**Project Goal**: Our goal is to develop a modeling framework for the JPMC Investor Relations Team to prepare for upcoming earning calls. Our framework takes a set of analysts and outputs a set of predicted Tags and a set of generated questions for each analyst. The outputted Tags originate from an internal tagger, while we generate the questions from an existing data-set.

As it stands, our current progress and future work are below:

- **Exploratory Data Analysis**: We have surveyed the distribution of Tags[1] across analysts, and the top 10 critical terms per analyst, per Tag, per analyst given Tag, via TF-IDF.
- **Hot-Topic**: We are in the midst of developing a new feature, Hot-Topic. Hot-Topic identifies questions that refer to outstanding events.
- **Tag-Hot-Topic Model**: The Tag-Hot-Topic Model is a recent addition to our project. We surmised, given our initial goal was the development of questions for an upcoming earnings call, we can develop a preliminary model, to predict the Tag and Hot-Topic, whose output will be used in question generation. To that end, we are building a multilabel classifier[2] that outputs a Tag-Hot-Topic combination[3] for the given inputs.
- **Question Generation Model**: For our question generation model, we have developed an LSTM which produces a question for a specific analyst by only using the prior questions of the analyst. Our next iteration will include other inputs[4] like the Tag-Hot-Topic Model output, the company, the date and the previous question asked.
- **Analyst Profile**: Alongside our question generation model, our shareholders asked for our framework to display the predicted Tags per each analyst. Therefore we will save the output of the Tag-Hot-Topic Model and output those results alongside the predicted questions.

-

[1] The Tags are generated by an internal tagger that is currently in use at JPMC

[2] We've done some cursory work with sklearn, but are not satisifed with the results. Are there other methodologies or techniques for doing the multilabel classification?

[3] EX. Given two tags, A, B, the labels are: A0, A1, B0, B1

[4] While we've spoken about other input into neural nets for text generation, we have never worked with such data. Are there examples we can work with or a professor we can speak with about this topic?