1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.
    a. Ji Ma jima@illinois.edu (Team Leader compton)
    b. Mingqing Teng mt52@illinis.edu
    c. Chien-Wei Chao cwchao4@illinois.edu
2. What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?
    a. What is the task?
        i. NLP approach to analyze stock info on social media, discord channels relative to stock price actions
    b. Why is it important or interesting?
        i. Provide social media level insight on a particular stock.
    c. What is your planned approach?
        i. Social media like reddit, twitter and discord contains a lot of information on insight toward a potential stock price action movement. We want to build a system to parse those natural language and provide information retrieval features like
            1. providing correlation scores for historical price movement
            2. predictions for future price movement based on social sentiment
    d. What tools, systems or datasets are involved?
        i. Dataset
            1. https://github.com/Tyrrrz/DiscordChatExporter
                a. This is for downloading corpus of data for a dedicated discord channel
            2. Atlas Trading Channel 700077347251945472
        ii. NLTK
            1. For nlp analysis
        iii. Mplfinance
            1. High performance plotting
        iv. Alpaca and alphavantage
            1. Pricing data
            2. Historical OHLC
        v. Natural Language AI
        vi. Firebase
            1. NoSQL backend and object storage
    e. What is the expected outcome?
        i. A UI tool (if time is permitted) or a CLI tool or discord bot (if little time is permitted for working on UI)
        ii. With the tool, user can query and analyze previous social sentiment, and subscribe to predictions on future sentiment
3. Which programming language do you plan to use?
    a. Python

4. Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.
    a. Scraping discord channels
    b. Apply sentiment analysis for each document
    c. Connect historical OLHC data
    d. Document tokenization.
    e. Provide a rationale / embedding (maybe one hot if we know the total vocab size) for each text for potential price action
    f. Preparing Training data – <sentiment_score, sentiment_confidence, array<token>, price_delta>
        i. Each document's impact to price_delta
    g. Preparing Training data – <aggregated_sentiment_score, array<tokens>, ticker, price_delta>
    h. Connecting realtime data
    i. Realtime model inference for social signals