Home Work 2, due Sunday, Nov. 3, 11:59 pm

The file parsed.7z contains transcripts of 1385 Earnings Calls from this earnings season (2nd quarter 2019).

The data is in csv files, the actual text is in the contents.csv file

Download the data set and store each of the texts in a DataFrame. Collate the texts, eliminating the name of the speakers and run an LDA process on this data set. Perform the same analyses and graphs as in the notebook in Lecture 3.

Try several numbers of topics and see if the topics make sense as a way to cluster the documents. Compute perplexity and do the cluster graphs using PyLDAvis

Next treat each comment (i.e. each row in the DataFrames) as a separate document and again run an LDA analysis with several different numbers of topics.