



Deriving Alpha from News: GDELT

Ty Painter
Capstone Update #3
2/10/21



GDELT Data

- Can now read in monthly data
- Each month is ~12 GB
 - *Storage...*
 - *Zipped files are ~3 GB*
- January 2021 value counts
 - *Websites*
 - *Keywords*
 - *Topics*

	website	counts
0	iheart.com	161187
1	msn.com	81981
2	medium.com	48707
3	reuters.com	37931
4	yahoo.com	28885
5	indiatimes.com	26354
6	prnewswire.com	22544
7	dailymail.co.uk	18859
8	texasguardian.com	18031
9	apnews.com	17731
10	sandiegosun.com	16727
11	newyorktelegraph.com	16668
12	ianslive.in	13846
13	sfgate.com	12787
14	chron.com	12676
15	marketwatch.com	12166
16	washingtontimes.com	11970
17	onenewspage.com	11206
18	thenews.com.pk	10850
19	thehindu.com	10523
20	tmcnet.com	10394
21	finanznachrichten.de	9590
22	business-standard.com	9512
23	upstrack.com	9483
24	investigate.co.uk	9434
25	chinadaily.com.cn	9295



Stock Data

- Gathered dates of addition/removal from S&P 500, sector, and industry
 - *Market cap*

	Symbol	Security	SEC filings	GICS Sector	GICS Sub-Industry	Headquarters Location	Date first added	CIK	Founded
0	MMM	3M	reports	Industrials	Industrial Conglomerates	Saint Paul, Minnesota	1976-08-09	66740	1902
1	AOS	A. O. Smith	reports	Industrials	Building Products	Milwaukee, Wisconsin	2017-07-26	91142	1916
2	ABT	Abbott	reports	Health Care	Health Care Equipment	North Chicago, Illinois	1964-03-31	1800	1888
3	ABBV	AbbVie	reports	Health Care	Pharmaceuticals	North Chicago, Illinois	2012-12-31	1551152	2013 (1888)
4	ABMD	Abiomed	reports	Health Care	Health Care Equipment	Danvers, Massachusetts	2018-05-31	815094	1981
...
500	YUM	Yum! Brands	reports	Consumer Discretionary	Restaurants	Louisville, Kentucky	1997-10-06	1041061	1997
501	ZBRA	Zebra Technologies	reports	Information Technology	Electronic Equipment & Instruments	Lincolnshire, Illinois	2019-12-23	877212	1969
502	ZBH	Zimmer Biomet	reports	Health Care	Health Care Equipment	Warsaw, Indiana	2001-08-07	1136869	1927
503	ZION	Zions Bancorp	reports	Financials	Regional Banks	Salt Lake City, Utah	2001-06-22	109380	1873
504	ZTS	Zoetis	reports	Health Care	Pharmaceuticals	Parsippany, New Jersey	2013-06-21	1555280	1952

505 rows × 9 columns

	Date	ticker	security	Reason	action
0	February 3, 2022	GPS	Gap	Market capitalization change.[6]	remove
1	February 2, 2022	CEG	Constellation Energy	S&P 500 and 100 constituent Exelon Corp. spun ...	add
2	December 20, 2021	LEG	Leggett & Platt	Market capitalization change.[7]	remove
2	December 20, 2021	SBNY	Signature Bank	Market capitalization change.[7]	add
3	December 20, 2021	SEDG	SolarEdge	Market capitalization change.[7]	add
...
286	December 5, 2000	ABK	Ambac Financial	Market Cap changes.	add
287	July 27, 2000	JDSU	JDS Uniphase	Market Cap change.[223]	add
287	July 27, 2000	RAD	RiteAid	Market Cap change.[223]	remove
288	December 7, 1999	LDW	Laidlaw	Market capitalization change.[224]	remove
288	December 7, 1999	YHOO	Yahoo!	Market capitalization change.[224]	add

561 rows × 5 columns

Web Scrapping

- List of 100 most popular websites for January 2021
 - Foreign websites
 - Website subscriptions
 - Created a template for yahoo.com URLs
 - Some dead URLs

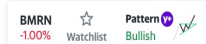
SHAREHOLDER ALERT: Pomerantz Law Firm Reminds Shareholders with Losses on their Investment in BioMarin Pharmaceutical Inc. of Class Action Lawsuit and Upcoming Deadline – BMRN



Pomerantz LLP

November 30, 2021 · 4 min read

In this article:



NEW YORK, Nov. 30, 2021 (GLOBE NEWSWIRE) -- Pomerantz LLP announces that a class action lawsuit has been filed against BioMarin Pharmaceutical Inc. ("BioMarin" or the "Company") (NASDAQ: BMRN) and certain of its officers. The class action, filed in the United States District Court for the Northern District of California, and docketed under 21-cv-08254, is on behalf of a class consisting of all persons and entities other than Defendants that purchased or otherwise acquired BioMarin securities between January 13, 2020 and September 3, 2021, both dates inclusive (the "Class Period"), seeking to recover damages caused by Defendants' violations of the federal securities laws and to pursue remedies under Sections 10(b) and 20(a) of the Securities Exchange Act of 1934 (the "Exchange Act") and Rule 10b-5 promulgated thereunder, against the Company and certain of its top officials.



Enjoy an ad-free
and 24/7 tech su
[Get it now >](#)

Quote Lookup

Related Quotes

Symbol	Last Pr
BMRN	91.

14 final.text

'NEW YORK, Nov. 30, 2021 (GLOBE NEWSWIRE) -- Pomerantz LLP announces that a class action lawsuit has been filed against BioMarin Pharmaceutical Inc. ("BioMarin" or the "Company") (NASDAQ: BMRN) and certain of its officers. The class action, filed in the United States District Court for the Northern District of California, and docketed under 21-cv-08254, is on behalf of a class consisting of all persons and entities other than Defendants that purchased or otherwise acquired BioMarin securities between January 13, 2020 and September 3, 2021, both dates inclusive (the "Class Period"), seeking to recover damages caused by Defendants' violations of the federal securities laws and to pursue remedies under Sections 10(b) and 20(a) of the Securities Exchange Act of 1934 (the "Exchange Act") and Rule 10b-5 promulgated thereunder, against the Company and certain of its top officials. If you are a shareholder who purchased or otherwise acquired BioMarin securities during the Class Period, you have until December 22, 2021 to ask the Court to appoint you as Lead Plaintiff for the class. A copy of the Complaint can be obtained at www.pomerantzlaw.com. To discuss this action, contact Robert S. Willoughby at newaction@pomlaw.com or 888.476.6529 (or 888.4-POMLAW), toll-free, Ext. 7980. Those who inquire by e-mail are encouraged to include their mailing address, telephone number, and the number of shares purchased. [Click here for information about joining the class action] BioMarin develops and commercializes therapies for people with serious and life-threatening rare diseases and medical conditions. The Company is developing, among other product candidates, BMN 307, an AAV5 mediated gene therapy, which is in a phase 1/2 clinical trial to normalize blood phenylalanine concentration levels in patients with phenylketonuria ("PKU"). The Company's Phearless Phase 1/2 study is evaluating BMN 307 in adults with PKU. On November 7, 2018, BioMarin shared pre-clinical data of BMN 307, which demonstrated lifetime Phe corrections in mouse models, and announced that the Company was planning to file an investigational new drug application ("IND") for BMN 307.

NLP

- [distilRoberta-financial-sentiment](#)

- *Entire article is too big to pass in*

Input is too long, try to truncate or use a parameter to handle this: The expanded size of the tensor (566) must match the existing size (514) at non-singleton dimension 1. Target sizes: [1, 566]. Tensor sizes: [1, 514]

- [PEGASUS for Financial Summarization](#)

Next Steps

- GDELT data storage
- Additional web scraping templates
- Adjust HuggingFace API parameters
- [GitHub](#)