

Determining Whether Press Release Language Has Changed Since the Start of the Pandemic

By Daniel Berry

Introduction

PRESS RELEASE

October 13, 2020

**Apple introduces
HomePod mini: A powerful
smart speaker with amazing
sound**

**Walmart Launches COVID-19
Vaccine Drive to Reach More
Residents in Vulnerable
Communities**

**How Walmart is Responding to
the Coronavirus**

Jan. 31, 2020

Press release

Salesforce and Amazon Web Services (AWS)
Expand Global Strategic Partnership

November 19, 2019 at 8:03 AM EST

**CVS Health announces
COVID-19 resources for
Aetna members**

Friday, March 6, 2020

Problem Statement

Use NLP and binary classification models to determine whether or not the top five Fortune 100 companies have changed how they communicate through press releases.

Data Collection & Data Cleaning

Gather Information from Fortune's Website

FORTUNE RANKINGS ▾ MAGAZINE NEWSLETTERS PODCASTS COVID-19 MORE ▾								SEARCH	SIGN IN
RANK ^	NAME	REVENUES (\$M)	REVENUE PERCENT CHANGE	PROFITS (\$M)	PROFITS PERCENT CHANGE	ASSETS (\$M)	MARKET VALUE — AS OF MARCH 31, 2020 (\$M)	CHANGE IN RANK (FULL 1000)	• NEW
1	Walmart	\$523,964	1.9%	\$14,881	123.1%	\$236,495	\$321,803.3	-	
2	Amazon	\$280,522	20.5%	\$11,588	15%	\$225,248	\$970,680.1	3	
3	Exxon Mobil	\$264,938	-8.7%	\$14,340	-31.2%	\$362,597	\$160,696.3	-1	
4	Apple	\$260,174	-2%	\$55,256	-7.2%	\$338,516	\$1,112,640.8	-1	
5	CVS Health	\$256,776	32%	\$6,634	-	\$222,449	\$77,375.8	3	
6	Berkshire Hathaway	\$254,616	2.7%	\$81,417	1924.8%	\$817,729	\$442,897.1	-2	
7	UnitedHealth Group	\$242,155	7%	\$13,839	15.5%	\$173,889	\$236,555.2	-1	
8	McKesson	\$214,319	2.9%	\$34	-49.3%	\$59,672	\$21,845	-1	
9	AT&T	\$181,193	6.1%	\$13,903	-28.2%	\$551,669	\$209,387.5	-	

Fortune 500	RANK
Walmart	1
<h2>Company information</h2>	
As of 2/1/21	
Country	U.S.
Headquarters	Bentonville, Ark.
Industry	General Merchandisers
CEO	C. Douglas McMillon
Website	https://www.stock.walmart.com
Company Type	Public
Ticker	WMT
Revenues (\$M)	\$548,743
Profits (\$M)	\$19,742
Market Value (\$M)	\$411,690
Employees	2,200,000

Gather HTML from Newsrooms

All press releases

Use any of the filtering options below to narrow your search

Press release keyword search

Year

All years  

Showing 16 of 1133 items



[CVS Health President and CEO reacts to President Biden's COVID-19 update](#)

Press Release - March 11, 2021



[CVS Health now offering COVID-19 vaccines in 29 states](#)

Press Release - March 11, 2021

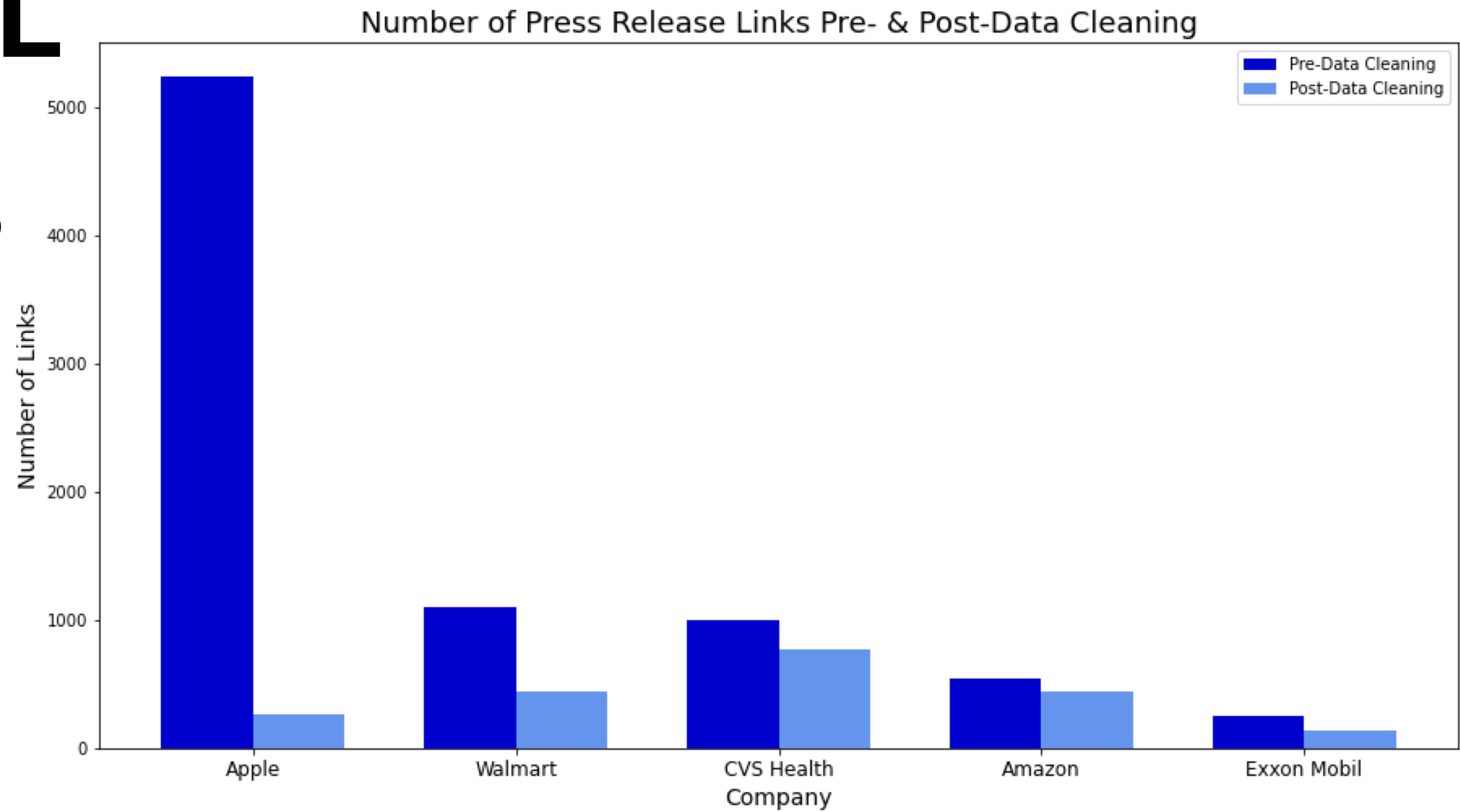
CVS Health announces availability of COVID-19 antibody testing in all MinuteClinic locations in Massachusetts



[CVS Health announces availability of COVID-19 antibody testing in all MinuteClinic locations in Texas](#)



Parse HTML from Files



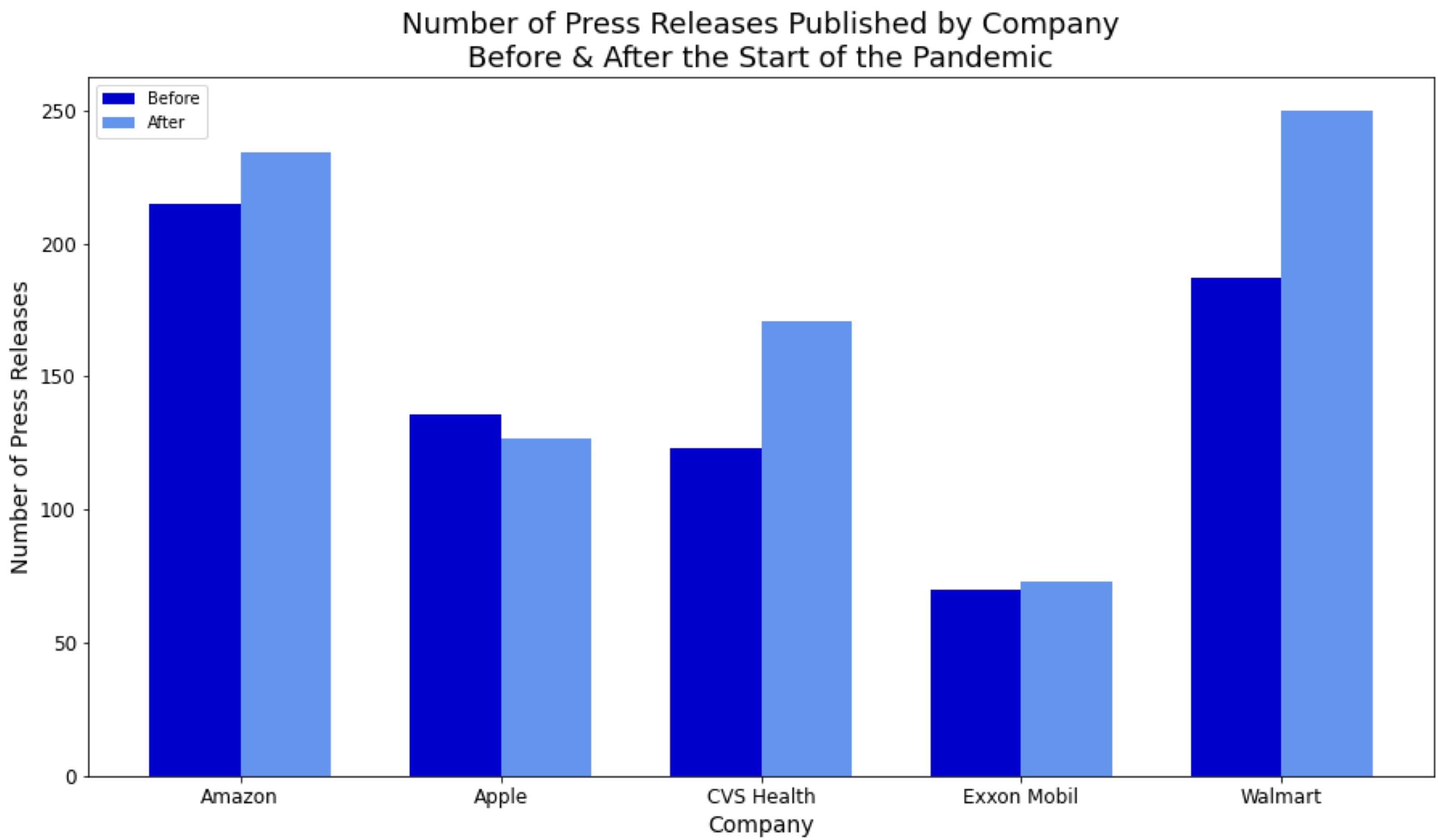
Gather & Clean Press Release Text

List of words removed from press releases:

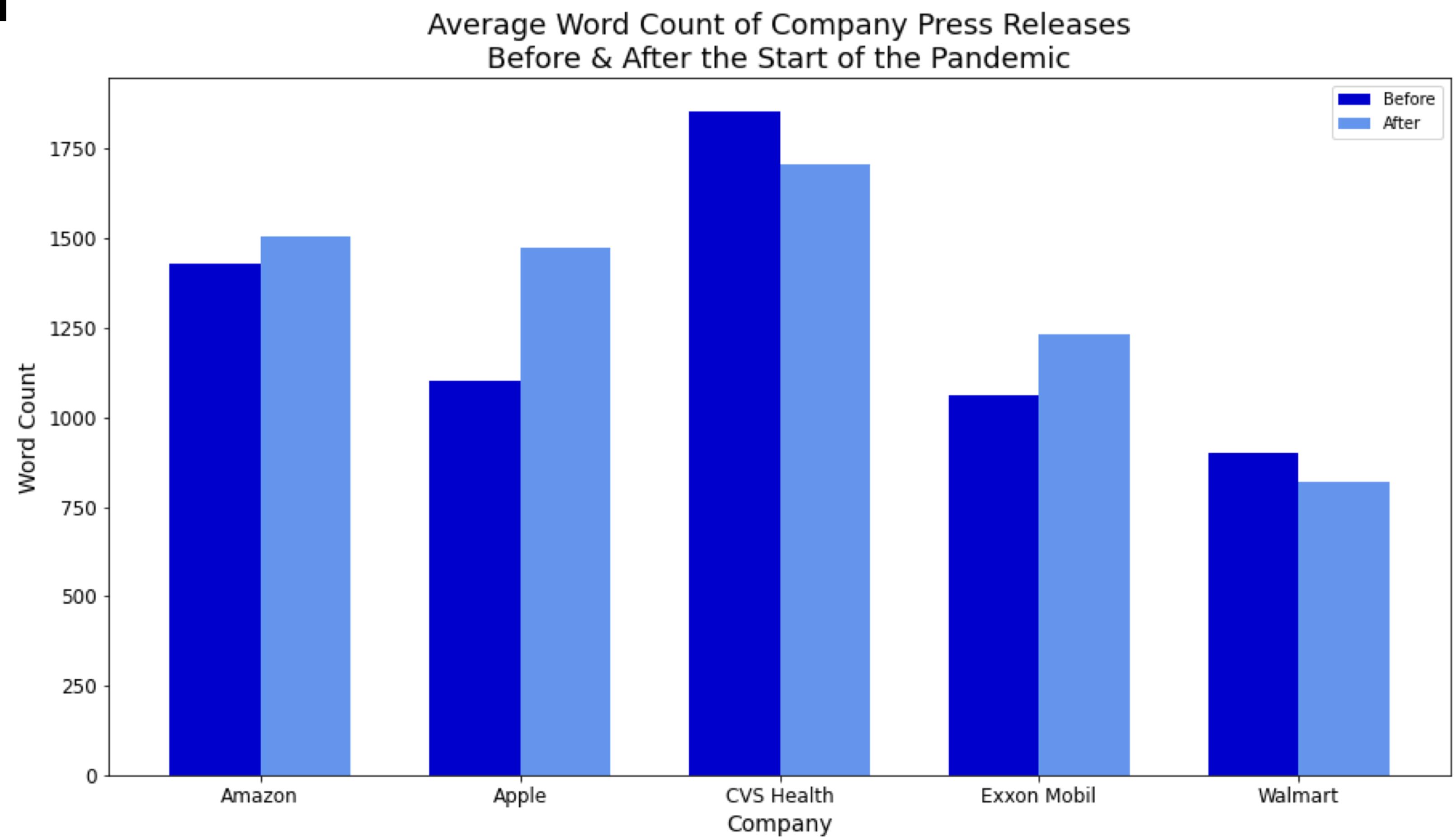
- 2019
- 2020
- 2021
- Covid-19
- Covid
- COVID-19
- COVID
- Coronavirus
- coronavirus
- pandemic
- Amazon
- Apple
- Walmart
- CVS Health
- CVS
- Exxon Mobil
- Exxon
- menu
- newsroom
- blog
- press
- amazon.com
- \n
- \r
- \t

EDA

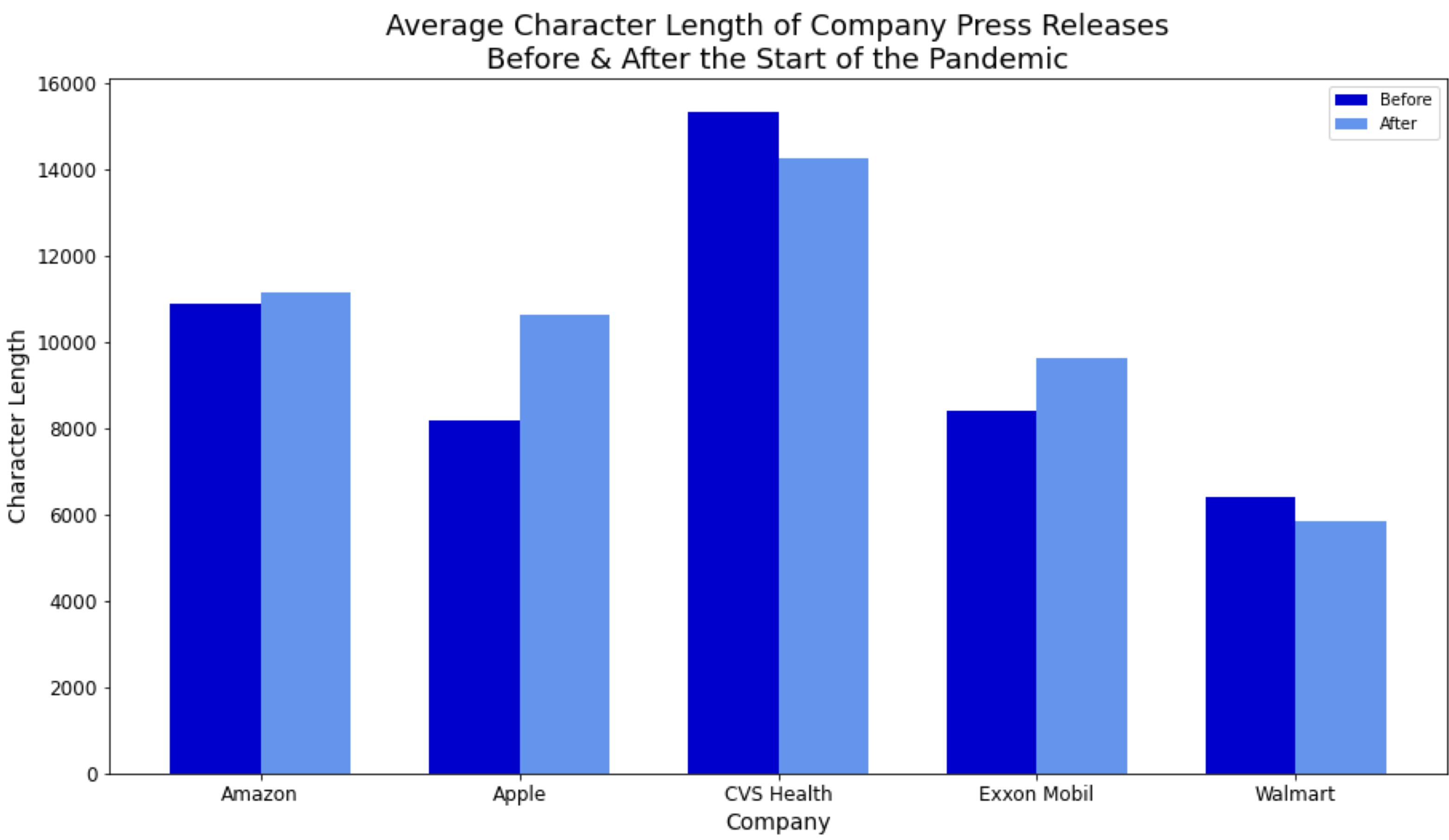
**There have
been slightly
more press
releases
published
after the start
of the
pandemic than
before.**



**On average,
most
companies
have published
longer press
releases by
word count
after the start
of the
pandemic than
before.**

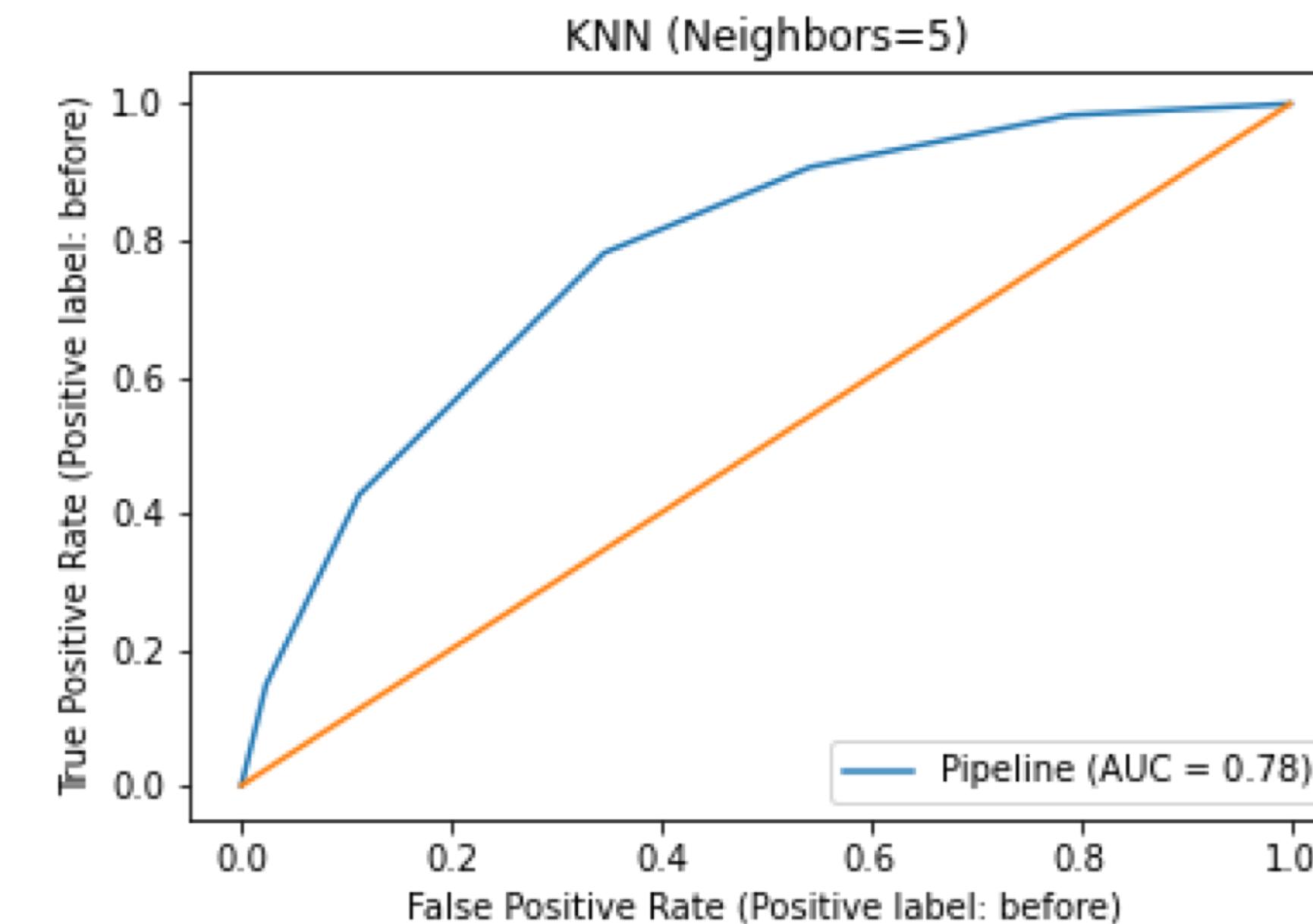
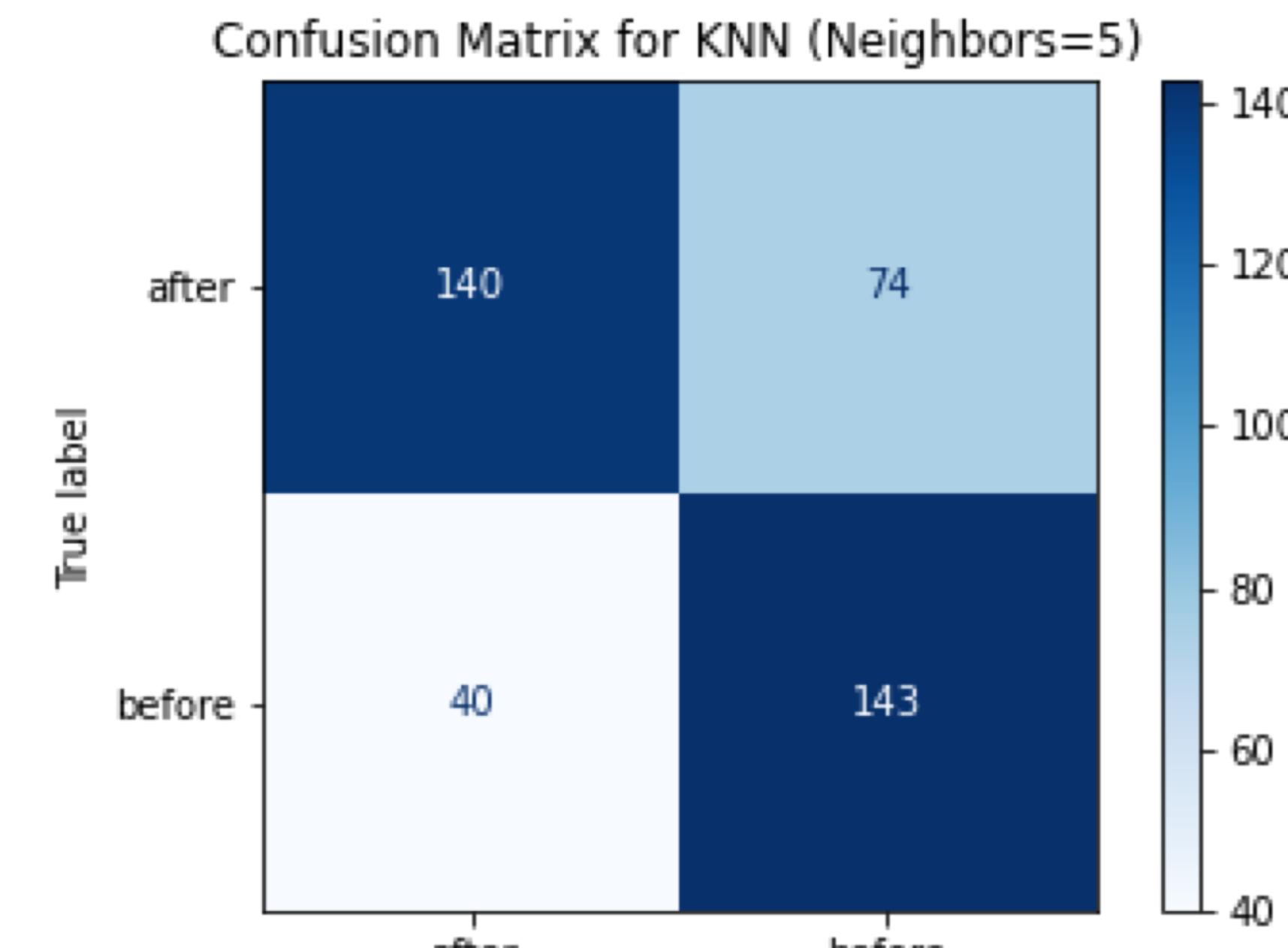


The same
holds true
for
company
press
releases in
terms of
character
length.



Modeling

**With a baseline
of 0.54, KNN
with
n_neighbors
set to 5
performed the
best, achieving
an accuracy
score of 0.71.**



Full Model Results

Model	Hyper-parameter	Train score	Test score	ROC-AUC Score
Logistic Regression	penalty: 'l2'	0.759461	0.654911	0.736581
Logistic Regression	penalty: 'none'	0.921783	0.695214	0.735687
KNearest Neighbors	n_neighbors: 5	0.782169	0.712846	0.776658
KNearest Neighbors	n_neighbors: 8	0.749369	0.672544	0.763188
KNearest Neighbors	n_neighbors: 3	0.831791	0.707808	0.753536

Conclusion

Following a robust data collection and modeling process, and due to all of the models beating the baseline score, companies *have* changed the way they communicate through press releases in significant enough ways that machine learning can detect the difference.

Next Steps

- Explore *how* the press release language changed between the two time periods by exploring the most-used words in each category
- Test additional classification models to see if the model performance improves
- Use GridSearchCV to test more hyperparameters and determine if there are any other changes that could improve the models' performance