1. Fill out this form to help us evaluate you better: https://forms.gle/3Jx1BUtNaFKWU8Dt6
2. Go through the website: https://capx.live/
   a. There are three pages on the site. You will have to **sign up via your google account** to access the last one i.e. Profile Page. Pages are <u>Home page</u>, <u>Join waitlist</u> and <u>Profile</u>.
   b. Please fill this survey after that to give your feedback on the site and improvements needed for the same. **Please use the same email to fill the survey as the one you used for signing up on CapX site.**
      i. Survey link: https://forms.gle/v481RSR3a97Dhs569
3. **Stock Movement Analysis Based on Social Media Sentiment**

   Objective: Develop a machine learning model that predicts stock movements by scraping data from social media platforms like Twitter, Reddit, or Telegram. The model should extract insights from user-generated content, such as stock discussions, predictions, or sentiment analysis, and accurately forecast stock price trends.

   ## Task Requirements:

   1. **Data Scraping:**
      ○ Select one platform: Twitter, Reddit, or Telegram.
      ○ Scrape relevant data from specific handles, subreddits, or Telegram channels focused on stock market discussions and predictions.
      ○ Clean and preprocess the data, ensuring it's ready for model input (e.g., removing noise, handling missing data).
   2. **Data Analysis(Optional):**
      ○ Perform sentiment analysis or topic modeling on the scraped data.
      ○ Extract key features such as sentiment polarity, frequency of mentions, or any other indicators relevant to stock movements.
   3. **Prediction Model:**
      ○ Build a machine learning model that takes the processed data as input and predicts stock movements.
      ○ Test the model's accuracy on historical data or known stock trends.
      ○ Provide a detailed evaluation of the model's performance, including metrics like accuracy, precision, recall, and any improvements that can be made.
   4. **Technical Skills Required:**
      ○ Proficiency in Python, with experience in web scraping (using libraries such as BeautifulSoup, Scrapy, or Selenium).
      ○ Knowledge of Natural Language Processing (NLP) techniques for sentiment analysis and text mining.
      ○ Experience in building and evaluating machine learning models using libraries such as scikit-learn, TensorFlow, or PyTorch.

   ## Submission Details:

   ● **Code:** Submit a GitHub repository link containing the following:

- ○ Well-documented code for the scraper and prediction model.
    - ○ Clear instructions on how to run the code, including dependencies and setup requirements (provide a README file).
    - ○ Jupyter notebooks or scripts demonstrating the data scraping, preprocessing, model training, and evaluation steps.
- **Report:** Submit a PDF report that includes:
    - ○ A detailed explanation of the scraping process, challenges encountered, and how they were resolved.
    - ○ Description of the features extracted and their relevance to stock movement predictions.
    - ○ Model evaluation metrics, performance insights, and any potential improvements.
    - ○ Suggestions for future expansions, such as integrating multiple data sources or improving prediction accuracy.
- **Demo Video(Optional):** Record a 5-10 minute demo video that walks through:
    - ○ The scraping process (e.g., how the data is collected).
    - ○ Model training and prediction.
    - ○ Results and any significant findings from the model's predictions.
    - ○ Upload the video to a platform (YouTube, Vimeo, etc.) and include the link in your submission.

Evaluation Criteria:

- Data scraping efficiency and relevance.
- Feature extraction and its impact on stock prediction accuracy.
- Model accuracy and evaluation metrics.
- Quality of code, documentation, and report.

How to scrape?

Here are few (but not limited) ways to scrape data from the aforementioned channels. Candidates can refer to these for quick implementation.

1. **Scraping Twitter with Tweepy (Using Twitter API)**
    a. To scrape data from Twitter using Tweepy, you'll first need to create a Twitter Developer account and get API keys. Then, you can use Tweepy to gather tweets.
2. **Scraping Reddit with PRAW (Python Reddit API Wrapper)**
    a. You can use PRAW to scrape data from subreddits like r/stocks or r/investing, where people often discuss stock market predictions.
3. **Scraping Telegram Channels with Telethon**
    a. To scrape messages from a Telegram channel or group, you can use Telethon, a Python library for Telegram API.