**Read the document carefully and follow as instructed. The behavioral survey (Link at the end of the document) is to be filled once the task is completed.**

**About the Task**

## Machine Learning Roles / Business Analyst (Time Series Modeling)

**Important Points:**
(1) Please use separate .py files to harvest ONLY the data from API or scrap from websites.
(2) Use different / alternative data for training your data, including technical features, or other country and financial features. Do not use just the history of the price as features.
(3) Use a combination of models of your choice to benchmark the accuracy, you can use both classification (up or down), or time series models. Our preferred models are MLP, RNN, LSTM, or GRU - for time series, and Logistic, Random Forest, Naive Bayes Classifier for classification. You are free to use other models
(4) Provide a report with succinct visualization of results and all your different .py scripts (class object oriented good scripting practices) and final Python notebook.

Please submit a **PDF of google slides** or a **document** presenting your findings. Upload the PDF onto your Google Drive and share the link as follows:
Google drive link with general access set as 'Anyone with a link' and role set as 'Editor'
***Share the link in the Behavioral Survey.***
Your evaluation criteria is partially technical and partially the ability to explain meaningful results in a presentable manner.

**Time-Series Analysis**

- Use any daily time series from Investing.com or similar source with a strong sample of covariates. Target commodities price like: Oil, Natural Gas, Resin, or Metal Prices.
    - Please make sure to get an extensive list of feature space, think through structural other external factors.
    - Feature Importance. Dynamic Time Warping and/or XGBoost/Shapley Value hybrid model approach to quantify which factors influence the target positively or negative
        - **Keep the analysis focused on the feature selection and feature importance aspects**
    - LSTM derivatives on day ahead prediction with confidence bounds
        - How would you improve and present your results with more time and resources
        - References:
            - Run multiple time series ML models
                - https://lazypredict.readthedocs.io/en/latest/
            - Uncertainty quantification
                - https://uncertainpy.readthedocs.io/en/latest/
                - https://github.com/scikit-learn-contrib/mapie

- Related thinking and documentation in a short report

**Evaluation is based on the following parameters:**

| Stages | Evaluation Criteria |
|---|---|
| 1 | Extensiveness of the dataset and understanding of projects and tenders data structure |
| 2 | Modular, DRY Code |
| 3 | Config Params, Unit Tests & Logging Standards |
| 4 | Data Visualization presentation of results and understanding of the problem statement |
| 5 | Explainability, the ability to experiment, do creative relevant problem solving, and structure your analytical thinking with documentation |

**Behavioral Survey**