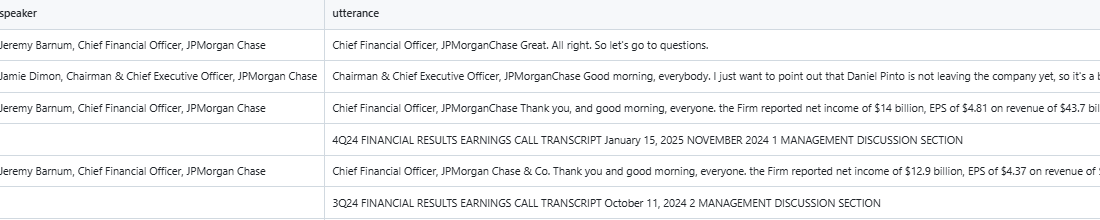
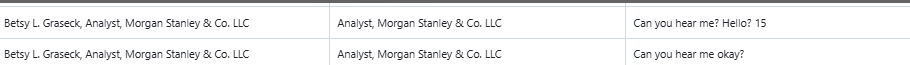
**JPMorgan management discussion Data** new requests for preprocessing:



* we need to move the Q424 row next to the speaker column and delete that row
* we then need to remove any text that contains "Chief Financial Officer JPMorgan.... thank you, good morning etc". These appear mulitple times and generate noise

**JPMorgan Q&A Data** new requests for preprocessing:

* remove any row that comes from operator
* "lagging" the analyst name and role next to the CEO, CFO, answer and add two more columns next to the answer with analyst name for the question and analyst role
* Remove any row that has less than i.e. 25 words (example below of redundant text)



* Add a question number and answer number column to each row
* Chunk the answers only and make sure that analyst name is carried across all chunks of the answer number linked to the question number
* Carry across all chunks the allocated answer number