S&P500 Component changing daily

Create a time-varying list available on CRSP

Link the stock market data from CRSP with the company identifier from Raven Pack

Non-matching

Headlines

LDA, pick a subset of the words.

Specify the vocab of LDA

Data cleansing

Computing-complexities:

BERT: embeddings

Model fitting

**Do: Research support**

**Parallel processing**

**Running over-night**

Or: Data sampling

Paid-RA?