
17	2020-02-27	62	25	37	0.02
18	2020-02-28	67	22	45	-3.65
19	2020-03-02	53	21	32	0.55

20	2020-03-03	Buy	Sell	Delta	Daily Return in %
21	2020-03-04	59	20	39	3.14
22	2020-03-05	63	19	44	2.16
23	2020-03-06	51	24	27	0.04
24	2020-03-09	75	21	54	0.17
25	2020-03-10	68	21	47	-2.11
26	2020-03-11	54	15	39	-0.36
27	2020-03-12	70	24	46	-3.99
28	2020-03-13	77	20	57	-3.05
29	2020-03-16	84	18	66	-1.14
30	2020-03-17	100	16	84	1.36
31	2020-03-18	90	16	74	-1.99
32	2020-03-19	81	21	60	-1.89
33	2020-03-20	87	15	72	1.50
34	2020-03-23	80	21	59	4.42
35	2020-03-24	87	18	69	4.85
36	2020-03-25	91	18	73	-1.37
37	2020-03-26	74	17	57	1.28
38	2020-03-27	66	15	51	-0.65

In [20]:

```
plt.figure(figsize=(7,5))
plt.scatter(x='Delta',y='Daily Return in %',data=gDelta)
plt.xlabel("Relevance Delta")
plt.ylabel("Daily Return in %")
plt.show()
```



In [13]:

```
gt = pd.read_excel(r'C:\Users\timod\Desktop\Gold_GoogleTrends.xlsx')
gt['Daily Return in %'] = ts['Daily Return in %'].abs()
gt = gt.drop(gt.index[[0]])
gt
```

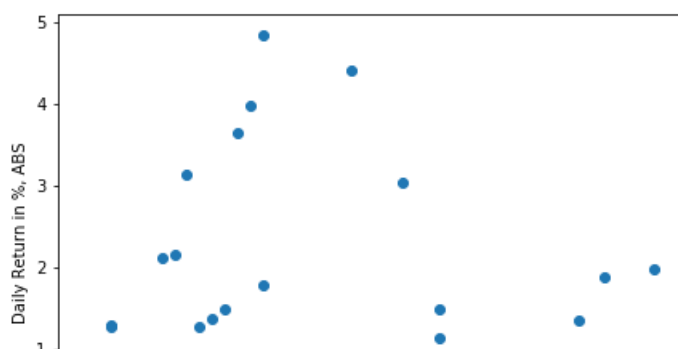
Out[13]:

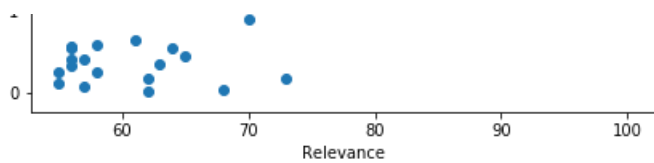
	Date	Gold	Daily Return in %
1	2020-02-04	57	1.30
2	2020-02-05	55	0.12
3	2020-02-06	56	0.54
4	2020-02-07	58	0.26

	Date	Gold	Daily Return in %
5	2020-02-10	55	0.26
6	2020-02-11	56	0.34
7	2020-02-12	57	0.08
8	2020-02-13	56	0.57
9	2020-02-14	56	0.42
10	2020-02-18	57	1.28
11	2020-02-19	58	0.58
12	2020-02-20	57	0.41
13	2020-02-21	66	1.50
14	2020-02-24	70	0.90
15	2020-02-25	69	1.79
16	2020-02-26	65	0.44
17	2020-02-27	62	0.02
18	2020-02-28	67	3.65
19	2020-03-02	64	0.55
20	2020-03-03	63	3.14
21	2020-03-04	62	0.18
22	2020-03-05	62	2.16
23	2020-03-06	68	0.04
24	2020-03-09	73	0.17
25	2020-03-10	61	2.11
26	2020-03-11	63	0.36
27	2020-03-12	68	3.99
28	2020-03-13	80	3.05
29	2020-03-16	83	1.14
30	2020-03-17	94	1.36
31	2020-03-18	100	1.99
32	2020-03-19	96	1.89
33	2020-03-20	83	1.50
34	2020-03-23	76	4.42
35	2020-03-24	69	4.85
36	2020-03-25	65	1.37
37	2020-03-26	64	1.28
38	2020-03-27	61	0.65

In [14]:

```
plt.figure(figsize=(7,5))
plt.scatter(x='Gold',y='Daily Return in %',data=gt)
plt.xlabel('Relevance')
plt.ylabel('Daily Return in %, ABS')
plt.show()
print(linregress(gt['Gold'],gt['Daily Return in %']))
```



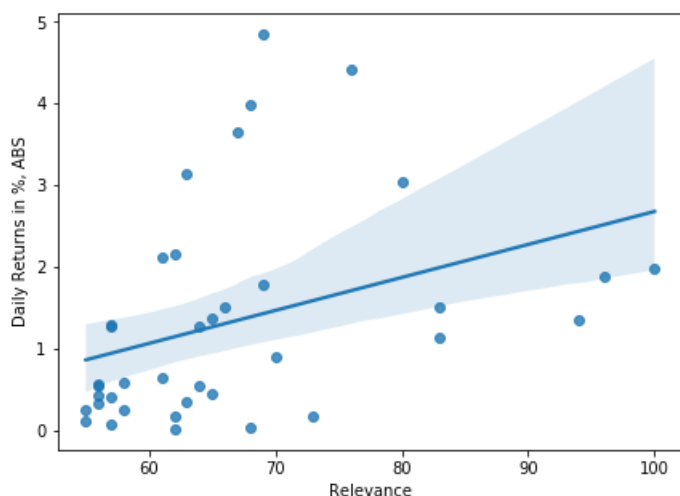


```
LinregressResult(slope=0.040353339075268384, intercept=-1.3556488036468302,
rvalue=0.36347108479010737, pvalue=0.02489190531778809, stderr=0.0172381396096906)
```

In [30]:

```
plt.figure(figsize=(7,5))
sns.regplot(x='Gold',y='Daily Return in %', data=gt)
plt.xlabel('Relevance')
plt.ylabel('Daily Returns in %, ABS')
print(linregress(gt['Gold'],gt['Daily Return in %']))
```

```
LinregressResult(slope=0.040353339075268384, intercept=-1.3556488036468302,
rvalue=0.36347108479010737, pvalue=0.02489190531778809, stderr=0.0172381396096906)
```



In [12]:

```
gt2 = pd.read_excel(r'C:\Users\timod\Desktop\GLD_GoogleTrends.xlsx')
gt2['Daily Return in %'] = ts['Daily Return in %'].abs()
gt2
```

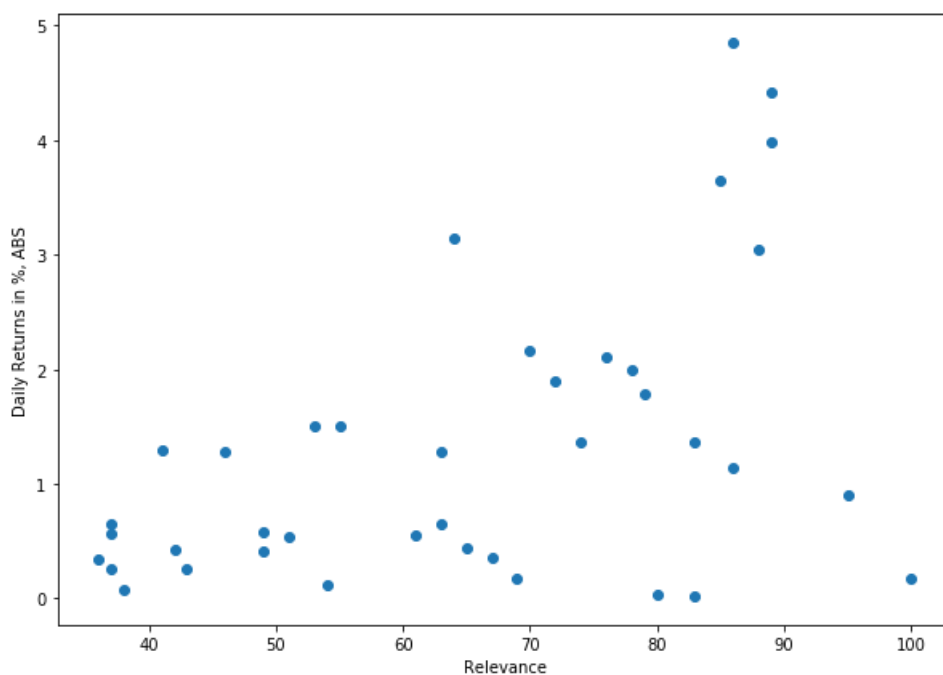
Out[12]:

	Date	GLD	Daily Return in %
0	2020-02-03	37	0.65
1	2020-02-04	41	1.30
2	2020-02-05	54	0.12
3	2020-02-06	51	0.54
4	2020-02-07	43	0.26
5	2020-02-10	37	0.26
6	2020-02-11	36	0.34
7	2020-02-12	38	0.08
8	2020-02-13	37	0.57
9	2020-02-14	42	0.42
10	2020-02-18	46	1.28
11	2020-02-19	49	0.58
12	2020-02-20	49	0.41
13	2020-02-21	53	1.50
14	2020-02-24	95	0.90

15	2020-02-25	79	1.79
16	2020-02-26	65	0.44
17	2020-02-27	83	0.02
18	2020-02-28	85	3.65
19	2020-03-02	61	0.55
20	2020-03-03	64	3.14
21	2020-03-04	69	0.18
22	2020-03-05	70	2.16
23	2020-03-06	80	0.04
24	2020-03-09	100	0.17
25	2020-03-10	76	2.11
26	2020-03-11	67	0.36
27	2020-03-12	89	3.99
28	2020-03-13	88	3.05
29	2020-03-16	86	1.14
30	2020-03-17	74	1.36
31	2020-03-18	78	1.99
32	2020-03-19	72	1.89
33	2020-03-20	55	1.50
34	2020-03-23	89	4.42
35	2020-03-24	86	4.85
36	2020-03-25	83	1.37
37	2020-03-26	63	1.28
38	2020-03-27	63	0.65

In [46]:

```
plt.figure(figsize=(10,7))
plt.scatter(x='GLD',y='Daily Return in %',data=gt2)
plt.xlabel('Relevance')
plt.ylabel('Daily Returns in %, ABS')
plt.show()
print(linregress(gt2['GLD'],gt2['Daily Return in %']))
```



```
LinregressResult(slope=0.03329566634722044, intercept=-0.8468698168592146,
rvalue=0.4021187078527223, pvalue=0.0014610280445052623, stderror=0.00682772420556018)
```

```
rvalue=0.4921167079927225, pvalue=0.0014619360449035632, stderr=0.009662775420356916)
```

In []: