Background

Every quarter, public companies report earnings and company updates. Companies host a conference call in order to provide additional commentary and answer questions from participants.

Data

The dataset used consists of 400 earnings call transcripts for S&P 100 companies.

The team webscraped Seeking Alpha earning transcripts. The transcripts include a prepared remarks section, a question and answer section, and instructions from the call operator. For the purpose of my analysis, I analyzed only the prepared remarks from the company’s representatives.

I took the following steps to pre-process and clean the transcripts:

• Replaced instances of “n’t” with “ not”

• Removed numbers (unless part of hashtag or mention) and punctuation

• Removed capitalization

• Removed stop words (“a”, “we”, “the”, etc), stop phrases (“good morning”), and stop paragraphs (introductory or disclosure statements)

• Lemmatized words