

# What are Companies Thinking About

Can we pick it up from Earnings calls?





**Goal: Large Trends**

Inflation Privacy Advertisement  
COVID  
Port Congestion Vaccine



# Earnings Calls

A public company's presentation to investors and general market about their financial results of a reporting period and future plans

At the end of every quarter

Every public company does one

After publishing quarter results

Consists of:

- Prepared Statement
- Q&A





# Listening to Calls is Hard

“So last year was about putting a stake in the ground for where we're heading, and this year is going to be about executing.” – Meta

“We've provided reconciliations to the most directly comparable GAAP financial measures in the shareholder letter posted to our Investor Relations website. These non-GAAP measures are not intended to be a substitute for our GAAP results.” – Airbnb

“Except for historical information, the matters discussed in today's conference call may contain forward-looking statements within the meaning of the Private Securities Litigation Reform Act of 1995, including projections, estimates and descriptions of future events.” – S&P

# Goal

## What is the market thinking about?

- Provide common topics discussed in corpus of earning calls but **not specific to earnings calls**

## Are people worried or excited about this topic?

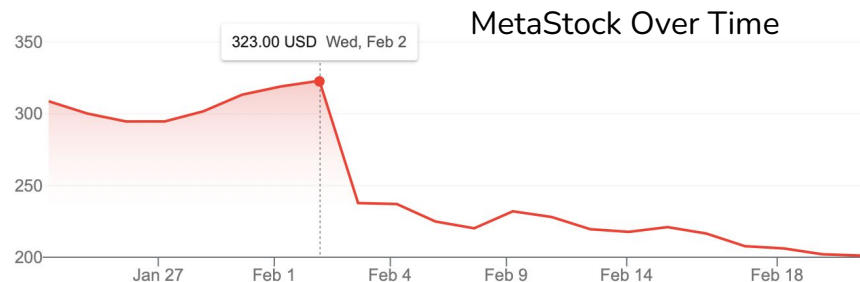
- Sentiment on topics

## Example:

Topic: **privacy measures**

Sentiment: **negative**

Facebook stock drop





## Data



**The  
Motley  
Fool**  
Transcripts

Image source: The Motley Fool.

1000 earnings calls prepared  
statement transcripts from Q4

> 150K sentences

3,309,825 words

84K unique words



# Challenges

Financial words not useful for our interest

1000 Companies - Everyone different

Putting the best foot forward



# Pipeline



## Sentiment per Sentence

- VADER
- Loughran McDonald
- FinBert

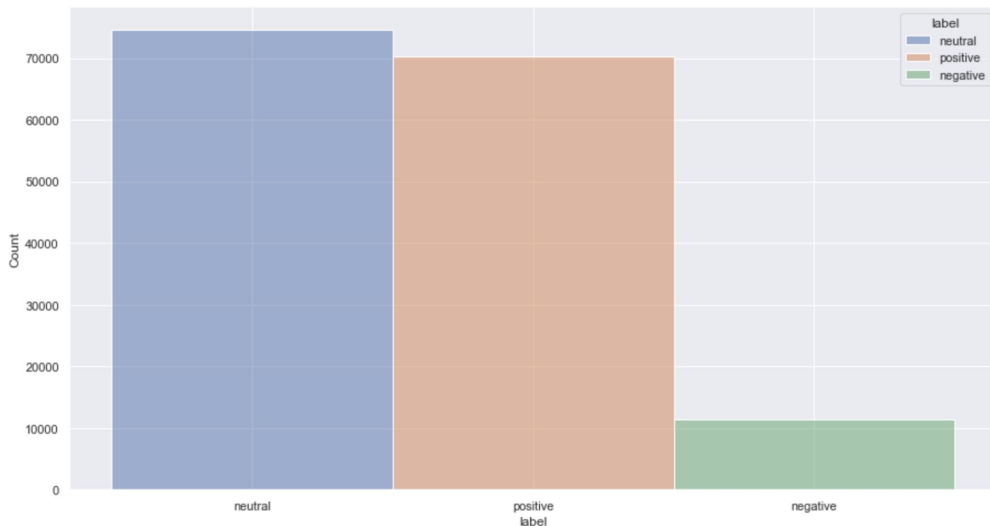
## Topic Modeling

- Tokenize
- Combine ENTs
- Filter Nouns
- Remove punctuation
- *Remove general earnings call words*





# Sentiment



- Neutral: “'Inventory turnover rate was 153 days”
- Positive: “In fiscal 2021, we delivered nearly 10,000 homes, the most in our history, and we grew home building revenue by 20% to a record \$8.4 billion.”
- Negative: “'First, our outlook assumes supply chain challenges and inflation persists, particularly in the first half.'”

I would want to put my best foot forward too!





# Sentiment Oddness

“Second, we expect price, plus productivity, to more than offset inflation for the full year.”

Inflation marked positive, actually negative

“The second quarter volume decline was more than offset by [..] inflation-driven pricing actions [...] driving an overall organic net sales growth of 2.6%.”

Inflation marked positive, actually positive

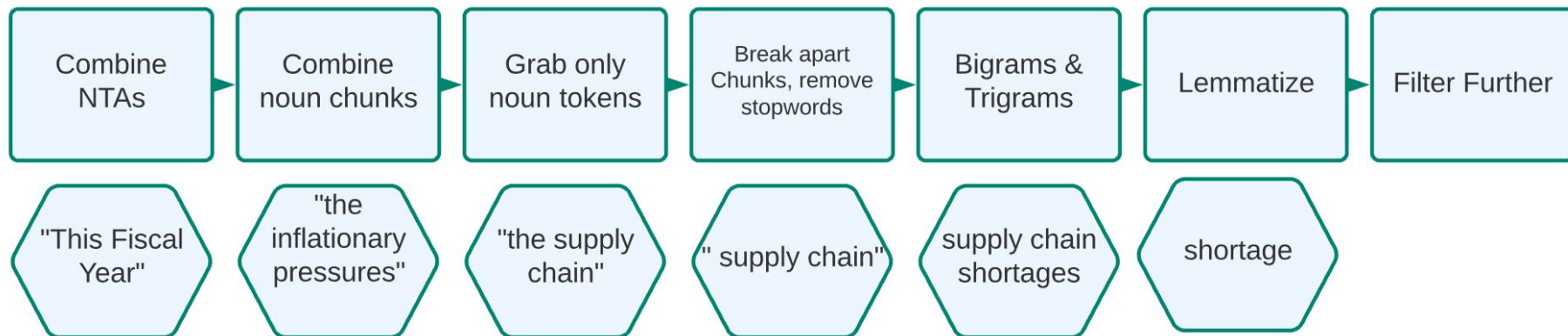
label	
term	
aviation	[positive, neutral, negative]
beef	[positive, negative, neutral]
housing	[neutral, positive, negative]
inflation	[neutral, positive, negative]
port_congestion	[negative, positive, neutral]
real_estate	[positive, negative, neutral]
vaccine	[positive, negative, neutral]



# Topic Modeling Pipeline

**Interest:** Find general topics discussed that are not specific to earnings calls

Ex: Inflation, semiconductor shortage, privacy





## Why?

Most terms are generic

Some terms are specific to the business being covered

## Filtering

## Remove rare terms

## Remove terms in **few** documents

## Remove **common** terms

## Remove terms in **many** documents

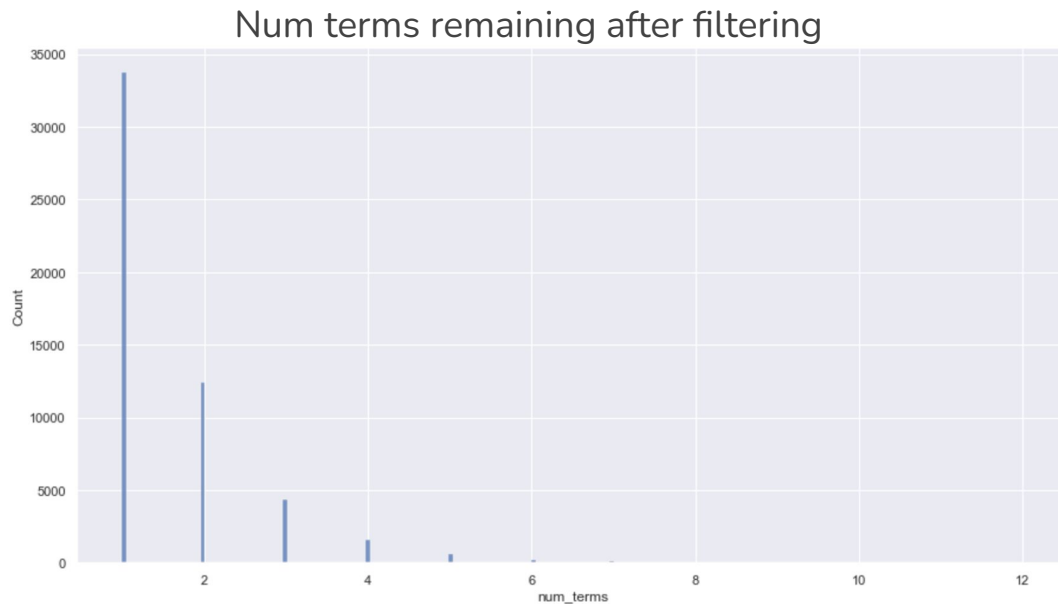


# Further Filtering Problem

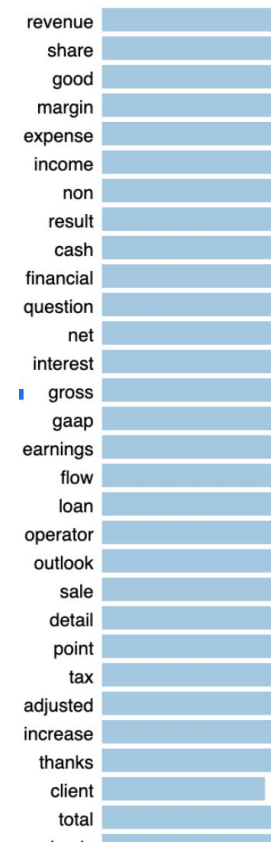
## Massive Reduction in Information

Reduces sentences by half

Reduces average terms to 1-2 per sentence



# Topics Inconclusive in LDA & LSA



## Filtering

Reduces sentences by half

Reduces average terms to 1-2 per sentence

## Not Filtering

Financial terms dominate

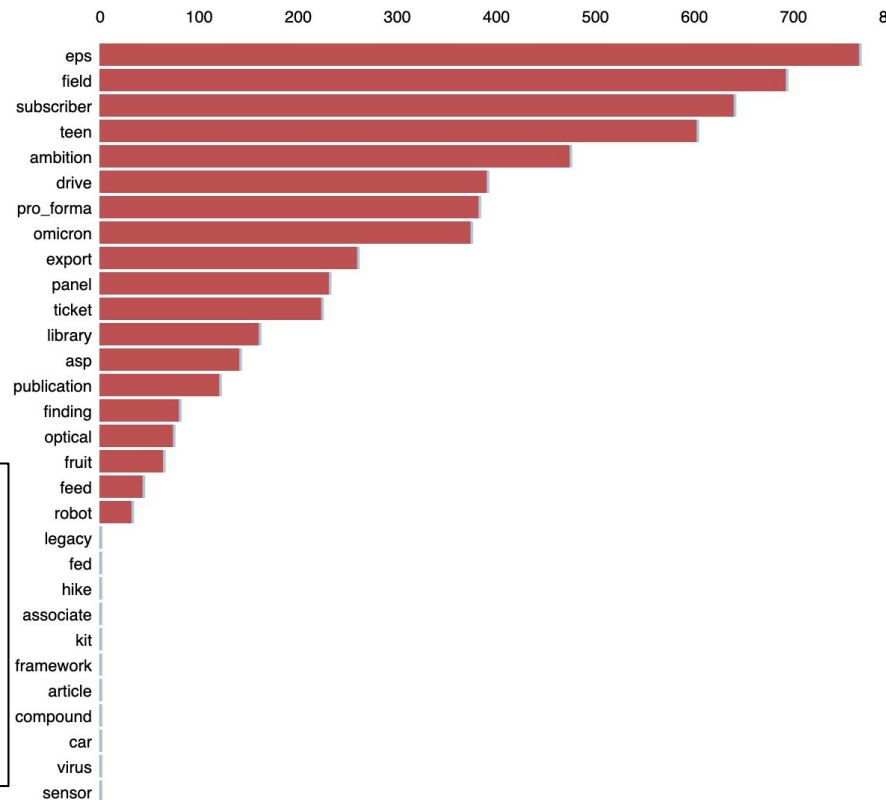
Tested 7 filtering themes

Tested multiple topic number

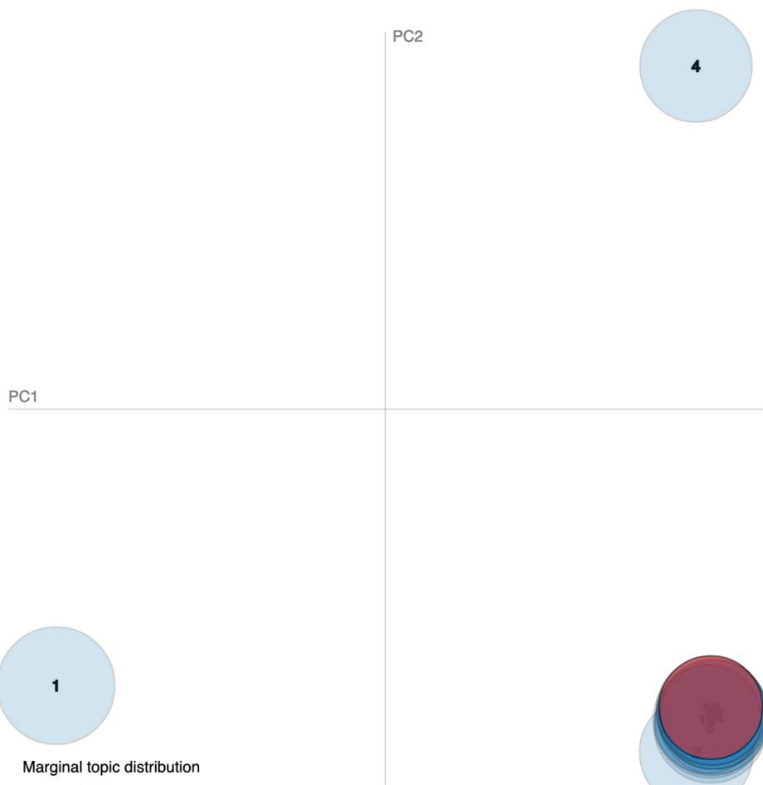
Chosen using coherence scores and print topics

Inspected manually via pyLDAvis

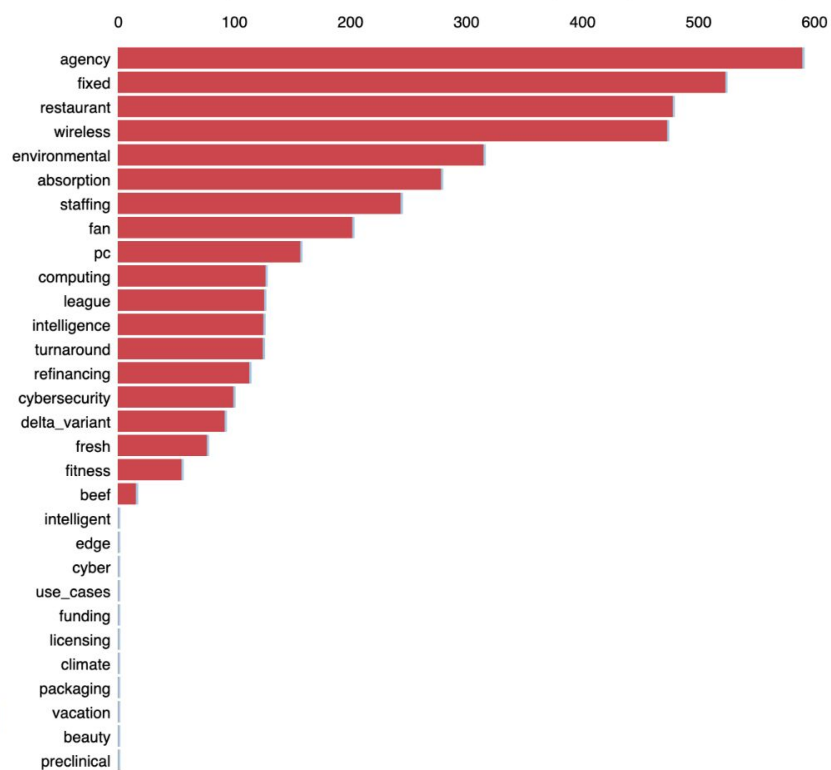
Top-30 Most Relevant Terms for Topic 15 (6% of tokens)



Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 15 (5.9% of tokens)



Overall term frequency

Estimated term frequency within the selected topic

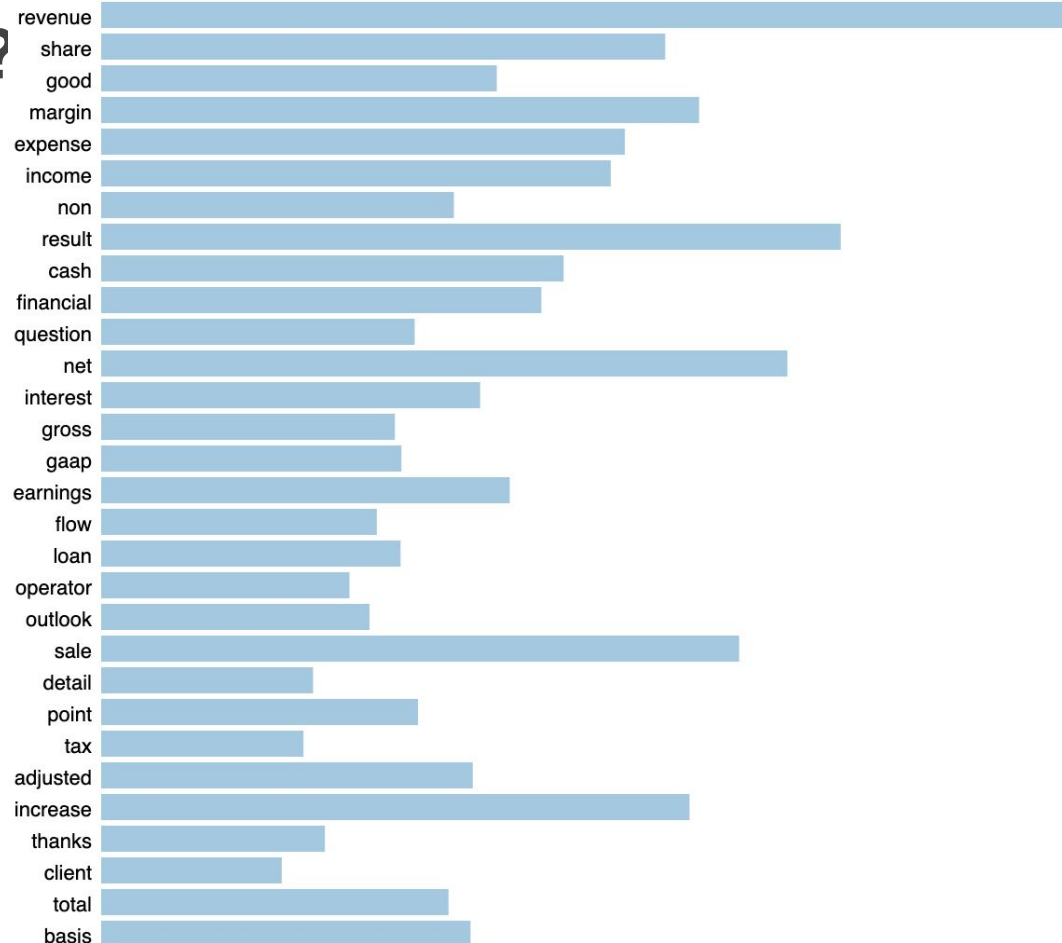
1.  $\text{saliency}(\text{term } w) = \text{frequency}(w) * [\sum_t p(t | w) * \log(p(t | w)/p(t))]$  for topics  $t$ ; see Chuang et. al (2012)
2.  $\text{relevance}(\text{term } w | \text{topic } t) = \lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$ ; see Sievert & Shirley (2014)



Top-30 Most Salient Terms<sup>1</sup>

0 500 1,000 1,500 2,000 2,500 3,000 3,500

# Not Filtering?



If filtering is limited

Financial terms dominate



# Project Status

**“I found 1000 ways how not to make a light bulb”**

I was unable to find topics of interest for me in this project

Likely due to:

- Overly broad scope of problem
- Too much variation in companies
- Cleaning pipeline issues
  - Noun chunks was a mistake
  - Lemmatization earlier
  - Filtering should be based on a financial terms dictionary