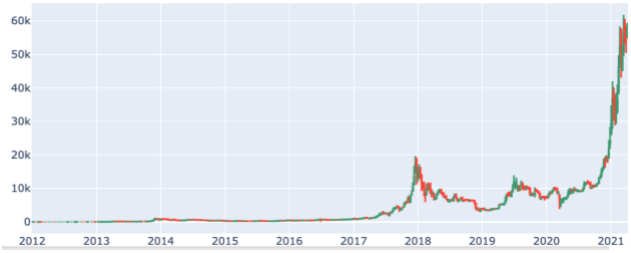
Final Report

Analyze and predict digital currency

Name: Meshal Fahad Almubarak

The dataset contains 4,857,377 row with 7 features for each. A few feature highlights include range of open price, close, low, and High. Nearly a third of the individual features could be grouped into more general categories, and an analysis of 4 of them was undertaken to inform baseline models and feature engineering.



Feature Engineering
-Combining particular Feature and ranges of numeric features to highlight strong signals and illogical values status identified during EDA
-The entire training dataset of 4,857,377 records was split into 0.1 train vs. holdout. Predictions on the 10% holdout were limited to the very end, so this split was only used and scores seen just once.

Final result: 50 layers on deep learning (`LSTM`)

Timestamp	Close	Predictions
2014-01-12 09:05:00	878.01	952.829529
2014-01-12 09:06:00	878.01	952.761353
2014-01-12 09:07:00	878.01	952.719849
2014-01-12 09:09:00	878.01	952.699463
2014-01-12 09:10:00	878.01	952.692810

Timestamp	Close	Predictions
2021-03-30 23:56:00	58686.00	19096.496094
2021-03-30 23:57:00	58685.81	19096.652344
2021-03-30 23:58:00	58723.84	19096.679688
2021-03-30 23:59:00	58760.59	19096.775391
2021-03-31 00:00:00	58778.18	19097.062500

Tools
Numpy and Pandas for data manipulation.
Keras and Scikit-learn for modeling.
Matplotlib and Plotly for plotting.
Matplotlib for connect with google drive.