

# True Beacon Assignment

The code that is provided does a few things:-

- 1) Read the the parquet File
- 2) Does some important Data cleaning parts
- 3) Made a very Basic Z-score based model to maximise Profit and Loss
- 4) Made the best model possible out of all the other models

## Project Overview

- Objective: To test a hypothesis regarding pairs trading strategy, leveraging volatility disparities between financial indices.
- Methodology: The project involves data importation, baseline model construction, application of machine learning algorithms for enhancement, and a comparative analysis of all models.
- Data Source: Utilised minute-level data spanning from January 1, 2021, to June 30, 2022, with various filtering techniques applied to ensure quality and relevance.

## Key Steps and Techniques

1. Data Preparation: The project begins with data cleansing, involving trade time filtering, weekend and non-trading days exclusion, and handling of anomalies in time-to-expiry (TTE) values.
2. Baseline Model - Z-Score: A foundational model was established using Z-Score for normalisation, which generates trading signals based on the spread between the indices
3. Performance Evaluation: Utilised a standardised evaluation function to compute key metrics like Sharpe Ratio, total PnL, and maximum drawdown, facilitating model comparisons.

## Assumptions and Commentary

- Data Integrity: Assumes the preprocessed data is representative of the market dynamics within the specified timeframe.
- Statistical Normalisation: Relies on the assumption that Z-Score normalisation effectively identifies trade opportunities by pinpointing deviations from the mean spread.
- Market Conditions: The models do not account for transaction costs, slippage, or regime changes, which could impact real-world applicability.

## Alternative Models and Machine Learning

- Approach: Explored machine learning models, including Linear Regression, Random Forest, and XG boost, to predict future spreads and refine trading signals.
- Feature Engineering: Involved creating lagged features and addressing missing values, alongside preparing datasets for model training and evaluation.
- Model Comparison: Conducted a thorough comparison of the machine learning models against the baseline, focusing on improving profitability and risk-adjusted returns.

## Conclusion and Deliverables

- Optimal Model: The XG Boosting model was identified as the most promising, offering an optimised balance between profitability and risk.
- Insights: The project illustrates the complexity of quantitative trading strategies and highlights the potential of machine learning to enhance trading decisions.
- Future Directions: Suggests further exploration into model robustness against market changes and inclusion of transaction costs in the evaluation.

## Final Insights

By focusing on a pairs trading strategy, the Z Score model still performed better, if i had more time i might have tried DL models as well