BTC_closing_prediction

Cesar Macedo Perez

21/4/2020

Introduction

On January 3, 2008, born a completely new system of *Electronic Cash* based in peer-to-peer exchange and, without any third parties thaa controlled it. This new *e-cash* was named **Bitcoin** and was created by a person, or group of people, under the pseudonym of Satoshi Nakamoto. You can read more about its origin in the Bitcoin Foundation page at http://www.bitcoin.org/bitcoin.pdf.

Since then, Bitcoin has been the target of many studies and controversies, mostly because the base Technology, called Blockchain, allows the users to remain completely anonymous and, unfortunately, this has been used for illegal activities.

Regardless of this, Bitcoin has become one of the most important speculative assets all over the world, mostly because of its huge **volatility**. Just to have an idea, the Bitcoin price in April 2010 was about USD 0.003 and, since then, its price started to grow, by January 2017, its price was USD 19,798, yes, an amazing grow of almost 660 million percent. While I'm writing this, you can buy a Bitcoin for USD 6,901, again, yes, a fall of 65% percent.

And that's why I decided to make this brief analysis, due to this extreme volatility many people have become rich and many more became poor, losing all their life savings. Just a few people understand how this Cryptocurrency works and what variables drive its price behavior.

The motivation of this study is to give a better understanding of how Bitcoin works and what's behind Investor psychology when deciding to move to Bitcoin depending on the behavior of other variables. I'm not trying to give any kind of Financial Advice and much less to offer an "easy money" approach to Cryptocurrencies Market.

Objective

Create an algorithm that predicts the Closing Price of Bitcoin using the opening Prices/Indexes of selected Exchange Rates, Cryptocurrencies and Stock Prices, and Indexes in the Mexican Financial Market.

The ***exchange_mxn*** database

For this project I created a database with the Bitcoin Opening and Closing Prices and compared them with another 17 variables (Cryptocurrencies and Stocks Prices, Commodities and Financial Indexes), the date range goes from 2018-02-15 to 2020-04-15. All prices are at the beginning of each session, from Monday to Friday and, in MXN(Mexican Pesos).

Source: https//mx.investing.com

DAY - Day of the week ranked from 1 to 7, starting Monday = 1.

QUARTER - Quarter of the year, Example: 1st quarter covers from January to March of the same year.

DATE - the date in format yyyy/mm/dd showed as number.

BTC_open - Bitcoin price at the beginning of the session.

BTC_close - Bitcoin price at the end of the session.

XRP - Cryptocurrency Ripple price.

ETH - Cryptocurrency Ethereum price.

USD - US Dollar/MXN exchange rate.

EUR - Euro/MXN exchange rate.

JPY - Japanese Yen/MXN exchange rate.

KRW - South Korean Won/MXN exchange rate.

SAR - Saudi Arabia Riyal/MXN exchange rate.

AED - Emirati Dirham/MXN exchange rate.

GBP - Sterling Pound/MXN exchange rate.

INR - India Rupee/MXN exchange rate.

RUB - Russian Ruble/MXN exchange rate.

OIL_WTI - West Texas Oil price per Barrel.

GOLD - Gold price per Ounce.

NASDAQ - Nasdaq Composite Index in Points.

DOW - Dow Jones Index in Points.

IPC - Indice de Precios y Cotizaciones (Mexican Prices and Quotations Index) in Points.

AMZN - Amazon Stock Price.

BABAN - Ali Baba Group Stock Price.

Exploring the data

Let's explore our data summary:

```
##       Day       Quarter          Date          BTC_open        BTC_close
##  Min.   :1   Min.   :1.000   Min.   :17577   Min.   : 65553   Min.   : 65553
##  1st Qu.:2   1st Qu.:1.000   1st Qu.:17774   1st Qu.:119985   1st Qu.:119985
##  Median :3   Median :2.000   Median :17973   Median :143382   Median :143889
##  Mean   :3   Mean   :2.414   Mean   :17972   Mean   :142177   Mean   :142142
##  3rd Qu.:4   3rd Qu.:3.000   3rd Qu.:18170   3rd Qu.:171554   3rd Qu.:170044
##  Max.   :5   Max.   :4.000   Max.   :18367   Max.   :251276   Max.   :251276
##       XRP              ETH              USD             EUR
##  Min.   : 2.971   Min.   : 1745   Min.   :18.01   Min.   :20.06
##  1st Qu.: 5.288   1st Qu.: 3083   1st Qu.:18.91   1st Qu.:21.40
##  Median : 6.121   Median : 3966   Median :19.14   Median :21.82
##  Mean   : 7.508   Mean   : 5270   Mean   :19.44   Mean   :22.13
##  3rd Qu.: 8.972   3rd Qu.: 5572   3rd Qu.:19.59   3rd Qu.:22.75
##  Max.   :20.938   Max.   :17406   Max.   :25.36   Max.   :27.19
##       JPY              KRW                SAR              AED
```

```
##  Min.   :0.1643   Min.   :0.01556   Min.   :4.803   Min.   :4.904  
##  1st Qu.:0.1721   1st Qu.:0.01634   1st Qu.:5.040   1st Qu.:5.148  
##  Median :0.1754   Median :0.01681   Median :5.105   Median :5.212  
##  Mean   :0.1774   Mean   :0.01699   Mean   :5.182   Mean   :5.292  
##  3rd Qu.:0.1791   3rd Qu.:0.01742   3rd Qu.:5.224   3rd Qu.:5.334  
##  Max.   :0.2303   Max.   :0.02030   Max.   :6.753   Max.   :6.905  
##       GBP             INR              RUB             OIL_WTI      
##  Min.   :23.18   Min.   :0.2551   Min.   :0.2731   Min.   : 476.8  
##  1st Qu.:24.42   1st Qu.:0.2696   1st Qu.:0.2934   1st Qu.:1049.3  
##  Median :24.87   Median :0.2746   Median :0.2996   Median :1131.0  
##  Mean   :25.19   Mean   :0.2770   Mean   :0.3007   Mean   :1128.9  
##  3rd Qu.:25.78   3rd Qu.:0.2813   3rd Qu.:0.3049   3rd Qu.:1244.2  
##  Max.   :30.62   Max.   :0.3322   Max.   :0.3355   Max.   :1478.3  
##       GOLD           NASDAQ          DOW             IPC             AMZN      
##  Min.   :22203   Min.   :5969   Min.   :19028   Min.   :33435   Min.   :25020  
##  1st Qu.:24648   1st Qu.:6977   1st Qu.:24781   1st Qu.:42698   1st Qu.:32345  
##  Median :25739   Median :7442   Median :25728   Median :43787   Median :34478  
##  Mean   :26840   Mean   :7512   Mean   :25772   Mean   :44285   Mean   :34397  
##  3rd Qu.:28823   3rd Qu.:7850   3rd Qu.:26749   3rd Qu.:46771   3rd Qu.:35998  
##  Max.   :42049   Max.   :9697   Max.   :29440   Max.   :50518   Max.   :54448  
##     BABAN     
##  Min.   :2591  
##  1st Qu.:3189  
##  Median :3426  
##  Mean   :3483  
##  3rd Qu.:3731  
##  Max.   :4905
```

Bitcoin variability

A new column was added to the dataset, the column "var_rate" that is the daily variability in the Bitcoin Price, it was calculated with the next formula:

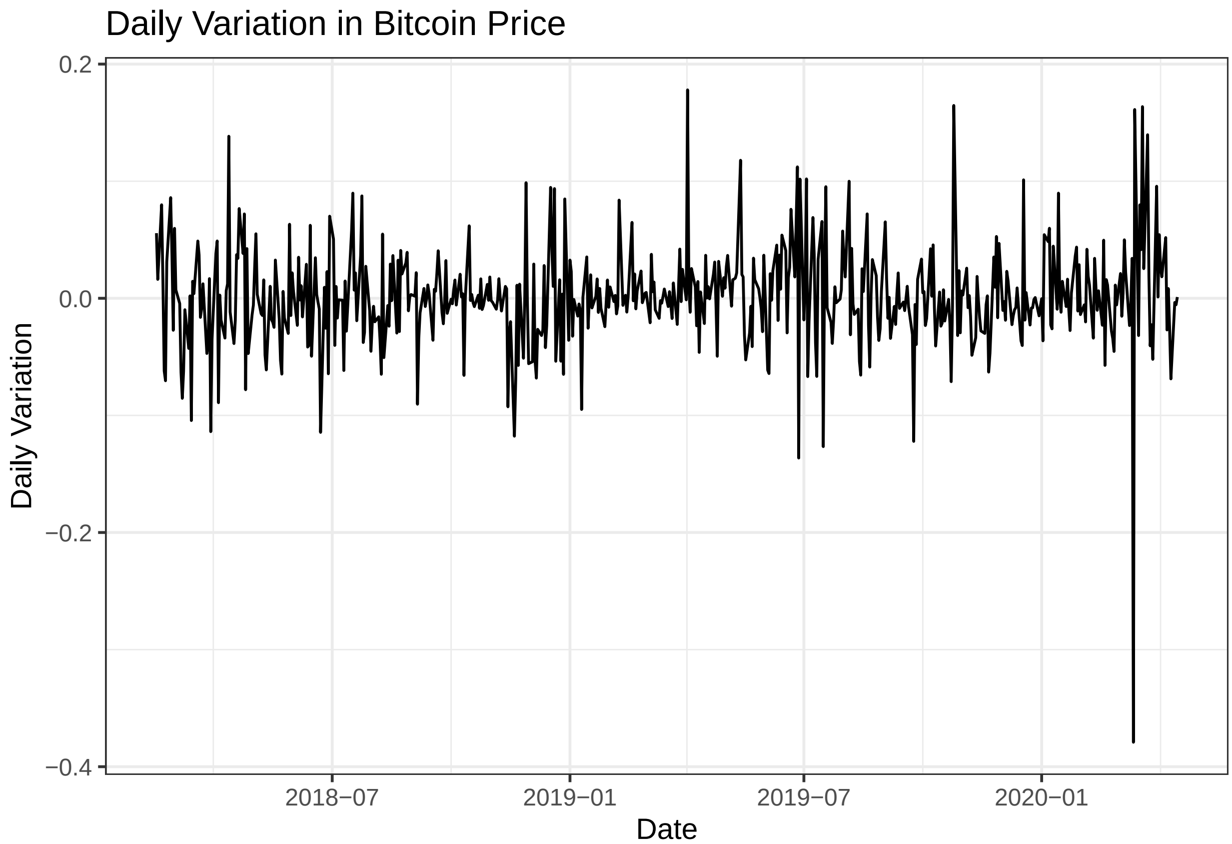
var_rate = (Bitcoin close / Bitcoin open) - 1

Let's look a the result:

```
summary(exchange_mxn$var_rate)
```

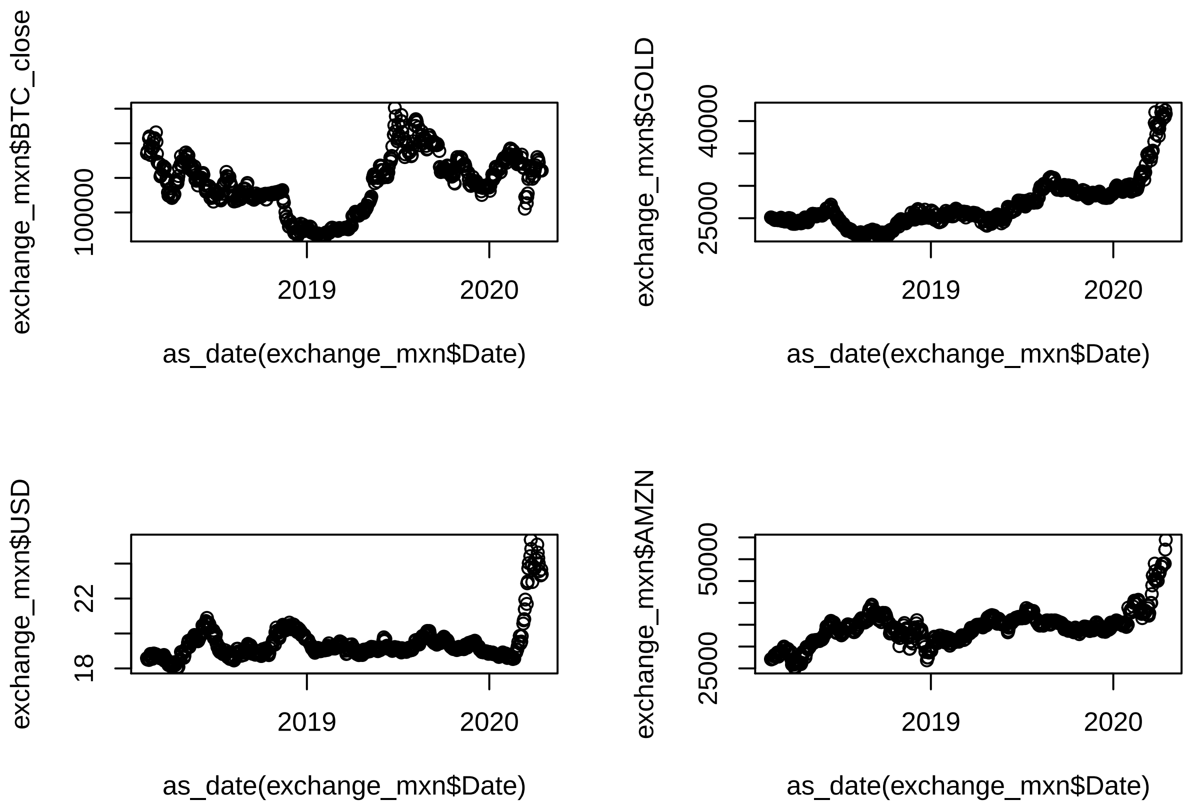
```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max. 
## -0.3791104 -0.0204171 -0.0006614  0.0006509  0.0202647  0.1779735
```

Apparently, the Bitcoin it's a very stable asset, with a Median and a Mean really close to zero, but the difference between the Min and the Max variation is 55.7%.



Now you can see the high variability of this asset. That's why, mean and median look so small, it's because of the extreme values that the mean is pushed towards zero.

Let's compare Bitcoin with the behavior of three more "Stable" Features: USD/MXN, GOLD and AMAZON Stock price.



There is a clear peak as of March 2020 and that is due to the COVID-19 effect, this effect is visible in all three Features compared but, again, Bitcoin contradicts this behavior and remains stable, affirming its reputation for unpredictability.

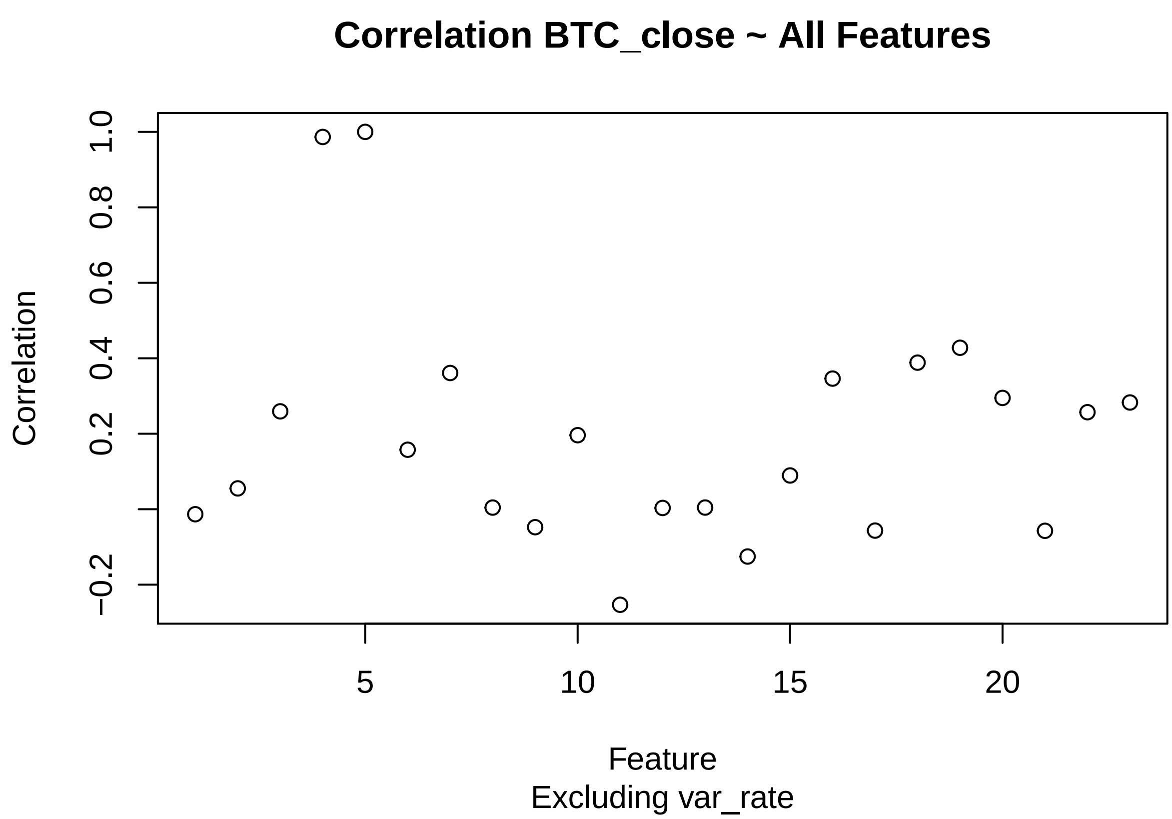
Correlations

So, with such an unpredictable asset, we need to determine which factors drive its value so, it's time to search for correlations.

First, let's look at the correlation of each feature with the Bitcoin closing price. The Feature "var_rate" is omitted since its correlation is useful for analysis.

```
##           Correlations
## Day       -0.013268702
## Quarter    0.055157966
## Date       0.259359518
## BTC_open   0.986534622
## BTC_close  1.000000000
## XRP        0.157614441
## ETH        0.361019709
## USD        0.004277706
## EUR       -0.047649730
## JPY        0.196149770
## KRW       -0.253283052
## SAR        0.003298389
```

```
## AED        0.004389287
## GBP       -0.125406325
## INR        0.089363537
## RUB        0.346152439
## OIL_WTI   -0.056706969
## GOLD       0.388378587
## NASDAQ     0.427865769
## DOW        0.294932066
## IPC       -0.057269609
## AMZN       0.257084111
## BABAN      0.282857548
```



The observed correlations are relatively low, go from -0.253 (KRW) to 0.4279(NASDAQ). We will have to prove if lower correlations are significant, to omit those features that won't help us to predict Y.

Calculating p-value.

We are going to calculate the p-value for every Correlation between -0.1 and 0.1 and, if this value is bigger than our significance level of 0.05, then we can conclude than there is no correlation between Y(BTC_close) and our Feature X.

Correlations between +/- 0.10:

```
##   Feature      corBTC
## 1     Day -0.013268702
## 2 Quarter  0.055157966
```

```
## 3     USD  0.004277706
## 4     EUR -0.047649730
## 5     SAR  0.003298389
## 6     AED  0.004389287
## 7     INR  0.089363537
## 8 OIL_WTI -0.056706969
## 9     IPC -0.057269609
```

Computing the p-value:

```r
cor.test(exchange_mxn$BTC_close, exchange_mxn$Day)[3]
```

```
## $p.value
## [1] 0.7529827
```

```r
cor.test(exchange_mxn$BTC_close, exchange_mxn$Quarter)[3]
```

```
## $p.value
## [1] 0.1904723
```

```r
cor.test(exchange_mxn$BTC_close, exchange_mxn$USD)[3]
```

```
## $p.value
## [1] 0.9191891
```

```r
cor.test(exchange_mxn$BTC_close, exchange_mxn$EUR)[3]
```

```
## $p.value
## [1] 0.2581582
```

```r
cor.test(exchange_mxn$BTC_close, exchange_mxn$SAR)[3]
```

```
## $p.value
## [1] 0.9376464
```

```r
cor.test(exchange_mxn$BTC_close, exchange_mxn$AED)[3]
```

```
## $p.value
## [1] 0.9170887
```

```r
cor.test(exchange_mxn$BTC_close, exchange_mxn$INR)[3]
```

```
## $p.value
## [1] 0.03369581
```

```r
cor.test(exchange_mxn$BTC_close, exchange_mxn$OIL_WTI)[3]
```

```
## $p.value
## [1] 0.1783001
```

```r
cor.test(exchange_mxn$BTC_close, exchange_mxn$IPC)[3]
```

```
## $p.value
## [1] 0.1740249
```

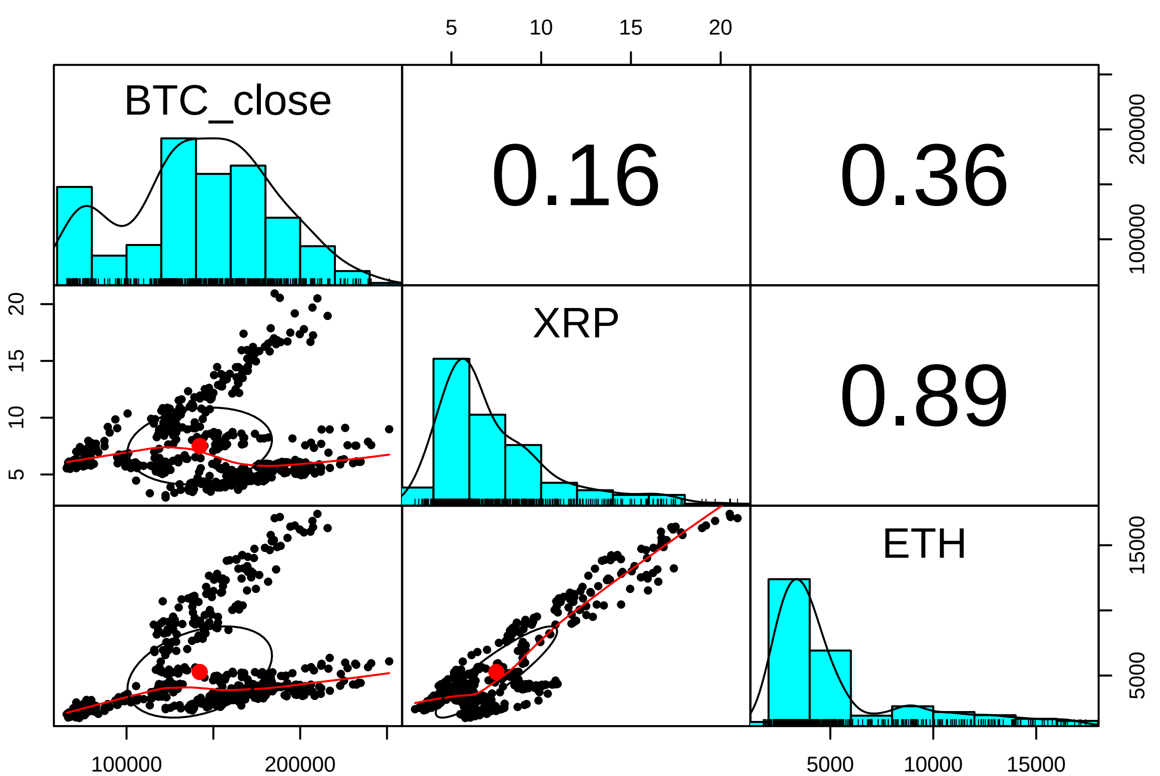
Based in the past results, we can conclude that BTC_closing is uncorrelated to almost all the detected features, with excepmtion of the Feature INR whose p-value is 0.0337.

Now we have to adjust our dataset to include our valid features.

```
## # A tibble: 6 x 15
##    Date BTC_open BTC_close   XRP    ETH   JPY    KRW   GBP   INR   RUB   GOLD
##   <dbl>    <dbl>     <dbl> <dbl>  <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl>  <dbl>
## 1 17577   175532    185294  20.9 17073. 0.174 0.0174  26.0 0.290 0.328 25128.
## 2 17578   185294    188286  20.5 17166. 0.174 0.0174  26.1 0.290 0.328 25063.
## 3 17581   191746    207057  19.7 16866. 0.174 0.0173  26.0 0.288 0.328 24961.
## 4 17582   207057    209930  20.5 17406. 0.174 0.0174  26.0 0.288 0.328 24984.
## 5 17583   209930    196903  19.2 16514. 0.174 0.0174  26.2 0.288 0.331 24818.
## 6 17584   196903    183041  17.9 15782. 0.175 0.0175  26.2 0.291 0.332 24893.
## # ... with 4 more variables: NASDAQ <dbl>, DOW <dbl>, AMZN <dbl>, BABAN <dbl>
```

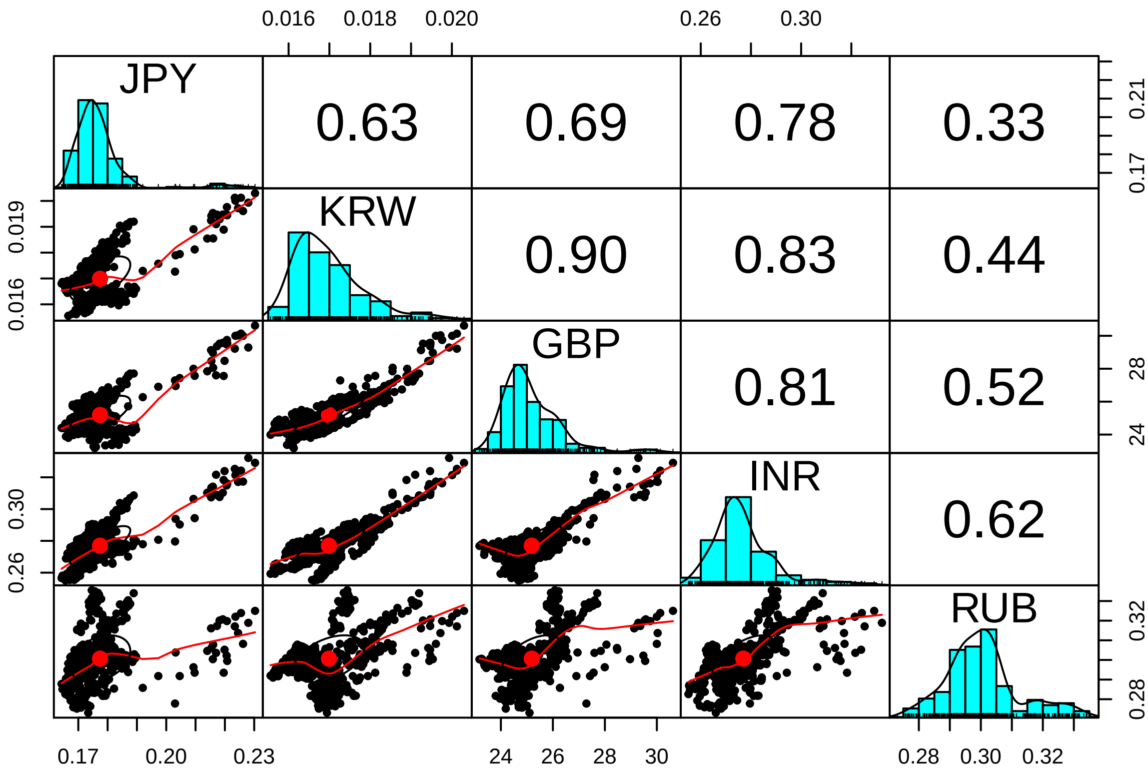
Correlations between features.

Cryptocurrencies



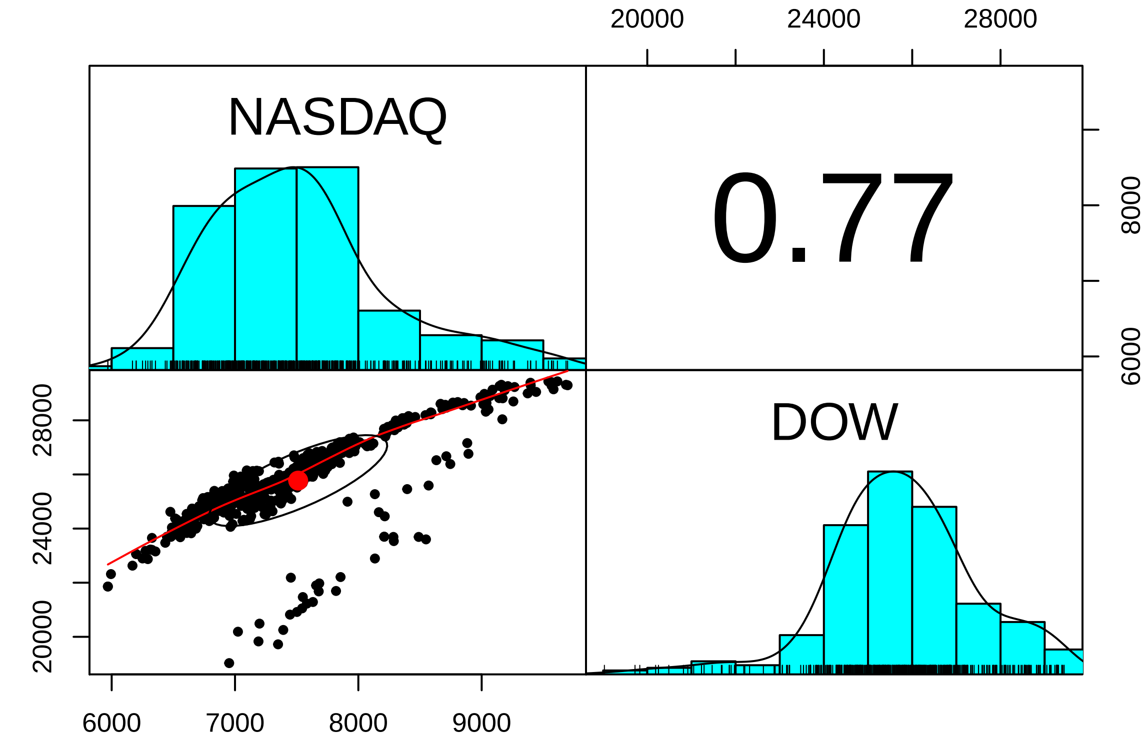
We can see a positive correlation between Cryptocurrencies, especially between Ripple and Ethereum but, not surprisingly, little correlation of both vs Bitcoin, even when the Bitcoin is Known because it moves the prices of the majority of Cryptocurrencies in the world.

Exchanges



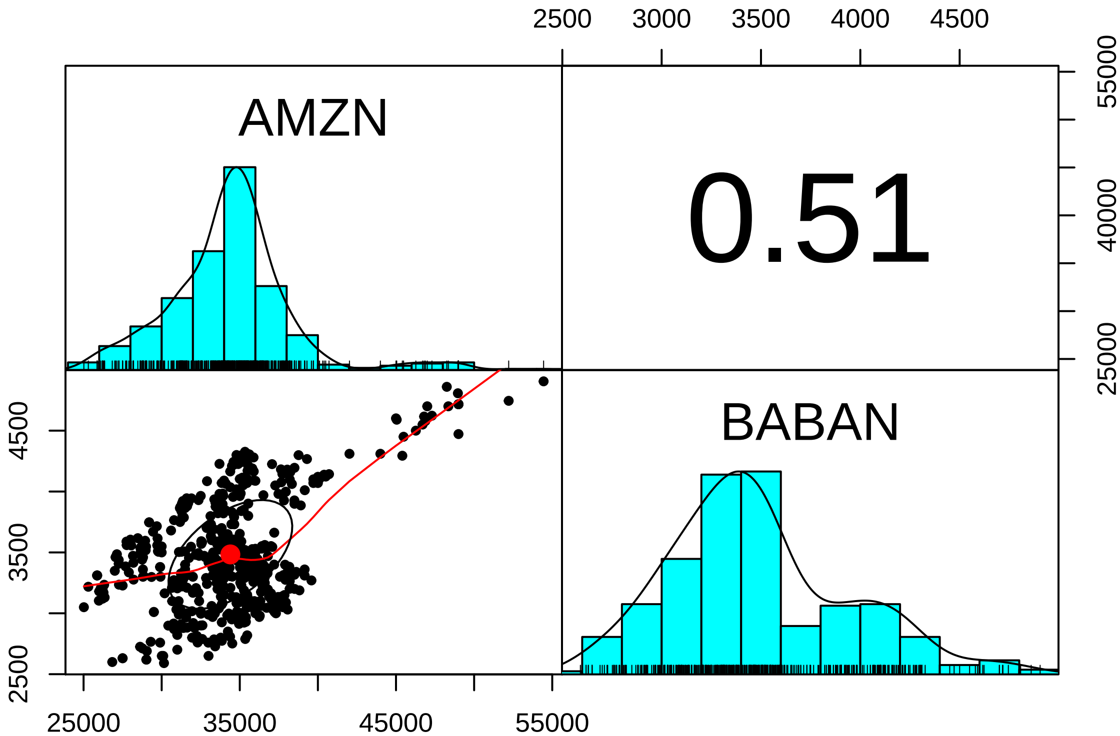
Here, a high positive correlation is seen across all exchange rates.

Indexes



As was expected, a high positive correlation between US Indexes.

Stocks



A positive correlation is seen but is much evident in lower and higher prices.

Fitting models

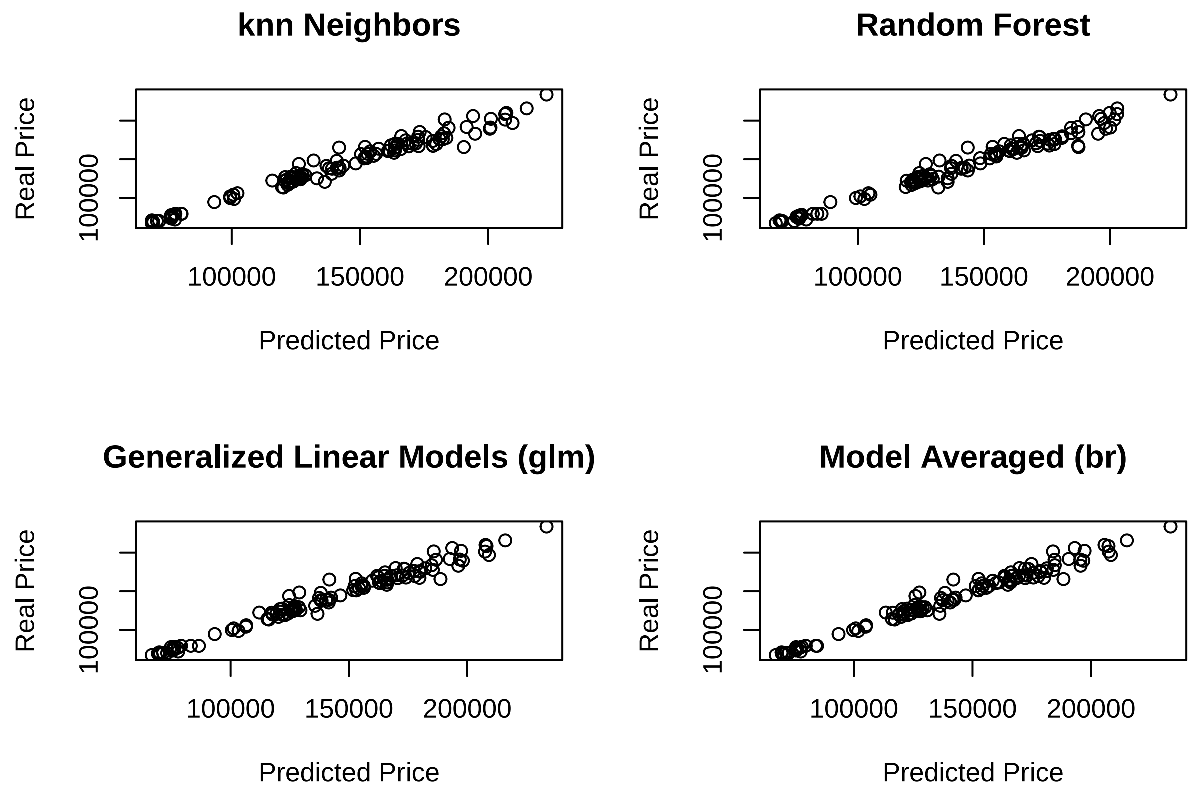
We are going to use a model-based approach, to do that, we are going to train four models *knn*, *glm*, *rf*, *blassoAveraged*. Then we will create an ensemble with all of them and make a table to compare their RMSE, and in the end, we will be able to choose the best model.

The first step is to split our dataset into the Train and Test sets.

```
set.seed(2049, sample.kind = "Rounding")

test_index <- createDataPartition(exchange_mxn$BTC_close, times = 1, p = 0.2, list = FALSE)

train_set <- exchange_mxn[-test_index,]

test_set <- exchange_mxn[test_index,]
```

Now we are ready to train our models.

Plotting - Predictions of each model vs Test_set data.



Results

Comparing the RMSE

Method	RMSE
knn	6762.081
glm	6463.010
rf	6839.016
br	6487.757

The RMSE indicates the averaged estimated error in our predictions, so, for example, in the *knn* model we can make a wrong prediction by MXN 6,762. That's a difference of 4.76% considering the averaged Bitcoin price.

Now that we have all the RMSE we can choose the one with the smallest loss function and that would be *glm* and then *br*, but before making an election, let's make an Ensemble of the four methods and see if an average makes better than all of them.

Creating an Ensemble.

With an Ensemble, our prediction will be the average of the predictions of all models for each case.

```
y_hat_ens <- (y_hat_knn + y_hat_glm + y_hat_rf + y_hat_br) / 4
```

And now we can compare its RMSE.

Method	RMSE
knn	6762.081
glm	6463.010
rf	6839.016
br	6487.757
Ensemble	6371.767

As we can see, the best option it's the Ensemble model. It has an RMSE equivalent to 4.48% of the averaged Bitcoin price.

Translating our Algorithm

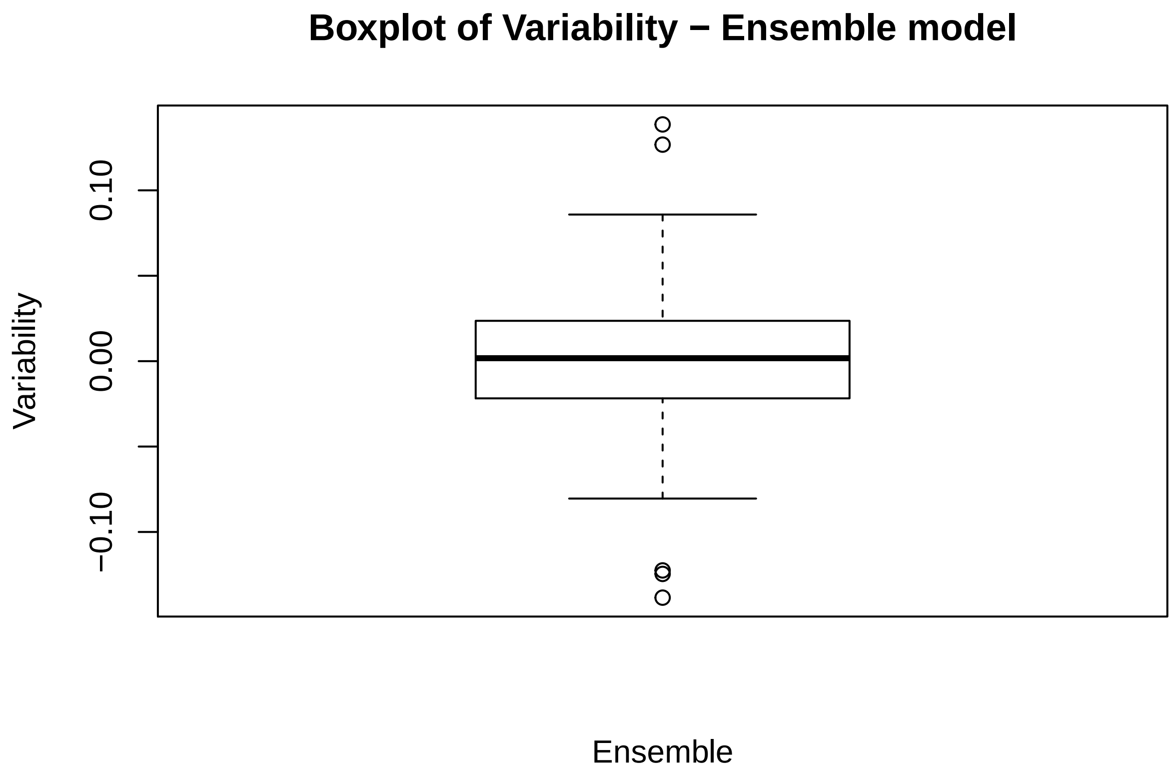
Now that we have a final model, besides the RMSE, we can summarize our performance when our algorithm was applied to the Test Set.

```
ens_var <- (y_hat_ens / test_set$BTC_close) - 1
summary(ens_var)
```

```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
## -0.1384326 -0.0215898  0.0017204  0.0009666  0.0230043  0.1385743
```

This is the summary, in terms of percentage, of how accurate our predictions were when compared them with the Test set. We can see that 50% of our predictions have variability between -2.12% and 2.41% (RMSE equivalent to 4.48%) and, our median and mean are close to zero.

In general, we can be satisfied with our model.



We have a compact box, but also we have some *outliers*, this is relatively normal in this kind of predictions, mainly for three reasons:

1. As every financial asset, Bitcoin is subject to sudden and, sometimes fake, news.
2. Its price is driven by supply and demand, thus, it depends on the variability in the prices of other assets.
3. Bitcoin is an asset with a very highly speculative use.

And these reasons provoke steep movements that are almost impossible to predict.

Conclusions

I expect that this analysis helps to understand a little better the Bitcoin behavior. As it was stated at the beginning of this document, and then proved with the analysis, even having good historic data, it's difficult to predict with extreme accuracy its price.

This kind of prediction is subject to the **"Ceteris Paribus"** condition, which means, "only if all the other things remain the same". It's a way of saying that we can predict the price but only if all the previously observed behaviors stay equal. This condition is common in Economics and Financial predictions and, in this conditioning, resides one of the biggest limitations of these models.

Further research needs to be made to add more variables to this model that helps us to make better predictions and, something important that has to be taken into account, is that Bitcoin isn't regulated yet in many countries, and in those in which already is, many changes are expected to come, of course, all of this will reformulate the way the Bitcoin is traded and valued.