DSC 550 Term Project

Jahedur Rahman

Bellevue University

DSC 550

Dr. Brett Werner

June 4, 2022

For my term project, I observed the relationships between mortgage interest rates, cryptocurrency prices, and Covid-19 rates to try and predict if the price of Bitcoin would increase or decrease on a certain day. From 2020 to 2021, we noticed some big differences in the mortgage interest rates and cryptocurrency prices. It was observed that mortgage interest rates were lower. In addition, cryptocurrency prices were at an all-time high. It seems like the pandemic had some sort of effect on the mortgage interest rates and cryptocurrency prices. The reason for finding the relationship between the three to try and predict if the price of Bitcoin would increase or decrease on a certain day is important is because it will help many people and organizations to invest in cryptocurrencies more confidently. If a pandemic similar to Covid-19 happens again, these investors will be more prepared and confident when investing in cryptocurrencies.

Preparing The Datasets

I used a Covid-19 dataset, a Bitcoin dataset, and a Mortgage Interest Rates dataset. These datasets had to be imported and cleaned in various ways.

Importing the Datasets

The Covid-19 dataset was in the form of a CSV file. I imported this CSV file using a pandas function, pd.read_csv(). Some of the data that were included in it were the number of daily covid cases, deaths, and other things. These data records also included the location and date they were recorded. The Bitcoin dataset was also in the form of a CSV file. I also imported this CSV file using a pandas function, pd.read_csv(). Some of the data that was included in it were current and past daily Bitcoin prices. This dataset was a very simple dataset to read and understand because it includes the opening price, the low price, the high price, and the volume of

the day it was recorded. The Current Mortgage Interest Rates were in the form of a table on a website. I used the BeautifulSoup and requests libraries to import this table as a data frame. Some of the data that was included in it were current and past daily records of mortgage interest rates.

Cleaning the Datasets

After importing the datasets I had to clean each one. The Covid-19 dataset had a lot of data that was not needed. I removed any record that did not have the location as the United States. This is because the other two datasets are based in the United States. Also, I removed columns that I wouldn't be using. In addition, I renamed the columns to simpler column names and added 'covid' in front of all of them so that when I merged the datasets I can easily see which columns belonged to which dataset. Additionally, I had to change the type of the Date column so it can be the same type as the Date columns in the other datasets. For the mortgage interest rates dataset, I first changed the type of the Date column so it can be the same type as the Date columns in the other datasets. Also, I renamed the columns to simpler column names. The initial column names were too long. Also added "rate" in front of each column name to distinguish these columns from other columns in the other datasets. In addition, I changed the types of the data in the columns from object to float. However, to do that I had to remove the % sign from all the columns that included it. Some rows had %. instead, so had to remove that as well. For the bitcoin dataset, I first changed the Date column type to datetime64. Then I added a new column where it would give information on whether there was an increase or decrease in Bitcoin's price on that day. If the closing price is more than the opening price then there is an increase and there is a 1 for that row otherwise there is a 0.

Merging the Datasets

After importing and cleaning the datasets, I had to merge all three of them into one data frame. I used the pd.merge() function from the pandas library. First I merged the mortgage interest rates data frame and the bitcoin data frame using an inner join on the Date column. I called this data frame semi_merged_data. Then I merged the semi_merged_data data frame and the Covid-19 data frame using an inner join on the Date column. I called this data frame full_merged_data. The full_merged_data data frame consisted of 548 rows and 22 columns.

Visualizations

After preparing and merging the data, I did various graphical analyses on it. I basically compared different columns and tried to find some sort of pattern between the two columns I was comparing. First, I compared the columns, "covid_total_cases", and, "rate_30yr(%)".
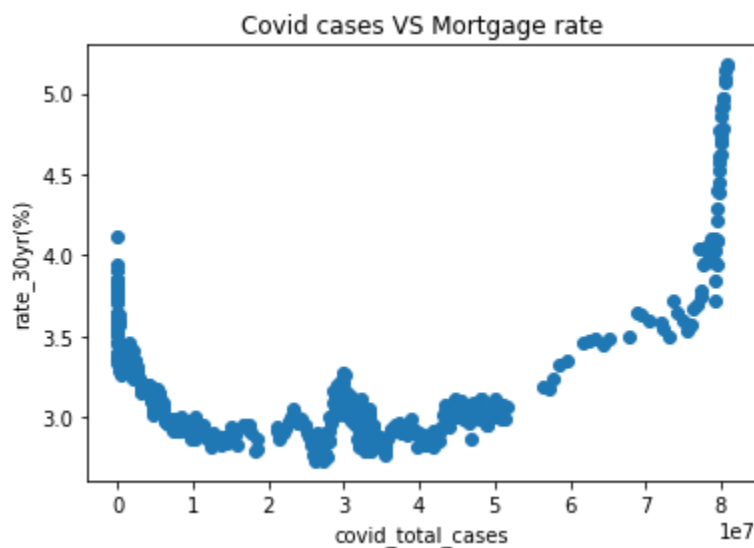
By looking at this graph we can observe that when covid-cases were around 10 million to 50 million, mortgage rates were very low. However, it started rising after 60 million cases. Also, most of the points are plotted between these values, so it was probably the time when the virus was spreading the most. Next, I compared the columns, "covid_total_cases", and, "Open". The "Open" column is the opening price of Bitcoin on a certain day.

Covid cases VS BTC Price
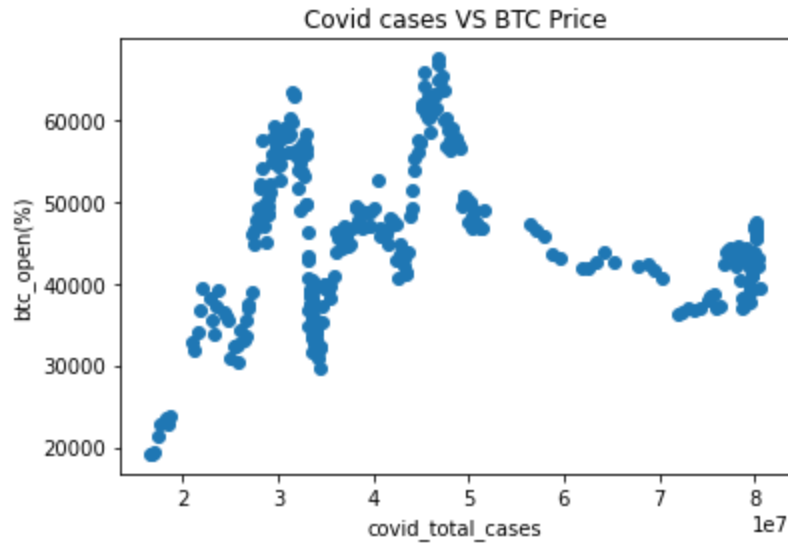
By looking at this graph we can observe that when covid cases were around 30 million to 50 million, Bitcoin reached high prices. However, the price fell after 50 million cases. Similar to the last graph, most of the points are plotted between these values, so it was probably the time when the virus was spreading the most. Then, I compared the columns, "Open", and, "rate_30yr(%)".

BTC Price VS Mortgage rate
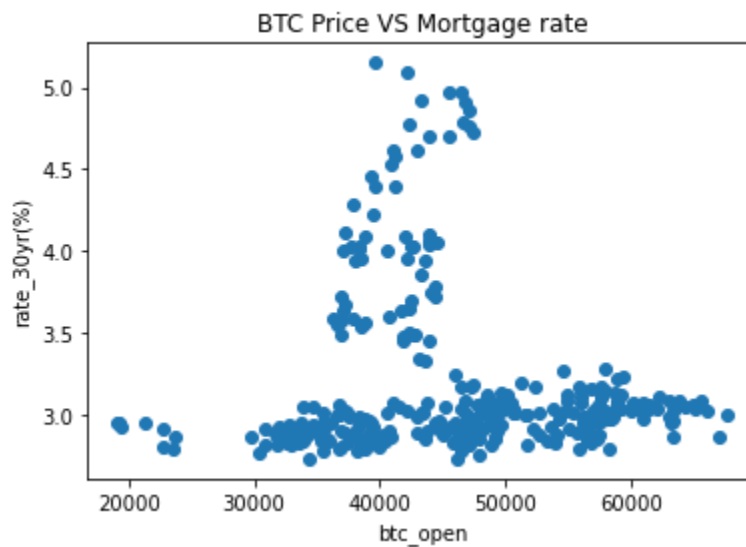
In this graph, most of the points are between the prices of $30,000 and $70,000. Also, the mortgage rate was around 3%. This could mean that at the time when many people were buying

Bitcoin, mortgage rates were lower. Finally, I compared the columns, "rate_30yr(%)", and, "Volume".
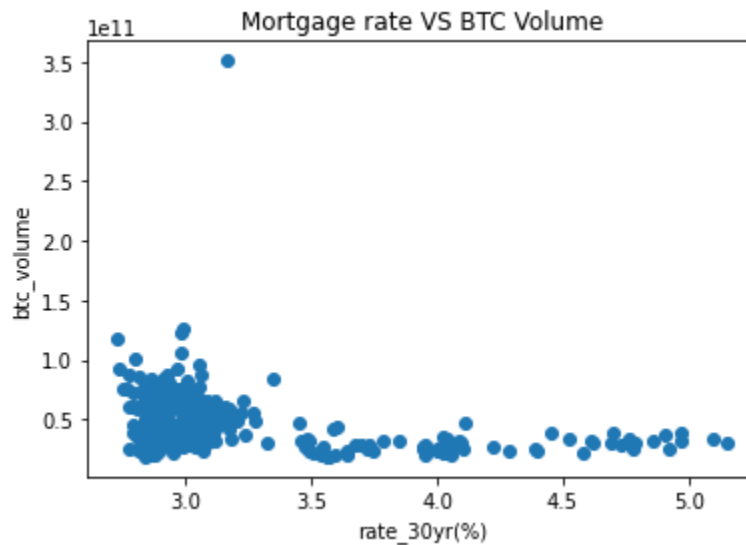
This graph can help explain the previous graph better. It looks like when there was a higher volume of Bitcoin, mortgage rates were low. As the volume decreased mortgage rates increased. By looking at the graphs I can see where most of the points in each graph are more closely packed together. This helps explain the graphs and variables more. When covid was spreading the most, mortgage rates went down, and the price of Bitcoin went up. When the spread started slowing down mortgage rates were going back up, and the price of Bitcoin started going down.

Model Building and Evaluation

After creating the visualizations I had to fit a model to my data. I fit and compared multiple different models. I tried fitting multiple different models to compare the accuracies and see which model gave the best fit. The models I tried are linear regression, decision tree classifier, and logistic regression. Linear regression gave very low accuracy, about 30%. The decision tree classifier and logistic regression gave similar accuracies. However, logistic regression was a little better with an accuracy of 68.75%.

This analysis/model building tells me that it is possible to predict if Bitcoin's price will increase or decrease on a certain day. However, I don't think the model is ready to be deployed. This is because there are other factors that can affect the price of Bitcoin. In addition, there are factors that can affect mortgage interest rates. My recommendation would be to explore all these other factors and make a better model using\the data collected from those other factors. An additional opportunity that can be explored is researching companies that are using Bitcoin as payment. Looking at this data can help decide how often Bitcoin is used when purchasing products. This in turn can affect the price of Bitcoin since more people will be investing in the cryptocurrency. In conclusion, there are many factors that can affect the Bitcoin price that should be researched to create a better and more accurate model.