Digital Investment Analysis and Predictive Modelling Capstone Project Report

Ramón.Torres

Northwest Missouri State University, Maryville MO 64468, USA
S556497@nwmissouri.edu

Abstract. This article presents a preliminary proposition for investment analysis for a digital asset. The analysis is based on historical data of the bitcoin price for the years 2015 to 2023. Correlation analysis is performed, and seasonality decomposition is completed to identify trends, price movements and low and high buying seasons for this digital asset. Based on correlation and residuals analysis an initial predictive model for the Bitcoin price is proposed using ARIMA methodology. In summary, general findings showed seasonality of the buying and selling trends at the end of each year analyzed and correlation was reached, setting up the scenario for the presentation of a bitcoin price predictive model. The tools used for this project included Jupiter Notebook, Latex using Oleaf, MS Word, Acrobat, Kaggle and CoinmarketCap Data Sets and Crypto and modelling was design using Pyhton and its libraries. All details, steps, files and code for the development of this project is available in https://github.com/rtmtorresmorales/44688-80-Capstone-Project

Keywords: natural language processing · investment analysis· digital assets and bitcoin · ARIMA · price predictive modelling· correlation· residual analysis

1 Introduction

The planned domain is Natural language processing and visualization to demonstrate trends on investment options. The goal is to analyze sentiment analysis of experts, investors and institutional investment firms. Based on current world events and the generalization of the cryptocurrencies, particularly the bitcoin, I would like to analyze investment trends to preserve capital. I will use Kaggle data sources to obtained data sets to analyze financial trends and compare to financial reports and determine the sentiment or opinions of experts on this field . The process will include the following steps:

1. Explore for possible investment analysis scenarios 2. Search for data sources and save it to computer in CSV and MS Excel formats 3. Use python to organize and clean 4. Select appropriate data sources and columns 5. Design and run analysis model 6. Use Tableau to present visualization of relevant results 7. Present project report

Subsequent paragraphs, however, are indented.

2   Goals of this Research

Based on current world events and the proliferation of the digital currencies, it is beneficial to analyze investment trends to preserve capital and enhance return on investments for such instruments. The project is intended to:

1. review forecasting and predictive modelling for the and investment instruments

2. review of historical trends of the bitcoin for the period from 2015 to 2023.

3. generate a predictive tool for future applications to other types of investments and asset classes

This tool may identified parameters that influence the market price of the digital currency. This exercise may be useful for the individual, institutional, or highly speculative investor.

3   Literature Review and methodology approach

Based on the historic data, an estimate and trade model will be developed to identify market value trends. In addition, a time series analysis, the forecasting model may provide sequencing analysis and determine any comprised intervals. In general, an ARIMA model or a method for forecasting and prediction of future values based on historical time series. The predictive model follows the approach of the WalletInvestor. This a prediction site tool that uses machine learning algorithms to analyze and forecast cryptocurrency prices. The site provides daily price predictions for over 800 digital instruments and offers a long-term forecast for up to five years. Studies shows ARIMA model may result in predictions with the smallest for a day to seven days. Therefore, is feasible to use as a predictive method of Bitcoin for that period. On the contrary, a long, short term memory deep learning algorithm with good predictive performance for time series data. This method refers to the data of multiple periods before predicting the price of Bitcoin. This approach is that the prediction accuracy of the model that uses previous period is the best.

4   Data Collection and processing

Data source: Cryptory library for Python, Coinmarketcap, and https://www.kaggle.com/datasets/adilbhatti/bitcoin-and-fear-and-greed. Data sources in CSV and MS Excel formats. Data scraping techniques: Use python to organize and clean. Select appropriate data sources and columns. Data attributes: Cryptograms bitcoin: seven rows, transaction date, open market value, low value, high value, total volume, closing values, market capitalization . Other specific data extraction details related to your project: Based on current world events, clean up items for dates before year 2019 and analyze effects after the COVID 19 Pandemic.

5   Data Cleaning

Using Python to upload the appropriate data sets. Review and check for null values, values before 2019 and possible duplicates. Made backups of original file and working file. Tools and techniques are used in the process: Pandas, Numpy Checking for missing values or deleting roes with missing results to avoid false or misrepresentation of results by improving the quality of the data. Seven columns and 2205 records. The definitions of important data attributes. date = date of transaction

Open = value of bitcoin when market open

Close = value of bitcoin when market close

High = highest value of bitcoin on a trading day

Low = lowest value of bitcoin on a trading day

Volume = total shares traded during a trading day

MarketCap = total value in dollars per trading day.

For analysis and forecasting: the independent variables volume and markeCAp by month. The dependent variable will be next day open.

6   Exploratory Data Analysis

Exploratory data analysis is a stage where data is organized and process to determine the main attributes and characteristics of the data set. Generally, data is organized using a visualization methods, tools, or techniques, like graphs, charts, and tables. It is important in any given data analytics projects to identify trends, outliers, tendencies and prepare the scenario for in depth analysis and discussion of findings. It helps us avoid making any assumptions that might overstated, understated or plain wrong. This will avoid adding additional errors to the forecast or predictive models, using Python or R for example. In research methodology it set the stage for an appropriate statistical summary to perform hypothesis testing. Some of the exploratory data analysis techniques are uni-variable non graphical and graphical analysis, multi-variable non-graphical analysis and graphical analysis. My project will include historical analysis of market values of the cryptocurrencies to consider the trends, frequencies, moving average and seasonality analysis. It may include time series analysis to determine variability rates and uni-variable correlation analysis among market values and prices. These tools are still under consideration. The techniques included analysis of market values of the cryptocurrencies to consider the trends, frequencies, moving average and seasonality analysis. The process is as follows:

Phase 1: Historical cryptocurrency market value analysis for the period of 2015 to 2023. This technique will provide market value trends.

Phase 2: Historical cryptocurrency trading volume analysis for the period of 2015 to 2023. This technique will provide volume trends.

Phase 3: Moving averages or mean values of cryptocurrencies main attributes, price and volume for the period of 2015 to 2023. This technique will facilitate the visualization of upward and downward trends.

Phase 4: Seasonality analysis on cryptocurrency market value for the period of 2015 to 2023. This technique will decompose the data by trends, seasonality, and noise. Therefore, providing the levels of variability on market value and volume.

Initial findings includes that the prices of cryptocurrencies showed a tendency of having market value under pressure for months of September and October of every year. This showed a downward trend on the values of the digital asset. Then after these periods of overselling, upward trends appear to present until the end of each year. By December of every year, a buying rally reached a peak, as well as the value of the asset.

7   Predictive Modelling and Analysis Initialization and Setup

For capstone project research the Crypto library for Python was used. It provides historical data from coinmarketcap.com and bitinfocharts.com. These will be the basis for working with making forecasting of the crypto currency over time. Based on the historic data, an estimate and trade model will be developed to identify market value trends. In addition, a time series analysis, the forecasting model may provide sequencing analysis and determine any comprised intervals. In general, an ARIMA model or a method for forecasting and prediction of future values based on historical time series. The predictive model follows the approach of the WalletInvestor. This a prediction site tool that uses machine learning algorithms to analyze and forecast cryptocurrency prices. The site provides daily price predictions for over 800 digital instruments and offers a long-term forecast for up to five years. Studies shows ARIMA model may result in predictions with the smallest for a day to seven days. Therefore, is feasible to use as a predictive method of Bitcoin for that period. On the contrary, a long, short term memory deep learning algorithm with good predictive performance for time series data. This method refers to the data of multiple periods before predicting the price of Bitcoin. This approach is that the prediction accuracy of the model that uses previous period is the best. To determine which previous data points, have more influence on the serial correlation of the crypto currency. Statistical tools will include the application of regression analysis, moving averages, time series and forecasting. To determine which previous data points, have more influence on the serial correlation of the crypto currency. The training process will be fed form the re-sampling of the data will be completed to consider different frequencies and visualization techniques. Using data method for time-series decomposition pf trends, seasonality, and noise. Analysis of movement of the toe series will be performed. The testing process will be looking for to alternatives stationary series or no stationary series based on a statistical p-value. For the predictive model, the series parameters will be plotted to visualize correlation or partial correlation. For this capstone project will be developed using Python programming language, statistical tools, packages, and libraries. Steps will include:

Step 0: Design, organization, and initial setup

Step 1: Installing and Setting up required analysis tools

Step 2 Accessing libraries, analytical, statistical tools

Step 3: Access Crypto Library

Step 4: Accessing Data sets - pull data from start of 2015 to present day

Step 5: Main object definition and accessing historical bitcoin prices from bitinfocharts

Step 6: Additional historical bitcoin data from coinmarketcap

Step 7: Graphs configurations for Cryto Historical trends 2015 to 2023

Step 8: Resampling frequencies on monthly price, annual and quarterly

Step 9: Graphs configurations for Crypto frequencies

Step 10: Graphs configurations for seasonal data decomposition

Step 11 Seasonality analysis parameters definition

Step 12 Seasonality analysis for bitcoin market price per month, decomposition and visualization of results

Step 13 Approximation of parameters using correlation plots.

Step 14 Parameters for approximation definition using correlation.

Step 15 Model Selection and fitting

Step 16 STL-correlational decomposition and visualization of correlational results and residuals

Step 17 Model prediction

The initial analysis included market price trend analysis from 2015 to 2023, moving averages of market prices by days, months, quarters, and years. Also, seasonality analysis decomposition was illustrated for the market price. Price decomposition and statistical analysis shows seasonality trends and residuals. Initial findings includes that the prices of cryptocurrencies showed a tendency of having market value under pressure for months of September and October of every year. This showed a downward trend on the values of the digital asset. Then after these periods of overselling, upward trends appear to present until the end of each year. By December of every year, a buying rally reached a peak, as well as the value of the asset. The following graphs depict the previous initial results analysis. Additional graph, illustrations, findings and conclusions will be presented in the following section of Oleaf submittal.

8 Discussion and findings

The project results and findings will be presented in this section. Various graph and chart were incorporated to aid visualisation of the results and facilitate the discussion. The charts used for this project included line charts, multi-line charts for historical price trends, pattern seasonality decomposition and visualizations, bar graph for correlation analysis and multi-line chart to represent the forecasting and predictive model during the period of 2015 to 2023.

After generating visualization charts for the obtained data sets, results included that from the historical chart, a high volatility was depicted for the years 2017 to 2023. As depicted on Figure 1.
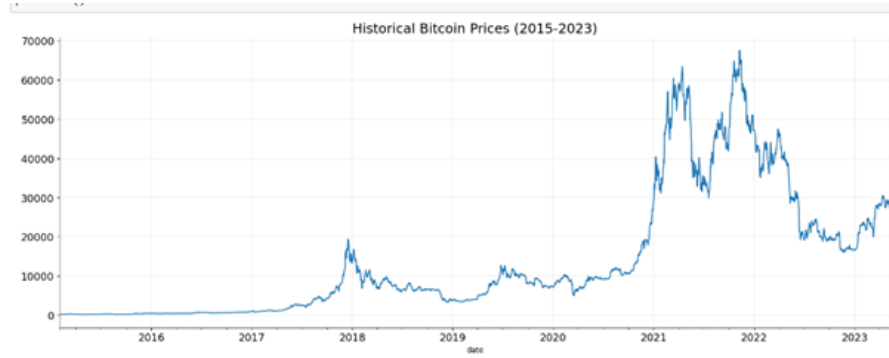
Fig. 1. Historical bitcoin market prices mean from 2015 to 2023.

The graph shows four high points in 2018, 2020, 2021 and 2022, showing values of 20,000, 10,000, 62,000 and 68,000, respectively. This is confirmed with moving average chart, showing higher price moving averages by years 2021 and 2022. AS shown on Figure 2.
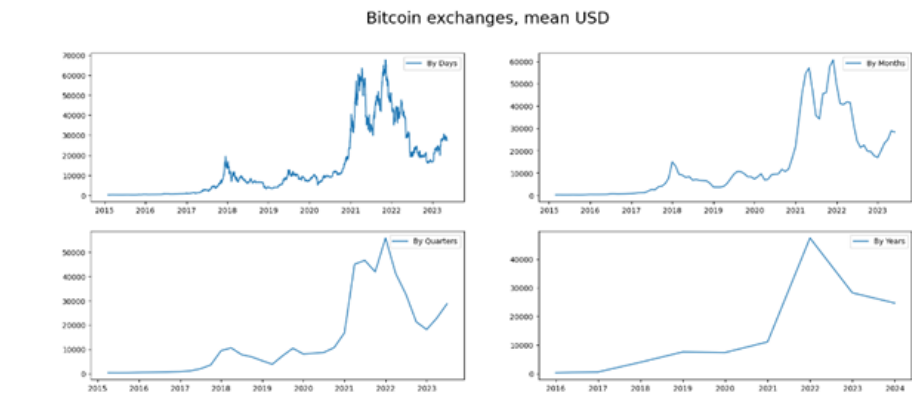

Fig. 2. Historical bitcoin market prices mean from 2015 to 2023 presented in days, months, quarters and years.

Prices trends move higher by 2022 and seasonality decomposition analysis resulted in a clear price seasonality by the end of each year. This is clearly shown on Figure 3.

The correlation analysis shows dependency and correlation on the results of the proposed model and the lagging values. Figures 4 and 5 illustrate this finding.
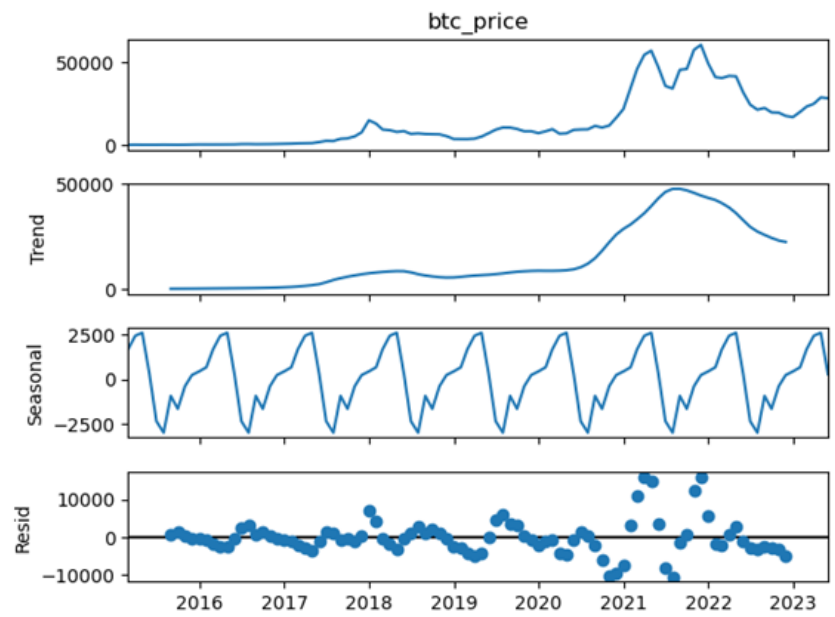
Fig. 3. Historical bitcoin market prices statistical seasonality analysis from 2015 to 2023 presented main parameters are trends, seasonality and residuals.
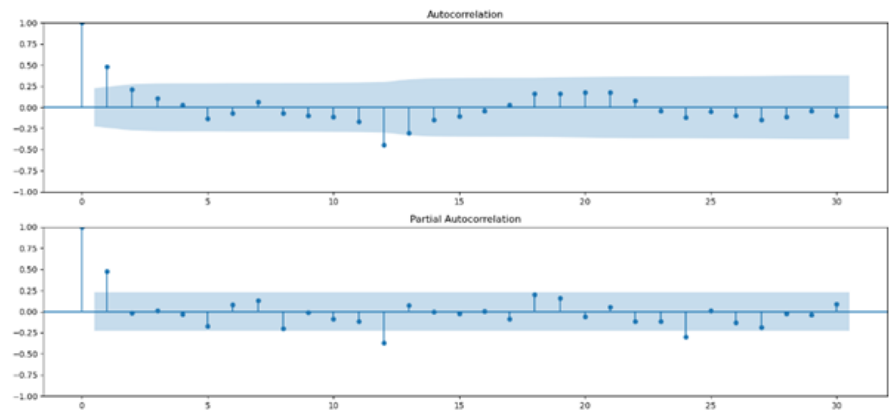

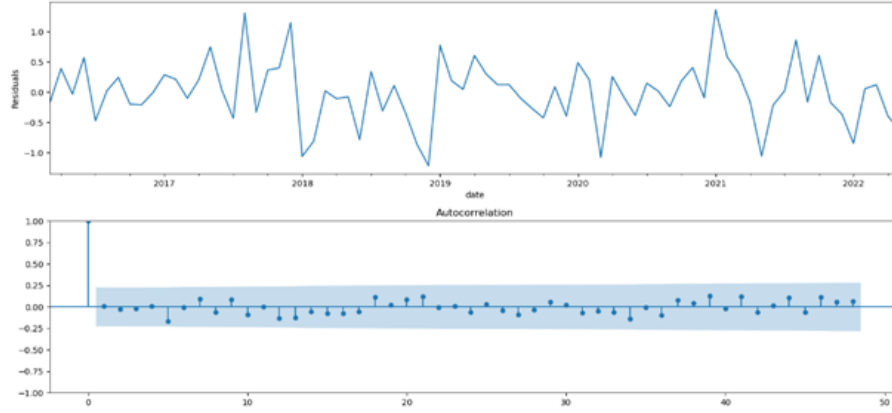**Fig. 4.** Bitcoin correlation and lagging values analysis based on historical data.

Fig. 5. Bitcoin correlation and residual analysis based on historical data.

As per the predictive model, the multi-line chart shows demonstrate the correlation among the actual bitcoin prices and the predicted values. The main result and comparison of the actual price and the predictive model is shown in Figure 6.
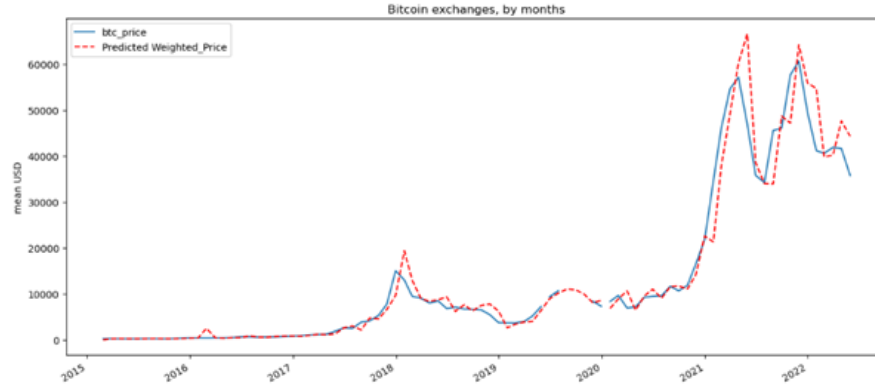

Fig. 6. Proposed Bitcoin market price predictive model.

9   Conclusions

From the historical data and inference analysis, four peak price points occurred in 2018, 2020, 2021 and 2022. The moving average showed the higher prices by years 2021 and 2022. The prices of cryptocurrencies showed a tendency of

having market value under pressure for months of September and October of every year. This showed a downward trend on the values of the digital asset. Then after these periods of overselling, upward trends appear to present until the end of each year. By December of every year, a buying rally reached a peak, as well as the value of the asset.

Based on seasonality analysis and statistical analysis, seasonality decomposition resulted in a clear bitcoin price seasonality by the end of each year. The correlation analysis shows dependency and correlation on the results of the proposed model and the lagging values, therefore promoting the proposition of a predictive or forecasting model for the bitcoin prices. As per the predictive model, the multi-line chart shows demonstrate the correlation among the actual bitcoin prices and the predicted values.

10   Ethical considerations and limitations

In general, financial investing may be a lucrative way of generating wealth with appropriate ethical compass. This tends to diminish when greed comes into play. Strategic investment planning is put aside for the quick and relentless return. Investors tendencies move toward no traditional assets, such as cryptocurrencies that has no clear economic underlying in an unregulated investment environment. This open up various ethical issues over this asset class, such as manipulation, unqualified investors, illegal uses of the asset and erratic investment behavior. Therefore correlation analysis needs a deeper review to determine an extract some of the previously mentioned ethical considerations, that may impose limitations on the predictive model, like some outliers and incongruities.

11   Recommendations

The investment asset shows a high level of volatility, a seasonal trading patterns that alter the price values, typically by the end of each year and perhaps showing some investor behavioral impulse to increase selling and buying the assets before and after the buying season hits. Enhancing the correlation analysis process to filter out outliers and erratic trends is highly recommended. In addition, further analysis and predictive modelling may support cryptocurrencies portfolio optimization techniques, combing various digital assets and their parameters, volatility and investment risk management based on seasonality, enhance performance and financial returns.

[]

References

1. https://notebook.community/dashee87/blogScripts/Jupyter/2018-02-11-analysing-the-factors-that-influence-cryptocurrency-prices-with-cryptory
2. Cryptocurrencies analysis and forecasting, https://www.kaggle.com/code/vovchuk/cryptocurrencies-analysis-and-forecasting

3. Chein, J.: Analysis of bitcoin price prediction using machine learning. AlphaSene (2023)
4. Kashettar,   S.:   5   best   crypto   prediction   site   tools, https://www.analyticsinsight.net/5-best-crypto-prediction-site-tools/
5. Sheynin, N.: 3 ways to apply natural language processing (nlp) in financial research. AlphaSene (2022)
6. Torres, R.: 44688-80-capstone-project, https://github.com/rtmtorresmorales/44688-80-Capstone-Project
7. Vimalachandran, A., Jayachandran, T.: The financial crash of 2020 and the retail trader's boon: a correlation between sentiment and technical analysis. SN Business Economics