Soc_Net_Data_Analysis

December 8, 2019

[65]:
```
# import relevant packages
from datascience import *
import numpy as np
import pandas as pd
import seaborn as sns
import statsmodels.api as sm
import statsmodels.formula.api as smf
import matplotlib as mp
from dateutil.relativedelta import relativedelta
%matplotlib inline

import matplotlib.dates as mdates
import matplotlib.pyplot as plt
import datetime
```

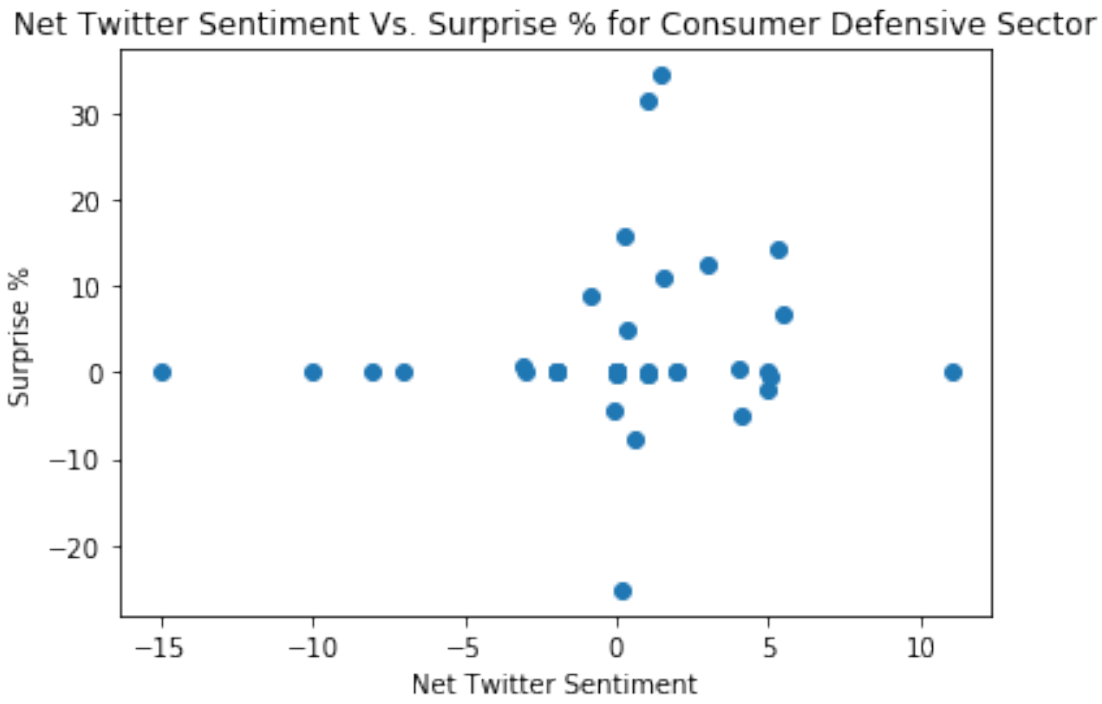
1 Consumer Defensive Sector

[66]:
```
data_cons_def = pd.read_csv('ConsumerDefensive.csv')
```

1.0.1 Consumer Defensive Scatter Plot

[67]:
```
plt.scatter(data_cons_def['Net_Twitter_Sentiment'],␣
 ↪data_cons_def['Surprise_Percent'])
plt.xlabel("Net Twitter Sentiment")
plt.ylabel("Surprise %")
plt.title("Net Twitter Sentiment Vs. Surprise % for Consumer Defensive Sector")
```

[67]:
```
Text(0.5, 1.0, 'Net Twitter Sentiment Vs. Surprise % for Consumer Defensive
Sector')
```



1.0.2 Consumer Defensive Correlation Coefficient

[68]:
```
data_cons_def.corr()
```

[68]:
```
                            Positive Twitter Sentiment  \
Positive Twitter Sentiment                    1.000000
Negative Twitter Sentiment                   -0.767709
Net_Twitter_Sentiment                         0.290984
EPS                                          -0.154315
Surprise_Percent                             -0.072681

                            Negative Twitter Sentiment  Net_Twitter_Sentiment  \
Positive Twitter Sentiment                   -0.767709               0.290984
Negative Twitter Sentiment                    1.000000               0.389678
Net_Twitter_Sentiment                         0.389678               1.000000
EPS                                           0.149162               0.000923
Surprise_Percent                              0.147098               0.115165

                                 EPS  Surprise_Percent
Positive Twitter Sentiment -0.154315         -0.072681
Negative Twitter Sentiment  0.149162          0.147098
Net_Twitter_Sentiment       0.000923          0.115165
EPS                         1.000000          0.118537
Surprise_Percent            0.118537          1.000000
```

Consumer Defense Sector Correlation Coefficient: 0.115165

This means that there's a very weak positive correlation between Net Twitter Sentiment and Surprise % in the Consumer Defense Sector

1.0.3 Consumer Defensive Ordinary Least Squares Analysis

[69]:
```
# Generate regression data
results_cons_def = smf.ols('Surprise_Percent ~ Net_Twitter_Sentiment',␣
 ↪data=data_cons_def).fit()

# Show summary of regression data
results_cons_def.summary()
```

[69]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:       Surprise_Percent   R-squared:                       0.013
Model:                            OLS   Adj. R-squared:                 -0.013
Method:                 Least Squares   F-statistic:                    0.5108
Date:                Sun, 08 Dec 2019   Prob (F-statistic):              0.479
Time:                        17:16:13   Log-Likelihood:                -146.32
No. Observations:                  40   AIC:                             296.6
Df Residuals:                      38   BIC:                             300.0
Df Model:                           1
Covariance Type:            nonrobust
================================================================================
=========
                               coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
---------
Intercept                 2.3990      1.523      1.575      0.124      -0.685
5.483
Net_Twitter_Sentiment     0.2455      0.343      0.715      0.479      -0.450
0.941
==============================================================================
Omnibus:                       19.767   Durbin-Watson:                   0.795
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               43.759
Skew:                           1.134   Prob(JB):                     3.15e-10
Kurtosis:                       7.594   Cond. No.                         4.44
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
"""
```

The coefficient is 0.2455, meaning that as the Net Twitter Sentiment increases by 1, the Surpise % increases by 0.2455% in the Consumer Defensive Sector.

The P value is greater than 0.05, meaning that we reject the alternative hypothesis, that Net Twitter Sentiment has an effect on the Surprise % in the agricultural sector. We accept the Null hypothesis, that Net Twitter Sentiment has an effect on Surprise % in the Consumer Defensive Sector.

The R^2 value is 0.013, meaning that only 1.3% of the variance in the Surprise % can be explained by the Net Twitter Sentiment in the Consumer Defensive Sector.

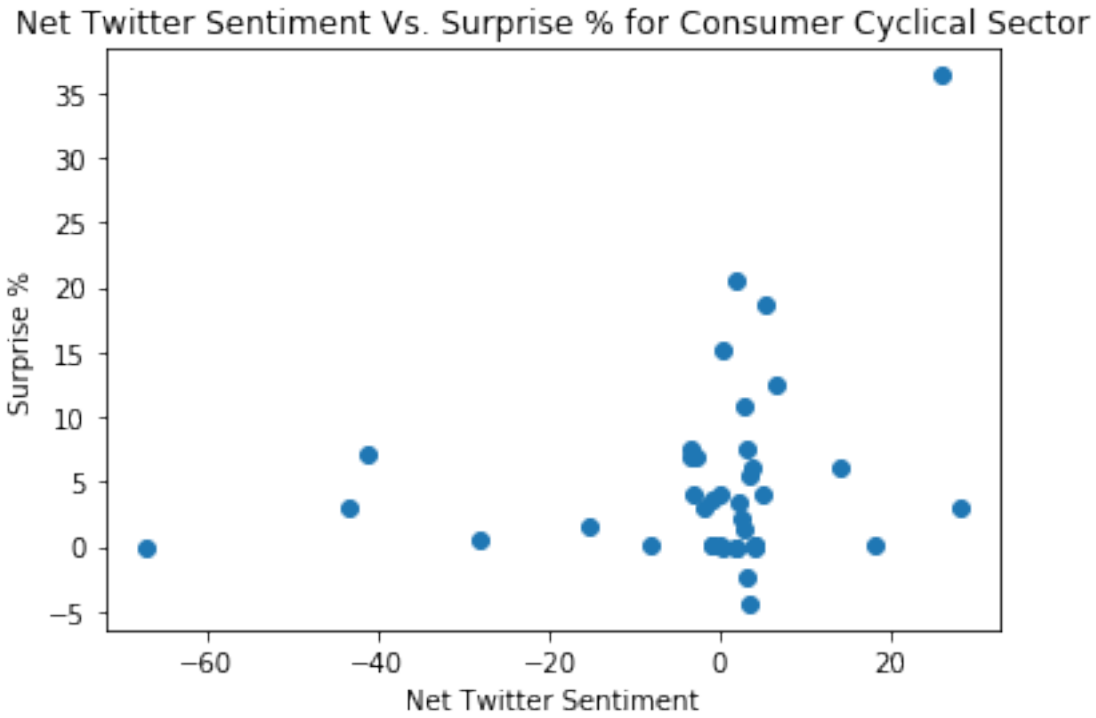
2 Consumer Cyclical Sector

[70]:
```
data_cons_cyc = pd.read_csv('ConsumerCyclical.csv')
```

2.0.1 Consumer Cyclical Scatter Plot

[71]:
```
plt.scatter(data_cons_cyc['Net_Twitter_Sentiment'],␣
 ↪data_cons_cyc['Surprise_Percent'])
plt.xlabel("Net Twitter Sentiment")
plt.ylabel("Surprise %")
plt.title("Net Twitter Sentiment Vs. Surprise % for Consumer Cyclical Sector")
```

[71]:
```
Text(0.5, 1.0, 'Net Twitter Sentiment Vs. Surprise % for Consumer Cyclical
Sector')
```



2.0.2 Consumer Cyclical Sector Correlation Coefficient

[72]:
```
data_cons_cyc.corr()
```

[72]:
```
                            Positive Twitter Sentiment  \
Positive Twitter Sentiment                    1.000000
Negative Twitter Sentiment                   -0.560862
Net_Twitter_Sentiment                        -0.042489
EPS                                           0.036988
Surprise_Percent                              0.004165

                            Negative Twitter Sentiment  Net_Twitter_Sentiment  \
Positive Twitter Sentiment                   -0.560862              -0.042489
Negative Twitter Sentiment                    1.000000               0.850992
Net_Twitter_Sentiment                         0.850992               1.000000
EPS                                           0.169854               0.228439
Surprise_Percent                              0.205911               0.251130

                                 EPS  Surprise_Percent
Positive Twitter Sentiment  0.036988          0.004165
Negative Twitter Sentiment  0.169854          0.205911
Net_Twitter_Sentiment       0.228439          0.251130
EPS                         1.000000          0.355174
Surprise_Percent            0.355174          1.000000
```

Consumer Cyclical Sector Correlation Coefficient: 0.25110

This means theres weak positive correlation between nnet twitter sentiment and surprise % in the Consumer Cyclical sector

2.0.3 Consumer Cyclical Ordinary Least Squares Analysis

[73]:
```
results_cons_cyc = smf.ols('Surprise_Percent ~ Net_Twitter_Sentiment',␣
 ↪data=data_cons_cyc).fit()

# Show summary of regression data
results_cons_cyc.summary()
```

[73]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:       Surprise_Percent   R-squared:                       0.063
Model:                            OLS   Adj. R-squared:                  0.038
Method:                 Least Squares   F-statistic:                     2.558
Date:                Sun, 08 Dec 2019   Prob (F-statistic):              0.118
Time:                        17:16:16   Log-Likelihood:                -135.09
No. Observations:                  40   AIC:                             274.2
Df Residuals:                      38   BIC:                             277.6
Df Model:                           1
```

```
Covariance Type:            nonrobust
==============================================================================
=========
                             coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
---------
Intercept                  5.0685      1.156      4.384      0.000       2.728
7.409
Net_Twitter_Sentiment      0.1093      0.068      1.599      0.118      -0.029
0.248
==============================================================================
Omnibus:                       29.144   Durbin-Watson:                   1.523
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               61.115
Skew:                           1.912   Prob(JB):                     5.36e-14
Kurtosis:                       7.696   Cond. No.                         17.0
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
"""
```

The coefficient is 0.1093, meaning that as the Net Twitter Sentiment increases by 1, the Surpise % increases by 0.1093% in the Consumer Cyclical Sector.

The P value is greater than 0.05, meaning that we reject the alternative hypothesis, that Net Twitter Sentiment has an effect on the Surprise % in the Consumer Cyclical Sector. We accpet the Null hypothesis, that Net Twitter Sentiment has an effect on Surprise % in the Consumer Cyclical Sector.

The R^2 value is 0.063, meaning that 6.3% of the variance in the Surprise % can be explained by the Net Twitter Sentiment in the Consumer Cyclical Sector.

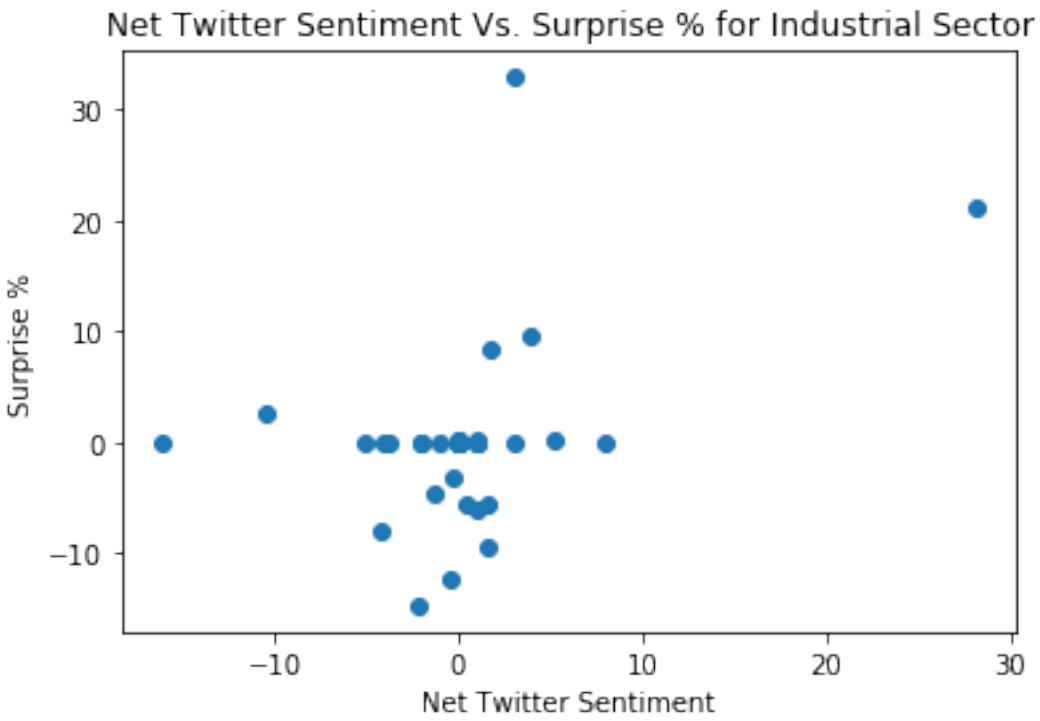
3 Industrial Sector

[74]:
```
data_industrial = pd.read_csv('Industrial.csv')
```

3.0.1 Industrial Scatter Plot

[75]:
```
plt.scatter(data_industrial['Net_Twitter_Sentiment'],␣
 ↪data_industrial['Surprise_Percent'])
plt.xlabel("Net Twitter Sentiment")
plt.ylabel("Surprise %")
plt.title("Net Twitter Sentiment Vs. Surprise % for Industrial Sector")
```

[75]:
```
Text(0.5, 1.0, 'Net Twitter Sentiment Vs. Surprise % for Industrial Sector')
```



3.0.2 Industrial Sector Correlation Coefficient

[76]:
```
data_industrial.corr()
```

[76]:
```
                            Positive Twitter Sentiment  \
Positive Twitter Sentiment                    1.000000
Negative Twitter Sentiment                   -0.332378
Net_Twitter_Sentiment                         0.672986
EPS                                           0.251735
Surprise_Percent                              0.507946

                            Negative Twitter Sentiment  Net_Twitter_Sentiment  \
Positive Twitter Sentiment                   -0.332378               0.672986
Negative Twitter Sentiment                    1.000000               0.473918
Net_Twitter_Sentiment                         0.473918               1.000000
EPS                                          -0.159911               0.109624
Surprise_Percent                             -0.063037               0.424807

                                 EPS  Surprise_Percent
Positive Twitter Sentiment  0.251735          0.507946
Negative Twitter Sentiment -0.159911         -0.063037
Net_Twitter_Sentiment       0.109624          0.424807
EPS                         1.000000          0.003291
Surprise_Percent            0.003291          1.000000
```

Industrial Sector Correlationn Coefficient: 0.424807

This means that there is a medium positive relationship between Net twitter sentiment and Surprise % in the Industrial Sector

3.0.3 Industrial Ordinary Least Squares Analysis

[77]:
```
results_industrial = smf.ols('Surprise_Percent ~ Net_Twitter_Sentiment',␣
 ↪data=data_industrial).fit()

# Show summary of regression data
results_industrial.summary()
```

[77]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:       Surprise_Percent   R-squared:                       0.180
Model:                            OLS   Adj. R-squared:                  0.159
Method:                 Least Squares   F-statistic:                     8.368
Date:                Sun, 08 Dec 2019   Prob (F-statistic):            0.00629
Time:                        17:16:20   Log-Likelihood:                -134.30
No. Observations:                  40   AIC:                             272.6
Df Residuals:                      38   BIC:                             276.0
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
=========
                             coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
---------
Intercept                  0.0427      1.128      0.038      0.970      -2.241
2.327
Net_Twitter_Sentiment      0.5561      0.192      2.893      0.006       0.167
0.945
==============================================================================
Omnibus:                       35.268   Durbin-Watson:                   1.491
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              135.338
Skew:                           1.946   Prob(JB):                     4.09e-30
Kurtosis:                      11.128   Cond. No.                         5.87
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
"""
```

The coefficient is 0.5561, meaning that as the Net Twitter Sentiment increases by 1, the Surpise

% increases by 0.5561% in the Industrial Sector.

The P value is less than 0.05, meaning that we accept the alternative hypothesis, that Net Twitter Sentiment has an effect on the Surprise % in the Industrial Sector. We reject the Null hypothesis, that Net Twitter Sentiment has an effect on Surprise % in the Industrial Sector.

The R^2 value is 0.180, meaning that 1.8% of the variance in the Surprise % can be explained by the Net Twitter Sentiment in the Industrial Sector.

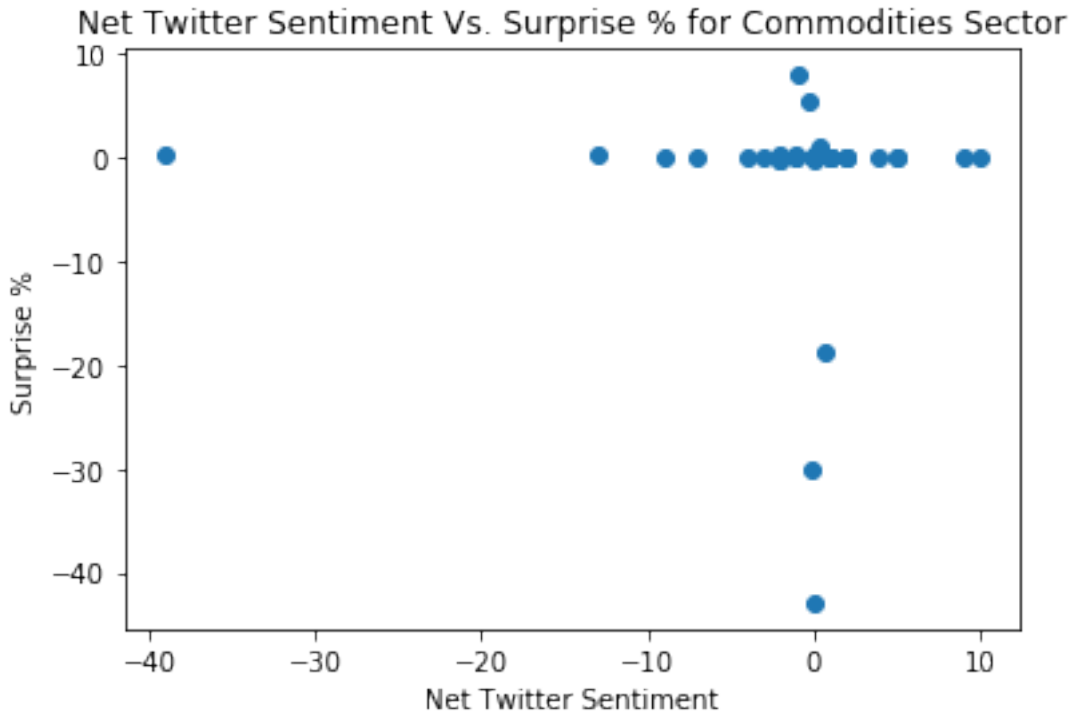
4 Commodities Sector

[78]:
```
data_commodities = pd.read_csv('Commodities.csv')
```

4.0.1 Commodities Scatter Plot

[79]:
```
plt.scatter(data_commodities['Net_Twitter_Sentiment'],␣
 ↪data_commodities['Surprise_Percent'])
plt.xlabel("Net Twitter Sentiment")
plt.ylabel("Surprise %")
plt.title("Net Twitter Sentiment Vs. Surprise % for Commodities Sector")
```

[79]:
```
Text(0.5, 1.0, 'Net Twitter Sentiment Vs. Surprise % for Commodities Sector')
```



4.0.2 Commodities Correlation Coefficient

[80]:
```
data_commodities.corr()
```

[80]:
```
                            Positive Twitter Sentiment  \
Positive Twitter Sentiment                    1.000000
Negative Twitter Sentiment                   -0.102177
Net_Twitter_Sentiment                         0.346322
EPS                                           0.227059
Surprise_Percent                              0.145027

                            Negative Twitter Sentiment  Net_Twitter_Sentiment  \
Positive Twitter Sentiment                   -0.102177               0.346322
Negative Twitter Sentiment                    1.000000               0.897820
Net_Twitter_Sentiment                         0.897820               1.000000
EPS                                          -0.196402              -0.084703
Surprise_Percent                             -0.118235              -0.047301

                                 EPS  Surprise_Percent
Positive Twitter Sentiment  0.227059          0.145027
Negative Twitter Sentiment -0.196402         -0.118235
Net_Twitter_Sentiment      -0.084703         -0.047301
EPS                         1.000000          0.150031
Surprise_Percent            0.150031          1.000000
```

Commodities Correlation Coefficient: -0.046301

This means that there a very weak negative correlation between Net twitter sentiment and surprise % in the Commodities sector

4.0.3 Commodities Ordinary Least Squares Analysis

[81]:
```
results_commodities = smf.ols('Surprise_Percent ~ Net_Twitter_Sentiment',␣
 ↪data=data_commodities).fit()

# Show summary of regression data
results_commodities.summary()
```

[81]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:       Surprise_Percent   R-squared:                       0.002
Model:                            OLS   Adj. R-squared:                 -0.025
Method:                 Least Squares   F-statistic:                    0.08297
Date:                Sun, 08 Dec 2019   Prob (F-statistic):              0.775
Time:                        17:16:23   Log-Likelihood:                -140.24
No. Observations:                  39   AIC:                             284.5
Df Residuals:                      37   BIC:                             287.8
Df Model:                           1
```

```
Covariance Type:            nonrobust
==============================================================================
=========
                             coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
---------
Intercept                -2.0001      1.466     -1.364      0.181      -4.971
0.971
Net_Twitter_Sentiment    -0.0572      0.198     -0.288      0.775      -0.459
0.345
==============================================================================
Omnibus:                       55.784   Durbin-Watson:                   1.593
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              307.982
Skew:                          -3.490   Prob(JB):                     1.33e-67
Kurtosis:                      14.866   Cond. No.                         7.48
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
"""
```

The coefficient is -0.0572, meaning that as the Net Twitter Sentiment increases by 1, the Surpise % decreases by 0.0572% in the Commodities Sector.

The P value is greater than 0.05, meaning that we reject the alternative hypothesis, that Net Twitter Sentiment has an effect on the Surprise % in the Commodities Sector. We accept the Null hypothesis, that Net Twitter Sentiment has an effect on Surprise % in the Commodities Sector.

The R^2 value is 0.002, meaning that 0.2% of the variance in the Surprise % can be explained by the Net Twitter Sentiment in the Commodities Sector.

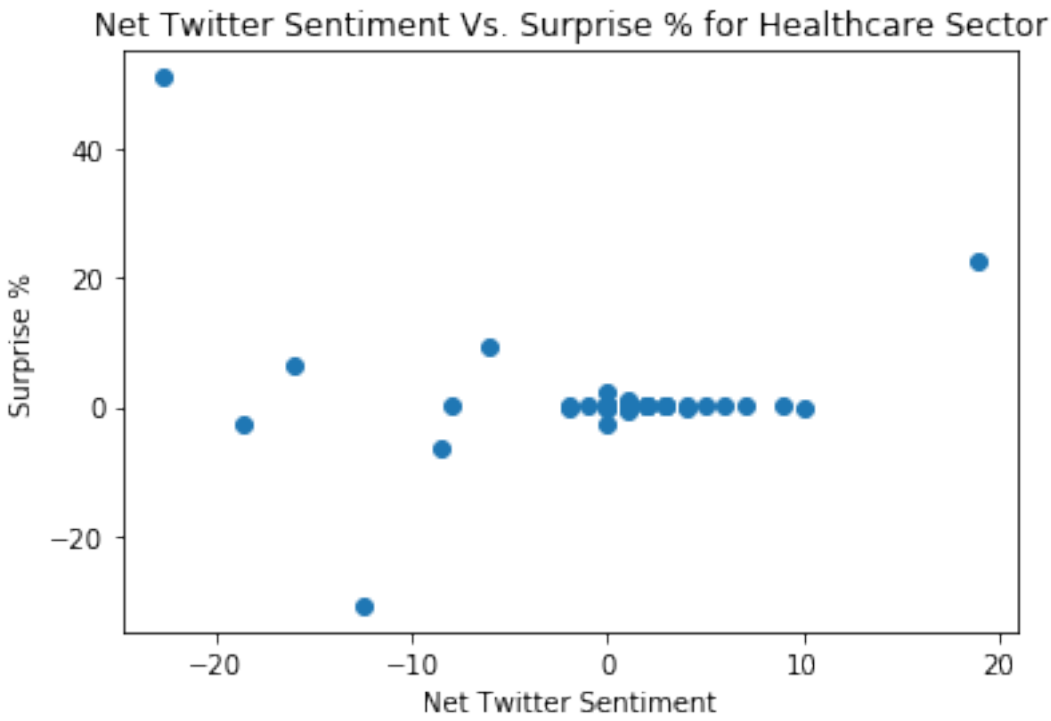
5 Healthcare Sector

[82]:
```
data_healthcare = pd.read_csv('Healthcare.csv')
```

5.0.1 Healthcare Scatterplot

[83]:
```
plt.scatter(data_healthcare['Net_Twitter_Sentiment'],␣
 ↪data_healthcare['Surprise_Percent'])
plt.xlabel("Net Twitter Sentiment")
plt.ylabel("Surprise %")
plt.title("Net Twitter Sentiment Vs. Surprise % for Healthcare Sector")
```

[83]:
```
Text(0.5, 1.0, 'Net Twitter Sentiment Vs. Surprise % for Healthcare Sector')
```



5.0.2 Healthcare Correlation Coefficient

[84]:
```
data_healthcare.corr()
```

[84]:
```
                            Positive Twitter Sentiment  \
Positive Twitter Sentiment                    1.000000
Negative Twitter Sentiment                   -0.693108
Net_Twitter_Sentiment                        -0.086013
EPS                                           0.283415
Surprise_Percent                              0.058537

                            Negative Twitter Sentiment  Net_Twitter_Sentiment  \
Positive Twitter Sentiment                   -0.693108              -0.086013
Negative Twitter Sentiment                    1.000000               0.777779
Net_Twitter_Sentiment                         0.777779               1.000000
EPS                                          -0.162849               0.022046
Surprise_Percent                             -0.128448              -0.126491

                                 EPS  Surprise_Percent
Positive Twitter Sentiment  0.283415          0.058537
Negative Twitter Sentiment -0.162849         -0.128448
Net_Twitter_Sentiment       0.022046         -0.126491
EPS                         1.000000          0.010507
Surprise_Percent            0.010507          1.000000
```

Healthcare Correlation Coefficient: -0.126491

This means that there's a weak negative correlation between Net twitter sentiment and Surprise % in the Healthcare sector

5.0.3 Healthcare Ordinary Least Squares Analysis

[85]:
```
results_healthcare = smf.ols('Surprise_Percent ~ Net_Twitter_Sentiment',␣
↪data=data_healthcare).fit()

# Show summary of regression data
results_healthcare.summary()
```

[85]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:       Surprise_Percent   R-squared:                       0.016
Model:                            OLS   Adj. R-squared:                 -0.011
Method:                 Least Squares   F-statistic:                    0.5854
Date:                Sun, 08 Dec 2019   Prob (F-statistic):              0.449
Time:                        17:16:26   Log-Likelihood:                -142.96
No. Observations:                  38   AIC:                             289.9
Df Residuals:                      36   BIC:                             293.2
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
=========
                             coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
---------
Intercept                  1.3512      1.737      0.778      0.442      -2.172
4.874
Net_Twitter_Sentiment     -0.1756      0.230     -0.765      0.449      -0.641
0.290
==============================================================================
Omnibus:                       33.929   Durbin-Watson:                   1.872
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              201.050
Skew:                           1.619   Prob(JB):                     2.20e-44
Kurtosis:                      13.793   Cond. No.                         7.57
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
"""
```

The coefficient is -0.1756, meaning that as the Net Twitter Sentiment increases by 1, the Surpise

% decreases by 0.1756 in the Healthcare Sector.

The P value is greater than 0.05, meaning that we reject the alternative hypothesis, that Net Twitter Sentiment has an effect on the Surprise % in the Healthcare Sector. We accept the Null hypothesis, that Net Twitter Sentiment has an effect on Surprise % in the Healthcare Sector.

The R^2 value is 0.016, meaning that 1.6% of the variance in the Surprise % can be explained by the Net Twitter Sentiment in the Healthcare Sector.

[ ]:

[ ]: