This assignment required the organization of hourly forex market data in both MongoDB and MySQL, and the creation of a vector of features for each currency pair every 6 hours. The vector included timestamp, VWAP price, liquidity, volatility, max and min prices, and fractal dimension (FD) calculated using Keltner Bands. The 70-30 train-test rule was used to classify each currency pair over a regression to determine if it was forecastable, non-forecastable, or partially forecastable. The report includes the classification and regression results along with a brief explanation of the methods used.

In my code:

I imported necessary libraries, read the CSV files of currency pairs using pandas, and resampled the data by calculating VWAP, liquidity, volatility, max, min, and FD values.

I then split the data into training and testing sets using the 70-30 rule and used PyCaret library to create a regression model on the training data. I compared multiple regression models and selected the best one based on R-squared score.

Finally, I evaluated the performance of the selected model on the test data by calculating mean absolute error, mean squared error, and R-squared score.

I choose these pairs: EURPLN,GBPCHF,HKDJPY,NZDUSD,SGDJPY,USDBRL,USDDKK,USDHKD,USDMXN,USDTRY

Based on the result, all of them are forecastable since all of my R^2 result are over 0.7 with different regression models.