**Problem Overview**

**🎯 The identified/potential problem statement?**

*How to predict future stock/options/commodifies’ prices based on historical data, and then can we use optimisation techniques to find determine an optimal portfolio to hedge? The objective is to develop a model that can predict the return (percentage change in price) over a given time horizon, providing actionable insights into potential future market behaviour.*  
  
**🔍 Why is it important to solve?**

*With many struggling with a higher-cost of living and the current job market causing major unemployment or lack of career flexibility, a greater understanding of stocks is needed so that we can make personal investments with more assurance, as these savings from investments often provide households with financial stability. However, predicting stock prices is difficult even more so in such volatile today’s geo-political environment but the next best thing, making a prediction based on previous data available is useful.*

**Data Overview**

**📊 Have you identified/sourced the required dataset?**

#### Alpha Academy Datasets: Time Series monthly adjusted

Economic and technical indicators like inflation, interest rates ,moving average, Bollinger bands, averages

Option prices historical data

Kaggle: <https://www.kaggle.com/competitions/stock-price-prediction-challenge/data>

Alpha Vantage: https://www.alphavantage.co/documentation/#

**If yes:**

* **📊Can you share a sample of data in Google Sheet [[sheet.new](https://docs.google.com/spreadsheets/d/1CyyMWW7TuaCwUk7b9c9T6_Hs9CtNZvpLZY6bfUcfXvM/edit)] (15 - 20 rows)?**
* **📏How many rows/columns?** *# Your response here.*

Example Alpha academy stock price monthly adjusted 309 rows × 8 columns

Kaggle datasets meet requirements

* **🔢 How many numerical/categorical columns?**
* **🎯 Which is the target column (for supervised learning)?**

Return on stock price

* **🤔What kind of ML problem (regression, classification etc)?**

Supervised ML Regression or Time Series and Optimisation

**Data Overview**

* **📙 Can you please add data dictionary (explanation of columns)?**

**Methodological Overview**

**🛠️ Have you considered any techniques, pre-processing etc., that you will use?**

*PCA maybe to reduce factors and dimensions , EDA to check for outliers , discount dates with missing prices or replace with average price monthly price ?, correlation matrix to check for relationships between stocks and indices, merging in technical and economic factors ?*

**⏳ Do you expect it to get solved within the time frame of the course?**

Yes

**💡 What potential challenges do you expect?**

*Dates from Kaggle and Alpha Academy stock data not aligning, model not giving good results, finding the right balance of difficulty and simplicity?*

**💬 What is the main thing you would like to discuss in the 1-1?**

Which ML model to use or should I use time series ?

How much maths is needed ?

Better to stick with Kaggle data then other sources such as alpha Academy API

The meeting form says to avoid Yahoo Finance data , why is this ?

Optimisation of stock portfolios after prediction ok to do if I have time ?

How to maximise employability with project ?

Use Geometric Brownian Motion model as comparison ?