Sentiment Analysis on Earnings Calls Transcripts

Group 2:

Loïc Nicollerat
Francesco Pio Numero
Teo De Xuan Justin
David Vencato

EPFL - Lausanne

29th May 2024

EPFL

Problem

Can one elaborate a profitable trading strategy based on earning calls by analyzing their sentiments?

Dataset - Summary Statistics

The dataset consists of earnings call transcripts.

Each row includes: date, exchange, ticker, quarter, full transcript.

- **Total Number of Transcripts:** 17.221.
- **Date Range:** 2019 to 2023.
- **Number of Unique Tickers:** 2.876.

Number of Unique Tickers

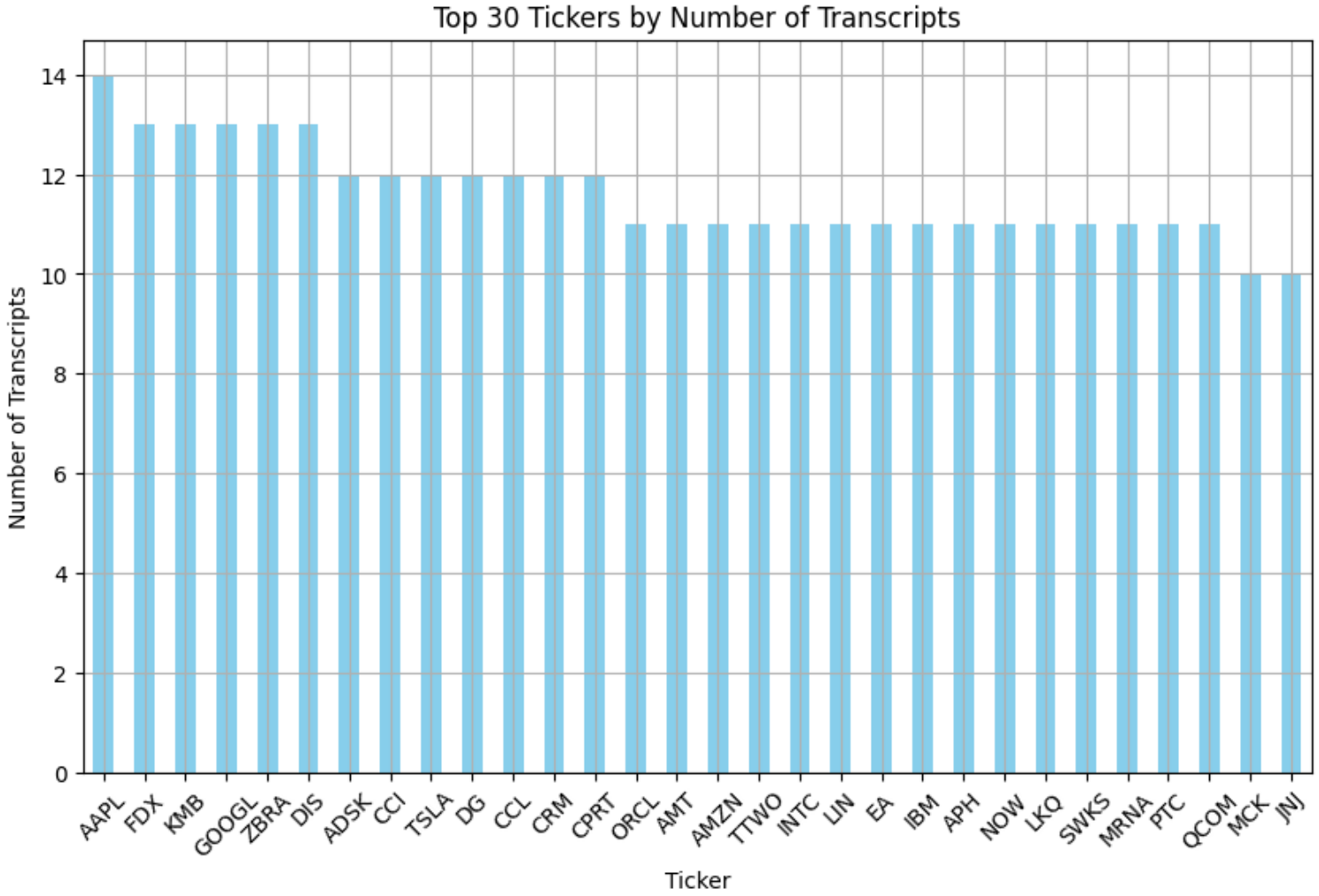


Figure: Distribution of Unique Tickers.

Data Preprocessing

- NYSE and NASDAQ are the most relevant exchanges by number of transcripts ($\approx$ 98% of transcripts).
- Split transcripts by year and extract calls belonging to companies in the S&P 500 during that year.

Advantages:

1) Stocks have higher trading volumes.
2) Computation time is quicker without loss of relevant information.

Date Range

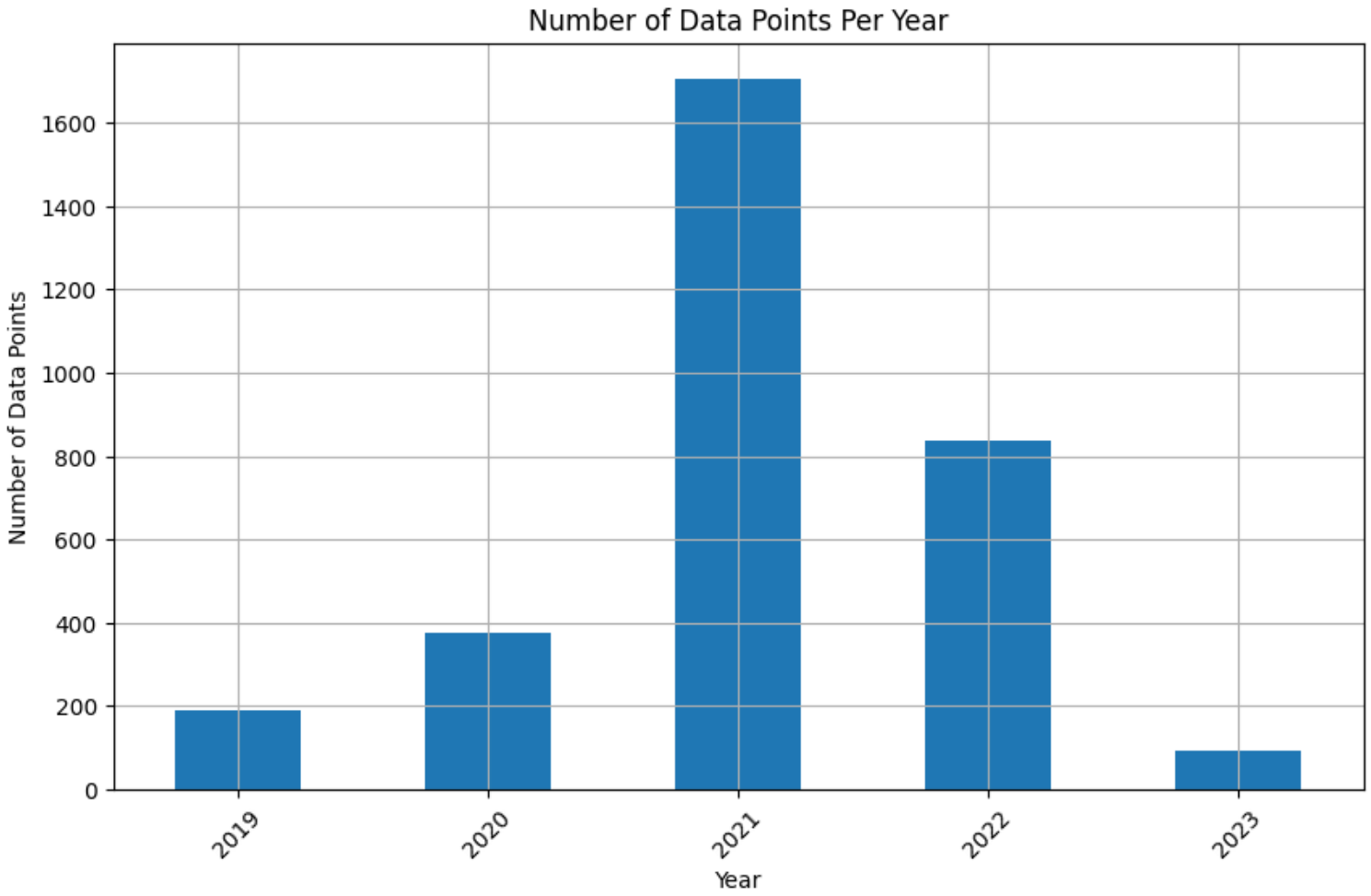


Figure: Distribution of Transcripts of S&P500's Companies.

Length of Transcripts

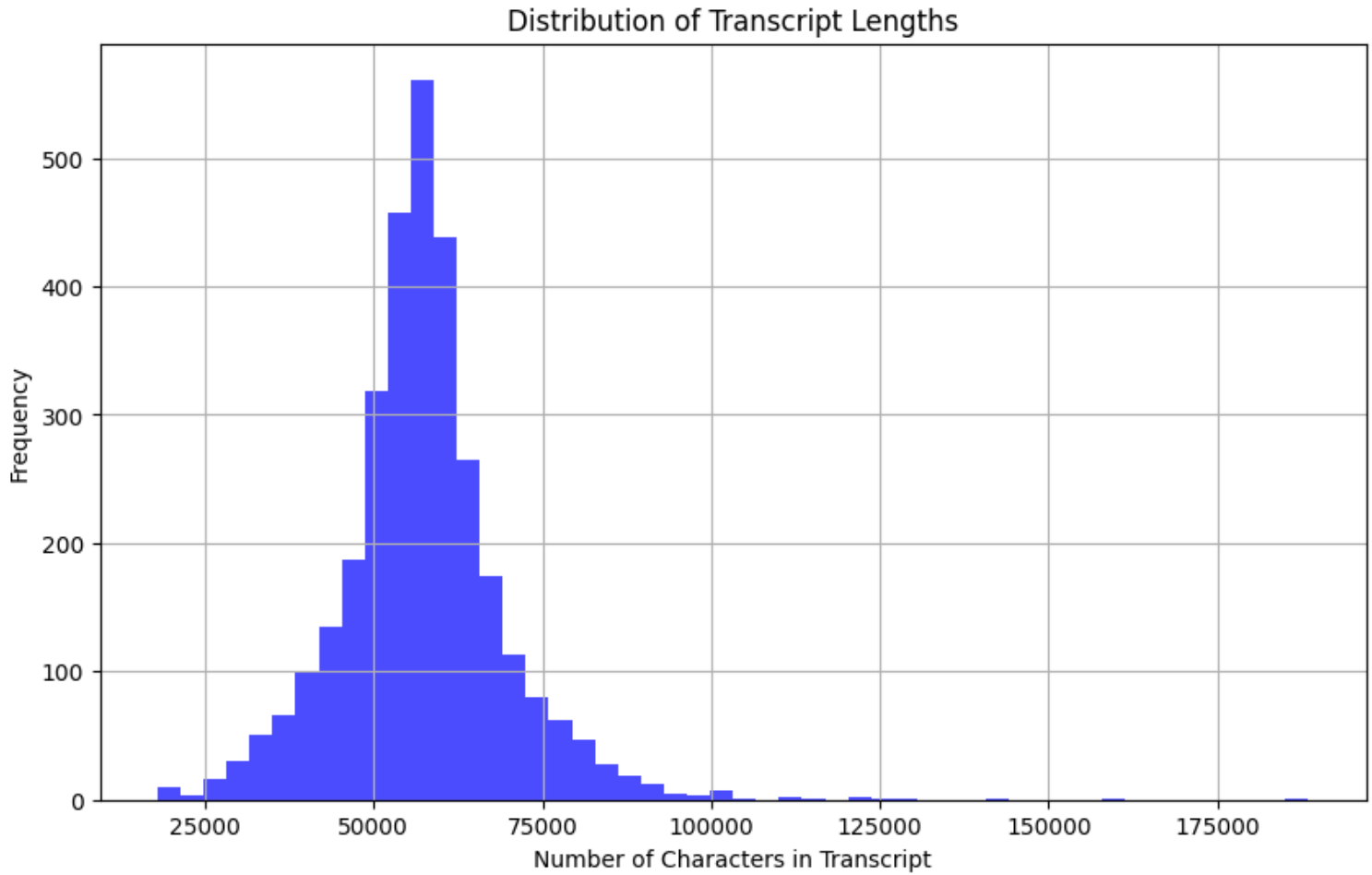


Figure: Distribution of Transcript Lengths.

Label the data

The methods used to label the data are:

- K-means clustering
- Lexicon Based Analysis
- Topic Modeling

K-means Clustering

K-means Clustering

- Vectorization by TF-IDF.
- Choose the hyperparameter K using the “elbow method”.
- Evaluate clusters with the Silhouette score.

Elbow Method

- **Elbow Method:** look for the "elbow" point where the rate of decrease slows.

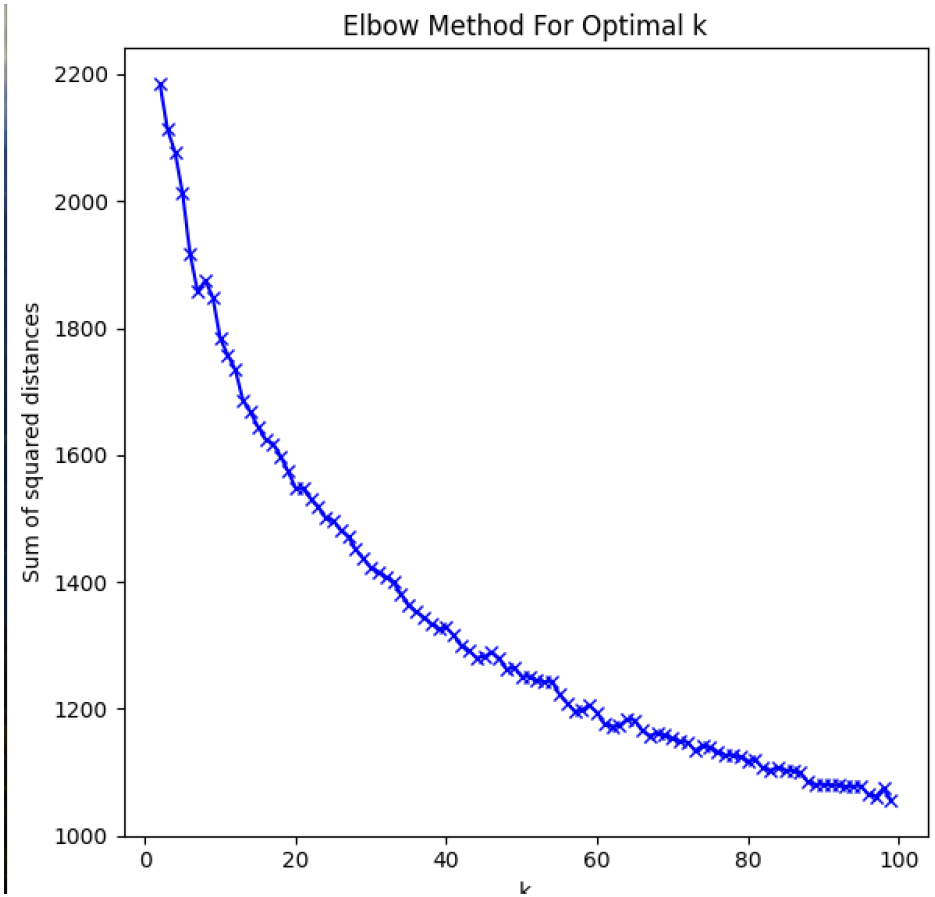


Figure: Sum of Squared Distances vs. K.

Silhouette Score

- **Silhouette Score:** higher scores indicate better-defined clusters.

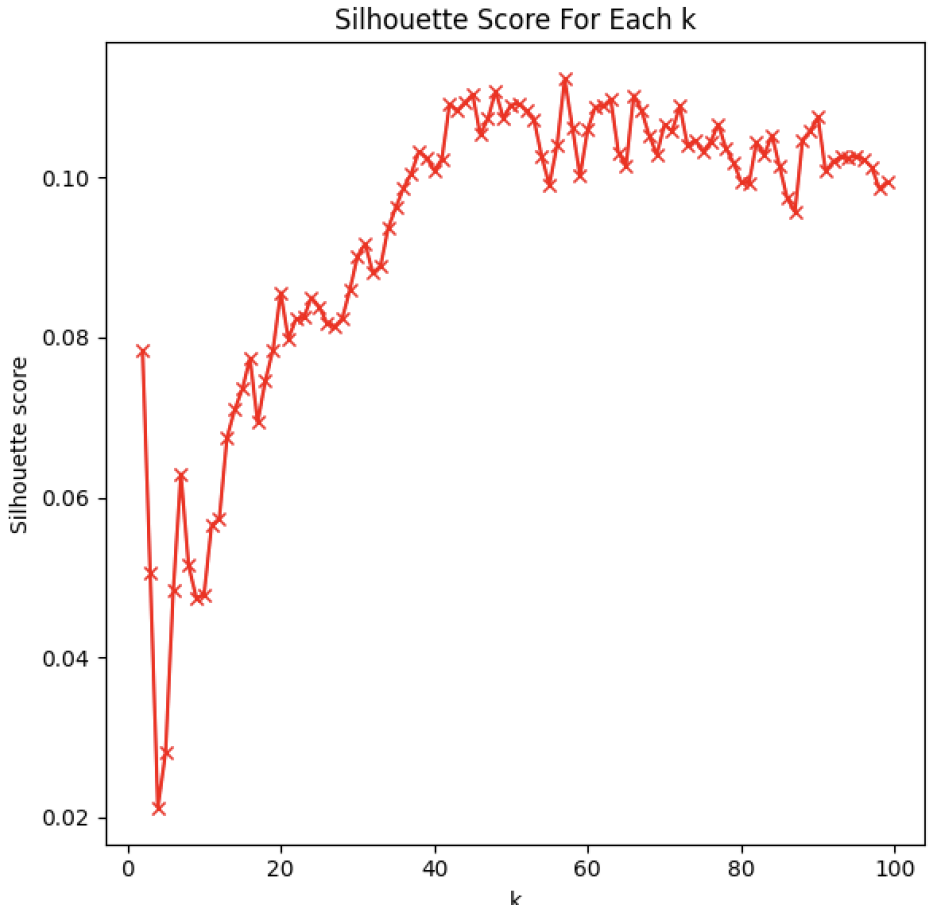


Figure: Silhouette Score vs. K.

Challenges of Clustering Earnings Call Transcripts

- **High Dimensionality:**
  - TF-IDF vectorization often results in a high-dimensional feature space.
  - Difficulty for K-means to find meaningful clusters.
- **Similar Vocabulary:**
  - Transcripts typically use a similar set of vocabulary and jargon.
  - Low variance across documents, making clusters less distinct.
- **Contextual Nuances:**
  - Transcripts contain nuanced language and context-specific meanings.
  - K-means relies on Euclidean distance, which does not capture semantic similarities well.

EPFL

Lexicon-based analysis

Lexicon-based analysis

- Use of Loughran-McDonald as dictionary.
- To each transcript is associated a "sentiment" vector.

	Pos.	**Uncert.**	**Neg.**	**Litig.**	**Constr.**
Doc1	110	55	48	12	6

Table: Sentiment vector of transcript 1.

E.g.

$$\text{Pos.} = \sum_{\substack{\omega \text{ word} \\ \omega \in \text{ Doc1}}} \mathbb{1}_{\{\text{pos. in L}-\text{M}\}}(\omega)$$

- Different weights for each sentiment:
  weights={ Neg: -1, Pos: 1, Uncert.: -0.5, Litig.: -0.7, Constr.: -0.8 }.
- Label = weighted mean of vector entries.

Example of sentiments for ten transcripts in 2019

	date	exchange	positive	uncertainty	constraining	litigious	negative	weighted_score
0	2019-10-23	NASDAQ: LRCX	143	79	10.0	15.0	71	0.044025
1	2019-12-04	NYSE: HRB	110	55	6.0	12.0	48	0.092208
2	2019-07-24	NASDAQ: XLNX	93	108	15.0	8.0	107	-0.258610
3	2019-06-21	NYSE: KMX	131	50	4.0	5.0	77	0.083521
4	2019-08-01	NASDAQ: QRVO	128	85	16.0	7.0	83	-0.047649
5	2019-09-25	NYSE: NKE	237	64	5.0	16.0	77	0.282707
6	2019-04-25	NASDAQ: MAT	169	82	7.0	15.0	95	0.045924
7	2019-12-12	NASDAQ: COST	111	94	2.0	9.0	76	-0.068151
8	2019-10-29	NASDAQ: MAT	191	62	6.0	11.0	87	0.169468
9	2019-08-28	NYSE: HRB	111	65	6.0	2.0	64	0.033468

Strengths and Weaknesses of LBA

- **Strengths**
  - Standard method to label, effective even with limited data.
  - Weighted scores between -1 and 1, facilitating comparison.
  - Different weights to different sentiments $\Rightarrow$ improving analysis granularity.
- **Weaknesses**
  - Struggles with understanding context.
  - Earning transcript calls are generally positive $\Rightarrow$ hard to distinguish finer sentiment nuances.
  - Equal weight to all words.

EPFL

Topic Modelling

Topic Modelling

Two steps:

1) LDA: get 1 topic with 20 words for each document. Example:

```
Document 1.
Topic:  year, think, analyst, well, richard, executive,
quarter, weve, thing, last, galanti, million,
financial, vice, officer, chief, president, basis,
good, like.
```

2) Vader on topics.

   Advantages:

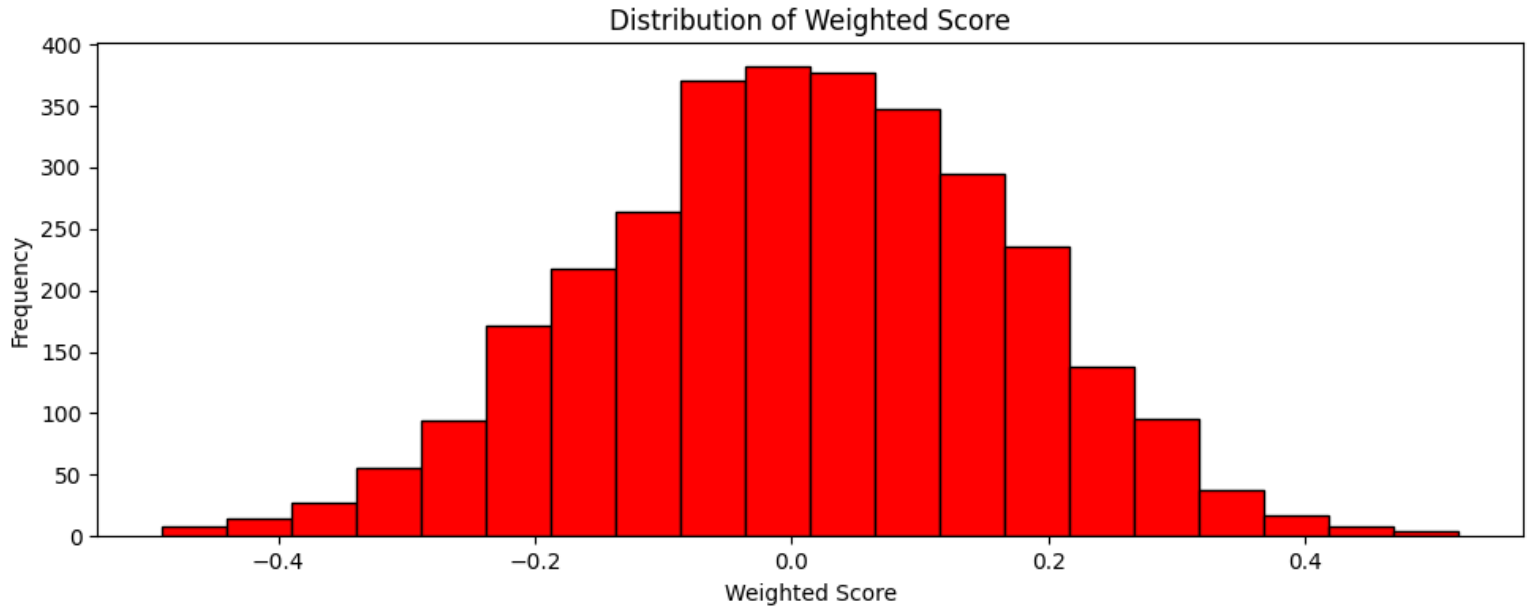
   - the previous weighted score does not work well with small number of words (sparse variance in labels);
   - Vader uses "intensifiers", "negations", "polarity" to obtaine more heterogeneity in documents with few words.

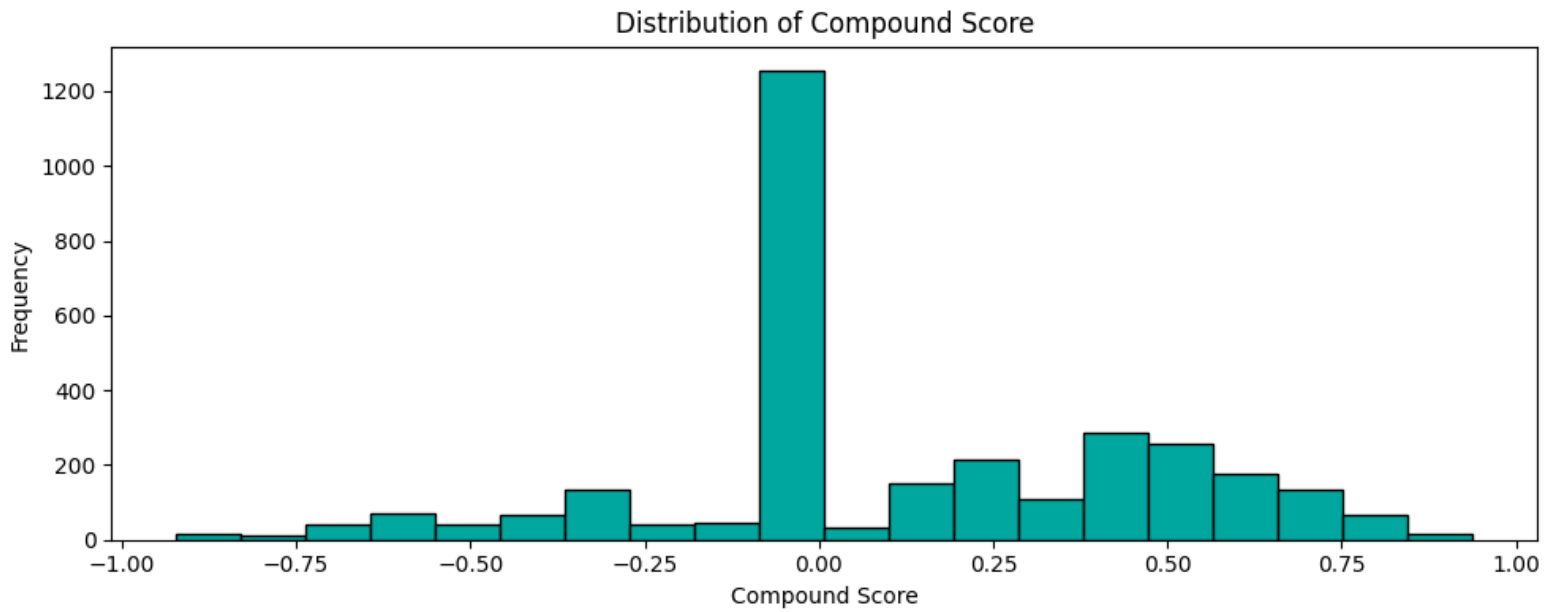
Strengths and Weaknesses of Topic Modelling

- **Strengths**
  - Captures underlying topics to understand document's content.
  - Heterogeneity of results.
  - Compound score $\in [-1, 1]$.
- **Weaknesses**
  - Vader is not specifically optimized for financial texts.
  - Complex and computationally expensive.
  - Harder to interpret the results, especially the topics.

EPFL

Weighted and Compound Score Distribution





Portfolio Strategy

Trading Framework

Starting point:

- earlier-collected dataset of tickers + weighted score, compound score;
- initial balance set at **25,000 USD**.

Setting

- Trade within available timeframe (2019 to 2023).
- Long-only and self-financing portfolio.
- Invest at most 10% of portfolio value in a long position by default.
- Compare strategies with *Boglehead Strategy*-buy and hold SPY.
- Test various strategies through multiple runs.

EPFL

Run 1 - Making Buy Decisions

Decisions using Weighted score/Compound score

- Strategy 1: Buy if the *weighted score* is above 0.3.
- Strategy 2: Buy if the *compound score* is above 0.5.
- Strategy 3: Buy only if the *weighted score* is above 0.3 and the *compound score* is above 0.5.

EPFL

Run 1 - Results

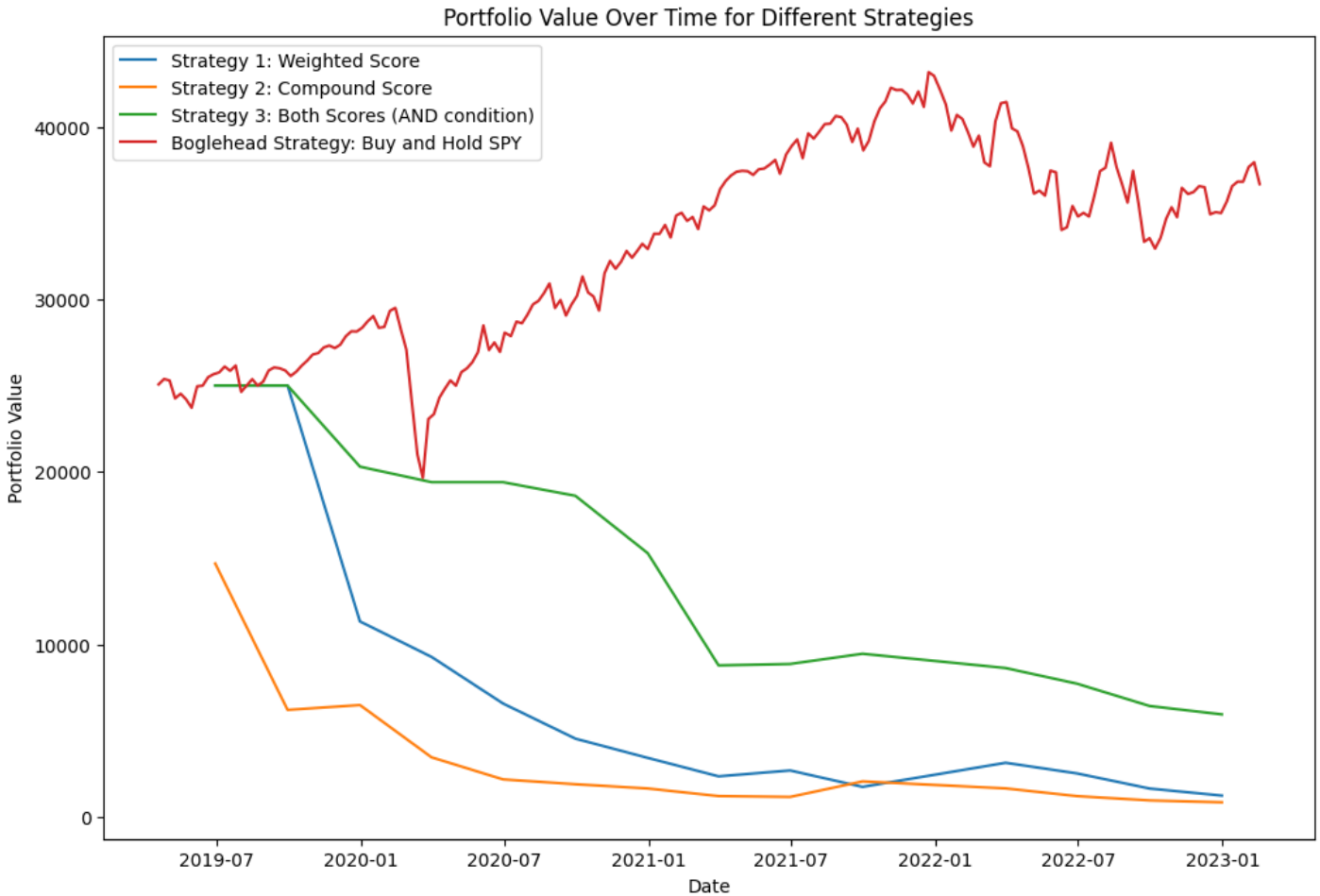


Figure: Strategies underperform SPY. 3 stronger than 1 and 2.

Other Trading Rules Added & Rationale

Dealing with Sentiment Turnaround

If *weighted_score* is below -0.2 or *compound_score* is below 0,

$\Rightarrow$ "Sell the stock" at close the day after a future quarter's earnings

Buying Stocks with more Positive Sentiment:

If a quarter has passed since owning the stock and another stock has an earnings report with more than $+0.2$ difference in *weighted score* or *compound score*:

$\Rightarrow$ Sell the earliest-held position's stock at close the day after the other stock's earnings call and buy the other stock at the same time.

EPFL

Run 2 & 3 - Other Trading Rules

Inclusion of Take Profit and Loss Orders

- Strategy 4 - 10: incorporate Strategy 3 + introduce different TP, SL values.

Lower SL values, Delayed Buying

- Strategy 11, 12: test lower SL values.
- Strategy 13, 14: test Hypothesis - let the market settle, then buy stocks with positive sentiment.

Run 2 - Result

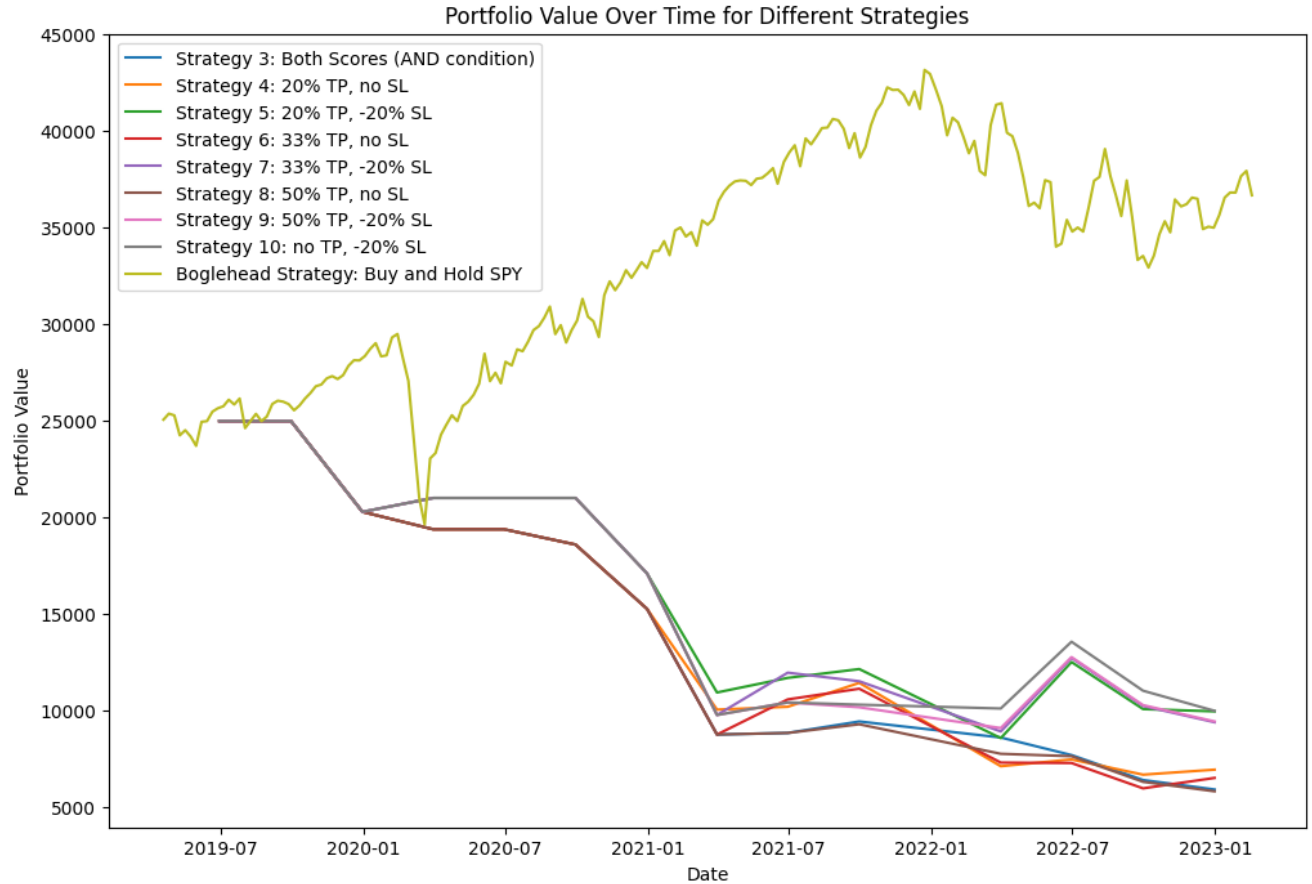


Figure: General underperformance. Strategies with TP/SL outperform baseline.

Run 3 - Result

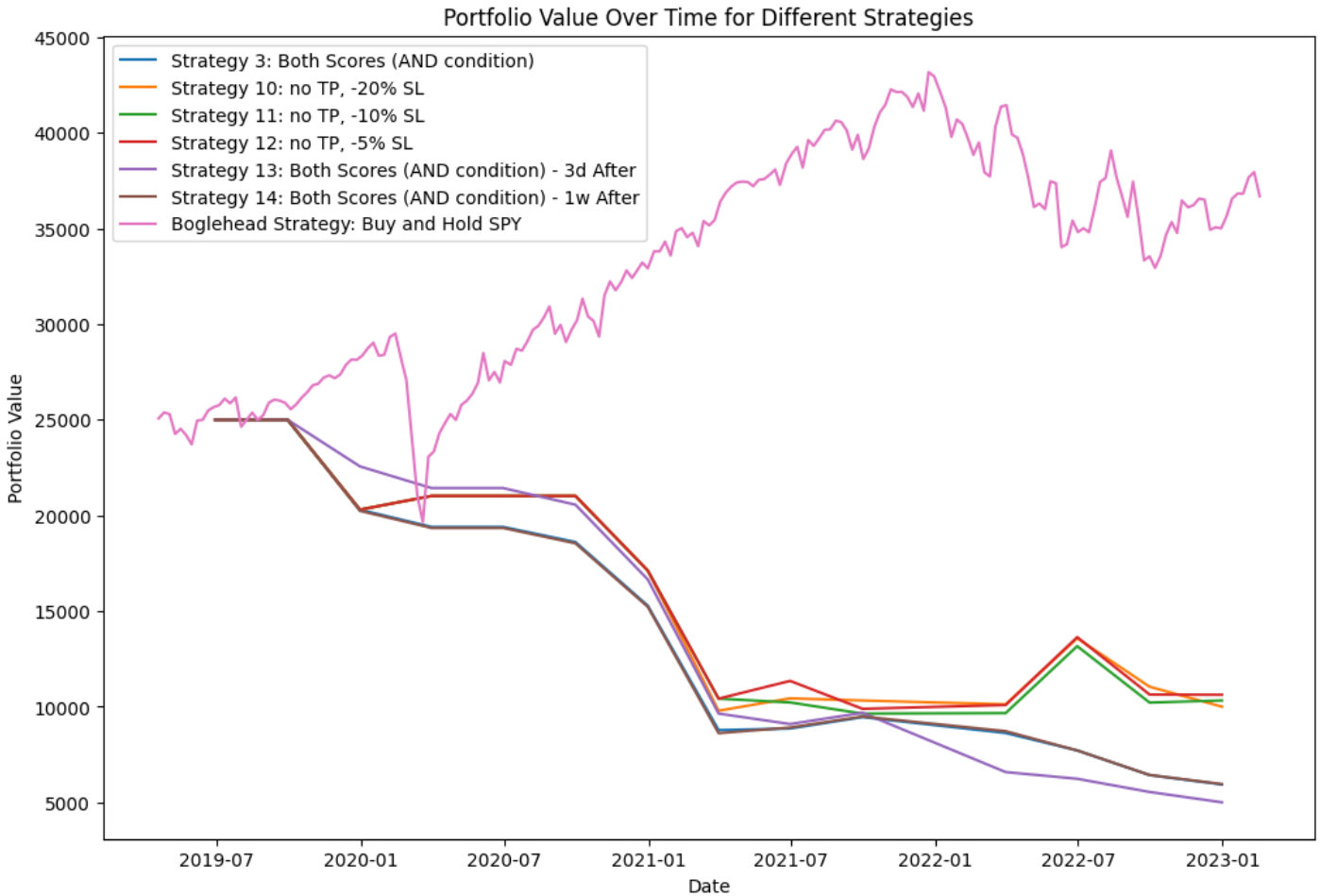


Figure: No substantial outperformance of strategies.

Remarks on Trading Strategies & Conclusion

Issues

- Positive earnings may not lead to positive price action and vice-versa.
- Basing strategy on Sentiment Analysis may not lead to robustness in trading.

Future Directions

$\Rightarrow$ Consider sector sentiment, whole market sentiment and factor into decision-making for trading.

$\Rightarrow$ Incorporate trading and investing concepts (e.g. volume, moving averages, valuation) into an overall strategy - better to buy undervalued stocks with positive sentiment.

$\Rightarrow$ Consider longer time-frames for holding stock (e.g. years).

$\Rightarrow$ Consider inversing sentiment (i.e. buying fear, selling greed).

EPFL

Thank you!