Predicting Cryptocurrency Prices

Om Patel   
*Wentworth Institute of Technology*

*Abstract*— This project investigates the trends and factors influencing cryptocurrency prices, with a focus on Bitcoin, and attempts to predict future price changes using a linear regression model. Data from trusted sources, including Kaggle and the Federal Reserve Bank, were utilized to analyze historical prices, market sentiment, and the Consumer Price Index (CPI). Key findings include a slight correlation between CPI and Bitcoin prices, suggesting that cryptocurrencies may serve as an inflation hedge. Market sentiment, however, showed limited impact on price trends, potentially due to dataset limitations. The regression model achieved a mean absolute error of $1075 and an R-squared value of 46.5%, indicating moderate predictive accuracy but room for improvement with additional factors. Despite its simplicity, the model highlights the complexity of cryptocurrency markets and provides a foundation for further exploration. Future work will aim to incorporate more economic variables, improved datasets, and advanced models for greater accuracy.

Keywords— BTC, CPI, Cryptocurrency, market, regression, inflation hedge

# Introduction

This project focuses on the trends of cryptocurrency prices throughout the years of different crypto coins and potentially attempts to predict future prices. This is an important and interesting topic because cryptocurrencies have been on the rise in recent years, especially with virtual currencies and virtual realities on the rise. There are many factors that influence cryptocurrencies prices similarly to how stock prices are influenced by many different economic factors. While it’s hard to always account for all the possible things that could happen, this project attempts to give a general idea of price trends and the potential future changes to the price.

# Datasets

## Source of dataset

The datasets used in this project are downloaded from Kaggle for the most part and one of them is from the Federal Reserve Bank as it is related to the CPI, or consumer price index. The sources that are from Kaggle have been scraped from trusted sites such as Coinbase using their APIs which allows them to be considered trustworthy for the purpose of this project. The datasets recorded the past prices of all of the cryptocurrencies that are included in the dataset using the API and one of the sources related to market/investor sentiment is collected from web scraping from multiple crypto news sources.

## Character of the datasets

An important note is that I was unable to access Statista for the datasets I wanted so I had to resort to Kaggle. The format of the datasets is all in csv or xlsx file format in order to access each row and column entry as desired by the user. The first source includes columns labeled timestamp, price, volume, and market cap [1]. The second source has the columns date, price, total volume, market cap, and coin name, and this was specifically used to look at Ethereum only [2]. The third source picks 50 popular cryptocurrencies and these were used to be merged into one singular file to track all of these values together and notice trends. The column names were labeled the same way as source 2 in order to manage similar values easily. To combine them, the files were simply merged together and appended to the end of the current edited file with the preferred columns needed for the purpose of this project with the resulting file being very large [3]. Source 4 was used to look at sentiment and polarity towards cryptocurrency in order to relate whether the views on crypto had an impact on prices and this was done through web scraping by the dataset owner. The columns are source, subject, sentiment, text, title, and url. This dataset was incredibly large due to the high volume of urls. The urls could have been cleaned because the visualization was more concerned with the sentiment column [4]. The final source came from a government website with accurate statistics of consumer price index. The dataset only has two columns which consist of the date and the CPI corresponding to that date. CPI, or consumer price index, measures the average change in prices paid by consumers for a representative basket of goods and services over time [5].

# Methodology

## Method A

The method/model that was chosen for this project was a linear regression model because as a beginner in data science, it is easier to understand and implement with a lack of experience in the field. The goal of this project was to potentially predict prices of cryptocurrencies based on past prices and other potential factors. The factors integrated in this research were economic factors such as inflation which was measured through looking at the consumer price index to see how prices have changed over time for goods. The assumption of this model is that there is a fairly linear relationship which can also be a drawback for outlying cases of events that may drastically impact the cryptocurrency market, but while maintaining normal economic circumstances, a linear regression representation is justifiable here compared to choosing a classification approach. The modules used in this model are the pandas library, matplotlib, sklearn modules, and the ast library.

Identify applicable funding agency here. If none, delete this text box.

# Results

The results of this project are split into multiple parts. First, there was attention given to the question of what economic aspects can impact cryptocurrency prices and in what way do they affect them? The two major topics identified were market/investor sentiment and consumer price index. Section A will talk about market/investor sentiment followed by Section B which will contain information about consumer price index. The final section C will contain the regression model and equation derived.

## Social Sentiment Correlation

A graph with blue dots

Description automatically generated

Fig. 1. Displays the change in sentiment over time towards cryptocurrency

This graph shows the change in sentiment over time of people and crypto news sources towards the market. According to these results, the polarity and sentiment does not really change over time towards crypto so it probably does not have a significant impact on cryptocurrency prices. I was expecting the sentiment to have highs and lows at different times but this graph says otherwise. However, I was unable to access Statista so maybe that dataset could’ve provided me better results.

## Cryptocurrency Price and Consumer Price Index

A graph showing the price of bitcoin

Description automatically generated

Fig. 2. Displays the price of bitcoin over time while graphing the change in consumer price index against it over the same period of time and an equivalent unit

This graph shows the CPI in relation to BTC price (Bitcoin price). The CPI consistently rises whereas BTC has a general upward trend, however it has a drastic dip around 2021 and 2022. This dip could be due to the Terra Luna crash or FTX situation which were both big situations that I won’t explain here, but it was bad for crypto markets because it hurt many traders’ confidence and lost the market lots of money as well, so it would be reasonable to say many people retracted their investments in crypto after that incident. However, looking at the graphs without the steep dips, it can also be said that Bitcoin or cryptocurrencies are seen as inflation hedges by some. This means that as the CPI increases sharply, more people may put their money in cryptocurrencies to prevent their money from losing value from inflation. This graph displays the volatility of the crypto market while slightly supporting a relation between CPI and cryptocurrency prices.

## Regression Model

#### A graph showing a green line Description automatically generated

Fig. 3. Linear regression model representation of cryptocurrency, specifically Bitcoin, with respect to time.

This graph is a display of the linear regression model created through this project. The green is the predicted price whereas the red is the actual price and the model tends to underestimate the real price and also goes below 0 at times as well. This may be due to overemphasis on small market caps/volumes. The model may tend to assume steep negative relationships upon small drops and once again the nature of this model is that it is a simple linear regression model that doesn’t account for all of the total possible variables impacting it.

Two important statistics from this model are that the mean absolute error was around $1075 which is fairly accurate on the grand scale of Bitcoin value which is very large. The R-squared value, proportion of variance, is 46.5% which means there are still a lot of factors not accounted for by this model.

# Discussion

The shortcomings of this project is the fact that I am new to data science and I simply do not have the knowledge currently to make a fully integrated model that consists of all the possible variables to create a fully accurate model. I also would have liked to use the Statista datasets but unfortunately it was inaccessible when I needed it because I believe those sources would have given me a lot better datasets to visualize with. I think that was the biggest setback for this project because they had some really good datasets for market sentiment. Due to the lack of trustworthy datasets I could use, I wasn’t able to visualize as many correlations and connections between economic factors and cryptocurrencies. In the future, I would definitely use the Statista datasets and find a few more economic aspects to compare cryptocurrency prices to while also looking at a couple more coins because I focused heavily on Bitcoin for this project. With more knowledge, I could also implement a better regression model, too.

# Conclusion

The results of this project showed that there was definitely a slight correlation between CPI and BTC prices because Bitcoin is one of the more well-known cryptocurrencies, and as CPI increases, people may look to cryptocurrencies to preserve the value of their money through investment. Market sentiment didn’t show much impact or change over time, but this may simply be due to poor datasets since I was restricted access to Statista. With the research that I did conduct, I created a model that was fairly close to real life prices of Bitcoin. It would still need a lot of work on it before it could practically be used as a predictor of crypto prices, and even so, there is always going to be some unpredictability because different world events can have a variety of impacts on crypto prices such as the FTX incident or Terra Luna crash. The effect of these results on the real world are simply to show a few of the factors that influence cryptocurrency price and how they impact it.

##### Acknowledgment

I’d like to acknowledge my Data Science professor Weijie Pang at Wentworth Institute of Technology for providing me the teaching and opportunity to create this project.

##### References

1. U. Shah, “Bitcoin USD Historical Data,” Kaggle, [Online]. Available: <https://www.kaggle.com/datasets/utkarshx27/bitcoin-usd-historical-data>. [Accessed: Dec. 2, 2024].
2. The Devastator, “Cryptocurrency Price Market Data,” Kaggle, [Online]. Available: <https://www.kaggle.com/datasets/thedevastator/cryptocurrency-price-market-data>. [Accessed: Dec. 2, 2024].
3. S. Rajkumar, “Cryptocurrency Historical Prices (CoinGecko),” Kaggle, [Online]. Available: <https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrency-historical-prices-coingecko>. [Accessed: Dec. 2, 2024].
4. O. van Hoorn, “Crypto News,” Kaggle, [Online]. Available: <https://www.kaggle.com/datasets/oliviervha/crypto-news>. [Accessed: Dec. 8, 2024].
5. Federal Reserve Bank of St. Louis, “Consumer Price Index for All Urban Consumers: All Items in U.S. City Average (CPIAUCSL),” FRED, Federal Reserve Bank of St. Louis, [Online]. Available: <https://fred.stlouisfed.org/series/CPIAUCSL>. [Accessed: Dec. 8, 2024]..