Case Study - Banks' earnings calls

Introduction

Earnings conference calls are regular quarterly meetings between the senior management teams of a company and their equity analysts. Compared to firm's other types of disclosures, earnings calls are indeed very informative for market participants, as they contain more forward-looking details about the firm's expected performance and direction. Earnings calls usually begin with the management presenting the previous quarter's earnings results followed by a Q&A session between the management and participants.

In this task, you are provided with around two thousands transcripts of earnings calls for 89 US banks between 2005 and 2010[1], in xml format. In addition, there is a Mapping file that maps the filenames to the banks annual fundamentals. Fundamentals includes book value of assets (at), book value of common equity (ceq), total dividends (dvt), gross profit/loss (gp), net income/loss (nt), total revenue (revt), total market capitalisations (mkvlt), all in million USD, and Book Value Per Share (bkvlps) in USD. Furthermore, in the csv file named, "Culture dict", you will find words that signals the corporate culture based on the Li et al. (2021).

Questions

1. Do you see a need for data cleaning?

2. Try to parse the textual data from the xml files.
3. Separate the management presentation section from the Q&A section.

   (a) Do you see a pattern in length of presentations and Q&As based on the banks' profitability?

4. Do you see a pattern in the tone/sentiment for the presentations before and during the 2008 financial crisis?
5. Do you find any associations between the banks' culture (e.g., a dictio- nary analysis based on Li et al. (2021)) and banks' failure risk[2] during the crisis?
6. Which management teams have the most similar language to the man- agement of Lehman Brothers?
   (a) What is the best way to measure similarity? Discuss.

Remarks:

1. Use Beautiful Soup to read xml files.
2. Try to use multiprocessing (i.e. Pandarallel).