



My First Crypto Project Report

Best and Brightest Analysis Team:

Daniella Mayoral Tyler Cutrer Celenia Chapa

Request

Our client, MegaCoin Co., contracted our team to analyze and create a time series forecasting model that will predict cryptocurrency. The client also had a second request to design a website for first-time buyers of cryptocurrencies (also known as cryptos). Our team, Best & Brightest Analysis, immediately began discussing what this type of project would require to give a broader understanding to new users and bring fortune to the world!

As our team began the project, we discussed cryptocurrency and realized our knowledge of cryptocurrency varied wildly with some having a broader understanding than others. We started our project with initial research into the history of crypto and how it came to be such a huge component in today's market. We then found our data set and began developing our project: My First Crypto!

The Data

We downloaded our data from Yahoo Finance. The site provided data for a variety of cryptocurrencies and allowed us the ability to select the time frame and frequency. The project team decided that we would create the demo for the project with four currently popular cryptos: Bitcoin, Litecoin, Ethereum, and XRP. We decided to pull the data for the last five years and set the time frame as 6 December 2016 - 6 December 2021. The data consisted of daily data for each crypto. The data columns were of the opening and closing price, the adjusted closing price, the total volume of crypto sold, and the high/low price for each day. Fortunately, this was the exact format that we needed and no preprocessing or cleaning of the data was required.

Machine Learning

Arima Model for Time Series Forecasting

ARIMA is short for Auto-Regressive Integrated Moving Average which explains that a given time series based on its past values creates its equation to be used to forecast future values.

An ARIMA model can be understood by outlining each of its components as follows:

- Autoregression (AR): refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.

- Integrated (I): represents the differencing of raw observations to allow for the time series to become stationary such as data values being replaced by the difference between the data values and the previous values.
- Moving average (MA): incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Each component in ARIMA functions as a parameter with a standard notation of p, d, and q. Where the integer values substitute for the parameters to indicate the type of ARIMA model used.

The parameters can be defined as:

- p: the number of lag observations in the model; also known as the lag order.
- d: the number of times that the raw observations are different; also known as the degree of difference.
- q: the size of the moving average window; also known as the order of the moving average.

Linear regression models work best when the predictors are not correlated and are independent of each other.

The value of d, therefore, is the minimum number of differences needed to make the series stationary. And 'q' is the order of the 'Moving Average' (MA) term. It refers to the number of lagged forecast errors that should go into the ARIMA Model. ARIMA models are, in theory, the most general class of models for forecasting a time series. Time series can be classified by transformations such as differencing and logging. The easiest way to think of ARIMA models is as fine-tuned versions of random-trend models with a bit of fine-tuning consisting of adding lags of the differenced series and/or lags of the forecast errors to the prediction equation. This is needed to remove any last traces of autocorrelation from the forecast errors.

To identify the appropriate ARIMA model for time series data, we need to begin by identifying the orders. We used SARIMAX to evaluate if the order would fit the model because there is more than one order that is used. SARIMAX stands for Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors.

The basic information is pretty self-explanatory:

- Dep. Variable - What we're trying to predict.
- Model - The type of model we're using. AR, MA, ARIMA.
- Date - The date we ran the model
- Time - The time the model finished
- Sample - The range of the data

After generating the SARIMAX summary(figure 1), it's important to understand what this information is in tales and the best way to understand is to break down each section into 3 steps:

- General Information(black and Red Box)
- Analysis (Green Box)
- Assumptions(Blue Box)

Dep. Variable:	y	No. Observations:	57			
Model:	SARIMAX(0, 1, 1)x(2, 1, 0, 12)					
Date:	Sun, 12 Dec 2021	AIC:	872.117			
Time:	17:22:32	BIC:	879.253			
Sample:	0 - 57	HQIC:	874.763			
Covariance Type:	opg					
coef	std err	z	P> z	[0.025	0.975]	
ma.L1	0.6329	0.108	5.871	0.000	0.422	0.844
ar.S.L12	-0.9021	0.419	-2.155	0.031	-1.723	-0.082
ar.S.L24	-0.5307	0.177	-3.005	0.003	-0.877	-0.185
sigma2	1.813e+07	1.88e-08	9.64e+14	0.000	1.81e+07	1.81e+07
Ljung-Box (L1) (Q): 0.02 Jarque-Bera (JB): 21.96						
Prob(Q): 0.88		Prob(JB): 0.00				
Heteroskedasticity (H): 7.84		Skew: -0.79				
Prob(H) (two-sided): 0.00		Kurtosis: 6.08				

Figure 1: SARIMAX Results Bitcoin

General Information(black and Red Box)

This section provided the basics needed such as

- Model in which it's been tested and best suited
- Date that the model was ran
- Time that the model was finished
- Number of Observation
- Sample which the range of the data

With this in mind, we want to dive into the detail of the model. There are many models that are explored by many data analytics however we are going to only focus on the ones we know from our output results. The following list below shows the types of models that can be formed:

ARIMA (0,0,0) = white noise model: mostly used in engineering models and will not be focused on in this report.

ARIMA (1,0,0) = first-order autoregressive model: if the series is stationary and autocorrelated, perhaps it can be predicted as a multiple of its previous value, plus a constant.

ARIMA (2,0,0) second-order autoregressive model can be described as a system whose mean reversion takes place in a sinusoidally oscillating fashion, as the motion of a mass on a spring that is subjected to random shocks.

ARIMA (0,1,0) = random walk: a random walk is a mathematical term known as a stochastic or random process, that describes a path that consists of a succession of random steps. An example of a random walk is the random walk on the integer number which starts at 0 and at each step moves

+1 or -1 with equal probability. Other examples include the path traced by a molecule as it travels from a liquid state to a gas state.

ARIMA (1,1,0) = differenced first-order autoregressive model: one in which the current value is based on the immediately preceding value.

ARIMA (0,1,1) Single Exponential smoothing: The model uses a technique for smoothing time series data using the exponential window function. Whereas in the simple moving average the past observations are weighted equally, exponential functions are used to assign exponentially decreasing weights over time. It is an easily learned and easily applied procedure for making some determination based on prior assumptions by the user, such as seasonality.

ARIMA (0,2,1) or (0,2,2) Double Exponential Smoothing: a Model that an exponential moving average that takes into account the tendency of data to either increases or decrease over time without repeating

Note: The difference between single and double exponential is that single exponential smoothing uses a weighted moving average with exponentially decreasing weights.

Analysis (Green Box)

In order to make sure the model is confirmed we need to evaluate it statistically. For this, we need to test the data using P-test and Z-test to verify the data is significant. In order for the test to be valid, the P-test must be less than 0.5 to prove the theory behind the model.

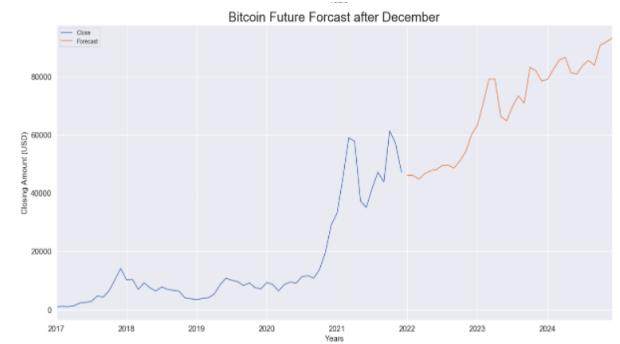
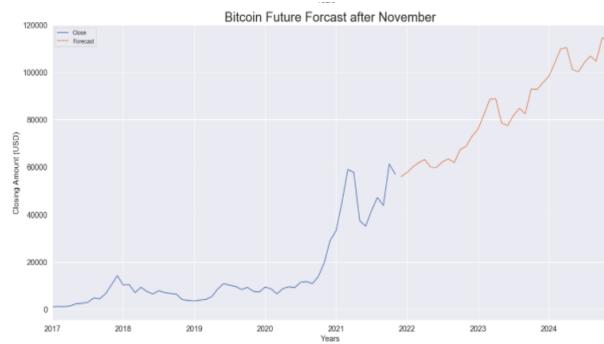
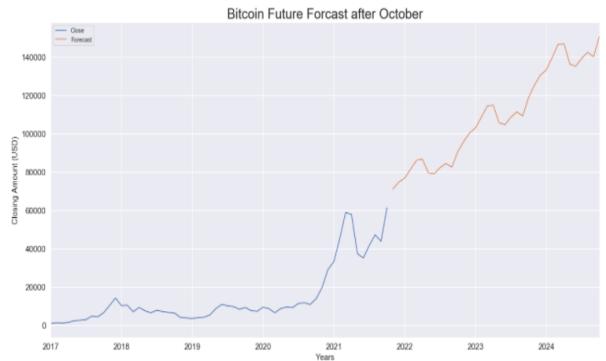
Assumptions(Blue Box)

Now we understand the two boxes let's put everything. The final box is known as the Ljung Box and it is a test that provides any errors that could be white noise. From Figure 1 we see that the 0.02 which is close to no white noise showing the data can be formed to an exponential regression model to test the data.

With this in mind, we can run a test model and actual model to predict the data.

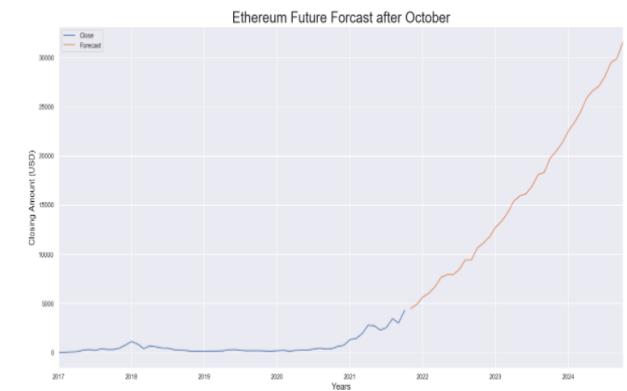
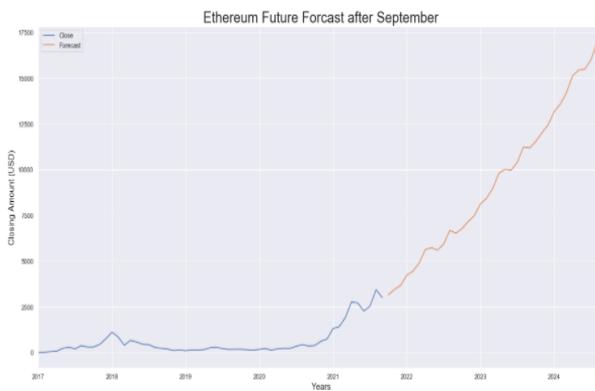
Bitcoin Analysis:

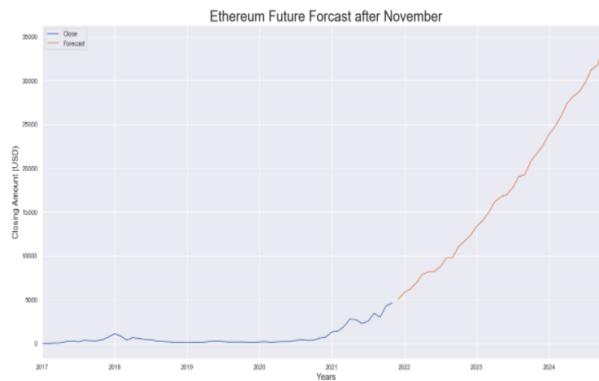
In these graphs, you can see that Bitcoin continues with a steady rise with slide decreases in price. The largest dip in price is predicted to be in mid-2023 with the highest peak being in early 2024. Our predictions show that Bitcoin will rise to roughly \$90,000 - \$110,000 in the next three years. Our recommendation: Buy now to maximize your profits!



Ethereum Analysis:

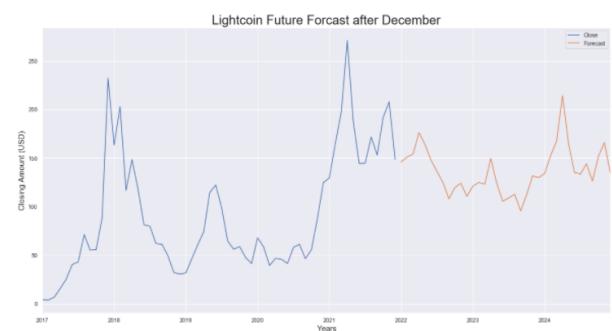
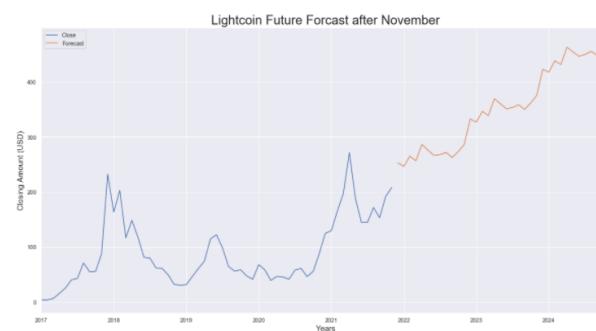
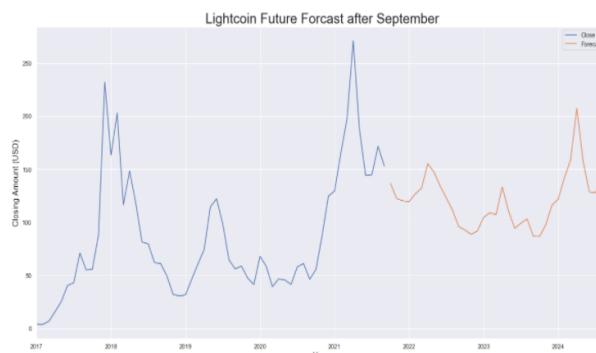
In these graphs, you can see that Ethereum is predicted to have a meteoric rise in the next three years. It will be on the continuous rise until 2024 and we predict that the price will be somewhere between \$25,000 - \$35,000. Our recommendation: Buy now to maximize your profits!





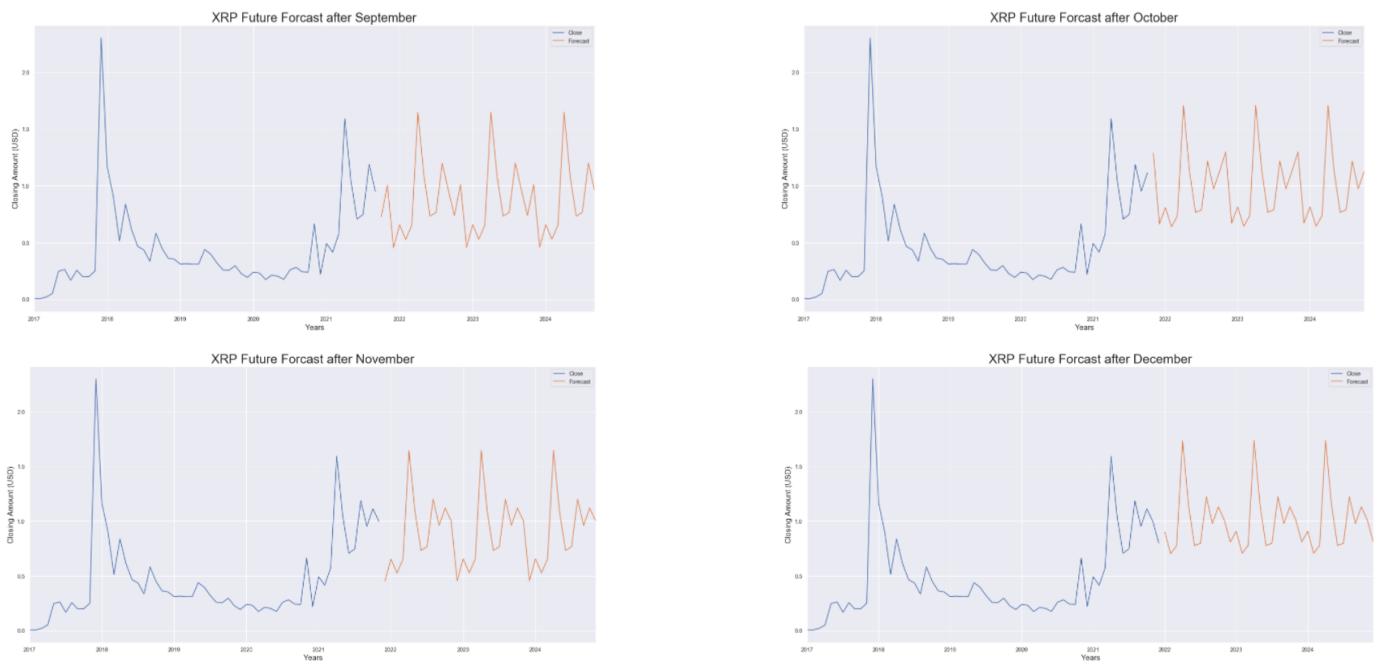
Litecoin Analysis:

In these graphs, you can see that Litecoin has many dips and peaks until 2024. We predict that the lowest dip will be in late 2023 and the highest peak will be in early 2024. Our recommendation: Be Cautious and keep your eye on the price changes! Remember to buy low and sell high!



XRP Analysis:

In these graphs, you can see that XRP is extremely inconsistent over the next three years with wildly contrasting changes in price. According to our predictions, the pattern does hold consistent with the dips occurring early every year and the peaks occurring a few months afterward. Our recommendation: Be cautious and time your purchases! Remember to buy low and sell high!

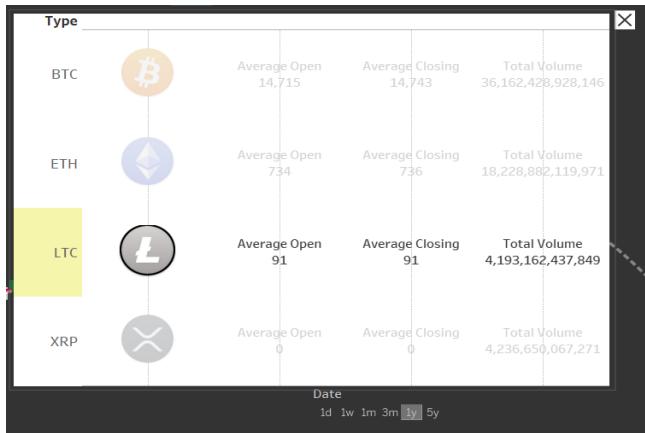


Tableau

This section was designed for the users to be able to play with the historical data of the four main cryptocurrencies that we decided to use in our model. The historical data was from 6 December 2016 - 6 December 2021. The dashboard was compiled using two different line graphs and a Japanese candlestick graph. The Japanese candlesticks are visualizing the tick data with ample information that tells us a story about the price movement. The four components of a candlestick are the price for open, high, low, close. Line Charts use points connected by line segments to demonstrate changes in values. The two-line charts represent the Market Cap Value (total dollar market value) and Total Volumes of crypto sold.



The Dashboard was configured using parameters that connected the charts to the chosen cryptocurrency. This allows the user to choose the desired crypto in a popup menu which will alter the values in the dashboard. Additionally, the user has more control by utilizing the date selector and selecting if they'd like to see 1 day, 1 week, 1 month, 3 months, 1 year, or 5 years.



Web Design

After initializing a Flask-powered API, we used Heroku to deploy the web page to the cloud. We decided to use Bootstrap to create the themes for the website. This theme provided large imagery and text space for our visualizations and analysis. The theme's imagery was updated with cryptocurrency photography provided by pexels.com to express the intricacies of cryptocurrency. When it came to understanding cryptocurrency, we wanted to approach our web design cleanly and simply. We created an easy-to-use menu that provided the user the ability to see recommendations from the SARIMA model, an analysis page that included the Tableau of our historical data, our report from the project, works cited page, and the project teams information.

Model Limitations

Our team had several limitations during the development of our demo page:

- Our main limitation was time constraints
- Cryptocurrency is highly volatile. Outside variables affect the price which could cause the model to be incorrect.
- Our model used historical data and not live data. This renders our model unable to adjust to unforeseen variables affecting the price.

Future Goals

- Link the model to live data so that it can predict more accurately.
- Update the model so that it can predict using daily fluctuations in data instead of monthly.
- Increase the number of cryptocurrencies used.

Resources

Bootstrap & Code Snippets:

Theme by Start Bootstrap

<https://startbootstrap.com/theme/clean-blog>

Video Header Styles:

<https://startbootstrap.com/snippets/video-header>

Images:

pexels-david-mcbee-730564.jpg

<https://www.pexels.com/photo/round-silver-and-gold-coins-730564/>

pexels-alesia-kozik-6765372.jpg

<https://www.pexels.com/photo/close-up-shot-of-gold-coins-6765372/>

pexels-olya-kobruseva-7873548.jpg

<https://www.pexels.com/photo/cryptocurrency-application-on-a-cellphone-7873548/>

pexels-darlene-alderson-7971337.jpg

<https://www.pexels.com/photo/man-people-woman-desk-7971337/>

pexels-marta-branco-1263324.jpg

<https://www.pexels.com/photo/three-round-silver-and-gold-colored-coins-1263324/>

pexels-nataliya-vaitkevich-6120169.jpg

<https://www.pexels.com/photo/marketing-blue-internet-technology-6120169/>

Coin Images

Catz Coin

<https://www.catzcoin.io/>

Sloth Coin

<https://coinmarketcap.com/currencies/slothcoin/>

Minion Coin

<https://thecoincompany.com.au/2020-minion-made-lunar-year-of-the-mouse-1oz-silver-proof-coin/>

Pexels Video:

<https://www.pexels.com/video/close-up-video-of-a-coin-7450203/>

Data:

Bitcoin

<https://finance.yahoo.com/quote/BTC-USD/history?p=BTC-USD>

Ethereum

<https://finance.yahoo.com/quote/ETH-USD?p=ETH-USD&.tsrc=fin-srch>

Litecoin

<https://finance.yahoo.com/quote/LTC-USD?p=LTC-USD&.tsrc=fin-srch>

XRP

<https://finance.yahoo.com/quote/XRP-USD?p=XRP-USD&.tsrc=fin-srch>

Model Inspiration/Sources:

Arima Model for Time Series Forecasting

<https://www.geeksforgeeks.org/python-arima-model-for-time-series-forecasting/>

Tableau for Cryptocurrency Dashboard

<https://tableau.toanhoang.com/cryptocurrency-dashboard-full-tutorial/>

Understanding ARIMA Models:

https://faculty.fuqua.duke.edu/~rnau/Decision411_2007/411arim.htm

History of Crypto

<https://www.thesoftwareguild.com/blog/a-brief-history-of-cryptocurrency/#:~:text=Many%20think%20cryptocurrency%20is%20a%20concept%20developed%20and,traced%20back%20to%20one%20man%3A%20cryptographer%20David%20Chaum.>

Glossary:

<https://www.finder.com/cryptocurrency-glossary>