Text analysis on Data from Twitter using API Calling

Healthcare in US

Project Definition

1. Target

3 stocks in sector of Healthcare in US

2. Period

2 week of December (from 02 Dec 2019 to 16 Dec 2019)

3. Data collection

Using API calling to collect data from Twitter

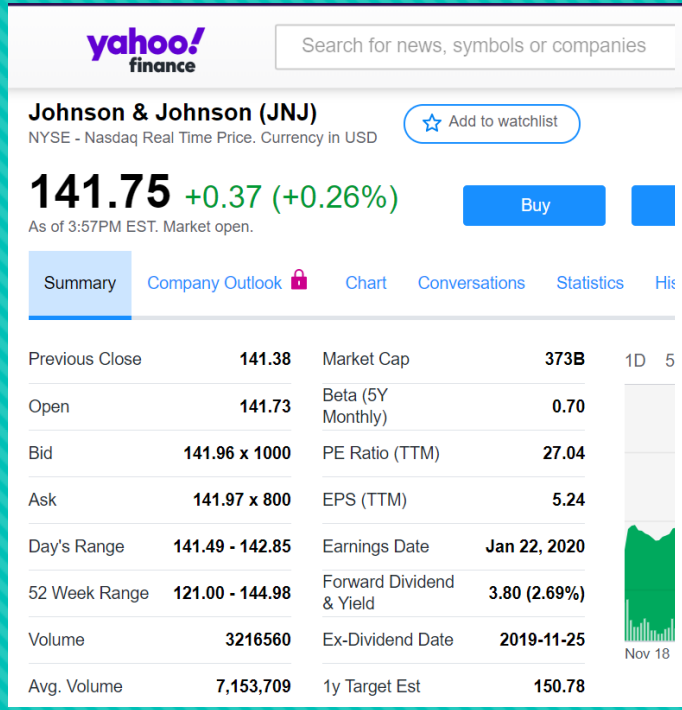
4. Text analysis

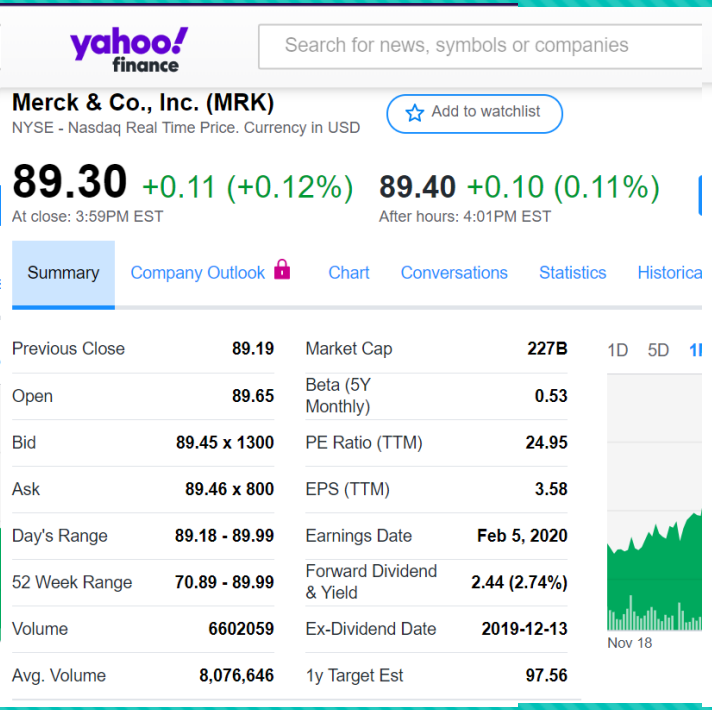
most common words
wordcloud
sentiment detection

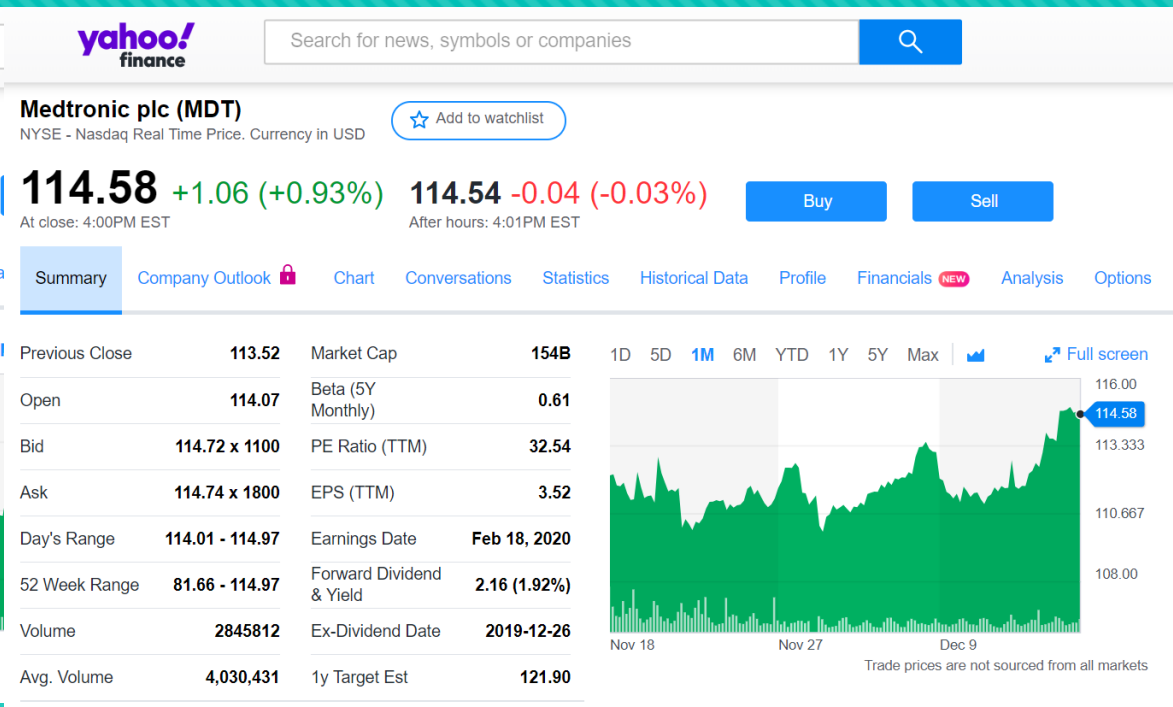
Johnson & Johnson

MERCK
INVENTING FOR LIFE

Medtronic







Target

JNJ twitter profile

MRK twitter profile

MDT twitter profile

1. API Calling : tweepy

consumer_key, consumer_secret, access_token, access_token_secret

2. Get User information: get_user()

Id
name
Screen_name
Location
url
Description…

3. Exact tweets: user_timeline()

Data Collection

Tweets collected

0	On #WorldAIDSDay, #JNJ is proud to share the powerful story of employee Fernando Salinas, who shortly after adoptinâ€¦ https://t.co/oXRGNrI06e
1	"I did what many wouldn'tâ€"I removed my gloves and touched the patients.â€ â€" @5BFilmâ€™s Alison Paolercio, a nurse whoâ€¦ https://t.co/zefy5jBu23
2	"The way to deal with the shame and trauma, in part, is to talk about it and find ways to heal."â€"@Lavernecox, the fâ€¦ https://t.co/3X179mO4i7
3	Meet Dr. @SchuitemakerH, the trailblazing HIV researcher who leads viral vaccine programs at @JanssenGlobal. Aheadâ€¦ https://t.co/aNBAKCWRhG
4	#DYK before BAND-AIDÂ® Brand adhesive bandages, home cooks treated cuts & nicks with strips of gauze they had to tieâ€¦ https://t.co/RmhGGi6OOB
5	Ahead of #WorldAIDSDay, #JNJ is proud to share the story of @5BFilm's Cliff Morrison & champion nurses on the frontâ€¦ https://t.co/VQy7Mxr0lW
6	RT @Ethicon: Strategic acquisitions & partnerships are crucial to our mission to advance lung surgery. Through our acquisition of Auris Heaâ€¦
7	The Lung Cancer Initiative at #JNJ is working to develop solutions that aim to prevent, intercept, and cureâ€¦ https://t.co/m3JEB4Mvk1
8	#JNJ is proud to join Emmy-nominated actress & activist @LaverneCox in the crusade to #makeHIVhistory. In honor ofâ€¦ https://t.co/UtmY3DUWkU
9	When four millennial #JNJ scientists noticed they were experiencing common skin dilemmas like fine lines & dark spoâ€¦ https://t.co/Rab13bH9tq
10	As families across the U.S. prepare for #Thanksgiving, go back in @JNJHistory and learn about the Laurel Club, an aâ€¦ https://t.co/h54C3IR28p
11	@HeartRiteHealth Hi there. Thanks for your interest in doing business with Johnson & Johnson. We ask that you visitâ€¦ https://t.co/SxI0R4HVzd
12	RT @JNJHistory: #JNJ has always inspired its employees to volunteerâ€"from our Laurel Clubâ€™s local efforts in the early 1900s, to todayâ€™s Gloâ€¦
13	Empowering nurses to develop & move their innovative ideas to propel patient care is just one way #JNJ champions nuâ€¦ https://t.co/6QnfKzFBJq
14	From generous work-family time to subsidized childcare expenses, #JNJ is helping ensure families of all shapesâ€¦ https://t.co/poCM49glTF
15	#DYK trillions of microorganisms called the microbiome co-exist in and on your bodyâ€"and could transform skin in surâ€¦ https://t.co/WEZ2l97kRD
16	RT @USGLC: Thrilled to have @JNJCares as the sponsor for tonightâ€™s #USGLCtribute dinner musical guests, the @ACChoir! These dynamic musiciaâ€¦
17	@JanssenGlobal Learn more about the many innovative ways #JNJ is working to build a diverse workforce: https://t.co/csT5mn1dVi #BelongatJNJ
18	"We can only achieve gender parity by having men & women partnering in the journey. We need to take bolder actionsâ€"â€¦ https://t.co/XEQy1FGAQL
19	RT @JanssenEMEA: "Be intentional about your decisions - every time you make one ask yourself, is this a chance to advance diversity & incluâ€¦

NLTK

Provides a set of diverse natural languages algorithms. Consists of the most common algorithms such as tokenizing, part-of-speech tagging, stemming, sentiment analysis, topic segmentation, and named entity recognition.

1. Tokenization

- Sentence Tokenization
- Word Tokenization

2. Frequency Distribution

- most_common()

3. Remove 'noise' words

- Set stopwords
- Remove Stopwords

3 Most Common Words

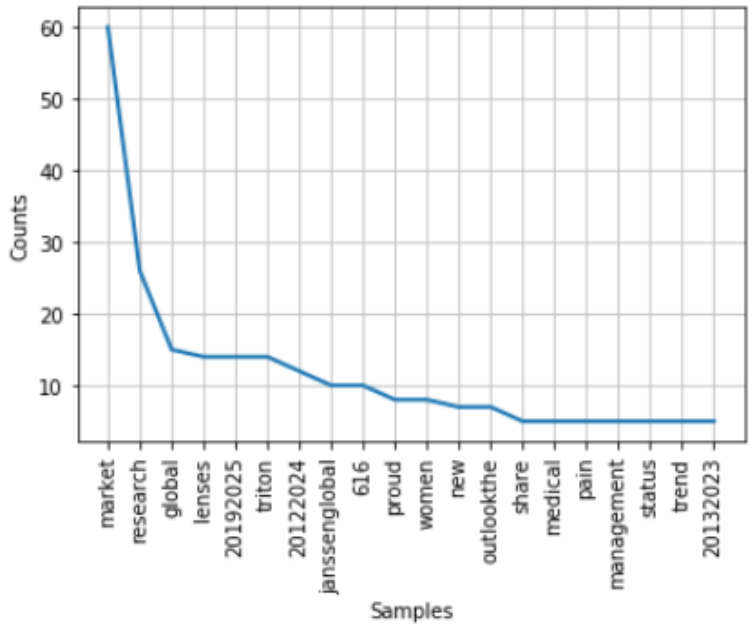
JNJ
3 Most Common Words

1. Market
2. Research
3. Global

```
#Most common words in JNJ tweets
JNJ_fdist.most_common(5)
```

```
[('market', 60),
 ('research', 26),
 ('global', 15),
 ('lenses', 14),
 ('20192025', 14)]
```

```
#Plot frequency distribution of tweets from JNJ
JNJ_fdist.plot(20,cumulative=False)
plt.show()
```



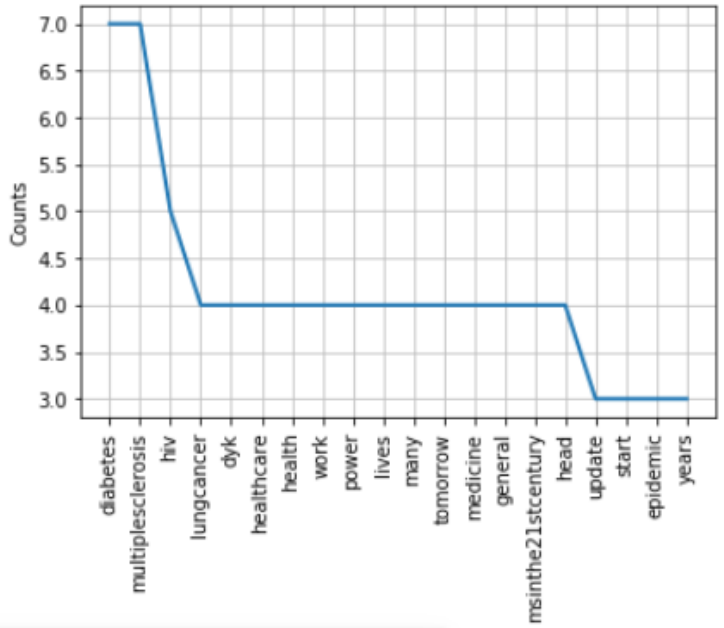
MRK
3 Most Common Words

1. Diabetes
2. Multiplesclerosis
3. HIV

```
#Most common words in MRK tweets
MRK_fdist.most_common(7)
```

```
[('diabetes', 7),
 ('multiplesclerosis', 7),
 ('hiv', 5),
 ('lungcancer', 4),
 ('dyk', 4),
 ('healthcare', 4),
 ('health', 4)]
```

```
#Plot frequency distribution of tweets from JNJ
MRK_fdist.plot(20,cumulative=False)
plt.show()
```



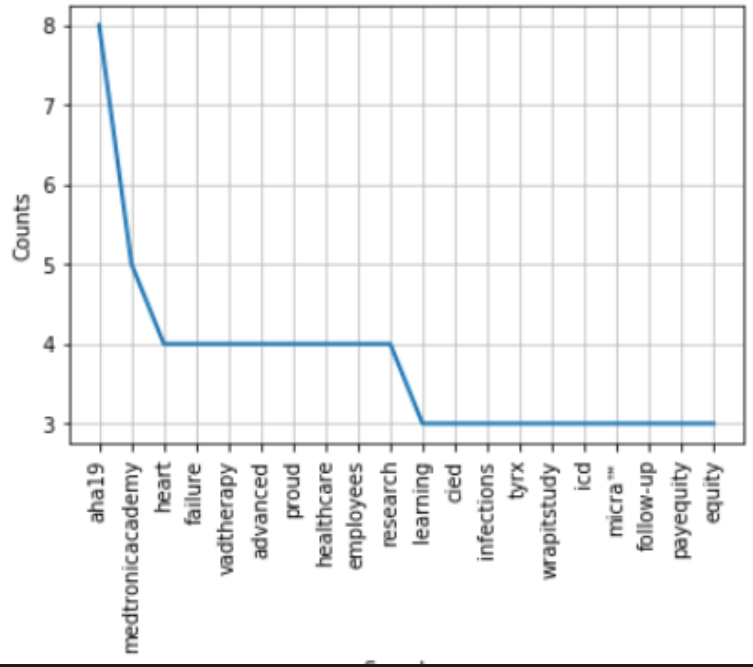
MDT
3 Most Common Words

1. AHA2019*
2. MedtronicAcademy
3. Heart failure

```
#Most common words in MDT tweets
MDT_fdist.most_common(5)
```

```
[('aha19', 8),
 ('medtronicacademy', 5),
 ('heart', 4),
 ('failure', 4),
 ('vadtherapy', 4)]
```

```
#Plot frequency distribution of tweets from JNJ
MDT_fdist.plot(20,cumulative=False)
plt.show()
```



* AHA19:American Heart Association's 2019 Annual Scientific Sessions

Wordcloud

Functionality to create pretty word clouds, visualize differences and similarity between documents, and avoid over-plotting in scatter plots with text.

1. **Prepare data**
   - Tokenize
   - Remove URLs, RTs, and twitter handles
2. **Set stopwords**
3. **Plot wordcloud**

Wordcloud

JNJ

3 Most Common Words

1. Market
2. Research
3. Global

MRK

3 Most Common Words

1. Diabetes
2. Multiplesclerosis
3. HIV

MDT
3 Most Common Words

1. AHA2019*
2. MedtronicAcademy
3. Heart failure

Vader Sentiment Analysis (Valence Aware Dictionary and sEntiment Reasoner)
A lexicon and rule-based sentiment analysis tool
Specifically attuned to sentiments expressed in social media. VADER uses a combination of A sentiment lexicon is a list of lexical features (e.g., words) which are generally labelled according to their semantic orientation as either positive or negative.

1. **Prepare data**
   - Tokenize
   - Remove URLs, RTs, and twitter handles
2. **Print polarity_scores**
3. **Summarize the scores**

Sentiment Analysis

```python
#Detect sentiment using Vader
sid = SentimentIntensityAnalyzer()
for sentence in MDT[1]:
    print(sentence)
    ss = sid.polarity_scores(sentence)
    for k in sorted(ss):
        print('{0}: {1}, '.format(k, ss[k]), end='')
        print()
```

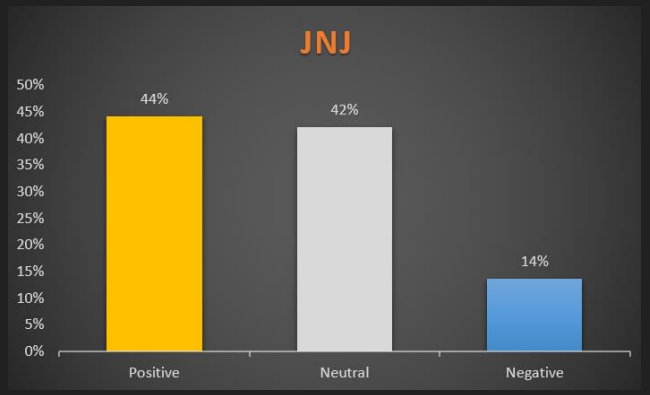
```
We're committed to working with public and private organizations to help address the full range of factors affectin…
compound: 0.5859,
neg: 0.0,
neu: 0.769,
pos: 0.231,
For more than 30 years, we have been committed to addressing the global challenges of the HIV epidemic but our work…
compound: 0.1779,
neg: 0.0,
neu: 0.876,
pos: 0.124,
We have the secret ingredient to formulating a great #Thanksgiving: a dash of gratitude for all our patients, emplo…
compound: 0.8126,
neg: 0.0,
neu: 0.67,
pos: 0.33,
```

```python
#summarize the score
MDT_summary = {"positive":0,"neutral":0,"negative":0}
for x in MDT[1]:
    ss = sid.polarity_scores(x)
    if ss["compound"] == 0.0:
        MDT_summary["neutral"] +=1
    elif ss["compound"] > 0.0:
        MDT_summary["positive"] +=1
    else:
        MDT_summary["negative"] +=1
print(MDT_summary)
```

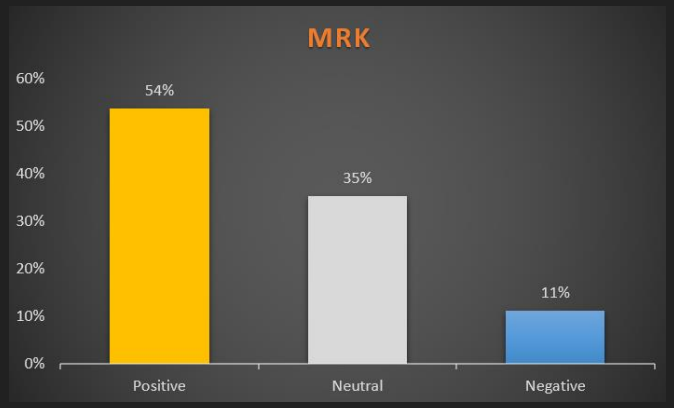
```
{'positive': 40, 'neutral': 20, 'negative': 6}
```

Sentiment Analysis

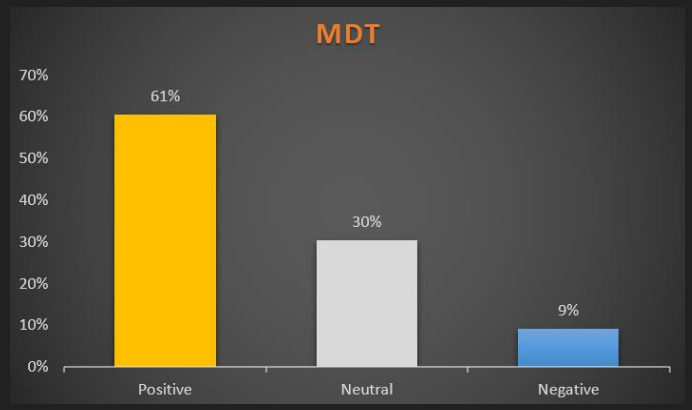
- JNJ
  - Positive : 42
  - Neutral : 40
  - Negative : 13



- MRK
  - Positive : 58
  - Neutral : 38
  - Negative : 12



- MDT
  - Positive : 40
  - Neutral : 20
  - Negative : 6



Stock volume and price of JNJ

SUM(High)

137.450 142.850

Volume

7M
6M
5M
4M
3M
2M
1M
0M

2 3 4 5 6 9 10 11 12 13 16

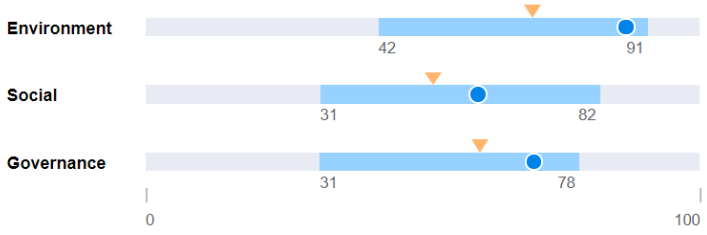
Environment, Social and Governance (ESG) Ratings

Total ESG score	Environment	Social	Governance
69 96th percentile	**86** 99th percentile	**59** 89th percentile	**69** 93rd percentile
Leader			

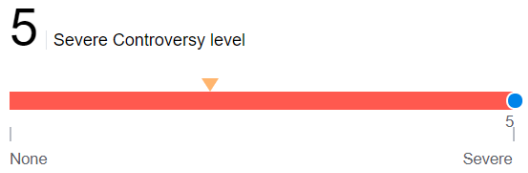
ESG Performance vs 82 Peer Companies

JNJ Peers Category Average

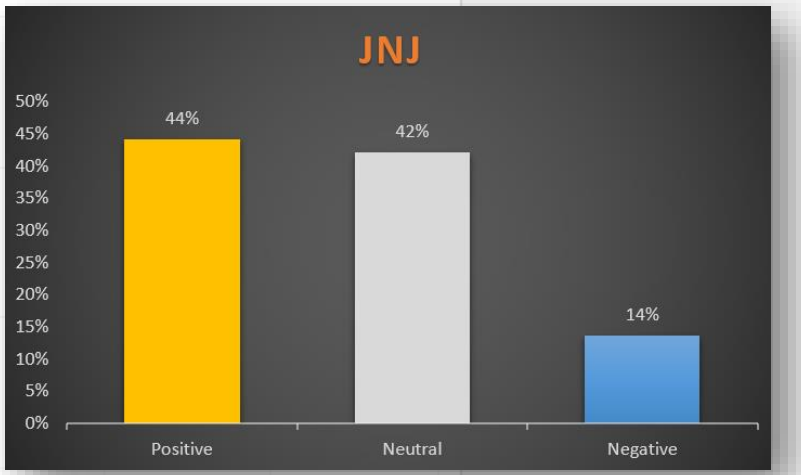
ESG PERFORMANCE



CONTROVERSY LEVEL



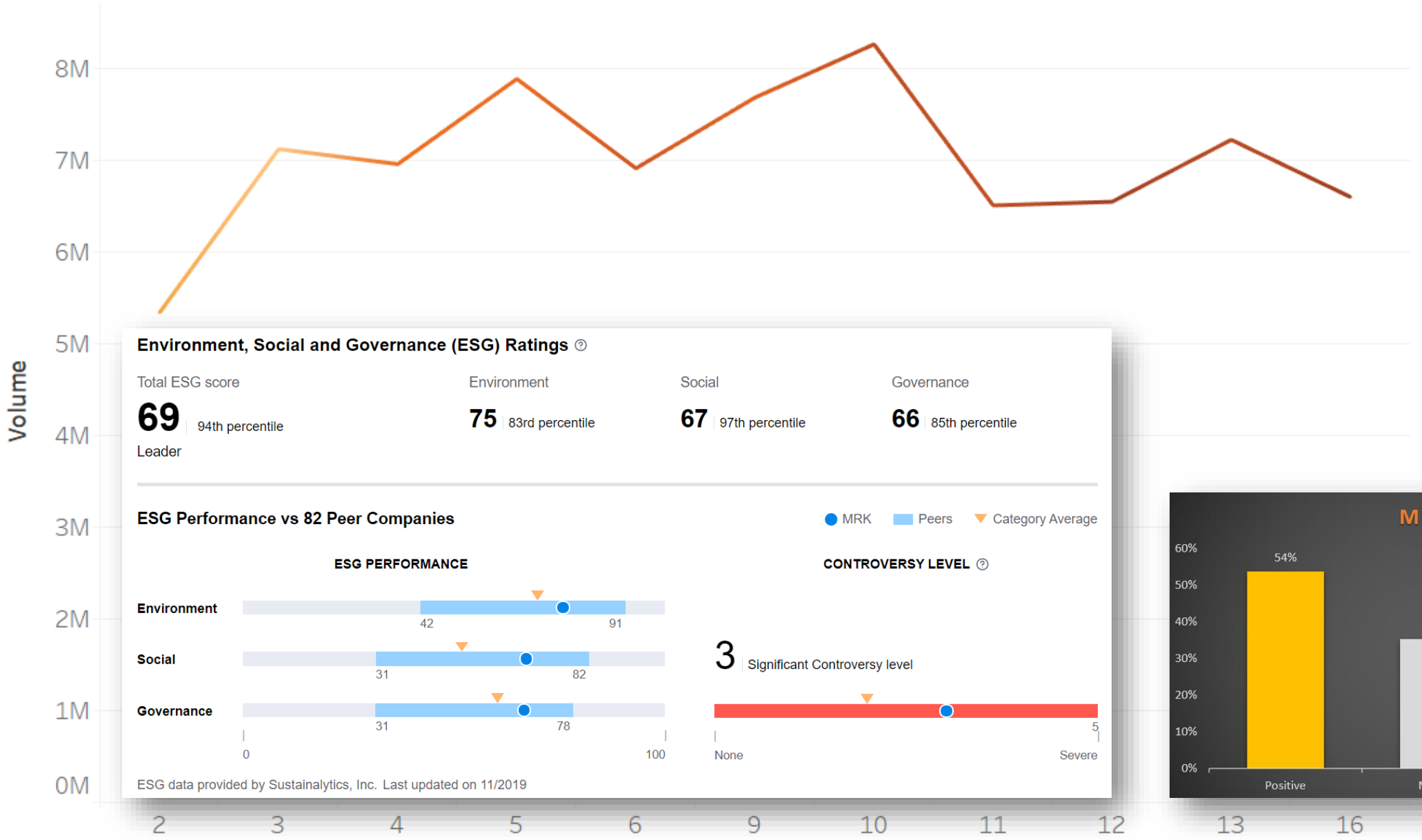
ESG data provided by Sustainalytics, Inc. Last updated on 11/2019



Stock volume and price of MRK

SUM(High)

87.470 89.990



Environment, Social and Governance (ESG) Ratings

Total ESG score	Environment	Social	Governance
69 94th percentile	**75** 83rd percentile	**67** 97th percentile	**66** 85th percentile
Leader			

ESG Performance vs 82 Peer Companies

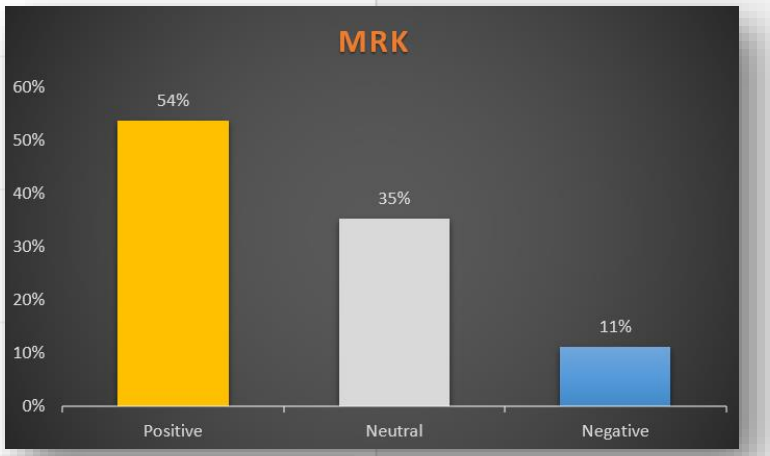
MRK Peers Category Average

ESG PERFORMANCE

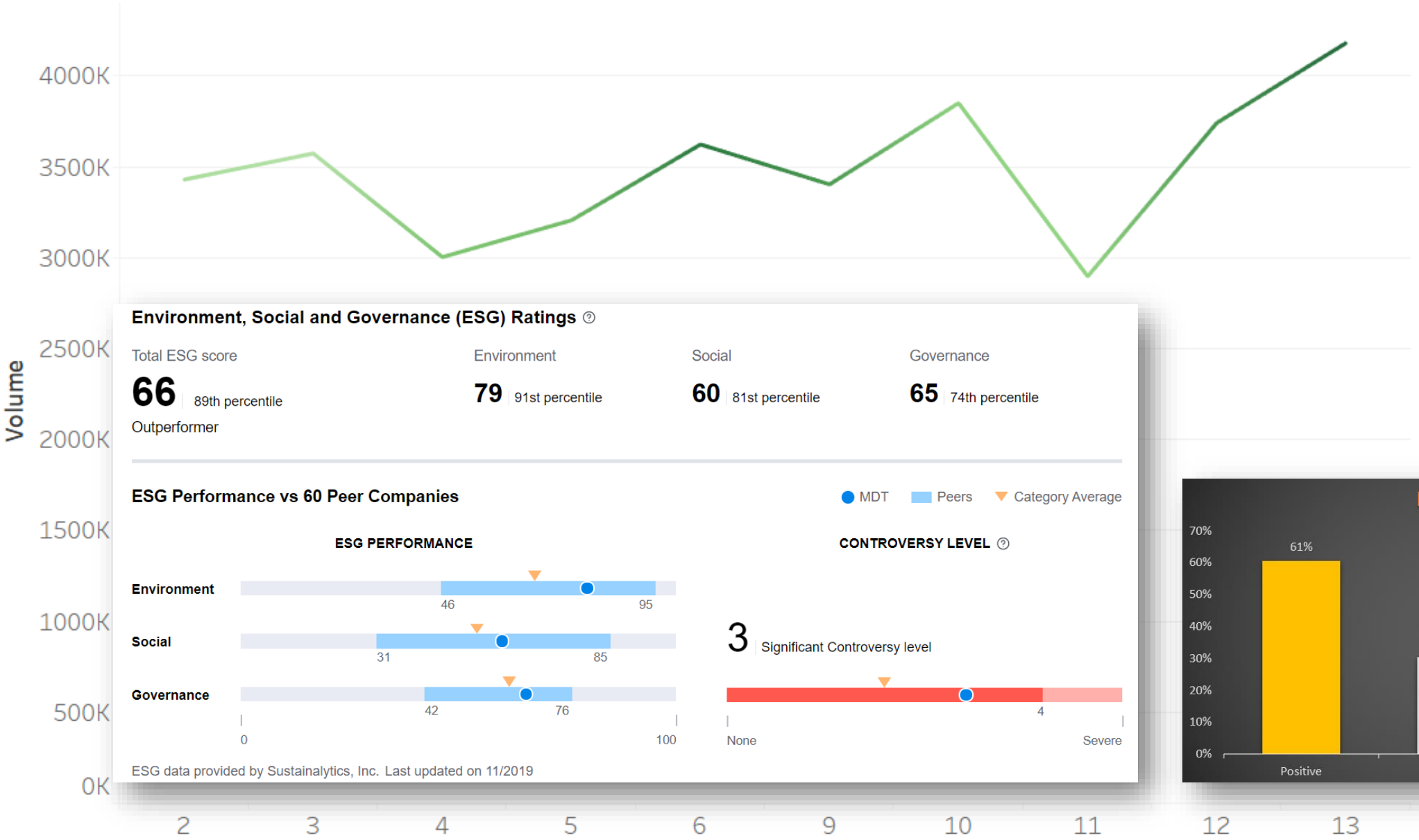
CONTROVERSY LEVEL

3 Significant Controversy level

ESG data provided by Sustainalytics, Inc. Last updated on 11/2019



Stock volume and price of MDT



SUM(High)

111.530 113.890

Environment, Social and Governance (ESG) Ratings

Total ESG score	Environment	Social	Governance
66 89th percentile	**79** 91st percentile	**60** 81st percentile	**65** 74th percentile
Outperformer			

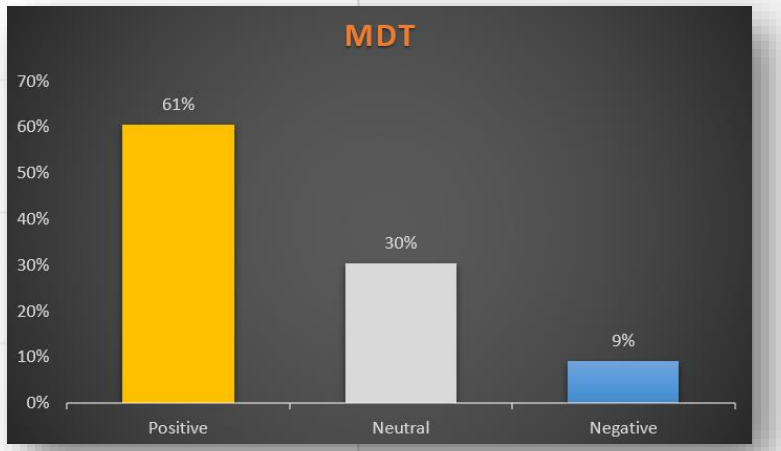
ESG Performance vs 60 Peer Companies

MDT Peers Category Average

ESG PERFORMANCE

CONTROVERSY LEVEL

ESG data provided by Sustainalytics, Inc. Last updated on 11/2019



Thank you

Chenxin Xie

076669433

chenxin.xie@ieseg.fr