Project 1:

First of all, I look for several factors that affect home prices and collect the history data from FRED. These factors are

GDP: Gross Domestic Product (quarterly data);

BORROW : Total Borrowings from the Federal Reserve (monthly data);

EMRATIO: Employment Ratio (monthly rate);

HOUSEHOLD: Household Estimates (monthly data);

HOUST: New Privately-Owned Housing Units Started: Total Units (monthly data);

INCOME: Real Disposable Personal Income (monthly data);

M2: A measure of the U.S. money supply that includes all components of M1 plus several less-liquid assets (weekly data);

MORTGAGERATE: 30-Year Fixed Rate Mortgage Average in the United States (weekly data);

PCE: Personal Consumption Expenditure (monthly data);

POP: Population (monthly data);

TDSP: Household Debt Service Payments as a Percent of Disposable Personal Income (quarterly data);

UNRATE: Unemployment Rate (monthly rate);

WAGE: Employed full time: Median usual weekly real earnings: Wage and salary workers: 16 years and over (quarterly data).

Secondly, quarterly or weekly data are converted into monthly data based on linear transformation or mean values. In addition, assuming the data is normally distributed, define data which is lager than $\mu + 3 * \sigma$ or smaller than $\mu - 3 * \sigma$ as outliers and then delete outliers, where μ is sample mean and σ is sample standard deviation.

By correlation matrix, I observe that Home Price Index has over 0.8 coefficients of correlation with several factors which are GDP, HOUSEHOLD, INCOME, M2, PCE, POP and WAGE and a high negative relationship with MORTGAGERATE. Hence, I prefer a linear model.

Thirdly, I select Ridge Regression which can solve the problems of overfitting and multicollinearity and give weights to variables. The data is split into training and validation sets which account for 80% and 20% of the data respectively. The training set is used to train the model and the validation set is used to evaluate the model. Additionally, I standardize the training and validation sets by the following formula:

$$x = \frac{x - \mu}{\sigma}$$

where μ is mean of training set mean and σ is standard deviation of training set.

The model is evaluated by RMSE, MAE and score which are 0.05, 0.04 and 0.99 respectively.

Finally, the model gives high weights on factors INCOME, MORTGAGERATE and PCE. It is true that with higher income and lower mortgage rate, people are more likely to buy houses and therefore home price index increases.

Project 3: Idea Generation for Price Prediction Signal for a Macro Asset or ETF

I would like to design a intraday price prediction signal for bonds with T+0 trade from technical analysis.

First of all, volume acts as a momentum factor which reflects the relation between supply and demand and therefore volume has a great influence on price. From the figure below, we can observe that when there is a huge growth in volume, the price will be more likely to increase or decrease dramatically. Meanwhile, under a explosive volume, the underlying price tends to reach maximum or minimum within recent several minutes. While the volume decreases after a explosion, there tend to be a reversal trend.

Secondly, volatility describes the speed and amplitude of price movement. Hence, volatility is an another indicator to predict price. When the price fluctuates steadily and then moves upward or downward dramatically, the volatility would increase.

In addition, I also observe that faced with extreme movements, there is a common pattern between different moving average curves. In detail, under the circumstances of growth, the relation is the underlying price > ma5 > ma10 > ma20 and the underlying price < ma5 < ma10 < ma20 on the contrary.

In summary, we can conclude the above analysis into the following conditions:
Condition 1: The volume exceeds the 90 percentile of history volumes.
Condition 2: The volatility of the underlying price increases.
Condition 3: (1) the underlying price > ma5 > ma10 > ma20; (2) the underlying price < ma5 < ma10 < ma20.
Condition 4: The current price reaches the maximum or minimum within recent several minutes when a explosive volume appears.

There would be an increase signal when Condition 1 &2 & 3(1) & 4 are satisfied and a decrease signal when Condition 1 &2 & 3(2) & 4 are satisfied.