Computation:

**GPU acceleration**

ID Mapping:

CUSIP – RavenPack

Mapping table lack of data: 12 companies

Look up the names of the companies

BERT theory:

**Topic number chosen**

**Different topic numbers each time**

**How to transfer embedding back to words**

LDA

**Gibbs sampling, check convergence tools or parameters.**

Grid: adding CUML to regression

Adding all S&P500, change index from CUSIP to RavenPack

Parallel processing

Data cleansing

Drop duplicates

Each step

One year first

**News time:**

**When during the day. After market close, the next day**

**Daylight saving, python program, UTC time to Eastern time**

**New York exchange is based on Eastern time, not Eastern Standard Time.**

**Daily saving time, daily standard time.**

**Weekend and holidays.**