

Исследование текстов пресс-релизов публичных компаний по квартальным результатам

ДПО КЛ 2025
Гаврилова Т. В.

21 июня 2025

Описание проекта

Цель – проанализировать тексты пресс-релизов, выявить используемые языковые конструкции

Гипотеза – подобные тексты пишутся в позитивном ключе

Этапы работы:

- Парсинг новостного сайта
- Предобработка текстов
- Извлечение цитат менеджмента (текст в кавычках)
- Токенизация, лемматизация, POS-tagging
- Позитив / Негатив анализ
- Коллокации / конструкции

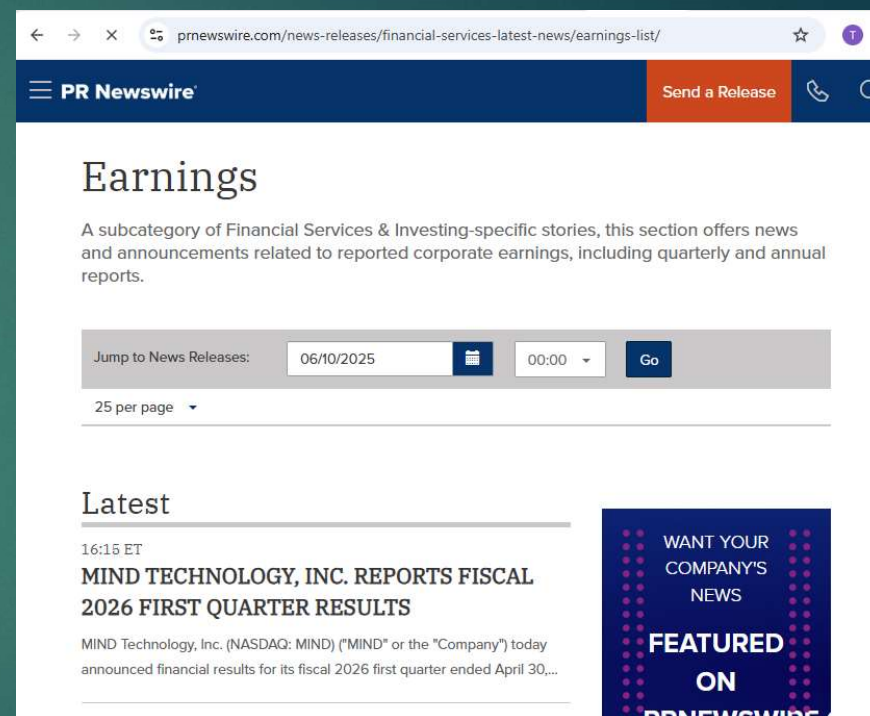
Использованы инструменты:

requests, time, BeautifulSoup,
re, nltk, Spacy,
pandas, WordCloud

Парсинг сайта, сбор корпуса

В результате парсинга сайта PR Newswire собран корпус:

- 2800 текстов – пресс-релизы публичных компаний по квартальным результатам
- опубликованные за период декабрь 2024 – июнь 2025 (отчётность за два квартала)



Проблемы

Парсинг - доступные ресурсы выдают историю только за год

Замусоренный корпус

Визуализация результатов

Подготовка текстов к анализу

1) Предобработка:

- удаление “мусорных” символов
- замена цифр на X

2) Извлечение цитат менеджмента

3) Корректировка списка стоп-слов

4) Применение библиотеки Spacy

- токенизация
- лемматизация
- удаление стоп-слов
- POS-tagging

```
# словарь стоп-слов скорректирован
stop_words = stopwords.words("english")

stop_words.remove('up')
stop_words.remove('down')
stop_words.remove('i')
stop_words.remove('we')
stop_words.extend(['first', 'second', 'third', 'fourth', 'quarter'])
```


100

2,800 пресс-релизов -----> 2,299 текстов цитат
(501 пустых – нет цитат в кавычках)

Облако слов построено из списка существительных, глаголов и прилагательных

	##
total tokens	467273.00
unique tokens	13894.00
ttr	0.03

```
tags_freqs = Counter(total_tags)
tags_freqs.most_common(20)
```

[('NOUN', 201375),
('VERB', 77551),
('ADJ', 73506),
('PUNCT', 29176),
('PROPN', 20856),
('ADV', 15446),
('PRON', 12254),
('SYM', 7935),
('NUM', 7660),
('SPACE', 5436),
('ADP', 5415),
('PART', 4230),
('X', 2145),
('AUX', 1592),
('SCONJ', 1099),
('DET', 588),
('CCONJ', 452),
('INTJ', 257)]



Выявление тональности комментариев менеджмента

с использованием *The Loughran-McDonald Sentiment Word Lists*

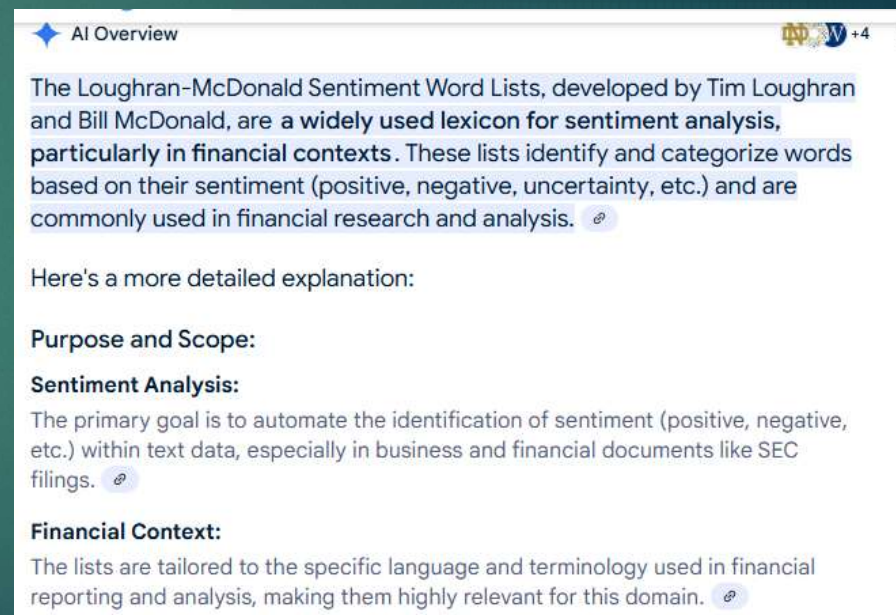
Словарь Loughran-McDonald содержит:
354 positive words
2,355 negative words

В проекте реализован простой счётчик
позитивных и негативных слов для
каждого пресс-релиза (в части цитат
менеджмента), маркировано:

17,257 позитивных токенов (181 уник, 1%)

6,479 негативных токенов (414 уник, 6%)

443,537 no-mood токенов (13,299 уник, 3%)



AI Overview

The Loughran-McDonald Sentiment Word Lists, developed by Tim Loughran and Bill McDonald, are a widely used lexicon for sentiment analysis, particularly in financial contexts. These lists identify and categorize words based on their sentiment (positive, negative, uncertainty, etc.) and are commonly used in financial research and analysis.

Here's a more detailed explanation:

Purpose and Scope:

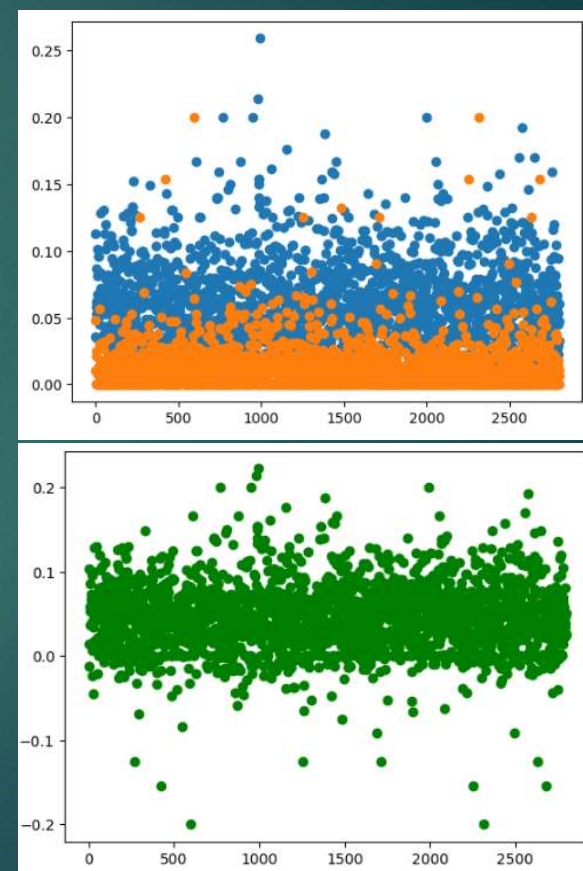
Sentiment Analysis:
The primary goal is to automate the identification of sentiment (positive, negative, etc.) within text data, especially in business and financial documents like SEC filings.

Financial Context:
The lists are tailored to the specific language and terminology used in financial reporting and analysis, making them highly relevant for this domain.

Соотношение ПОЗИТИВНЫХ И НЕГАТИВНЫХ СЛОВ

С использованием
The Loughran-McDonald Sentiment Word Lists

- позитивные слова ~ 6% токенов в сообщении
- негативные слова ~ 1% токенов в сообщении
- в 86% пресс-релизов количество позитивных слов (синие точки) превышает количество негативных (оранжевые точки)
- в 8% пресс-релизах количество негативных слов превышает количество позитивных (на зелёном графике точки в отрицательной области по оси y)



100

[illegible]

strong – 57%
improve – 31%
achieve – 23%

loss – 11%
decline – 7%
challenge – 9%

Результаты поиска N-грамм / Коллокаций

2-граммы

0	((x, x), 4466)
1	((x, million), 3601)
2	((company, s), 1585)
3	((non, gaap), 1383)
4	((forward, look), 1205)
5	((we, continue), 970)
6	((look, statement), 968)
7	((x, we), 926)
8	((long, term), 917)
9	((million, x), 911)
10	((cash, flow), 863)
11	((financial, measure), 777)
12	((x, billion), 740)
13	((gaap, financial), 736)
14	((year, year), 735)

`nltk.ngrams()`
`Counter`
`most_common`

((forward, look, statement), 966)

((gaap, financial, measure), 634)

((free, cash, flow), 385)

((earning, per, share), 207)

((long, term, value), 159)

3-граммы

((non, gaap, financial, measure), 547)

4-граммы

((cause, actual, result, differ, materially), 72)

5-граммы

`nltk.collocations`
`t-student`

`sorted_scores_raw[:30]`

```
[((('x', 'million'), 56.349013364535324),  
 (('x', 'x'), 48.12736794012535),  
 (('company', 's'), 38.77689696949056),  
 (('non', 'gaap'), 36.91591485825441),  
 (('forward', 'look'), 34.518501901754405),  
 (('look', 'statement'), 30.8811740254161),  
 (('long', 'term'), 30.17364850355293),  
 (('cash', 'flow'), 29.24091870041149),  
 (('we', 'continue'), 28.586102784894113),  
 (('financial', 'measure'), 27.443274623028287),  
 (('u', 's'), 26.65233829039698),  
 (('gaap', 'financial'), 26.581846408707865),  
 (('full', 'year'), 25.46715684910189),  
 (('x', 'billion'), 25.339050459514333),  
 (('net', 'income'), 24.77420920727387),  
 (('year', 'year'), 24.45458344681576),  
 (('adjusted', 'ebitda'), 24.411260417412688),  
 (('per', 'share'), 23.83222248786843),  
 (('december', 'x'), 23.09434604675527),  
 (('million', 'x'), 22.9074466802175),  
 (('we', 'believe'), 22.390734344640002),  
 (('we', 'remain'), 22.31218820027978),  
 (('increase', 'x'), 21.783848703441773),  
 (('compare', 'x'), 21.536014253929395),  
 (('balance', 'sheet'), 21.15649689251946),  
 (('approximately', 'x'), 20.99379920799983),  
 (('free', 'cash'), 20.45943713984459),  
 (('x', 'compare'), 20.18937725448386),  
 (('we', 'expect'), 20.084189264992713),  
 (('revenue', 'growth'), 19.56324992844218)]
```


Коллокации с учетом частеречной разметки

```
bigrams_with_adj[:50]
```

```
[(('financial_ADJ', 'measure_NOUN'), 38.50519032840984),  
 (('net_ADJ', 'income_NOUN'), 35.5189364002323),  
 (('financial_ADJ', 'result_NOUN'), 26.569286466874),  
 (('net_ADJ', 'loss_NOUN'), 26.449831431912816),  
 (('prior_ADJ', 'year_NOUN'), 25.148263526715223),  
 (('actual_ADJ', 'result_NOUN'), 23.379249433710953),  
 (('personal_ADJ', 'information_NOUN'), 22.14891216107219),  
 (('long_ADJ', 'term_NOUN'), 21.84392094084021),  
 (('gross_ADJ', 'profit_NOUN'), 21.493900415680223),  
 (('last_ADJ', 'year_NOUN'), 21.43421125115251),  
 (('gross_ADJ', 'margin_NOUN'), 21.41642739358515),  
 (('fiscal_ADJ', 'year_NOUN'), 20.023191739023794),  
 (('free_ADJ', 'cash_NOUN'), 18.83010440368343),  
 (('diluted_ADJ', 'share_NOUN'), 16.39417343985222),  
 (('financial_ADJ', 'performance_NOUN'), 16.224554577102705),  
 (('clinical_ADJ', 'trial_NOUN'), 15.87167321861407),  
 (('annual_ADJ', 'report_NOUN'), 15.80287690223825),  
 (('net_ADJ', 'sale_NOUN'), 15.203004424691699),  
 (('financial_ADJ', 'statement_NOUN'), 15.13406880228984),  
 (('full_ADJ', 'year_NOUN'), 15.104284214603291),  
 (('common_ADJ', 'share_NOUN'), 15.0180316351866),  
 (('executive_ADJ', 'officer_NOUN'), 14.821168135063365),  
 (('fair_ADJ', 'value_NOUN'), 14.775899503598227),  
 (('fiscal_ADJ', 'x_NOUN'), 14.66637319006129),  
 (('non_ADJ', 'gaap_NOUN'), 14.603291294011612),  
 (('adjusted_ADJ', 'ebitda_NOUN'), 14.57636544408712)]
```

Конструкции с учетом частеречной разметки

```
res = re.findall(r'we_PRON\scontinue_VERB\s\w+_VERB\b',  
res_freq = Counter(res)  
res_freq.most_common(20)
```

```
[('we_PRON continue_VERB see_VERB', 80),  
( 'we_PRON continue_VERB execute_VERB', 59),  
( 'we_PRON continue_VERB focus_VERB', 52),  
( 'we_PRON continue_VERB make_VERB', 44),  
( 'we_PRON continue_VERB invest_VERB', 35),  
( 'we_PRON continue_VERB drive_VERB', 29),  
( 'we_PRON continue_VERB build_VERB', 28),  
( 'we_PRON continue_VERB advance_VERB', 24),  
( 'we_PRON continue_VERB deliver_VERB', 20),  
( 'we_PRON continue_VERB expand_VERB', 17),  
( 'we_PRON continue_VERB work_VERB', 17),  
( 'we_PRON continue_VERB pursue_VERB', 15),  
( 'we_PRON continue_VERB expect_VERB', 15),  
( 'we_PRON continue_VERB experience_VERB', 14),  
( 'we_PRON continue_VERB monitor_VERB', 13),  
( 'we_PRON continue_VERB grow_VERB', 13),  
( 'we_PRON continue_VERB optimize_VERB', 11),  
( 'we_PRON continue_VERB improve_VERB', 11),  
( 'we_PRON continue_VERB believe_VERB', 11),  
( 'we_PRON continue_VERB prioritize_VERB', 10)]
```

```
res = re.findall(r'we_PRON\sremain_VERB\s\w+_ADJ\b',  
res_freq = Counter(res)  
res_freq.most_common(20)
```

```
[('we_PRON remain_VERB focused_ADJ', 150),  
( 'we_PRON remain_VERB committed_ADJ', 115),  
( 'we_PRON remain_VERB confident_ADJ', 98),  
( 'we_PRON remain_VERB optimistic_ADJ', 17),  
( 'we_PRON remain_VERB steadfast_ADJ', 16),  
( 'we_PRON remain_VERB dedicated_ADJ', 11),  
( 'we_PRON remain_VERB excited_ADJ', 6),  
( 'we_PRON remain_VERB vigilant_ADJ', 5),  
( 'we_PRON remain_VERB disciplined_ADJ', 5),  
( 'we_PRON remain_VERB bullish_ADJ', 5),  
( 'we_PRON remain_VERB diligent_ADJ', 5),  
( 'we_PRON remain_VERB mindful_ADJ', 4),  
( 'we_PRON remain_VERB agile_ADJ', 4),  
( 'we_PRON remain_VERB proactive_ADJ', 3),  
( 'we_PRON remain_VERB pleased_ADJ', 3),  
( 'we_PRON remain_VERB cautious_ADJ', 3),  
( 'we_PRON remain_VERB resilient_ADJ', 2),  
( 'we_PRON remain_VERB nimble_ADJ', 2),  
( 'we_PRON remain_VERB competitive_ADJ', 2),  
( 'we_PRON remain_VERB attentive_ADJ', 2)]
```


Словосочетания с самыми частотными поз/нег словами

```
res = re.findall(r'strong_ADJ\s\b\w+_NOUN')
res_freq = Counter(res)
res_freq.most_common(20)
```

```
[('strong_ADJ result_NOUN', 111),
 ('strong_ADJ performance_NOUN', 108),
 ('strong_ADJ balance_NOUN', 93),
 ('strong_ADJ growth_NOUN', 77),
 ('strong_ADJ start_NOUN', 76),
 ('strong_ADJ cash_NOUN', 66),
 ('strong_ADJ momentum_NOUN', 57),
 ('strong_ADJ foundation_NOUN', 51),
 ('strong_ADJ year_NOUN', 44),
 ('strong_ADJ execution_NOUN', 40),
 ('strong_ADJ demand_NOUN', 38),
 ('strong_ADJ revenue_NOUN', 33),
 ('strong_ADJ operating_NOUN', 30),
 ('strong_ADJ finish_NOUN', 28),
 ('strong_ADJ position_NOUN', 27),
 ('strong_ADJ liquidity_NOUN', 20),
 ('strong_ADJ earning_NOUN', 19),
 ('strong_ADJ capital_NOUN', 18),
 ('strong_ADJ return_NOUN', 16),
 ('strong_ADJ profitability_NOUN', 16)]
```

```
# поиск словосочетаний в тэгиrowанном тексте
res = re.findall(r'improve_VERB(?:\s\b\w+\b){2}', tagged)
res_freq = Counter(res)
res_freq.most_common(5)
```

```
[('improve_VERB operational_ADJ efficiency_NOUN', 23),
 ('improve_VERB gross_ADJ margin_NOUN', 12),
 ('improve_VERB financial_ADJ performance_NOUN', 8),
 ('improve_VERB cash_NOUN flow_NOUN', 8),
 ('improve_VERB operational_ADJ performance_NOUN', 7)]
```

```
res = re.findall(r'achieve_VERB(?:\s\b\w+\b){2}', tagged)
res_freq = Counter(res)
res_freq.most_common(10)
```

```
[('achieve_VERB significant_ADJ milestone_NOUN', 9),
 ('achieve_VERB full_ADJ year_NOUN', 8),
 ('achieve_VERB record_NOUN revenue_NOUN', 6),
 ('achieve_VERB sustainable_ADJ growth_NOUN', 5),
 ('achieve_VERB strategic_ADJ objective_NOUN', 5),
 ('achieve_VERB solid_ADJ result_NOUN', 5),
 ('achieve_VERB strong_ADJ financial_ADJ', 5),
 ('achieve_VERB remarkable_ADJ growth_NOUN', 4),
 ('achieve_VERB strong_ADJ result_NOUN', 4),
 ('achieve_VERB important_ADJ milestone_NOUN', 4)]
```

```
res = re.findall(r'\b\w+\b\schallenge_\w+\b')
res_freq = Counter(res)
res_freq.most_common(10)
```

```
[('despite_SCONJ challenge_VERB', 17),
 ('despite_SCONJ challenge_NOUN', 15),
 ('significant_ADJ challenge_NOUN', 11),
 ('remain_VERB challenge_VERB', 10),
 ('address_VERB challenge_NOUN', 9),
 ('market_NOUN challenge_NOUN', 9),
 ('face_VERB challenge_NOUN', 9),
 ('many_ADJ challenge_NOUN', 8),
 ('external_ADJ challenge_NOUN', 7),
 ('navigate_VERB challenge_VERB', 7)]
```

```
res = re.findall(r'challenging_\w+\b\s\b\w+\b')
res_freq = Counter(res)
res_freq.most_common(10)
```

```
[('challenging_ADJ environment_NOUN', 19),
 ('challenging_ADJ market_NOUN', 16),
 ('challenging_ADJ year_NOUN', 13),
 ('challenging_ADJ macro_ADJ', 11),
 ('challenging_ADJ economic_ADJ', 11),
 ('challenging_ADJ macroeconomic_ADJ', 9),
 ('challenging_ADJ condition_NOUN', 5),
 ('challenging_ADJ business_NOUN', 3),
 ('challenging_ADJ time_NOUN', 3),
 ('challenging_ADJ uncertain_ADJ', 3)]
```

Дальнейшие исследования



Анализ текстов пресс-релизов целиком

Более длинная история

Другие виды публикаций - отчетность

Другие виды новостных пресс-релизов

Анализ реакций / комментариев внешних агентов

Приложение. “STRONG” story

['our strong business model']
['was strong across our', 'were strong and consistent', 'and strong employee engagement', 'the strong cash flow']
['was strong across revenue']
['basis strong execution by']
['a strong finish to']
['a strong upward trajectory']
['a strong start to']
['a strong foundation for', 'reflects strong execution in']
['expected strong sequential expansion']
['from strong growth in']
['delivered strong results for', 'a strong emphasis on']
['a strong rebound in']
['delivered strong first quarter']
['s strong operating and', 'a strong fourth quarter']
['continued strong growth into', 'by strong ai demand', 'driving strong growth in']
['our strong start to', 'is strong nevertheless we']
['s strong wholesale margins']
['our strong brand partnerships']
['anticipate strong second half']
['our strong product portfolio']
['are strong and investing']
['a strong quarter which', 'with strong qx results']
['a strong sales funnel']

Приложение. “STRONG” story

['through strong operational performance']
['with strong revenue growth']
['a strong first quarter', 'and strong momentum we']
['our strong start to']
['a strong backlog and', 'remains strong as we', 'a strong track record']
['delivered strong revenue and']
['our strong operational performance', 'with strong fleet utilization']
['our strong operational performance', 'with strong fleet utilization']
['a strong start to']
['achieved strong results across']
['a strong global reception']
['demonstrating strong growth momentum']
['a strong cash position']
['very strong first quarter', 'our strong value proposition', 'alongside strong improvements on', 'see strong account growth']
['a strong quarter despite']
['our strong performance in', 'another strong quarter we']
['a strong start to', 'a strong bottom line', 'drive strong performance through']
['our strong pipeline activates']
['a strong financial foundation']
['our strong balance sheet']
['and strong cash flow']
['a strong close to', 'are strong and they', 'our strong cardiovascular team']

Приложение. “STRONG” story

['drove strong growth in', 'maintained strong momentum in', 'its strong upward trajectory']
['the strong performance of']
['provides strong earnings visibility', 'our strong balance sheet']
['delivered strong first quarter', 'continued strong execution of', 'remained strong at rmbx']
['driving strong returns and', 'another strong year in']
['see strong momentum in', 'with strong bookings momentum']
['remain strong despite a', 's strong qx in']
['experience strong demand for']
['another strong quarter of', 'delivered strong first quarter', 'our strong balance sheet']
['a strong cash position']
['demonstrated strong momentum through']
['of strong financial results']
['a strong performance driving']
['the strong performance was', 'to strong growth in']
['connection strong sales of', 'the strong performance reflects']
['by strong execution and', 'with strong visibility and']
['a strong first quarter']
['delivered strong financial results']
['with strong momentum advancing']
['and strong earnings in']
['a strong first quarter']
['generate strong growth momentum']
['a strong foundation for', 'ecosystem strong organizational ability']

Приложение. “STRONG” story

['through strong strategic execution']
['saw strong demand for']
['continued strong performance and']
['a strong start to']
['s strong performance driven']
['growth strong improvement in', 'receive strong endorsements from']
['a strong start april', 'with strong underlying orders']
['businesses strong balance sheet']
['see strong growth in']
['delivered strong revenue growth']
['by strong energy storage']
['the strong performance of']
['achieved strong and sustainable']
['driven strong margin performance']
['is strong and the']
['see strong growth in', 'see strong underlying growth', 'by strong growth in', 'as strong growth in']
['report strong gold production']
['the strong performance of', 'our strong confidence in', 'a strong balance sheet']
['a strong focus on']
['had strong quarterly results', 'generating strong cash flows']
['our strong first quarter', 's strong financial performance']
['s strong potential as']
['a strong balance sheet']

Приложение. “STRONG” story

['report strong x first', 'a strong us manufacturing']
['maintain strong quality control']
['maintain strong quality control']
['a strong start to', 'by strong project execution']
['make strong progress on']
['its strong global position', 'a strong growth market']
['very strong and encouraging', 'remains strong and we']
['a strong quarter operationally']
['a strong balance sheet']
['demonstrating strong sequential improvements']
['a strong gold price']
['our strong start to', 'have strong traction with']
['a strong capital base']
['a strong start to', 'made strong progress against']
['a strong start to', 'made strong progress against']
['with strong operational momentum', 'a strong position with']
['with strong client relationship']
['be strong although it']
['show strong growth due']
['saw strong revenue growth']
['a strong position to']
['very strong position to']
['stands strong amidst increasing']

Приложение. “STRONG” story

['very strong first quarter', 'recognition strong product velocities']
['demonstrated strong tolerability and']
['a strong backdrop to']
['drove strong sequential and']
['a strong emphasis on']
['delivered strong cash conversion']
['a strong liquidity position']
['a strong focus on']
['particularly strong adoption by']
['growth strong operating free', 'a strong finish to', 'margins strong free cash']
['a strong cash balance']
['and strong performance from', 'the strong residual value']
['our strong cash position']
['experience strong demand for']
['our strong start to']
['a strong financial balance']
['our strong first quarter']
['a strong balance sheet']
['delivered strong financial and']
['a strong foundation for', 'delivered strong fundraising results']
['these strong results for', 'and strong balance sheet', 'a strong balance sheet']
['our strong first quarter']
['our strong first quarter', 'reflecting strong interest in']

Приложение. “STRONG” story

['to strong enrollment trends']
['our strong balance sheet', 'a strong cash position']
['enjoyed strong results despite', 'a strong management team']
['see strong interest in', 'drive strong results with']
['of strong financial results']
['a strong return on', 'a strong liquidity position']
['our strong financial performance']
['seeing strong market acceptance']
['a strong capital position']
['reflected strong specialty product']
['results strong utilization across', 'a strong balance sheet']
['a strong second quarter', 'by strong economic growth', 'our strong second quarter', 'generate strong returns to']
['of strong growth and']
['a strong balance sheet']
['a strong start to']
['season strong demand for', 'the strong foundation we', 's strong operating performance']
['of strong revenue growth', 'through strong revenue growth']
['and strong production from', 'its strong deal pipeline']
['achieved strong top and', 'and strong sales and', 'our strong fundamentals while', 'remains strong with a']
['a strong start to']
['of strong operating results', 'continued strong results demonstrate', 'our strong first quarter', 'very strong liquidity and']
['s strong operating results', 'already strong capital and']
['another strong performance from', 'the strong pace set', 'exceptionally strong backlog in', 'our strong balance sheet']

Приложение. “STRONG” story

['see strong demand for', 'remains strong and flexible']
['delivered strong qx revenue']
['very strong selling environment', 'our strong pipeline low']
['to strong sales growth', 'the strong second quarter']
['a strong start to', 'with strong cash generation', 'a strong and flexible']
['a strong balance sheet']
['our strong first quarter']
['delivering strong results in']
['a strong balance sheet']
['experience strong growth in', 'experience strong organic growth']
['deliver strong top and']
['our strong balance sheet']
['s strong start to', 'drive strong profitable growth']
['with strong operating results']
['a strong balance sheet']
['a strong start to']
['a strong start to', 'deliver strong sustainable performance']
['and strong investment results']
['seeing strong progress across']
['see strong indicators of']
['with strong refinery reliability']
['delivering strong financial performance']
['with strong momentum delivering']

Приложение. “STRONG” story

['with strong product candidates']
['the strong volume performance']
['delivering strong adjusted ebitda', 'deliver strong adjusted ebitda', 'a strong position to']
['our strong positioning across']
['a strong year ahead']
['generate strong cash flow']
['a strong performance by', 'its strong book of', 'and strong balance sheet', 'a strong pace and', 'of strong free cash', 'network strong balance sheet']
['is strong positioning us']
['of strong gross margin', 'its strong cash flow', 'are strong and its']
['and strong balance sheet', 'a strong balance sheet']
['achieved strong organic net']
['and strong alignment with']
['a strong start to', 'platforms strong customer adoption']
['reported strong first quarter']