PRE-IPO SENTIMENT ANALYSIS
COS 397 S02: NLP

Daniel Hu, advised by Dr. Christiane Fellbaum

AGENDA

Motivation and Goal

Problem Background, Related Work

Approach

Implementation
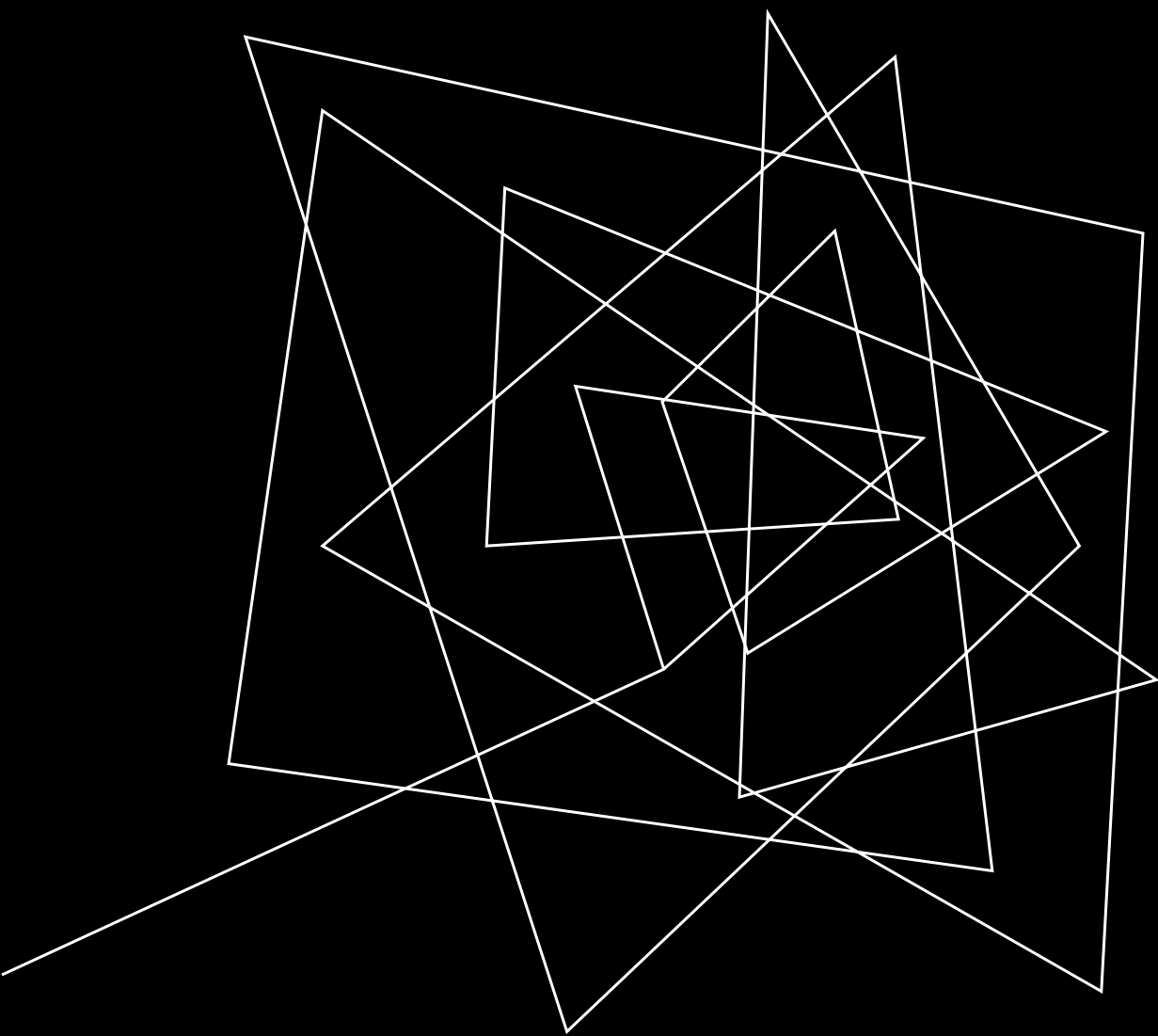
Results

Conclusion

MOTIVATION AND GOAL

- My interest in startups and IPO's specifically

- Applying sentiment analysis to the financial world

- Understand the relationship between social media and the financial world

PROBLEM BACKGROUND, RELATED WORK

- Cornelli, Goldreich, Ljungqvist: "Investor Sentiment and Pre-IPO Markets, looking at **European pre-IPO markets**

- Guilherme: "Investing in Stock IPO's with Sentiment Analysis from Twitter," looking at **text mining algorithms**

MY APPROACH

- Deep cleaning data: stemming, removing artifacts (stopwords, non-English words, URLs, non-alphanumerics, and non-empty tweets),

- Evaluating based on two models: FINBERT and VADER

- Comparing day gains and week gains to determine accuracy of pre-IPO sentiment after the IPO occurs

IMPLEMENTATION

Challenges, tools, and what remains to be done

CHALLENGES

RUNTIME

- It took hours to run cells: bugs, or long tasks?

- Importance of providing visual indicators for myself

CODE MODULARITY

- Began by running cells repeatedly

- Concluded by having methods for processing and analyzing data

DATA QUALITY

- Tweets are a flawed source of data: short and often corrupted

- Compensate with volume: runtime tradeoff

TOOLS

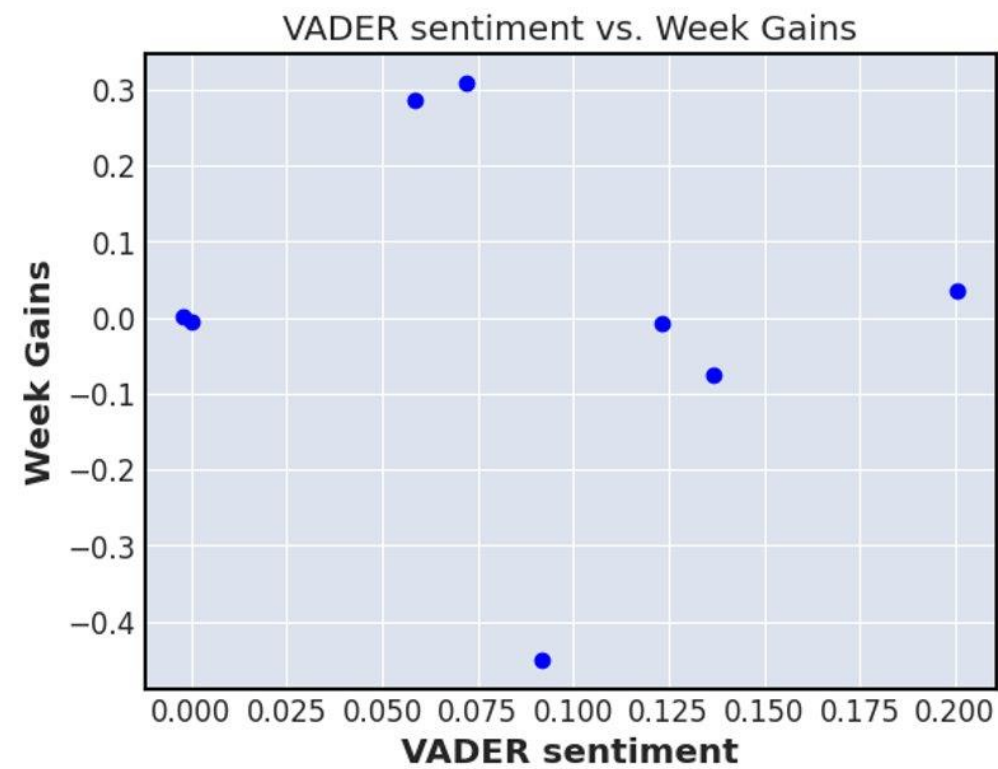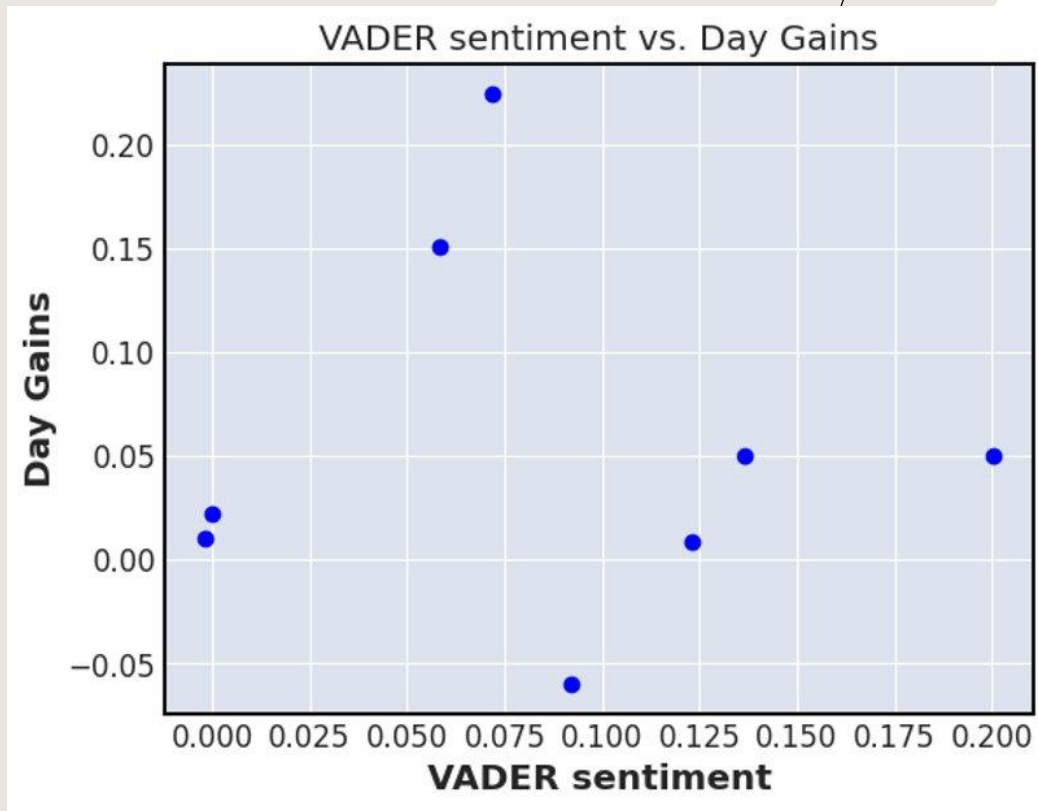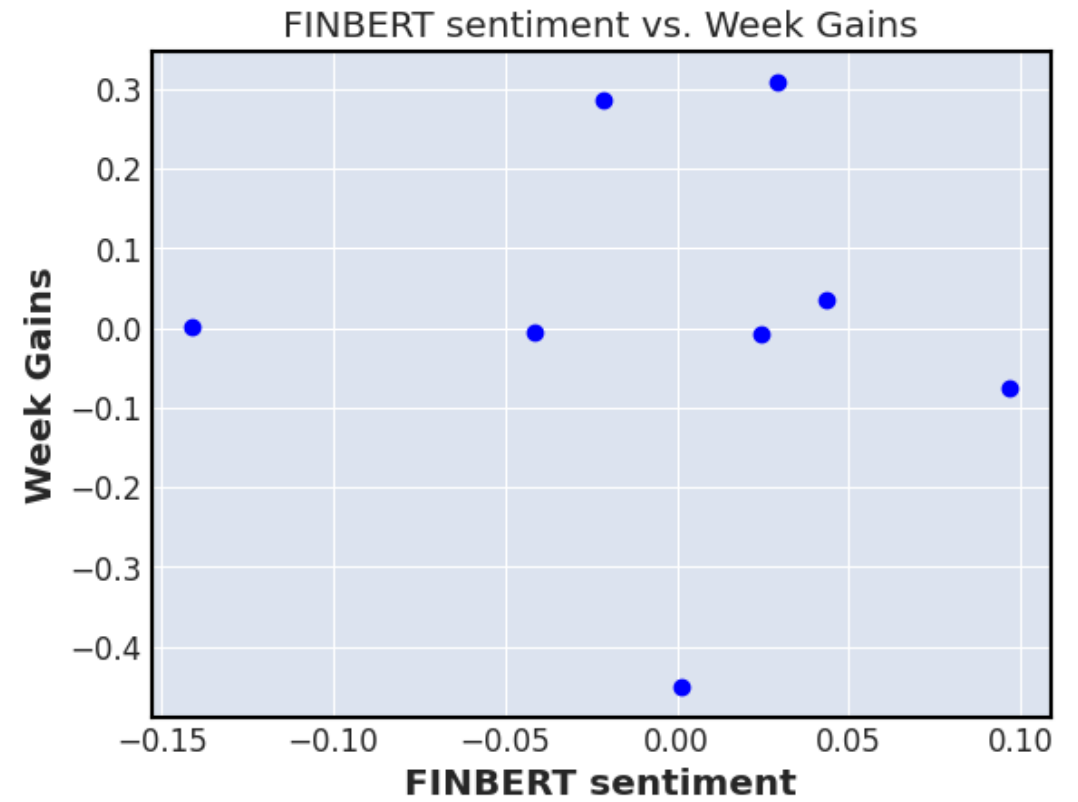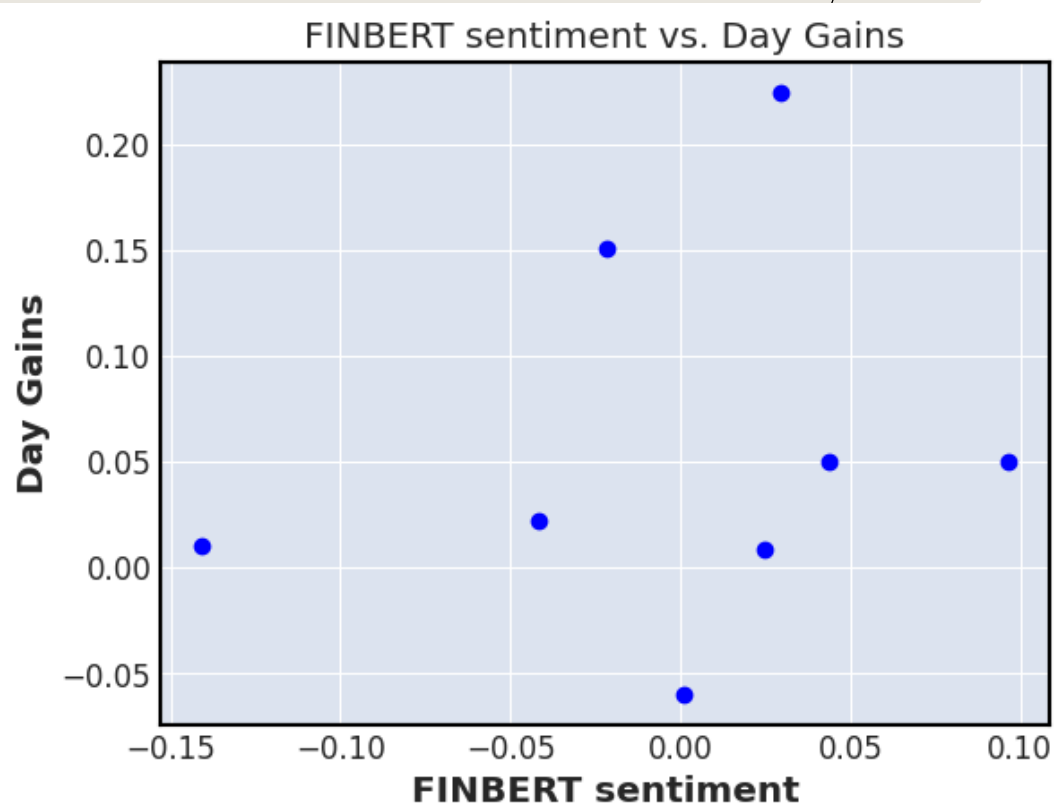| PANDAS	NLTK	VADER/FINBERT	MATPLOTLIB
Accessible and dynamic dataframes	Tools for processing data	Models with pros and cons	Go-to plotting library

WHAT REMAINS TO BE DONE

- Managing of large datasets

- Further examination of day-over-day trends in sentiment before the IPO

- Attempts at training my own model

RESULTS

DATA INSIGHTS

	FINBERT	VADER	DAY GAINS	WEEK GAINS
BANCO SANTANDER	-0.0417	0.0	2.18%	-0.6%
DIDI	-0.0215	0.0584	15.08%	28.46%
GM	0.0437	0.2005	5.00%	3.43%
UBER	-0.141	-0.0020	1.02%	0.214%

VADER sentiment vs. Day Gains

VADER sentiment vs. Week Gains

FINBERT sentiment vs. Day Gains

FINBERT sentiment vs. Week Gains

CONCLUDING THOUGHTS

- Nothing entirely conclusive, need to examine circumstances more closely

- Appreciation for the challenges of data analysis, with most of the effort coming from there

- A desire to build and train my own models

ACKNOWLEDGEMENTS

Thank you, Dr. Fellbaum, for all your help, and

my friends and family, for all your support!