Topic Idea #1: Predicting Stock Prices via Twitter Sentiment Analysis

1. **Problem Statement**
   - Stock price predictions are volatile and highly sensitive to external factors. The rise of "meme stocks" such as Gamestop, which gained huge social media traction and astronomical gains and losses within mere hours, begs the question of the degree of influence social media has on stock prices.

   - To learn the degree of this influence, this project will focus on collecting and analyzing tweets about public stocks and use a predictive model to estimate the closing price of these stocks the following day.

2. **Significance of the Problem**
   - The project will help to not only predict stock prices, but to recognize the accuracy to which social media has influence on these security prices in terms of volatility and volume.

   - Here is an academic research paper I found after formulating this idea similar to what we hope to accomplish: Kordonis, John & Symeonidis, Symeon & Arampatzis, Avi. (2016). Stock Price Forecasting via Sentiment Analysis on Twitter. 10.1145/3003733.3003787. The methodology to analyze and process the sentiment of the Tweets will be a guideline to our project. Furthermore, the study did indeed find a correlation between the sentiment of tweets and stock prices, so we hope to express findings to a similar degree in our analysis.

3. **Potential Datasets**
   - The Twitter API allows us to search for recent tweets or stream tweets. This will help us collect textual data in the form of tweets to use for our sentiment analysis and feed our predictive model.

   - To collect historical and closing market data for our predictive model, we will use Yahoo Finance data, provided by existing APIs and libraries such as yfinance.

Topic Idea #2: Identifying Incendiary Content

1. **Problem Statement**
    - Private social media companies are not beholden to the First Amendment, and therefore must make difficult decisions on what content should be permissible on their platform, whether morally or pragmatically.

    - The inflammatory nature of some content is one factor that companies consider in their censorship. It is useful to determine whether a post is capable of inciting hate or violence, regardless of the final decision.

    - We want to see if we can identify types of posts that have or may cause violence.

2. **Significance of the Problem**
    - Political unrest is on the rise and often leads to violent outcomes. Social media companies should be obligated to prevent violent groups from proliferating on their platforms.

    - Human moderators can only go through so many reports in one day. Automatic analytics might help make content processing more efficient.

    - https://link.springer.com/article/10.1007/s42979-021-00457-3

3. **Potential Datasets**
    - We can look through the Twitter api to go through many tweets as a data set.

    - What is challenging is finding a training set, i.e. posts that are clearly harmful, since this is highly subjective.

Topic Idea #3: Predicting High Risk Airlines

1. **Problem Statement**
   - People believe that large national airlines are reliable investments, however during COVID, these airlines experienced high volatility. They also aren't known for high dividend yields, with some such as JetBlue paying no dividends at all. Many issued more common stock and lowered common investor stakes in the company.

   - This project illustrates how covid has changed the reliability of airline investments, their financial status (debts, bankruptcies) and will predict trends in the sector in the near future.

2. **Significance of the Problem**
   - This project is significant as some airlines still cannot pay their debts and have extended leasing payments, so many investors were deceived. Stock prices were also super volatile.

   - Project insights would be useful to people interested in finance as well as investors, as it reveals trends for consideration other than the name and size of the company.

   - https://www.fool.com/investing/2021/09/23/why-investors-were-excited-even-though-airline-sto/

3. **Potential Datasets**
   - We can download historical market data for public U.S. companies in the airline sector e.g. American airlines, Alaska Airlines, Delta Airlines, Southwest Airlines, and United Airlines as well as large foreign airlines: KoreanAir, Aeroflot, Turkish Airlines etc for approximately 500 days (covers the pandemic timeframe).