# The Research Topic

The research topic focuses on the Kaggle Competition – Optiver – Trading at the Close. The main goal is to develop a model capable of predicting the closing price movements for hundreds of Nasdaq listed stocks using data from the order book and the closing auction of the stock.

# The SMART question

The main goal is to achieve a Mean Absolute Error (MAE) score between the predicted return and the observed target based on the test set of less than 10. The ambitious goal is to reach a MAE score of less than 6. The formula to compute the MAE is given by

Where n is the total number of data points, y\_i is the predicted value for data point I, x\_i is the observed value for data point i.

The main and ambitious goal are set based on the best MAE score achieved by the top team in the competition which is around 5.4 when drafting our proposal. As our team only try this kind of problem for the first time, we think less than 10 will be achievable and less than 6 will be ambitious.

# The source of the data set

<https://www.kaggle.com/competitions/optiver-trading-at-the-close/data?select=train.csv>

The training data set has about 5.2 million data points, each data point has value in 17 columns. One of the 17 column (the target column) is the result column that we can use to train the model. For the test set, the target column will be masked and we need to use our model to predict the value in this target column.

Further explanations on the value represented by each column can be found in the Kaggle competition site. https://www.kaggle.com/competitions/optiver-trading-at-the-close/data

# Link to team’s GitHub repo

# Modeling methods proposed

The main modeling method we propose to use will be Regression models under the Supervised Learning Category. We propose to start with Linear Regression and continue to try out other regression models such as Neural Network Regression, Decision Tree Regression etc if time permitted.

Some of the Exploratory Data Analysis questions we have in mind include:

* Whether the closing stock movements for the stocks are related to each other? (Ie, stocks tend to move together following the major announcements during the reporting period, or major announcements by the Fed for key interest decisions).
* Whether there is a pattern for the closing stock movements for the stocks in consecutive days.

# Reference

Tom Forbes, John Macgillivray, Matteo Pietrobon, Sohier Dane, Maggie Demkin. (2023). Optiver - Trading at the Close. Kaggle. https://kaggle.com/competitions/optiver-trading-at-the-close