# Predicting Tomorrow's Gold Price

Bhanu Yerra
Metis, Project 2

## Background:

Central banks across the world maintain gold reserves to guarantee the money of their depositors, foreign-debt creditors, and currency holders. Central banks also use gold reserves as a means to control inflation and strengthen their country's financial standing. In addition, gold is used by finance companies, global institutions, money managers and individual investors to hedge against inflation and to diversify their portfolios. As a precious metal, gold is popular for jewelry and ornamentation.

Given the standing of gold's popularity in the contemporary world, forecasting its price is a widely explored topic and interesting to multiple global institutions and small & large scale investors. It is also a complex problem given gold's price fluctuations are not entirely based on supply and demand, but also depends on a multitude of geo-political and financial factors.

From a modeling perspective, this means the independent variables need to consider *market factors* like S&P 500, Dow Jones Industrial average, treasury bond rates etc; *economic variables* like Consumer Price Index, inflation, interest rates; *forex variables* like US Dollar Index; and other *commodity indicators* like silver, platinum, and crude oil prices etc. Gold prices heavily depend on these factors, which in turn influence and correlate against each other. This complex relationship between various independent variables might pose structural and model specification issues, and perhaps an opportunity to apply various machine learning techniques.

## Data:

| Variable | Category, Type & Frequency | Description |
|---|---|---|
| Gold Spot Price: London Fix PM | Target variable, Float, Daily | PM price of gold as set by London Fix https://www.kitco.com/londonfix/gold.londonfix16.html |
| S&P 500 Index | Market Variable, Float, Daily | A composite index of Standard & Poors top 500 companies closing stock prices |
| ^DJI | Market Variable, Float, Daily | Closing Dow Jones Industrial Average |
| FTSE | Market Variable, Float, Daily | London FTSE, daily closing |
| 10 Year Treasury | Market Variable, Float, Daily | 10 Year US Treasury Notes, https://www.macrotrends.net/2016/10-year-treasury-bond-rate-yield-chart |

| | | |
|---|---|---|
| Federal Interest Rates | Economic Indicator, Float, Daily | The rate at which banks and creditors are lending money to each other. Based on the federal rate set by the Federal Open Market Committee. https://www.macrotrends.net/2015/fed-funds-rate-historical-chart |
| CPI-U | Economic Indicator, Float, Monthly | All Urban Consumers, monthly https://fred.stlouisfed.org/series/CPIAUCSL |
| Inflation (based on CPI) | Economic Indicator, Float, Monthly | Inflation. Use CPI or inflation but not both at the same time https://www.usinflationcalculator.com/inflation/historical-inflation-rates/ |
| US Dollar Index | Forex Indicator, Float, Monthly | A measure of USD value against a basket of foreign currencies: https://www.macrotrends.net/1329/us-dollar-index-historical-chart |
| Gold Futures | Commodity variable, Float, Daily | London Fix: https://www.kitco.com/londonfix/gold.londonfix16.html |
| Silver Spot Price | Commodity variable, Float, Daily | London Fix: https://www.kitco.com/londonfix/gold.londonfix16.html |
| Silver Futures | Commodity variable, Float, Daily | London Fix: https://www.kitco.com/londonfix/gold.londonfix16.html |
| Platinum Spot Price | Commodity variable, Float, Daily | London Fix: https://www.kitco.com/londonfix/gold.londonfix16.html |
| Crude Oil Spot Price | Commodity variable, Float, Daily | https://www.investing.com/commodities/crude-oil-historical-data |

The variables listed above will be scraped/downloaded from the above websites if listed. Further web search and scraping will be conducted if website is not associated with a variable above. Next step in the data assembly step is to stitch the above downloads into a cohesive data set. The variables identified have a combination of daily and monthly timeframes. The release of some of the economic indicators identified as variables have a built-in lag and when these indicators should play a role in influencing the gold price (GP) will be determined based on a set of rules identified after data exploration step.

## Model Building Approach:

Anticipated steps in model building process:

1. Data exploration step:
    a. A correlation analysis will be performed to check gold price (GP) against the above list of variables.
    b. A correlation analysis of GP against select variables for a number of time lags: 1 month, 2 months, 3 months etc.
    c. Train and test sets: Models that have time lag variables need to consider how dataset can be split so the training and test sets are truly independent. Random selection of data points might result in including data points in test set that partially overlap with data points in the training set.

2. A Benchmark Model: A simple linear regression model using the variables identified in Step 1a

3. Multiple exploratory models: Using regression statistics, analysis of residuals, and model performance, various transformations and combinations of the independent variables will be tested for inclusion in the regression models

4. Other Machine Learning Models: Time permitting.