

Textual Analysis of NVIDIA in Bloomberg News

Autumn 2024

NHH



Task 1: Corpus

Three corpora for the specific tasks:

Corpus 1 (Task 2)		Corpus 2 (Task 3)		Corpus 3 (Task 4 and 5)	
Metric	Value	Metric	Value	Metric	Value
Nr. of Documents	929	Nr. of Documents	897	Nr. Of Documents	898
Total Words	10137679	Total Words	109059	Total Words	76483
Unique Terms	30894	Unique Terms	11719	Unique Terms	9597
Avg. words per doc	10912	Avg. words per doc.	122	Avg. words per doc.	85
Top terms	Total count	Top terms	Total count	Top ten terms.	Total count
will	132207	nvidia	3710	nvidia	4693
us	85532	market	1007	market	678
think	76660	tech	742	tech	515
going	64302	stocks	653	chips	499
see	60450	company	636	earnings	487
get	57502	companies	618	apple	487
year	56949	earnings	599	stocks	461
market	55951	chips	582	companies	454
can	54488	lot	529	scanned	446
now	49588	china	525	company	429

Table 1. The three corpora.

Task 2

The most cosine similar companies during 2016-2024

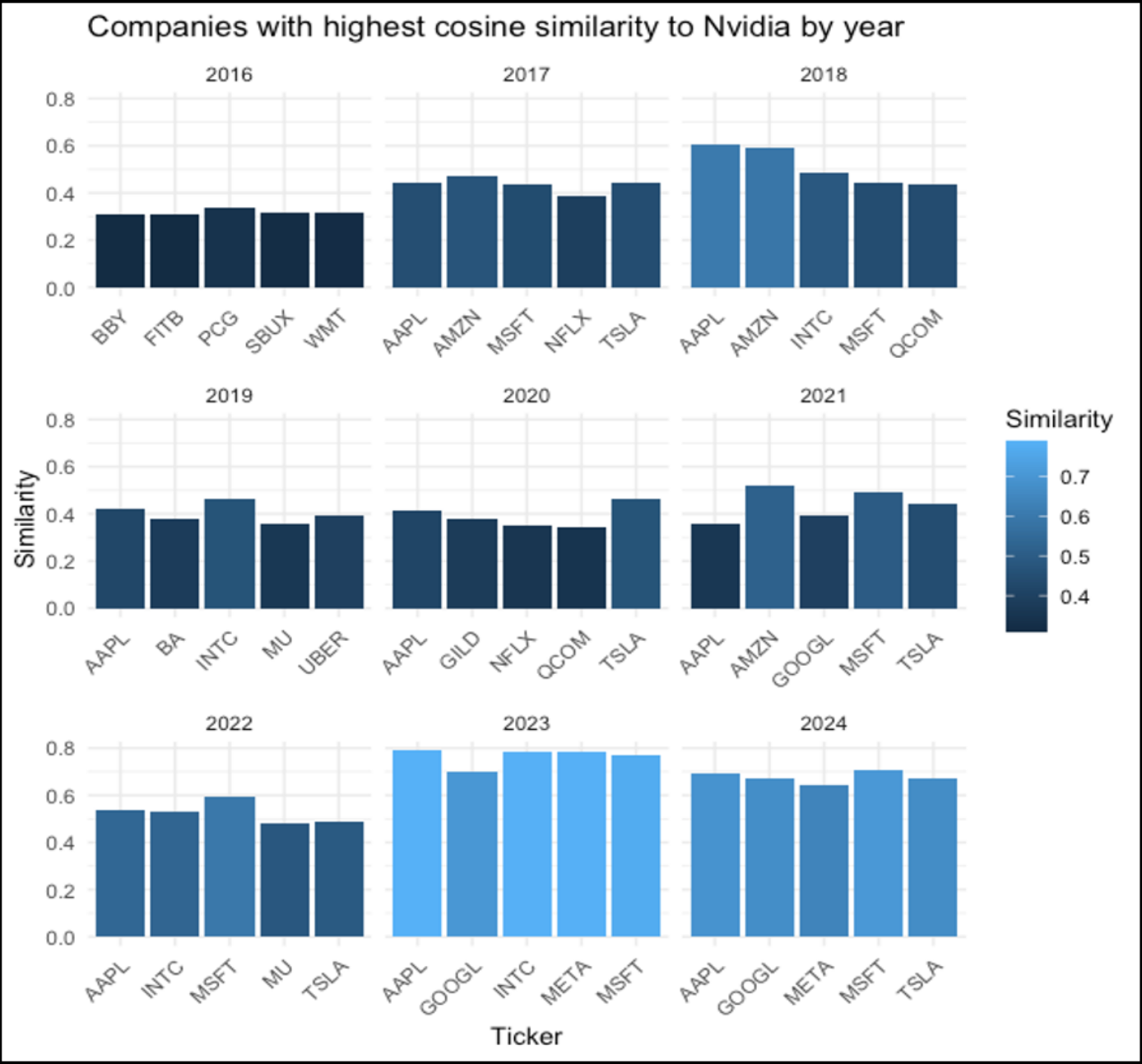


Term	Similarity	Ticker	Company
apple	0,836	AAPL	APPLE INC
intel	0,831	INTC	INTEL CORP
tesla	0,780	TSLA	TESLA INC
microsoft	0,776	MSFT	MICROSOFT CORP
amazon	0,712	AMZN	AMAZON.COM INC
micron	0,710	MU	MICRON TECHNOLOGY INC
meta	0,696	META	META PLATFORMS INC CLASS A
alphabet	0,696	GOOGL	ALPHABET INC CL A & CL C
qualcomm	0,620	QCOM	QUALCOMM INC
broadcom	0,610	AVGO	BROADCOM INC

Table 2. SP500 companies (Nov 2024) mentioned in the corpus ranked by cosine similarity. I with Corpus 1.

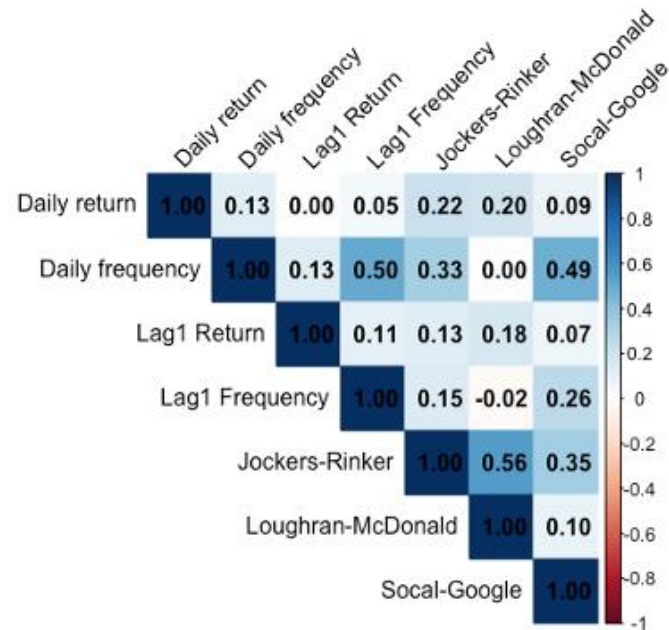
Task 2

The changes in most similar or related companies throughout the years 2016-2024





Task 2 NVIDIA's Daily Sentiment, Frequency and Stock Price and correlation



Daily return as a dependent variable of Sentiment and Frequency?

A linear regression with interaction terms.

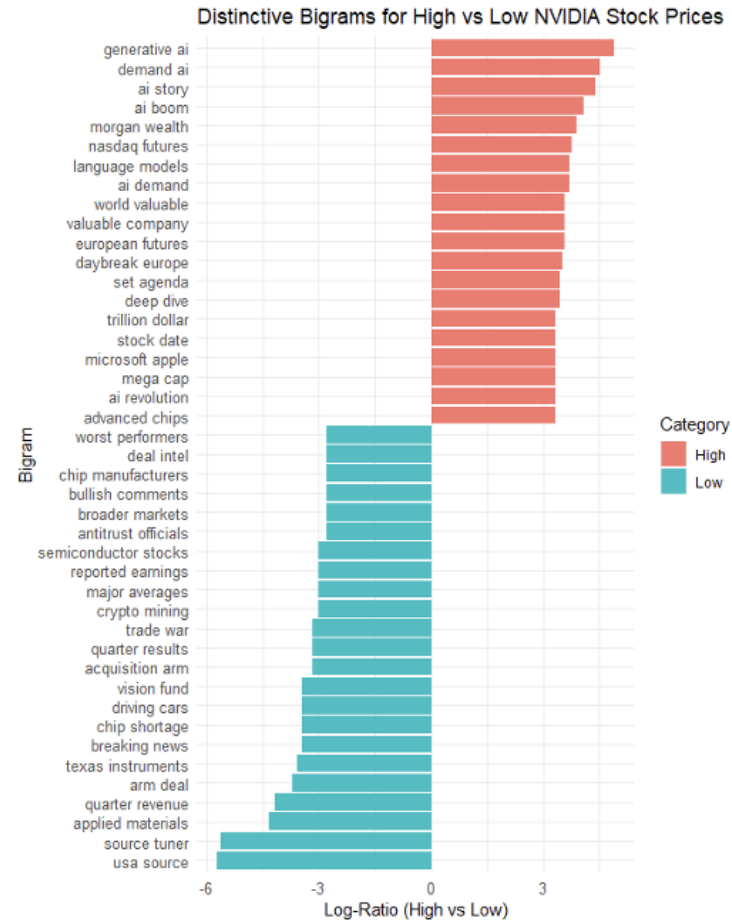
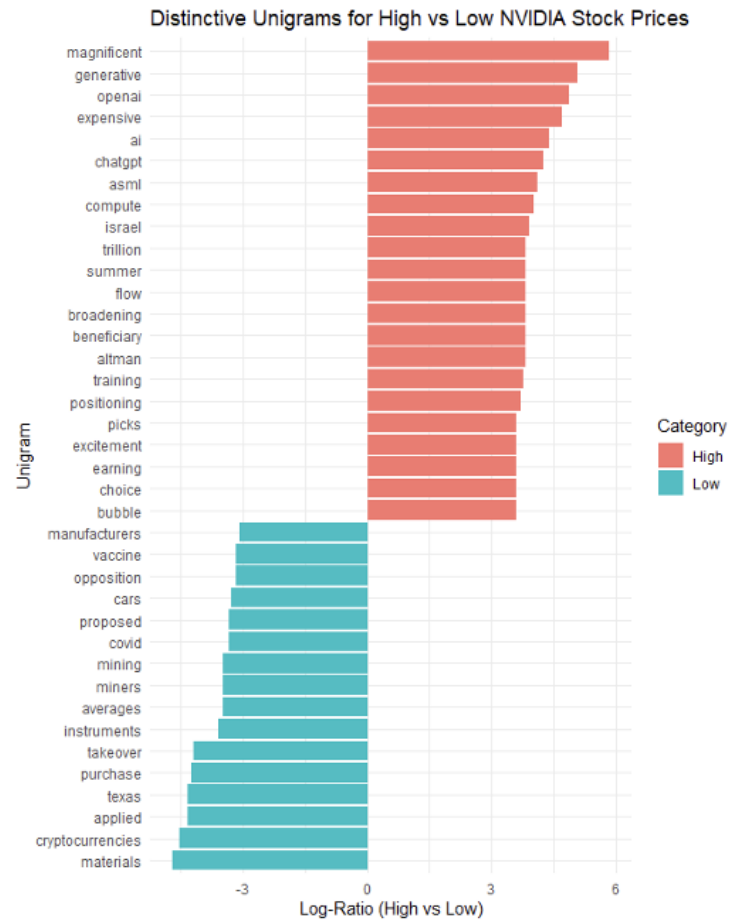
NHH



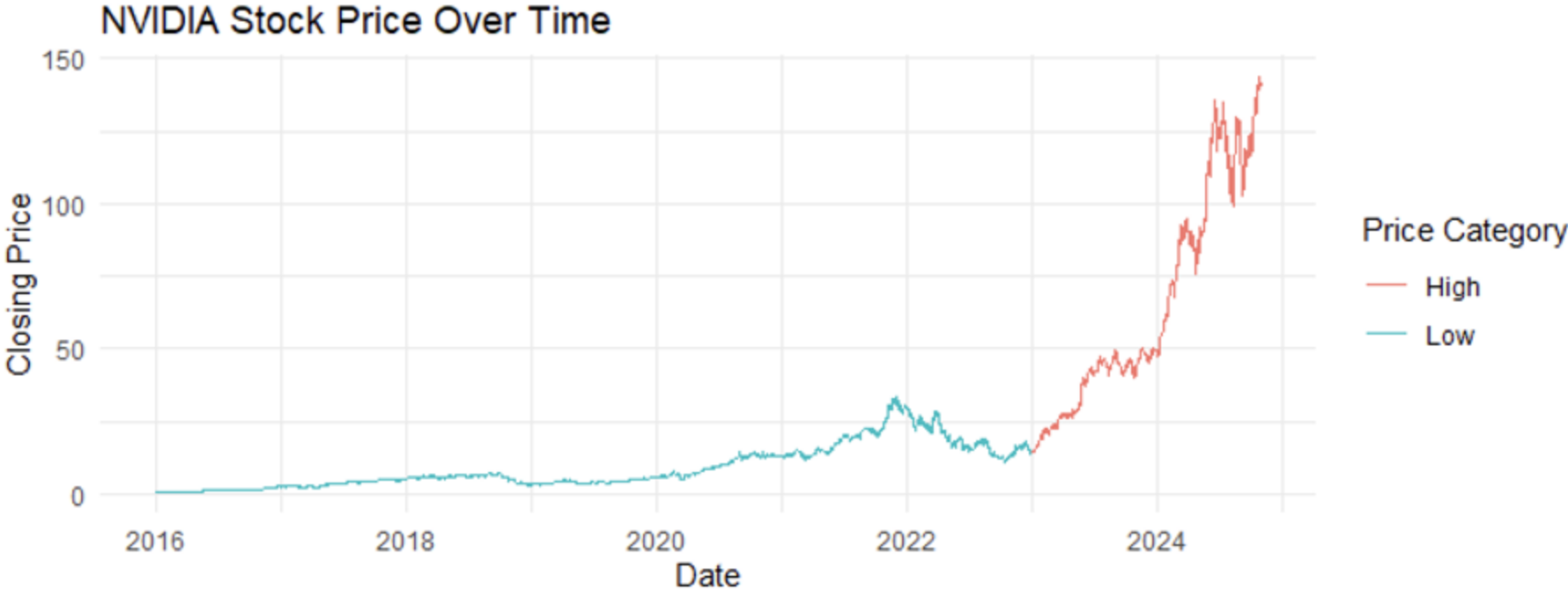
Linear regression on <i>Daily return</i> (dependent variable) using Corpus 2. Including interaction terms between daily frequency and sentiment.					
Term	Estimate	Std.error	t statistic	P value	Sign. Level
(Intercept)	0,00067	0,0030	0,23	0,82	
Daily frequency	0,00044	0,0007	0,64	0,52	
Jockers-Rinker	0,02834	0,0061	4,64	0,00	***
Loughran-Mcdonald	-0,00936	0,0070	-1,33	0,18	
Socal-Google	-0,00405	0,0023	-1,75	0,08	.
Daily frequency : Jockers-Rinker	-0,00269	0,0009	-2,95	0,00	**
Daily frequency : Loughran-Mcdonald	0,00456	0,0011	4,20	0,00	***
Daily frequency : Socal-Google	0,00067	0,0003	1,95	0,05	.

Table 2. Significance codes: '***' = 0.001, '**' = 0.01, '*' = 0.05, '.' = 0.1, ' ' = 1.

Task 4



Stock price



Task 5

Topics from LDA Topic Model (5 selected out of 8)

Shows clear

Topic 1



Topic 6



Topic 4



Topic 5



Topic 8



LDA Topic Distribution Across Years

