Cryptocurrency Drivers:

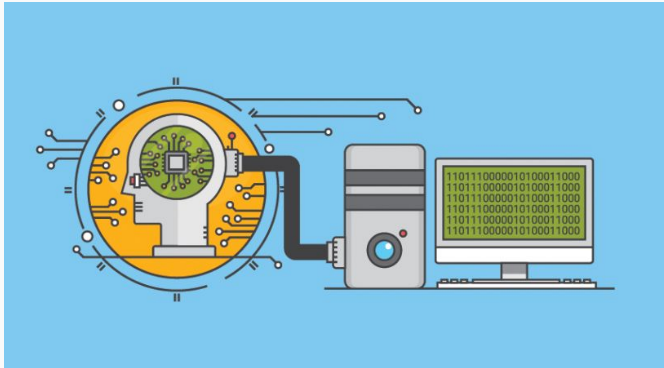
Implementing ML Techniques to discover drivers behind the price of Bitcoin

Team Members: Sean Simmons, Siegfried Vieluf, Santiago Pedemonte, Rawnak Mahjabib

Motivation & Summary

Our objectives:

- utilizing machine learning techniques to identify what the drivers behind the price of Bitcoin are, AND
- potentially attempt to build a trading model using insights gained from our model’s performance.



Data Collection, Cleanup, & Model Training

- We sourced our dataset using Bloomberg Export. Our dataset cleaning process was quite tedious. The dataset contained unmatched indexes & available dates.
- Some of our data was reported on a quarterly, monthly, weekly, basis and the vast majority was collected daily.

	Central Bank Balance Sheet Size	GRDI	Bloomberg Commodity Index	Gold	Bitcoin	Bloomberg Dollar Index	Citi Dollar Positioning	Bitcoin - Story Count	Inflation - Story Count	VIX Index	MOVE Index	CVIX Index	US 2 Year	US 2s10s	Sector Rotation
2021-04-01	35745.1094	119.426	83.8339	1729.31	58859.78	1148.66	-42.05	4119.0	2033.0	18.60	64.9900	6.56	0.1585	150.550	1.0654
2021-03-31	35703.9688	119.463	83.4445	1707.71	58960.20	1152.19	-42.24	4194.0	2858.0	19.80	71.2700	6.72	0.1603	157.624	1.0480
2021-03-30	35746.0195	119.242	82.7525	1685.20	58673.67	1154.39	-36.19	4892.0	2434.0	20.76	67.4400	6.79	0.1465	155.246	1.0434
2021-03-29	35814.6563	121.056	83.9328	1712.20	57234.35	1150.33	-33.09	4253.0	1829.0	20.40	65.9600	6.83	0.1407	156.351	1.0344
2021-03-26	35846.5664	120.157	84.1319	1732.52	54002.49	1148.11	-32.90	3730.0	2225.0	19.32	61.4900	6.86	0.1387	153.338	1.0463
2021-03-25	35450.7383	120.114	82.9932	1726.93	52001.01	1150.69	-36.79	4837.0	3467.0	20.80	59.5400	6.93	0.1367	149.057	1.0430

WTF?!

What('re) The Feature(s)?:

- **Central Bank Balance Sheet:** summarizes its financial position, and is made up of assets, liabilities and equity.
- US Budget Balance
- **US Consumer Price Index (CPI):** a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services.
- **Conference Board Inflation Expectations:** contains consumer expectations for inflation, stock prices, and interest rates.
- **Global Risk Demand Index:** gauges the change in risk demand from the price performance of risky assets relative to safe ones.
- **Dollar Index:** a measure of the value of the U.S. dollar relative to a basket of foreign currencies, often referred to as a basket of U.S. trade partner's currencies. The index goes up when the US dollar gains strength (value) compared to other currencies.
- **CITI Dollar Positioning**: this is a measure of dollar positioning, the thinking being as people get more short USD - they buy more Bitcoin.
- **Volatility Index (VIX):** looks at expectations of future volatility, also known as implied volatility.
- **Currency Volatility Index (CVIX):** created by computing a decentralized volatility index from cryptocurrency option prices.
- **Merrill Lynch Option Volatility Estimate (MOVE):** indicator of U.S. interest rate volatility. It measures the implied yield volatility of a basket of one-month over-the-counter options on 2-year, 5-year, 10-year, and 30-year Treasuries.
- **Sector Rotation Index:** designed to rotate sector allocations based on favorable characteristics regardless of the economic environments.
- **Bloomberg Commodity Index (BCOM)**: is calculated on an excess return basis and reflects commodity futures price movements.
- **Bitcoin Story Count:** global search volume for Bitcoin

Data Cleanup & Model Training (Challenges)

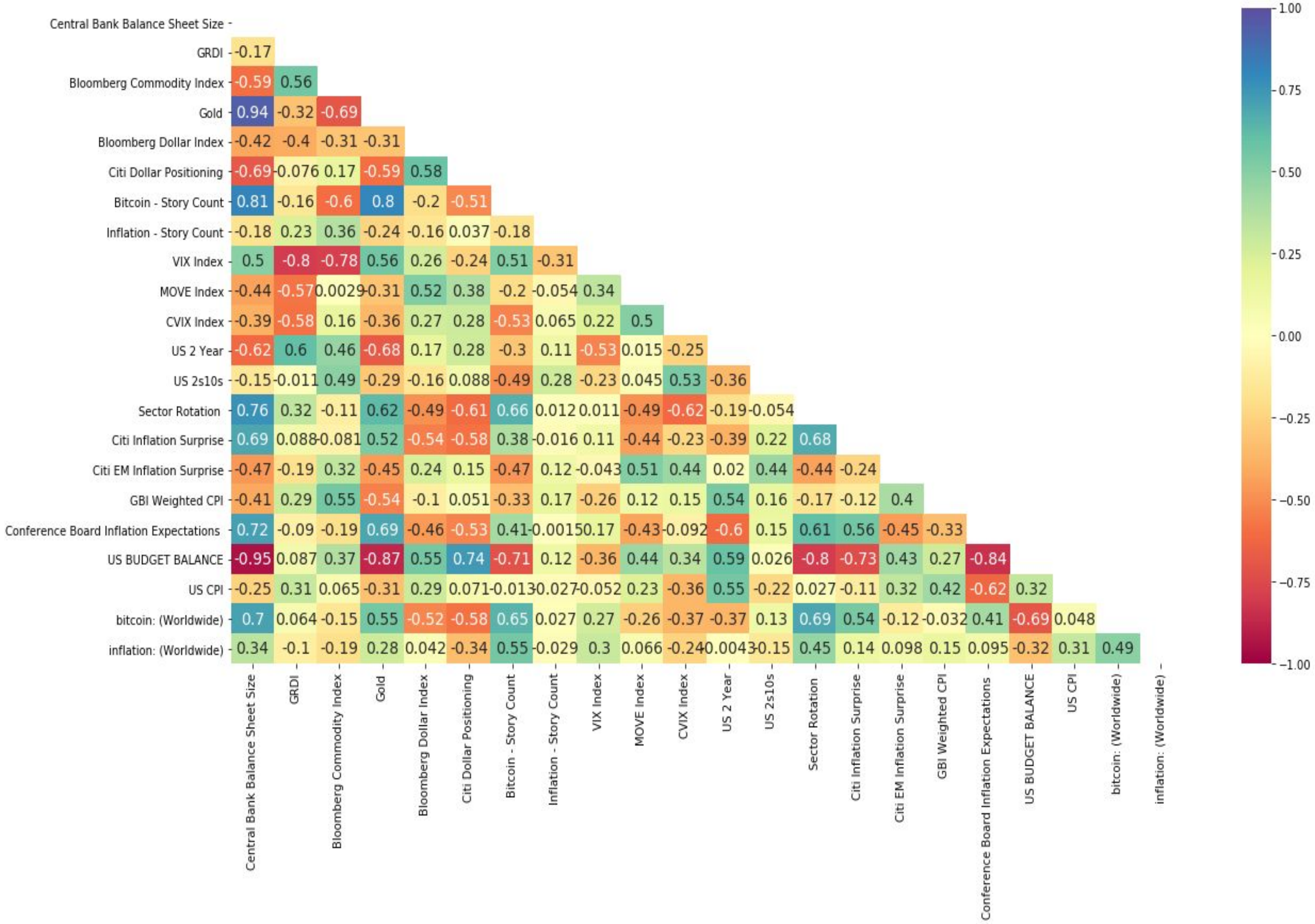
- **Monthly Reported Data & Daily Reported Data:** We needed to decide if we could mathematically transform a monthly reading into a daily reading by resampling it into the more frequently reported daily readings.
    - i.e. if we had a monthly CPI reading of 1% - could we run the ML algorithms assuming that 1% is reported daily across that time frame?
        - We discovered that we potentially can. This strategy however created additional problems. For example, if we used a % change on the daily data frame, or a first differencing, it would output many 0's.

- **Down-Sampling Our Data:**
    - Our models didn't appear to tell us much on a daily time frame as to what bitcoins price would be. Therefore, we thought maybe trying to look at drivers that were longer term and less 'noise driven' might make a difference.
        - We uncovered that did not appear to be the case as the models we tested using less frequent data also performed poorly.

Factor Correlation Heatmap

.80: BTC Story Count & Gold is positively correlated. This is indicative of a relationship between both variables in which both variables move in the same direction. Investors often look at both gold and BTC as a hedge against inflation.

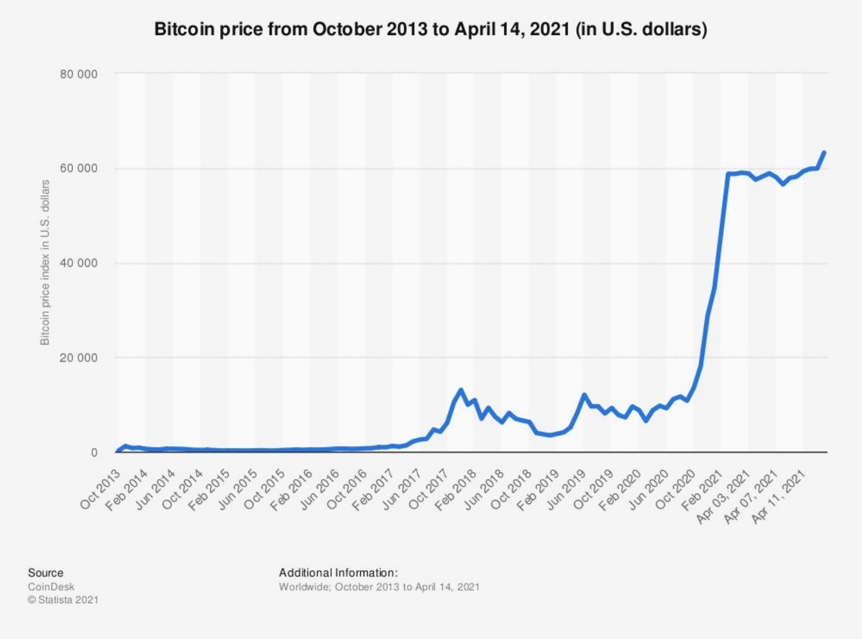
-.80: VIX Index & GRDI have a strong negative correlation



Our Core Hypothesis

Our core hypothesis was that bitcoin should respond to the following drivers:

- **Monetary & Fiscal Stimulus -** Central Bank Balance sheet, US Budget Balance, US Consumer Price Index (CPI), Inflation Expectations
- **Risk Seeking Behavior -** Global Risk Demand Index, Dollar Index, Dollar Positioning, Volatility Index (VIX), Crypto Volatility Index (CVIX), Merrill Lynch Option Volatility Estimate (MOVE), Sector Rotation Index
- **Inflation -** Bloomberg Commodity Index, Gold, US 2 Year, US 10-2 Year Treasury Yield Spread
- **Social Media Herding Biases & Hive driven thinking -** Bitcoin Story Count



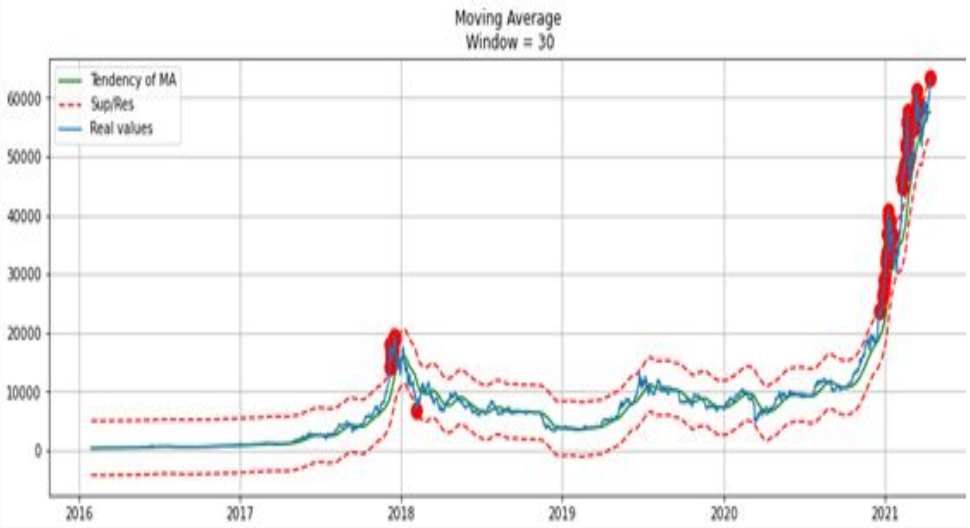
Model Summaries:

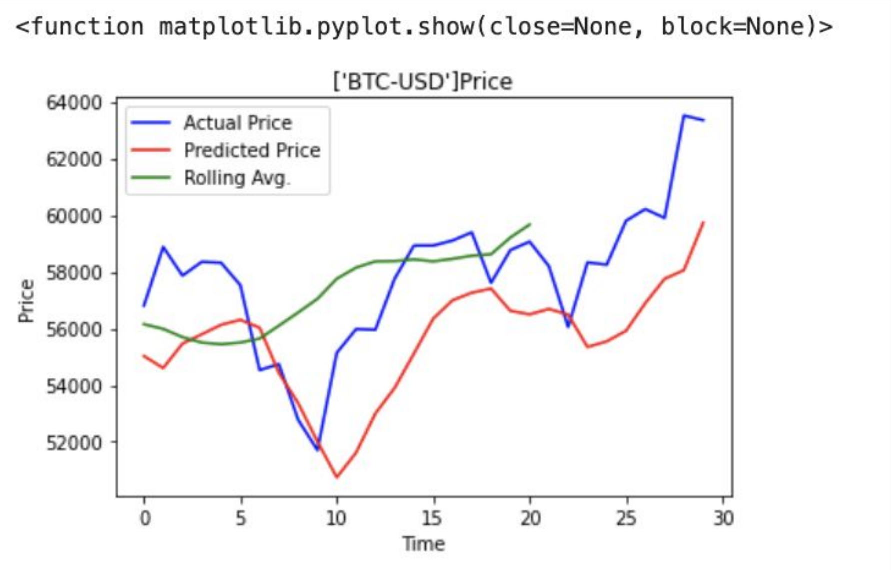
Models we used to try and determine a continuous variable & predict price of Bitcoin:

- Ridge Regression, Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net Regression, Principal Component Analysis (PCA and Regression), Backward Elimination, Ordinary Least Square (OLS), Random Forest, Neural Network - Keras

We have concluded that it might be better to try a classification path.

Neural Network Predictive Model:





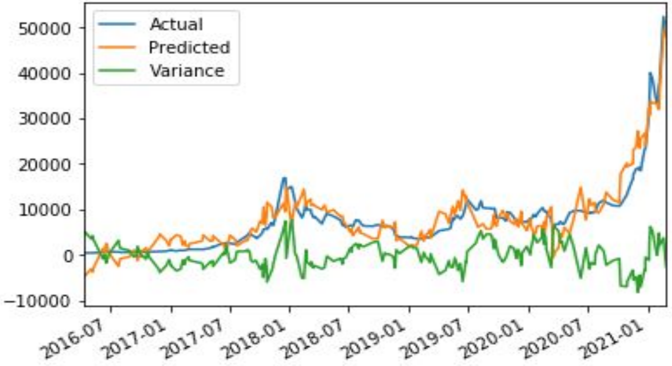
Predictive Model Using Least Absolute Shrinkage and Selection Operator (LASSO)

```
Training score for Lasso Regression:  89.93
R^2 score of Lasso Regression 88.18
Mean Absolute Error on test data for Lasso Regression: 2341.2619155809684
Mean Squared Error on test data for Lasso Regression: 8576689.860151675
Root Mean Squared Error on test data for Lasso Regression: 2928.5986171122313
```

```
10]: results_frame = pd.DataFrame({'Actual': y_test, 'Predicted': lasso_y_pred,
                                  'Variance': y_test - lasso_y_pred})

     results_frame.plot(kind = 'line')
```

```
10]: <matplotlib.axes._subplots.AxesSubplot at 0x1dc30556108>
```



R^2 of lasso regression: 88.18% which reveals that 88.18% of the data fits the model.

PCA Model Evaluation:

- We discovered that in the PCA factoring in 5 principal components that captured 80% of the data provided the same amount of accuracy of 11 components that had 91% of the data.
- This meant that with 80% of the data we were able to get the same accuracy in predictions.

```python
# Classification
# ---
from sklearn.ensemble import RandomForestClassifier
classifier = RandomForestClassifier(max_depth = 4 ,random_state=0)
classifier.fit(X_train,y_train)

# Predicting the Test set results
# ---
y_pred = classifier.predict(X_test)

# Confusion matrix & accuracy score
# ---
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
cm = confusion_matrix(y_test, y_pred)
print(cm)
print( accuracy_score(y_test, y_pred))
```

```
[[20 64]
 [20 98]]
0.5841584158415841
```

Next Steps:

Using our model accuracy that hit 58% we aim to create and test a trading model by identifying which factor(s) can become a signal to indicate profitable buy and sell positions.