Bitcoin Price Prediction Model

A Springboard Data Science Capstone Project

By: Jimmy Cheng

Problem

- Bitcoin is the longest surviving blockchain
- Many linear regression models have been applied to the field of price predictions but none have achieved much success
- Are there any particular factors that are instrumental in the price prediction?
- Can we use data to accurately predict daily bitcoin price changes to a 15% margin of error?

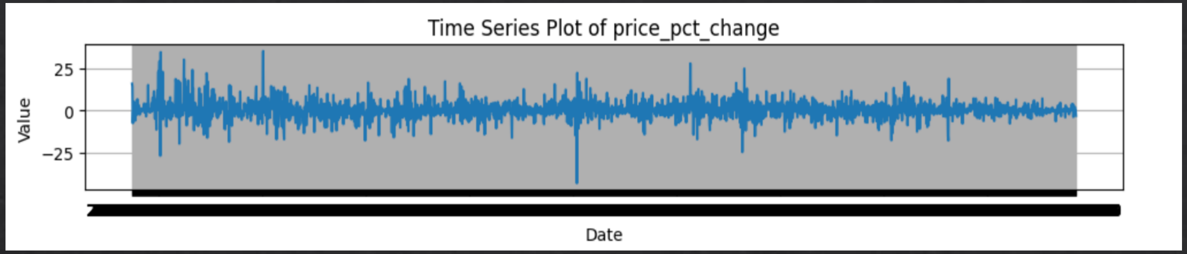
Data

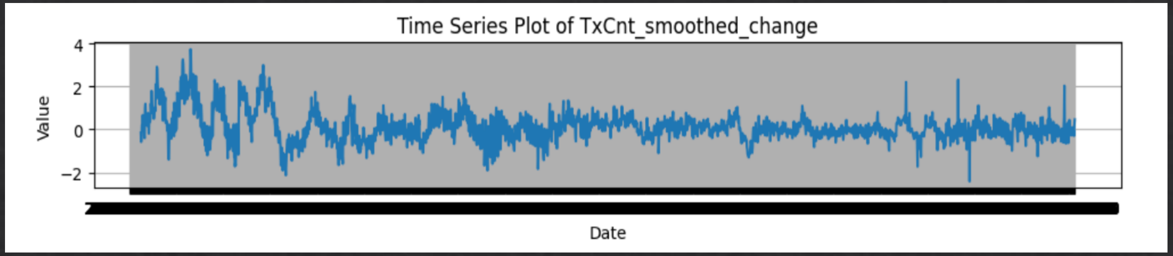
- Blockchain Data from CoinMetrics Python API
- Major Stock Indices historical data from Yfinance Python API
- US Gov historical macro data from Python FREDAPI
- Daily data from 2017 – Present (2471 observations) across 31 data series

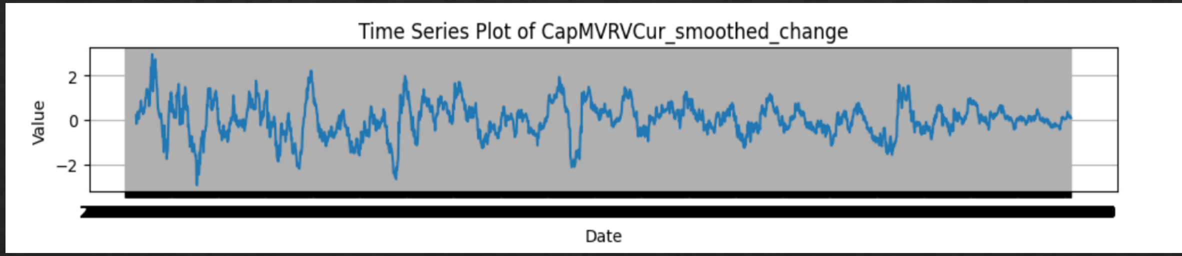
- **TARGET FEATURE: Daily Price Change**

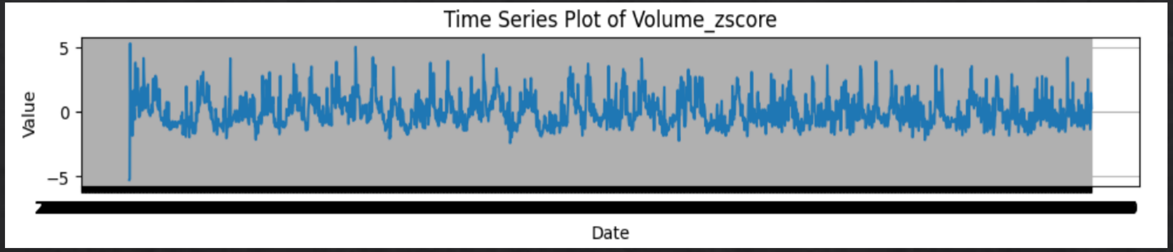
- Missing data on weekend was filled using forward fill method under the assumption that the future is always unknown.
- Data of different frequency is converted to daily frequency to match blockchain data and forward fill is used

Data Transformation





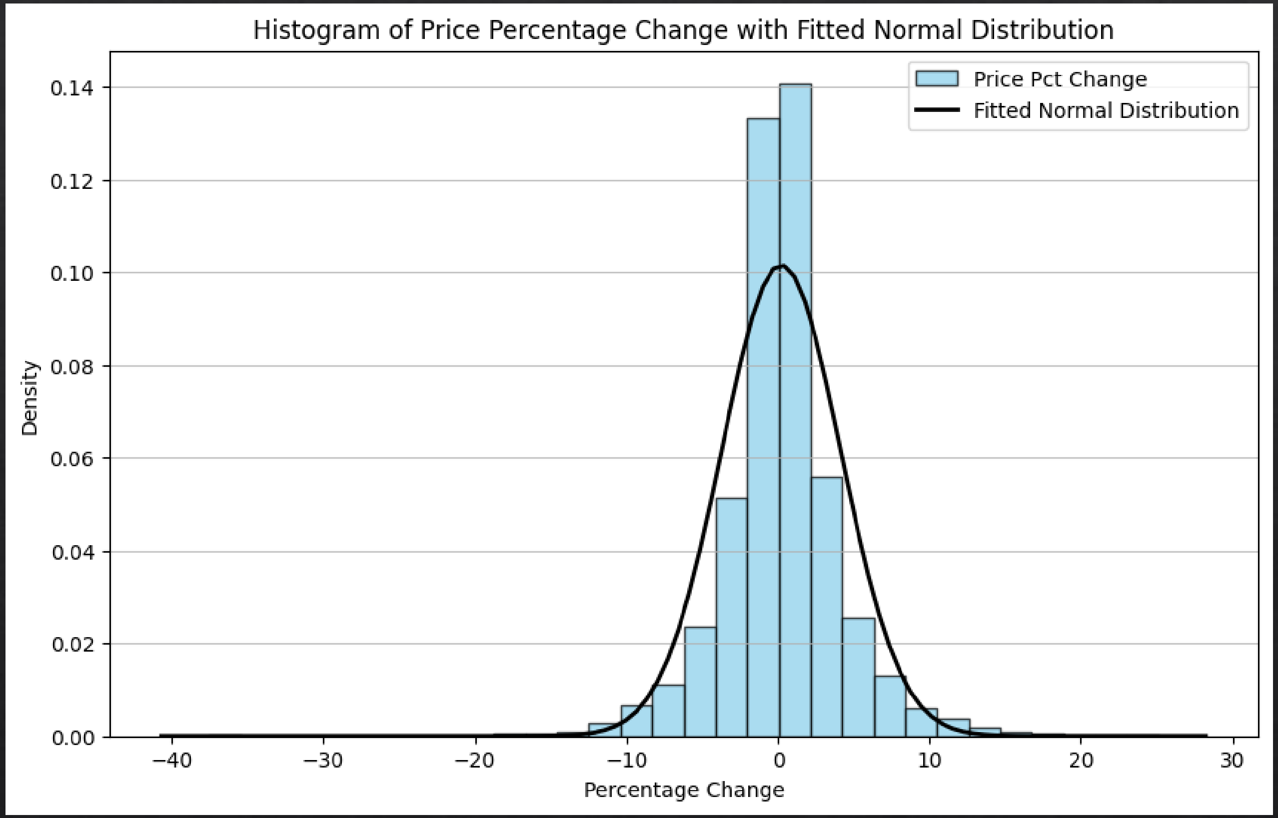




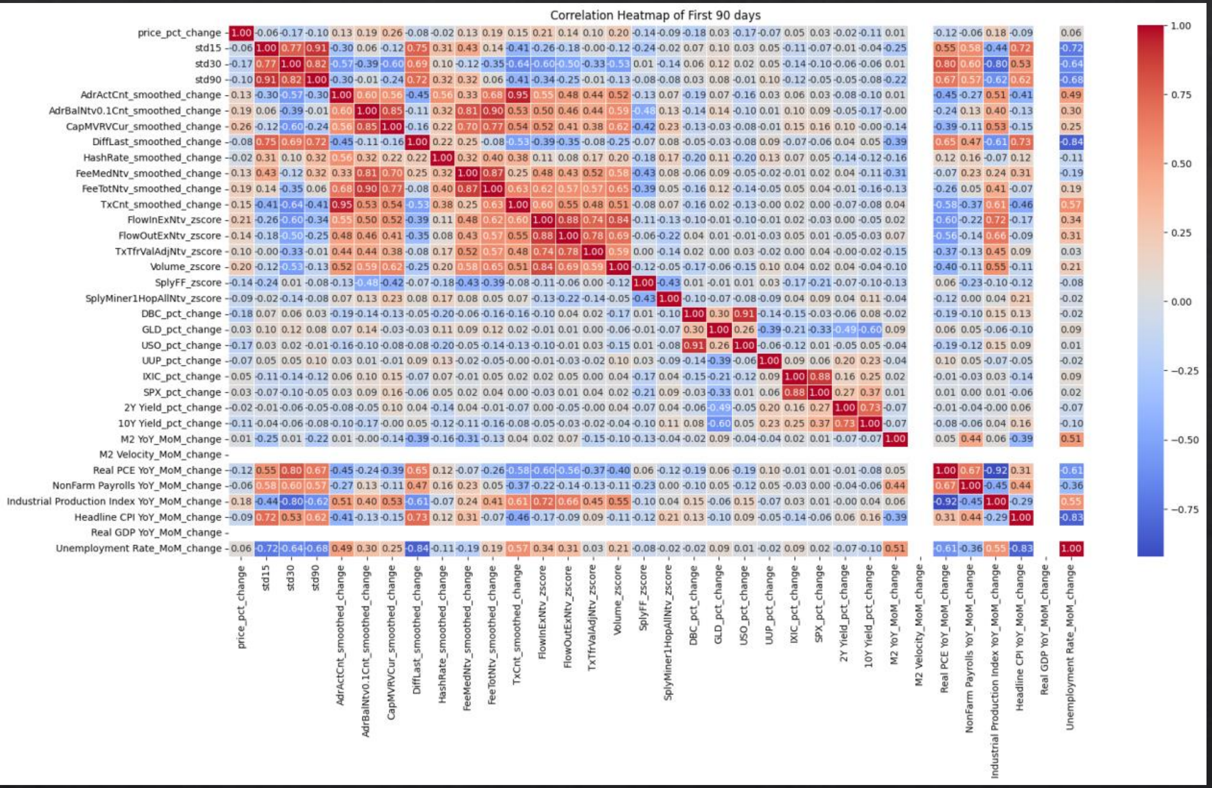
- Create Stationary Variables from non-stationary series
- ADFuller Test
- Create daily price percentage change series
- Smooth a few blockchain native metrics to reduce noise
- Create rate of change series for most time series data
- Create Z score series for volume related series

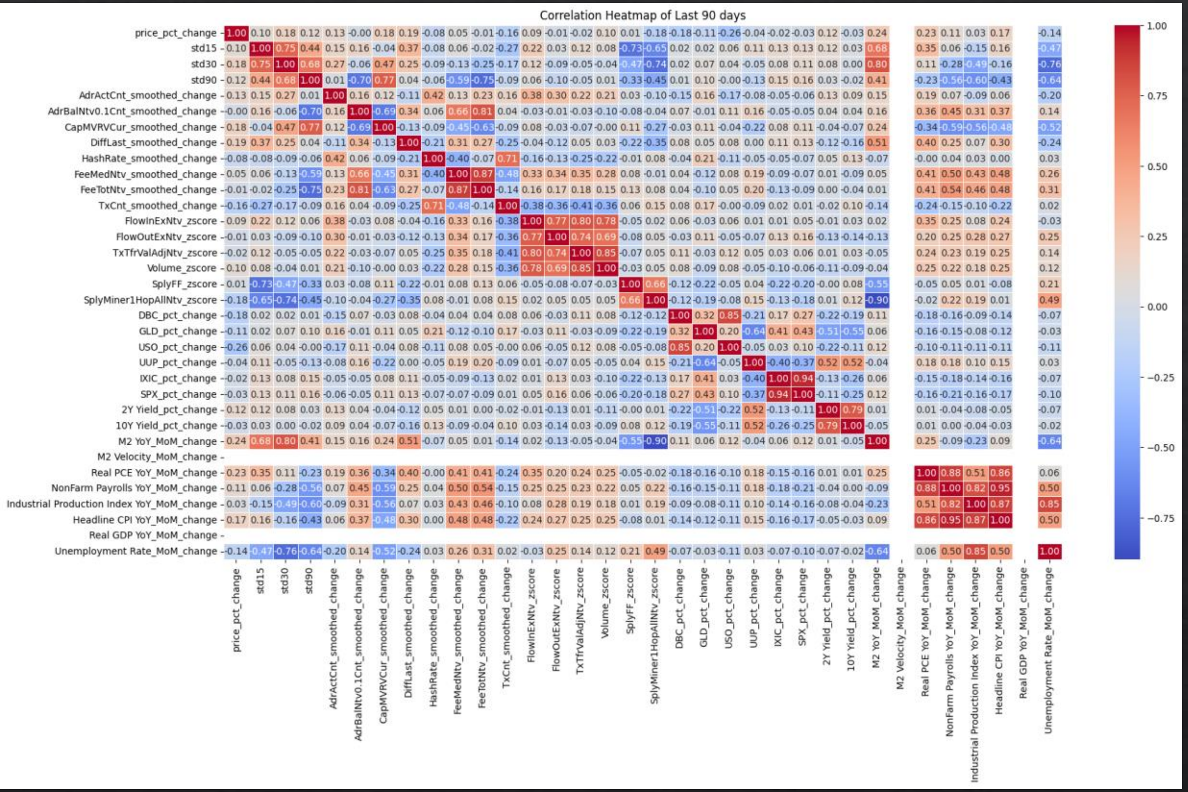
Exploratory Data Analysis

- What does our target feature – daily percentage price change looks like?
- We have a distribution that looks like a normal, with higher kurtosis and fatter tails



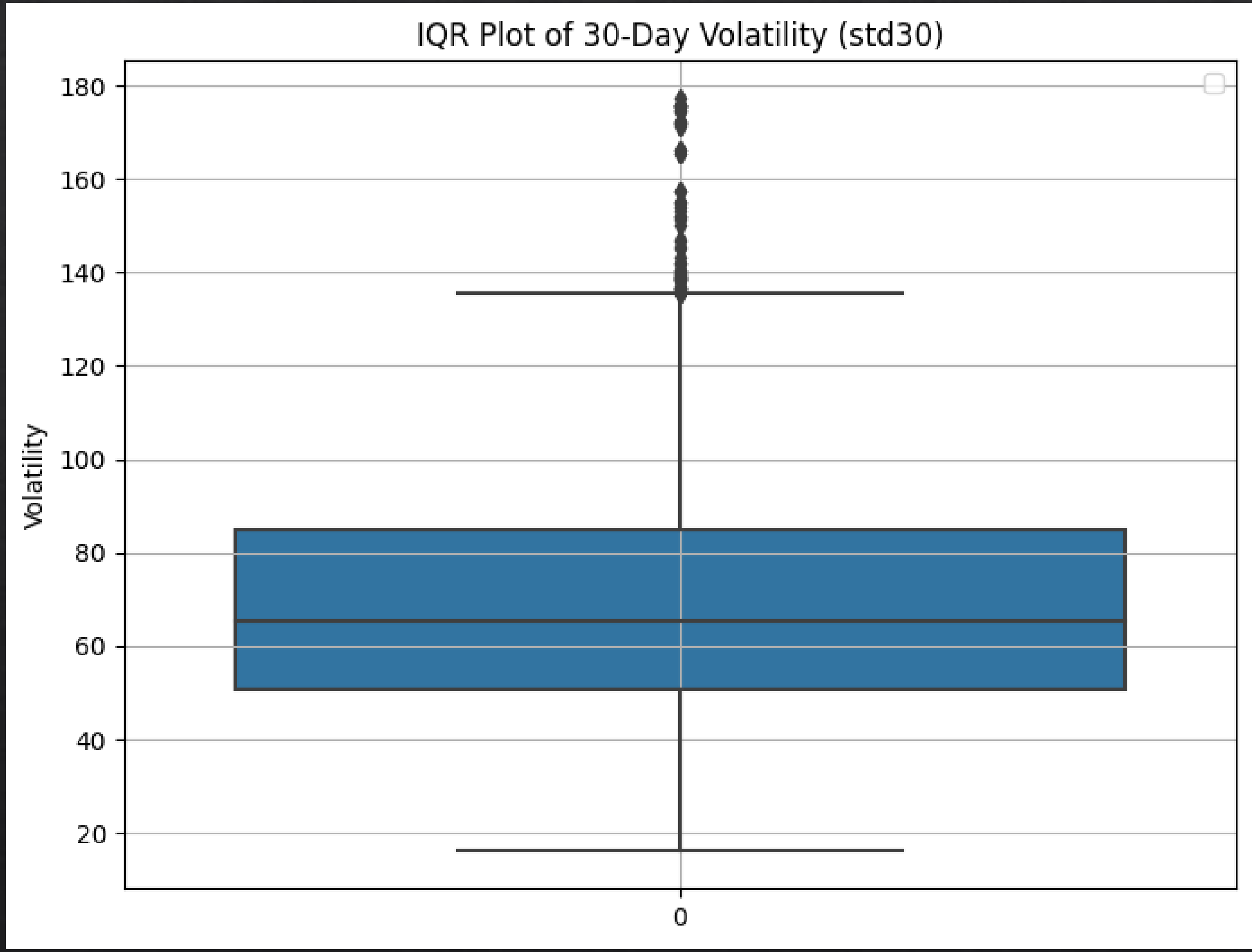
Correlation of factors over time



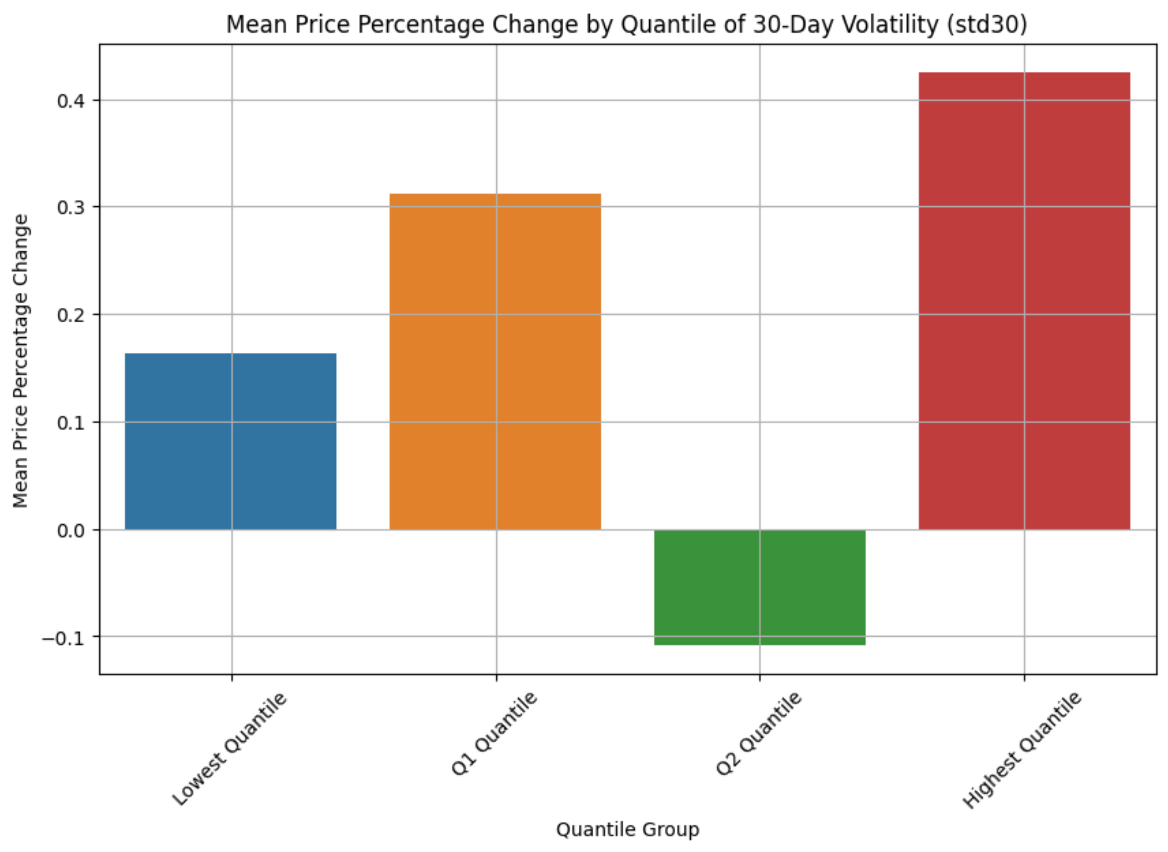


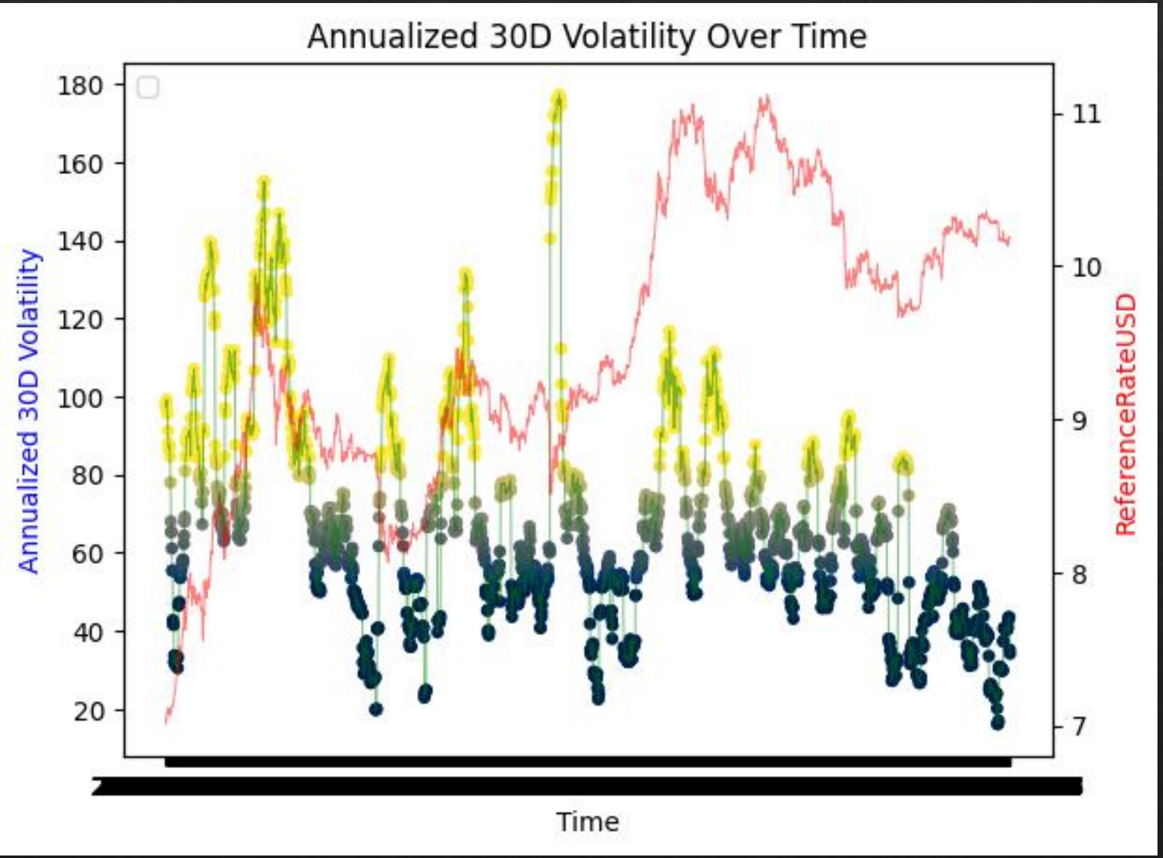
Volatility

- Historical volatility is an important factor in price predictions
- 15 day, 30 day, 90 day standard deviations have high correlations on days with big negative movement
- Create buckets for 30 day volatility for different investment regimes



Volatility


Mean Price Percentage Change by Quantile of 30-Day Volatility (std30)
Mean Price Percentage Change
0.4
0.3
0.2
0.1
0.0
−0.1
Lowest Quantile
Q1 Quantile
Q2 Quantile
Highest Quantile
Quantile Group


Annualized 30D Volatility Over Time
Annualized 30D Volatility
180
160
140
120
100
80
60
40
20
ReferenceRateUSD
11
10
9
8
7
Time

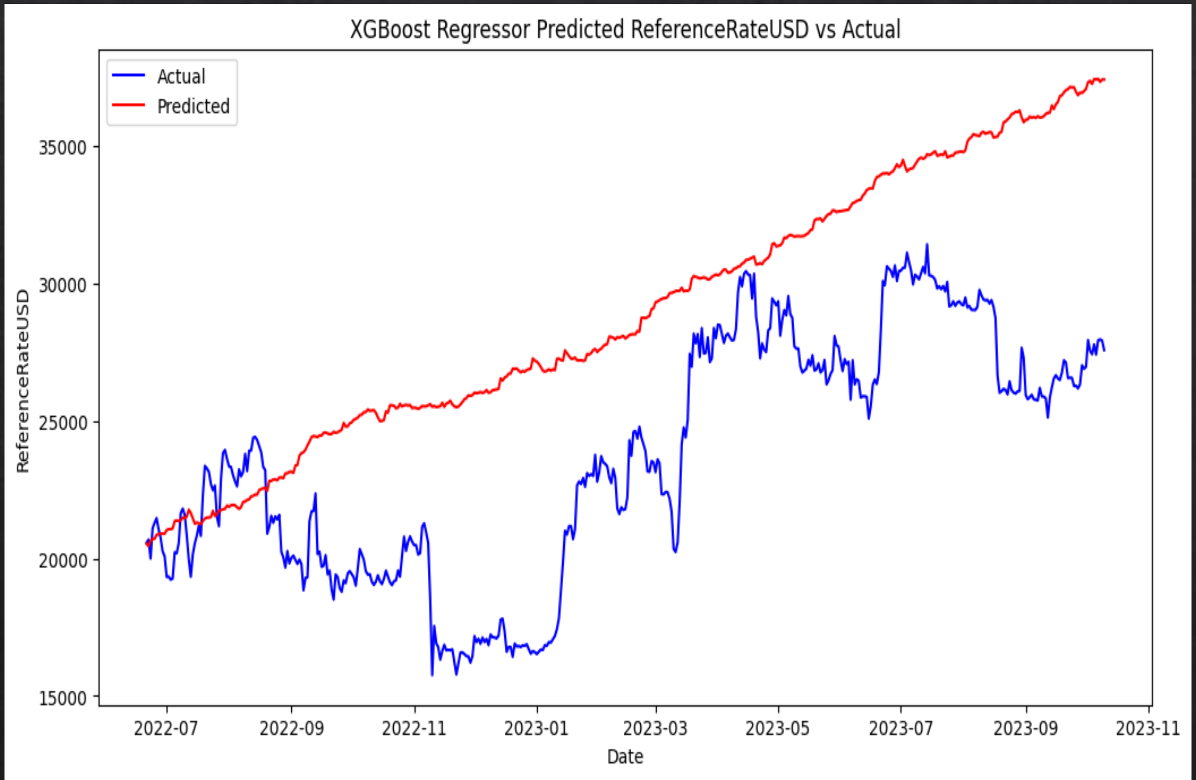
Preprocessing and Training Data Development

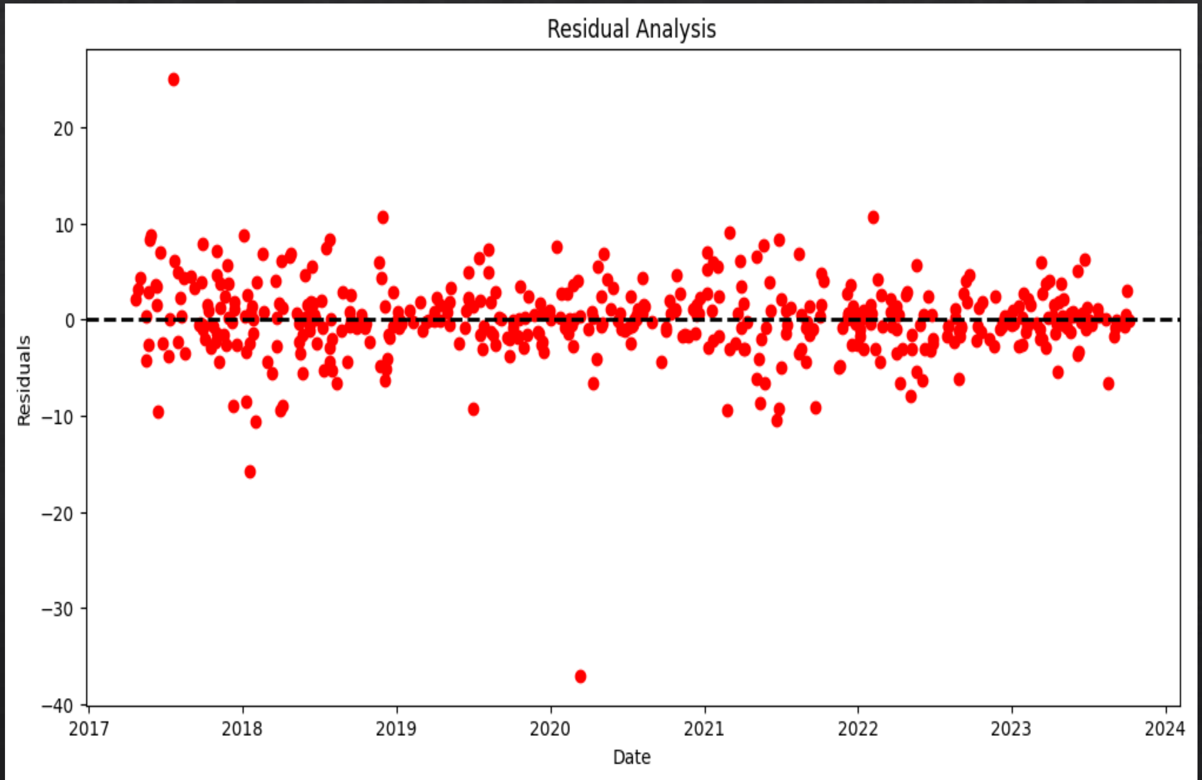
- Standard Scalar
- 80/20 Train Test Split
- Test for seasonality in data series using ACF, PACF plot
- **Model Metrics: RMSE, MAE, Residual Plots, Directional Accuracy**
- Directional Accuracy is defined by whether the predictions have the same sign as the actual daily change

Model Selection

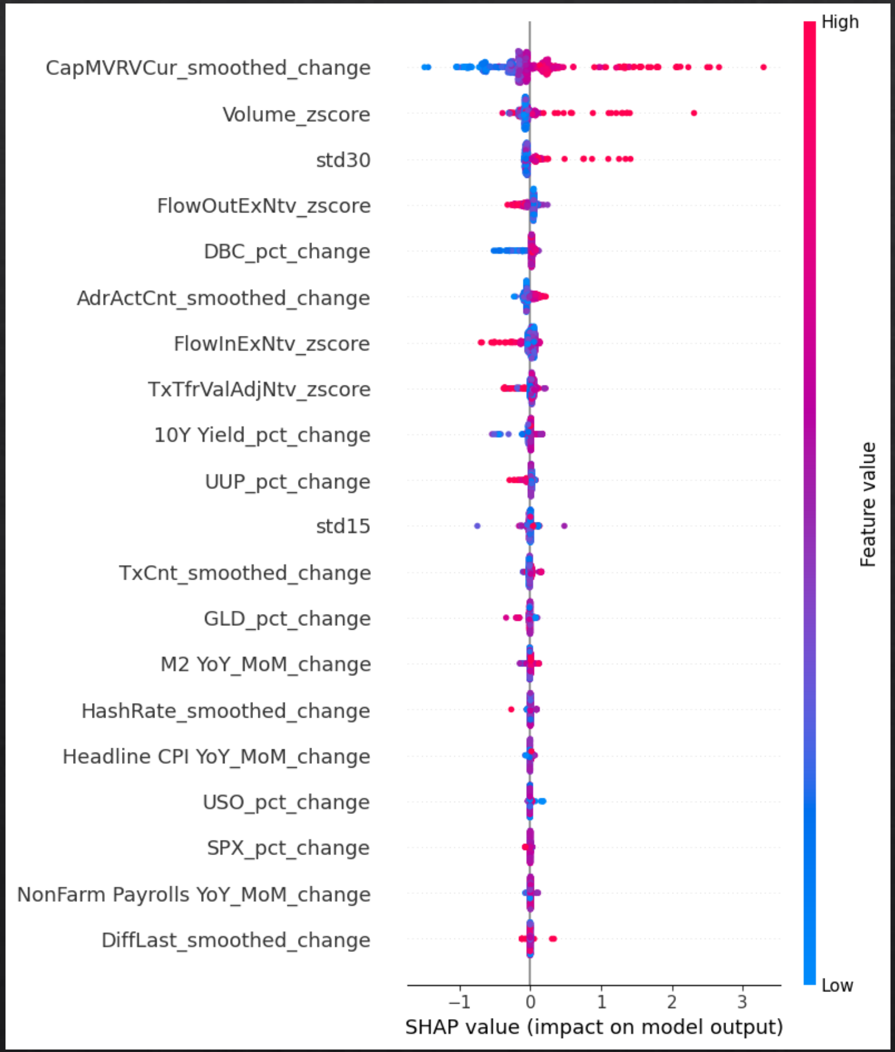
	RMSE	MAE	Directional Accuracy
LSTM with XGBoost	**1.924**	**1.4**	**69.72%**
XGBoost	3.969	2.585	56.6%
Random Forest	4.07	2.68	53.04%
LSTM 2 Layer	2.559	1.714	46.92%

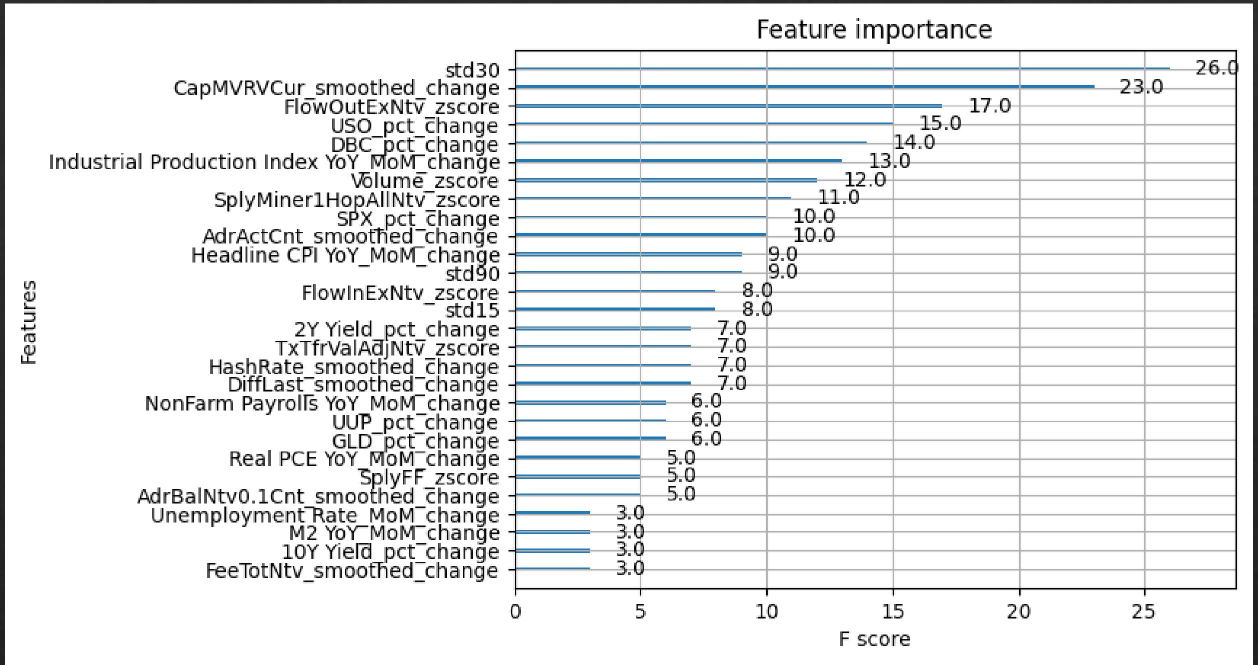
XGBoost Model





Feature Importance

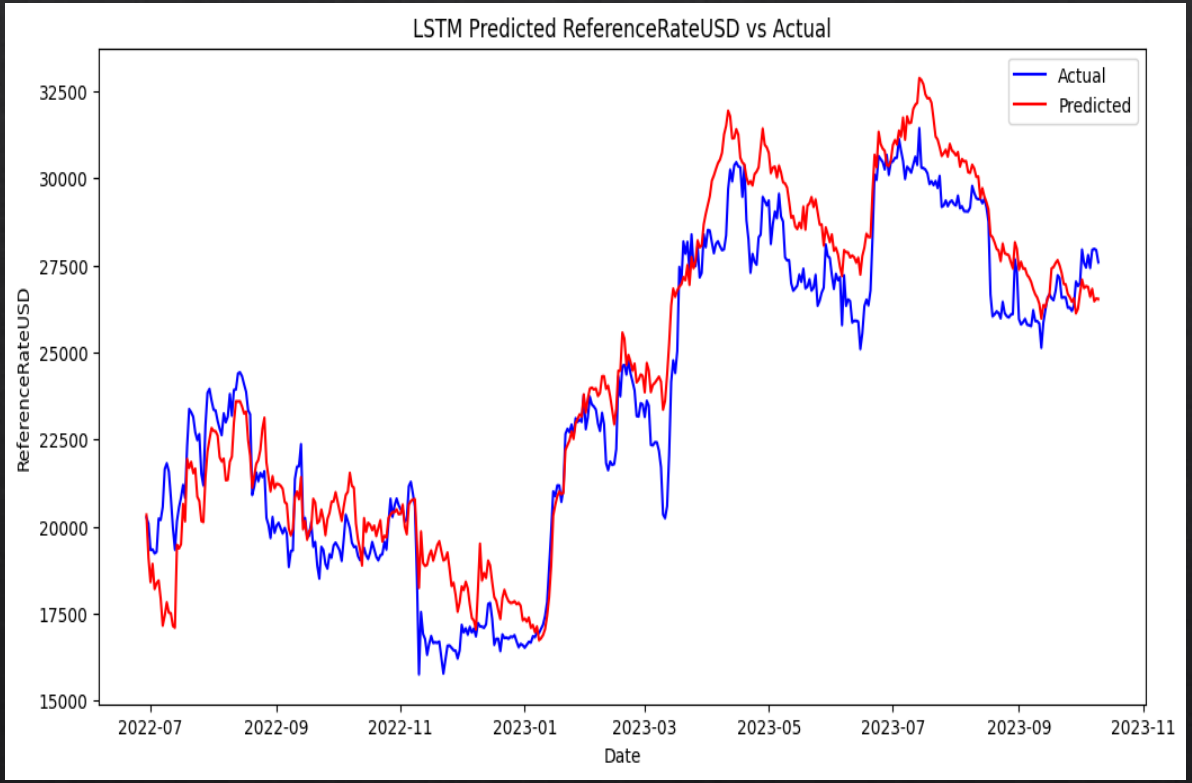


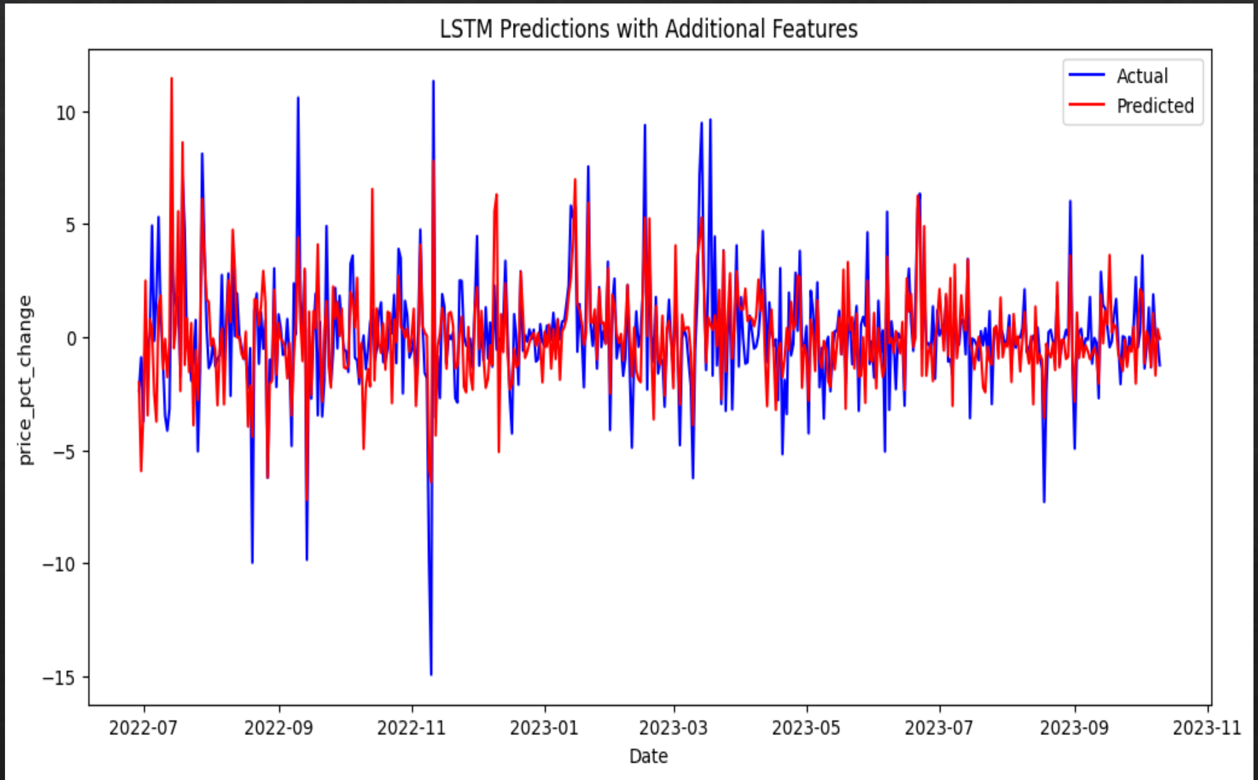


Final Model – LSTM with XGBoost top 5 features

- XGBoost converts the problem of predicting daily price percentage change into a time invariant problem
- Top ranked features are used as inputs into a Long Short Term Memory Neural Network
- There could be memory in the time series not captured by one factor ARIMA model

LSTM Model with XGBoost Features





Hyperparameter Tuning

Best Hyperparameters for XGBoost

Alpha: 10

n_estimators: 200

learning rate: 0.01

max_depth: 3

Best Sequence Length for LSTM Model: 40

Conclusion

- We have successfully built a LSTM Neural Network to predict bitcoin daily price changes with 70% accuracy.
- When transforming daily price changes to longer term price predictions, the model offers an error margin of within 15%.
- Model may need to be refitted and updated with new incoming data to keep up with accuracy