CRYPTOCURRENCY ANALYSIS

PROJECT REPORT:

TEAM : CRYPTO TRACKERS

1. ROHIT THIRUPASUR(190040095)
2. MOHD. TAHA ABBAS (200010045)
3. MAYANK KUMAR PANDA(200010043)
4. DEEPANSHU VERMA(200010020)
5. DHWANIL PATEL(200010023)

INDEX

PS: For running the Data Visualisation part of the Notebook please use the Datasets from the folder named “Data Visualization Datasets” and to generate the widget based graphs in the sampling section please use the datasets as provided in the folder titled “Datasets”. The link to the notebook and video are in the Contributions Section

INTRODUCTION

How often have you been intrigued by the buzz around about all these cryptocurrencies? With so much uncertainty, technical and legal jargon surrounding Cryptocurrencies, one wonders what this is all about. And is it even worth the hype? Why did we even need them? What about the legal issues surrounding them? Is it just like the stock market? What is this mining and what kind of mining do Computers do? Will they eventually lead to a big economic crash?

With such a novel concept of cryptocurrencies developing over the last decade, these questions have always puzzled us at one point or the other and even if the legal and technical jargon don't matter to you, one important and the most relevant question is: "Can I make profits by investing and which Cryptocurrency should I invest in?" Be it big trading firms or individual investors like me and you, the aspects of making profits is what drives us to explore and through this project we aim to get an insight into answering this pressing question through Data Analysis.

Team Crypto Trackers welcomes you to get an insight into the profitability of cryptocurrencies and help you make an informed decision backed with analysis.

For this project we have focussed on four cryptocurrencies: Bitcoin(BTC), Ethereum (ETH), Litecoin(LTC) and XRP(XRP). We have taken hourly data of all these cryptocurrencies from May 2018 to June 2021. The data has been taken from Bitstamp which is a bitcoin exchange based in Luxembourg and the prices are in US Dollars.

HISTORY OF CRYPTOCURRENCIES

Ever wondered why did all these digital currencies come into play? It is not that we need them for online transactions as we can make online transactions using our normal currency like Rupees, Dollars etc. The Cryptocurrencies arised due to a need for decentralised monetary and trading systems due to devastating effects of failure at the end of central institutions. Let me explain in slightly more familiar words. Most of the transactions we do nowadays is centralised which means we deal in currencies regulated by banks which are the central institutions. Failure of any such institution can affect a huge multitude of people and businesses due to all their finances being mediated by these institutions. Infact, the economic recession of 2007-2008 demonstrates to us how common people can be affected for no fault of their own due to failure of one or more central financial institutions. The 2007-08 crisis was caused due to the bankruptcy of Lehman Brothers- one of the biggest financial institutions of the USA.

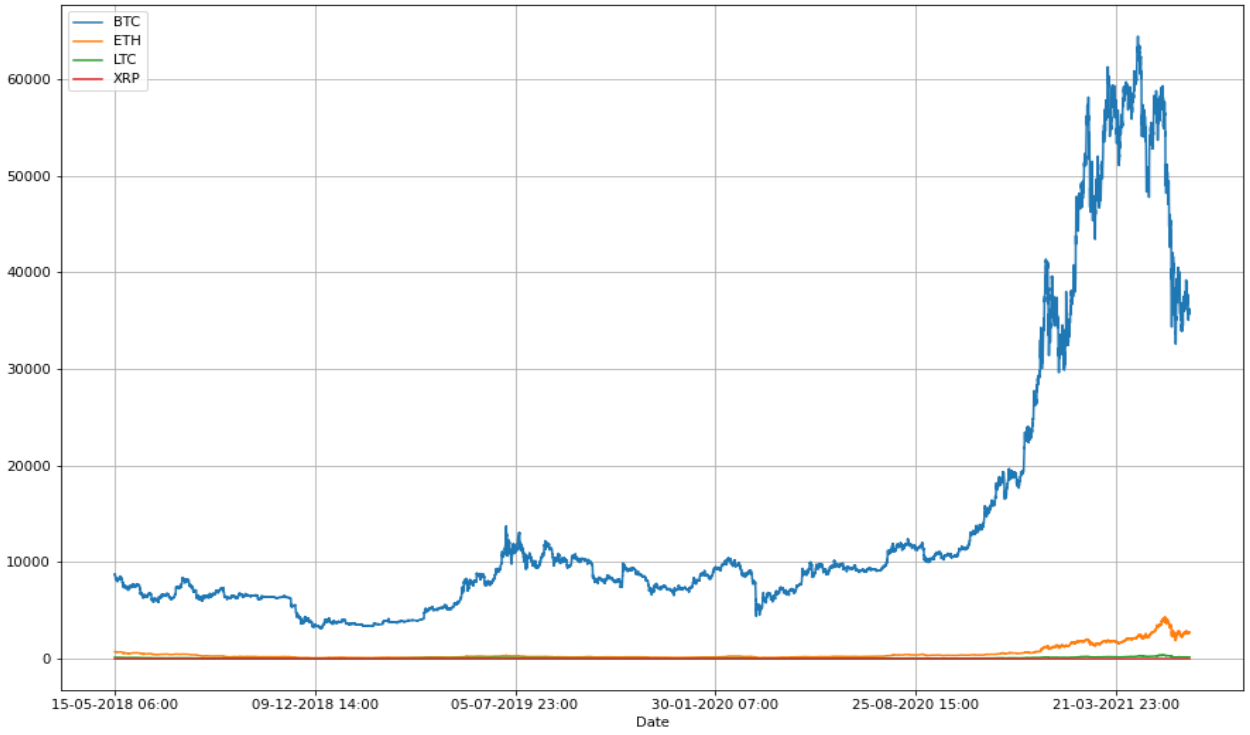
This crisis was what created a need for a Decentralized system which can be prone to failure of individual institutions. And with an anonymous paper on the idea of the first cryptocurrency Bitcoin, it all started with more and more people and firms trading in Cryptocurrencies. Ever since Bitcoin was created, there has been a tremendous increase in the number of cryptocurrencies and as of now there are somewhere between 4000-5000 different cryptocurrencies you can trade in.

Each Cryptocurrency serves a different purpose and each has its own value per coin which can be a few thousand dollars for a Bitcoin to a few cents for a Beam.So this was a brief history of Cryptocurrencies and the reason behind their existence.

DATA VISUALIZATION

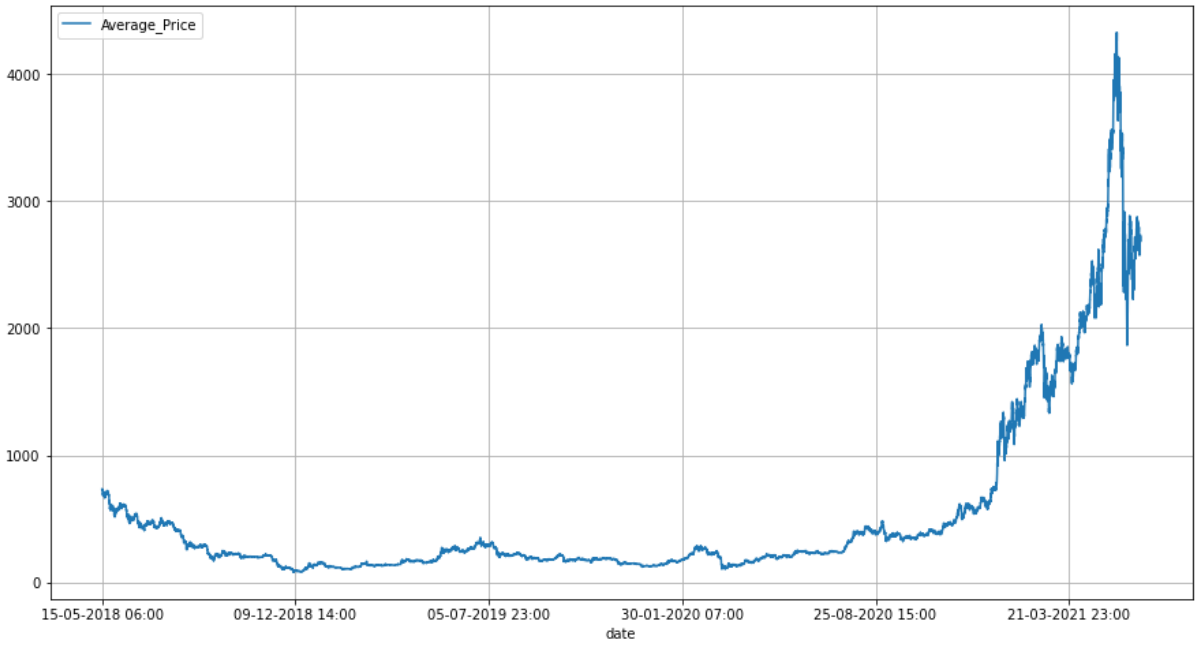
In this section, I have used pandas and matplotlib to draw some important visuals to get some useful insights from the data. We will compare how the different cryptocurrency prices vary , volume comparison, return ratio comparison, and many more. So let's get started without any further delay.

AVERAGE PRICE OF CRYPTOCURRENCIES FROM 2018 TO 2021:

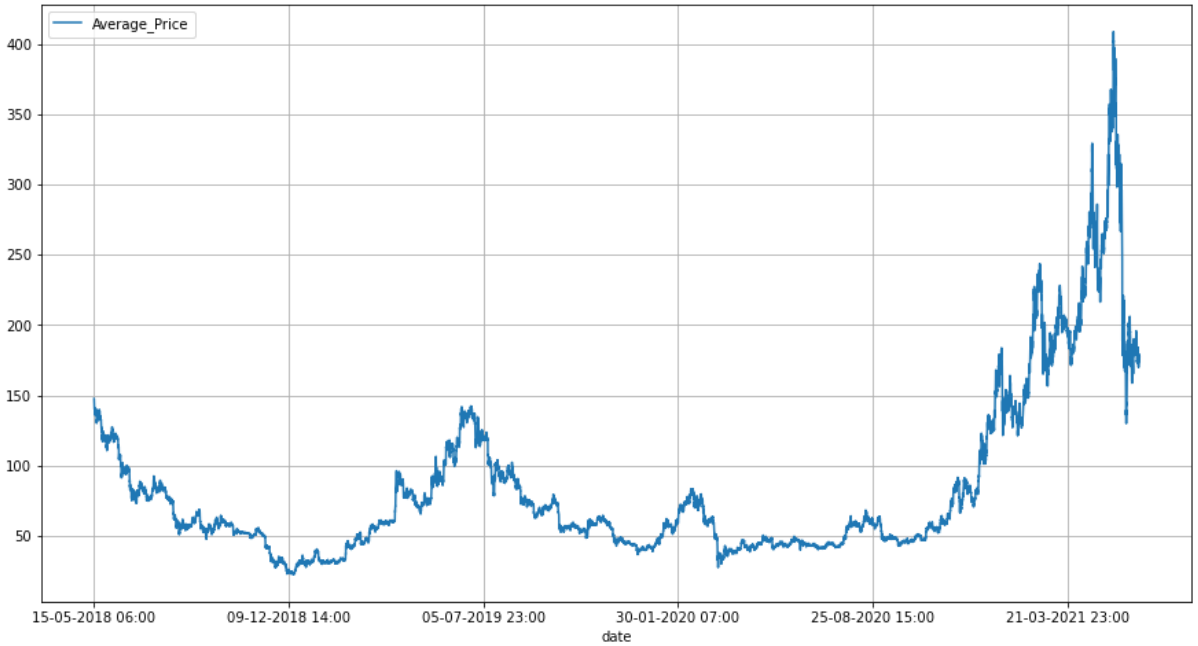


In the above graph, average prices of Ethereum, Litecoin and Ripple are much lower than Bitcoin and are not clear. So, separate graphs are as follows:

AVERAGE PRICE OF ETHEREUM FROM 2018 TO 2021:



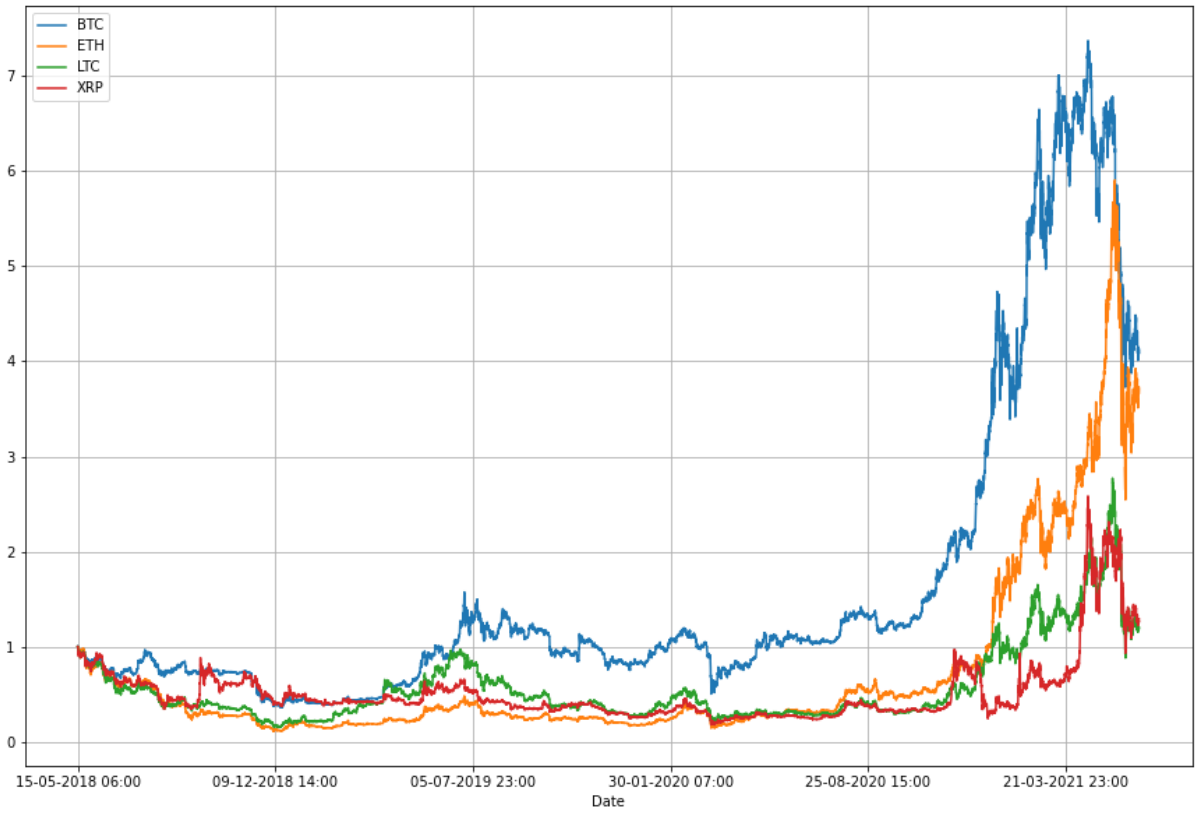
AVERAGE PRICE OF LITECOIN FROM 2018 TO 2021:



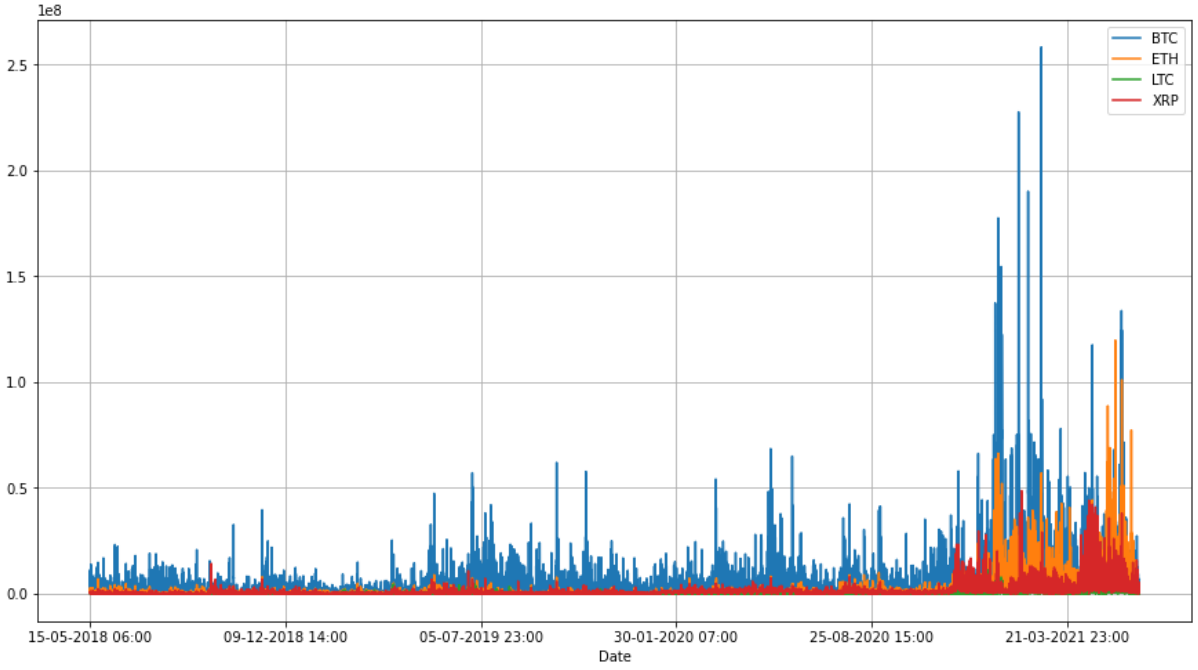
AVERAGE PRICE OF RIPPLE FROM 2018 TO 2021:

RETURN RATIO OF CRYPTOCURRENCIES FROM 2018 TO 2021:

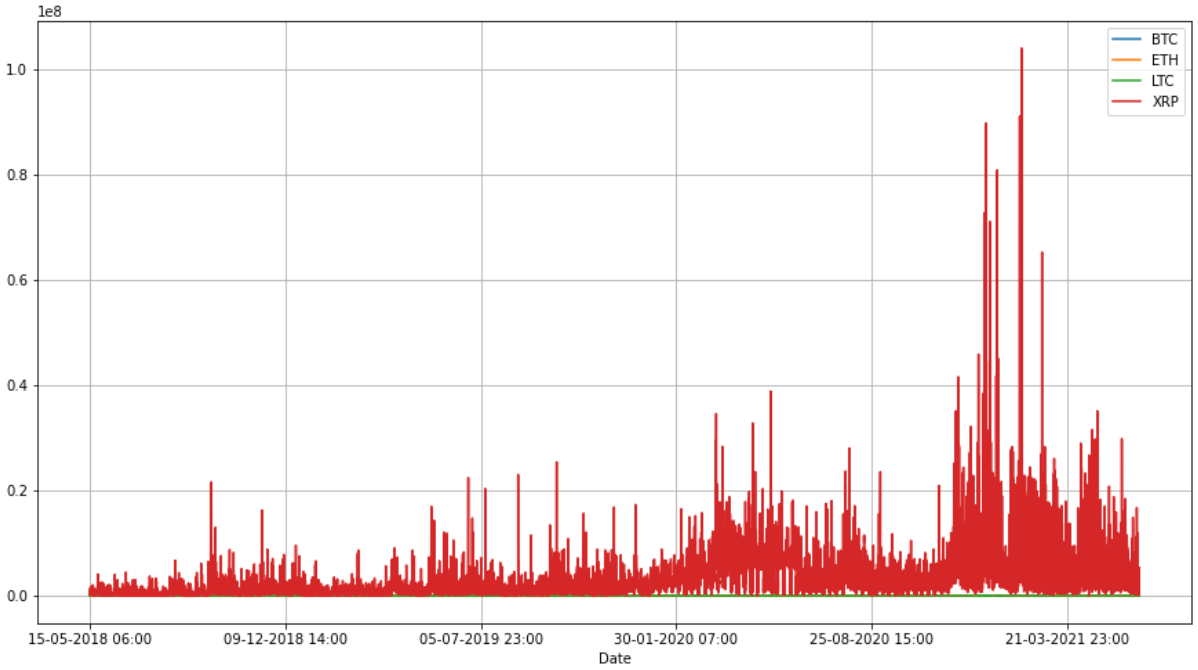
Return Ratio means the number of times invested money has become.



VOLUME TRADED IN USD OF CRYPTOCURRENCIES SINCE 2018:

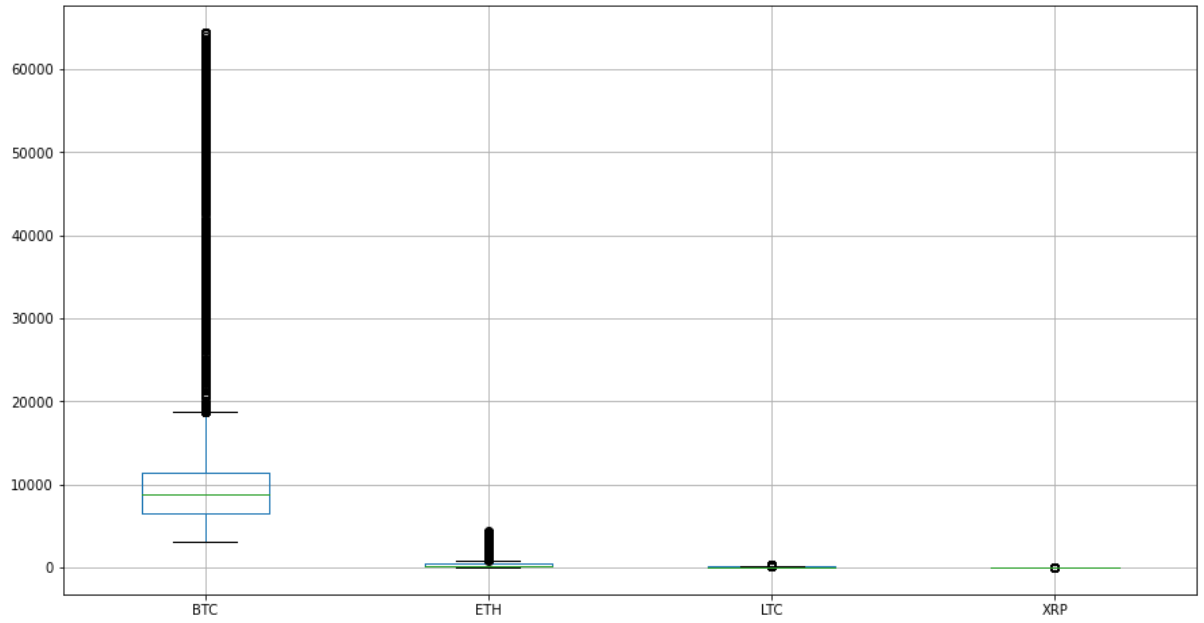


VOLUME TRADED OF CRYPTOCURRENCIES SINCE 2018:

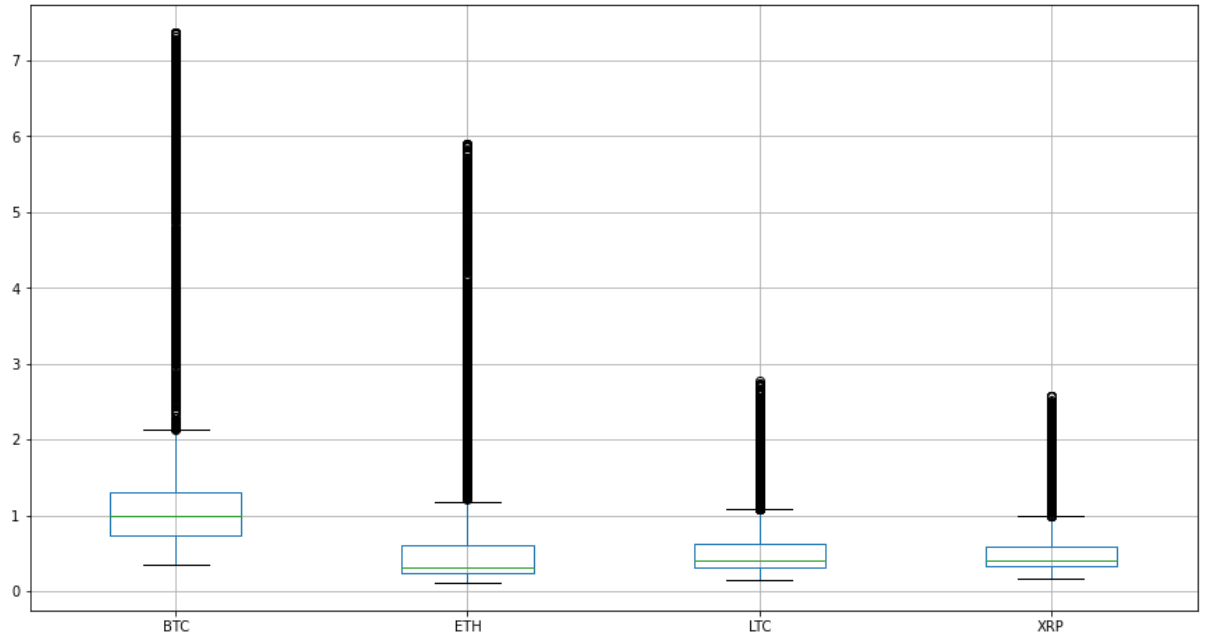


From the above graph it is clear that the number of Ripple coins traded from 2018 were much higher than the other cryptocurrencies because of its low price.

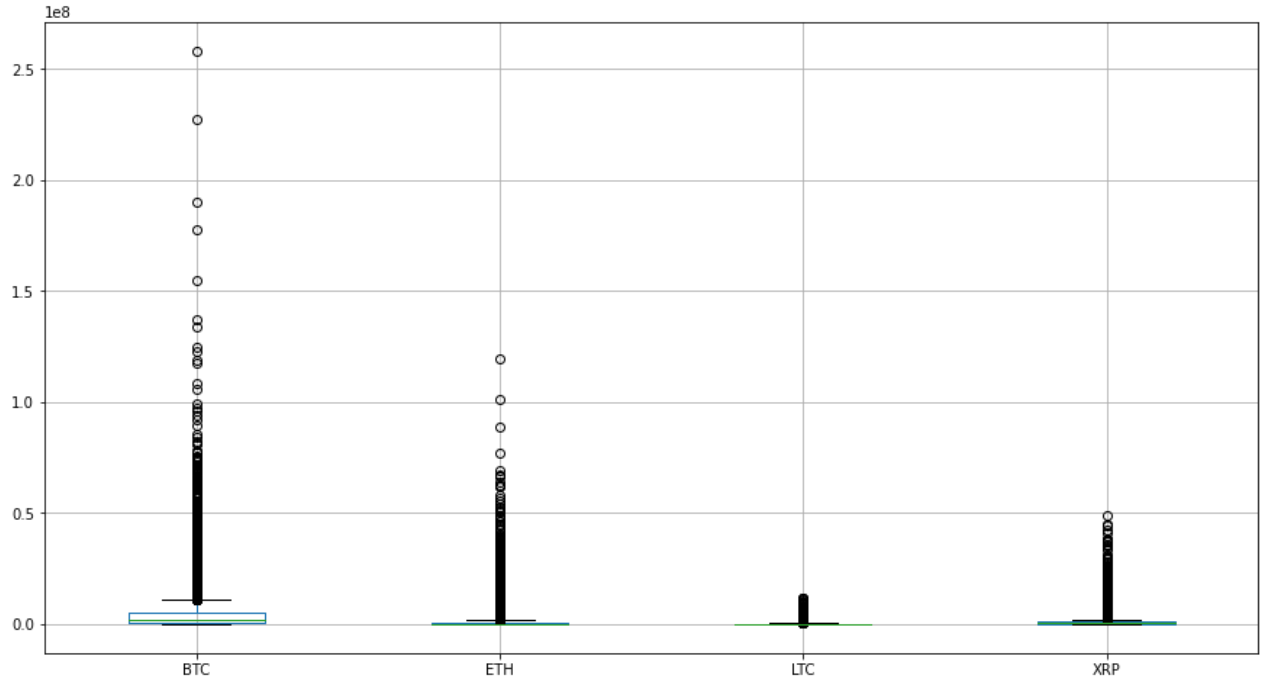
BOXPLOT OF AVERAGE PRICES OF CRYPTOCURRENCIES:



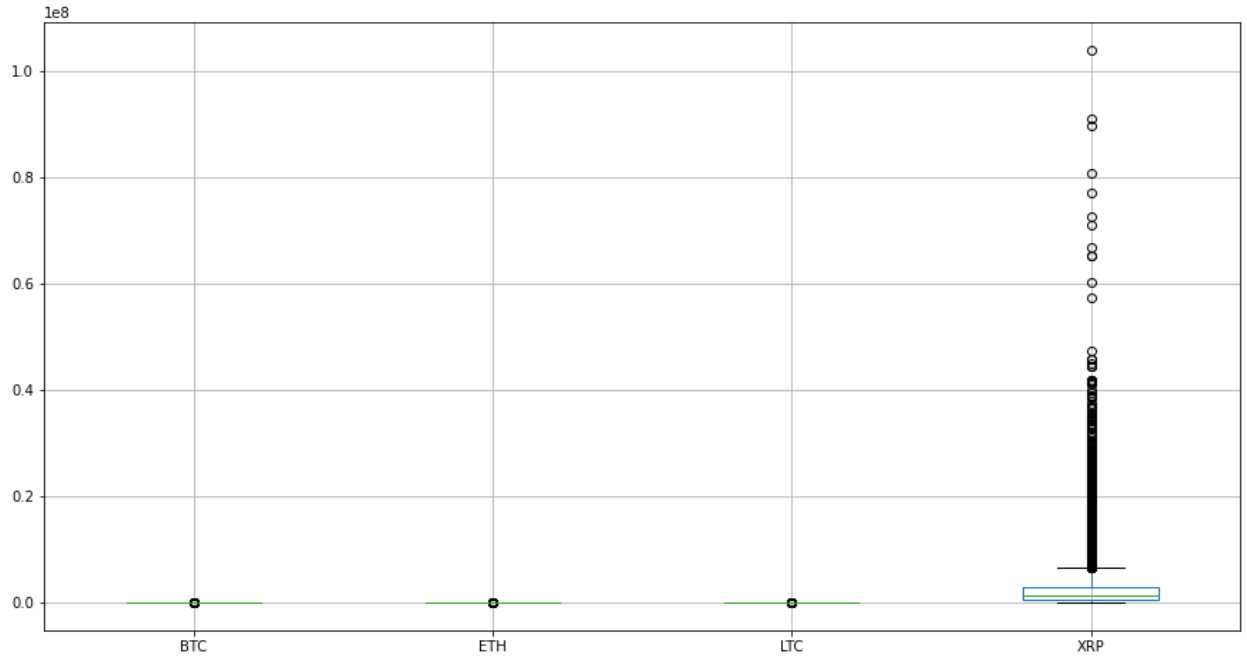
BOXPLOT OF RETURN RATIOS OF CRYPTOCURRENCIES:



BOXPLOT OF VOLUME TRADED IN USD OF CRYPTOCURRENCIES:



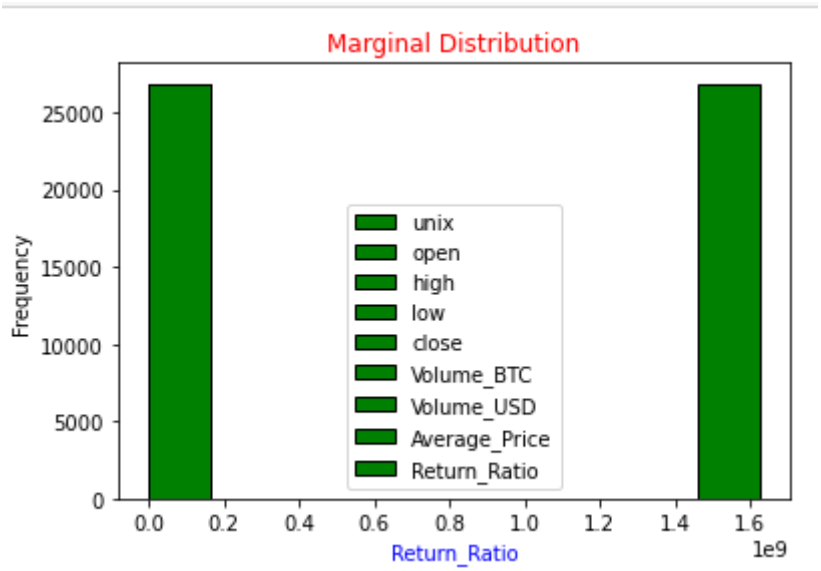
BOXPLOT OF VOLUME TRADED OF CRYPTOCURRENCIES:



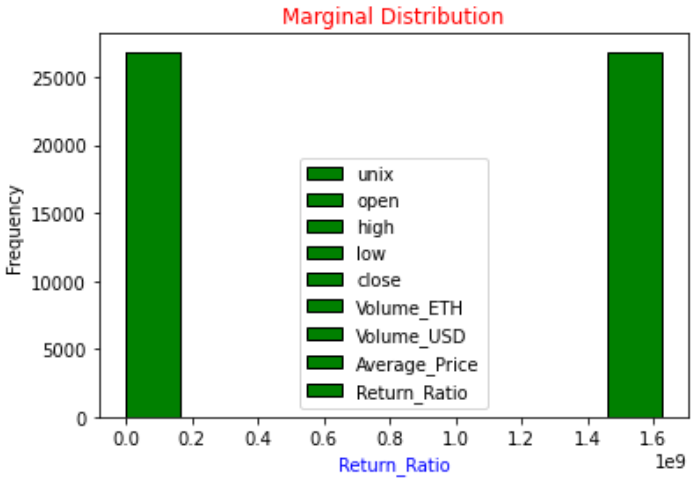
MARGINAL AND JOINT DISTRIBUTION

I have done Marginal and Joint distribution of Cryptocurrency and made a few statistical analysis.This analysis includes histogram of Return Rate for Bitcoin, Etherium, LTC and XRP respectively. Jupyter code for this analysis is provided in jupyter notebook. Main graphs of all four types of Cryptocurrency are as follows:

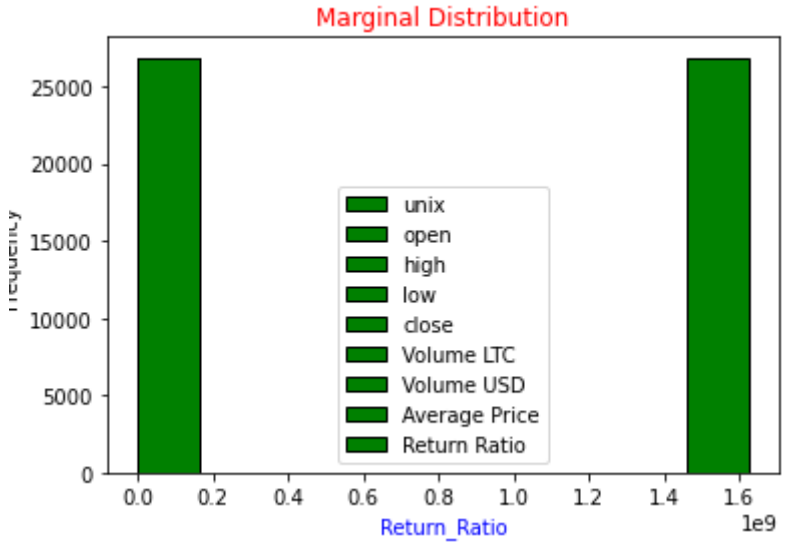
1. **BITCOIN:**



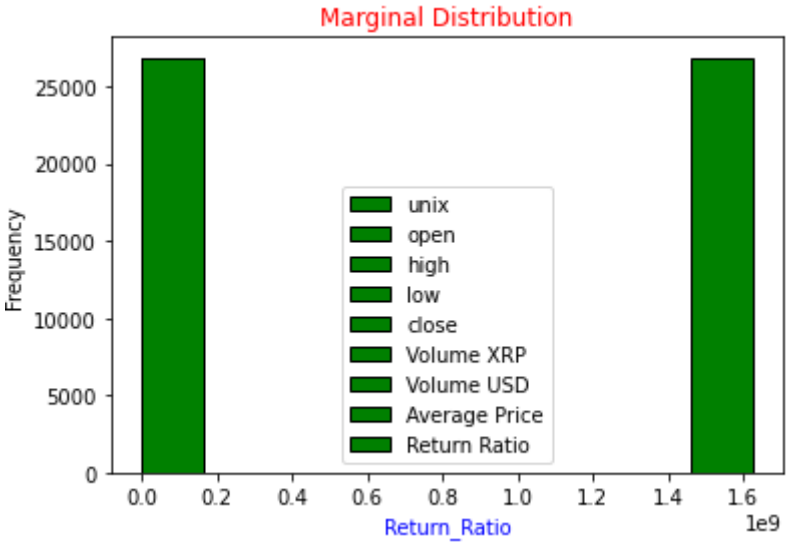
2. **ETHEREUM:**



3. LITECOIN:



4. RIPPLE:



SAMPLING AND CONFIDENCE INTERVALS

I did the Sampling Distribution and Confidence Interval part for the project, this included the analysis of the Volume USD transacted and the Return Ratio for four cryptocurrencies- BTCUSD, XRP USD, ETH USD , LTCUSD each containing 26850 data points ,Using the Sample mean and Sample Variance as the statistic , and mean of sample mean data and sample variance data as unbiased estimators for the true mean and true variance

BTC

Firstly, I imported the CSV file into a data frame using the read_csv command

```
df = pd.read_csv('D:\Bitstamp_BTCUSD_1h (1).csv')
```

Since in my analysis I required only data of Volume USD and Return Ratio , I then separated these two columns into separate series using

```
volume=df['Volume USD']
```

I calculated true Standard Deviation and True Mean of the Volume USD data using

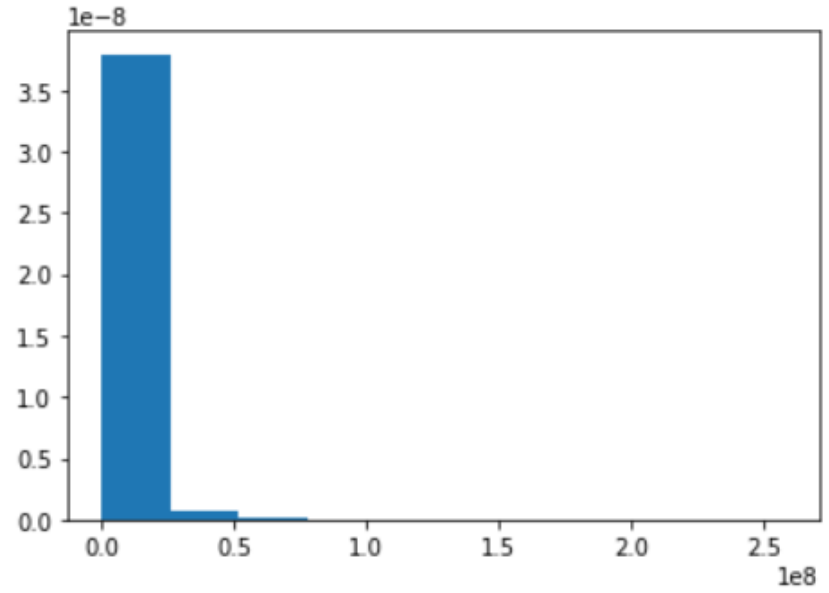
```
print('True Standard Deviation of Data is USD ', np.std(volume , ddof=1))
```

```
print('True mean of Data is USD ' ,np.mean(volume))
```

This true value will serve as a reference for comparison of my results in future.

Then I plotted the actual distribution of Volume USD distribution , the histogram looked like this

```
True Standard Deviation of Data is USD  7734536.902256386
```



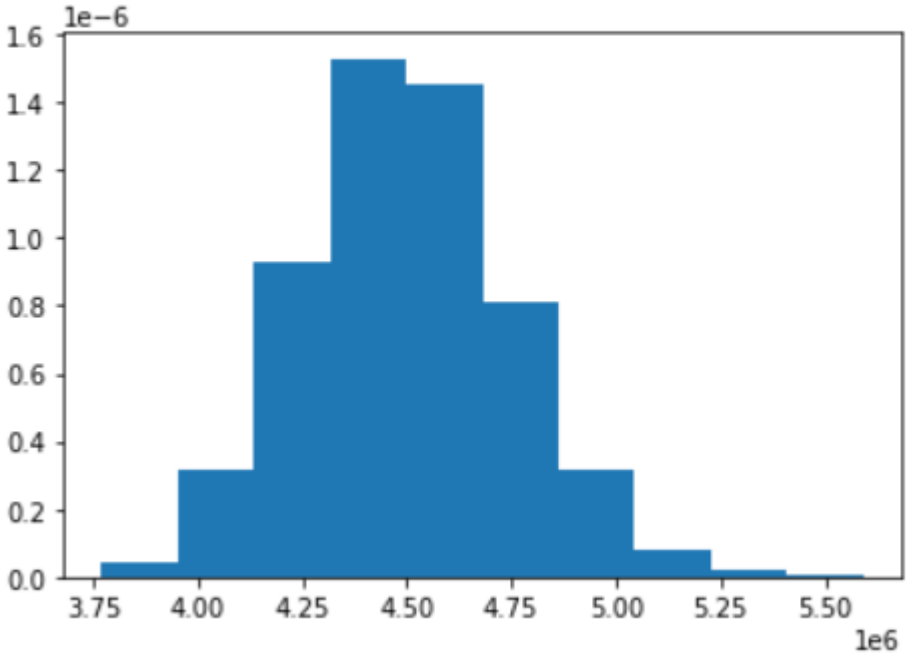
Then I calculated the sampling distribution of the mean of Volume of USD for this I chose k samples and calculated the mean , I did this process for N times and then obtained a histogram for this distribution , using this distribution I predicted the mean of the Volume USD

```
def s_distr(N=10000 , k=1000):

    mean=[]
    variance=[]
    for x in range(N):
        s_v=np.random.choice(volume , size=k)
        mean.append(np.mean(s_v))
        variance.append(np.var(s_v , ddof=1))
    print('Predicted Mean USD ',np.mean(mean))
    plt.hist(mean , density =True);


interact(s_distr , N=(10,100000) , k=(2,100000))
```

```
Predicted Mean USD  4497805.665642482
```

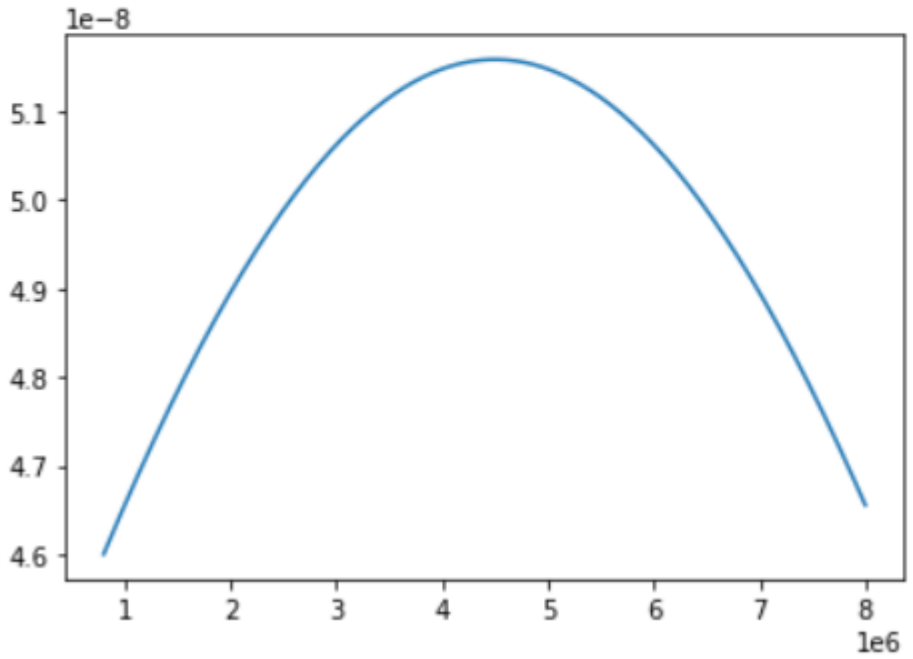


The predicted mean came as USD 4497805.66456 for k=1000 , N=10000 , which is fairly close to the actual mean of the data which is : 4499277.337603658

Also the distribution of mean resembles a normal distribution as expected

```
nrv=norm(4499277.337603658 , 7734536.902256386)
x=np.linspace(8e5,8e6, 100000)
plt.plot(x ,nrv.pdf(x))
```

```
[<matplotlib.lines.Line2D at 0x1e29de45670>]
```

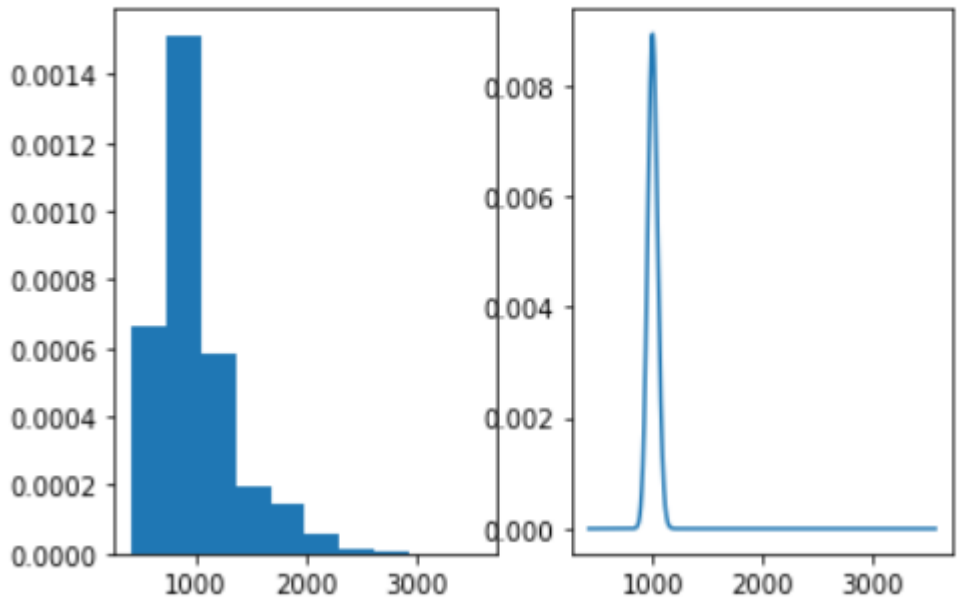


Then I calculated the distribution of variance using k samples N times to obtain a chi^2 distribution of k-1 degrees of freedom

```
def var_distr(N=10000 , k=1000):
    true_var=np.var(volume , ddof=1)
    v2=[]
    mean=[]
    variance=[]
    for x in range(N):
        s_v=np.random.choice(volume , size=k)
        variance.append(np.var(s_v , ddof=1)*(k-1)/true_var)
        v2.append(np.var(s_v , ddof=1))
    print('Predicted Variance  USD' ,np.mean(v2))
    print('Predicted Standard Deviation USD'  , np.sqrt(np.mean(v2)))
    plt.subplot(1,2,1)
    plt.hist(variance , density =True);
    plt.subplot(1,2,2)
    ch=chi2(k-1)
    x=np.linspace(min(variance) ,max(variance) ,10000)
    plt.plot(x , ch.pdf(x))


interact(var_distr , N=(10,100000) , k=(2,1000))
```

```
Predicted Variance  USD 59912447502512.26
Predicted Standard Deviation USD 7740313.139822721
```



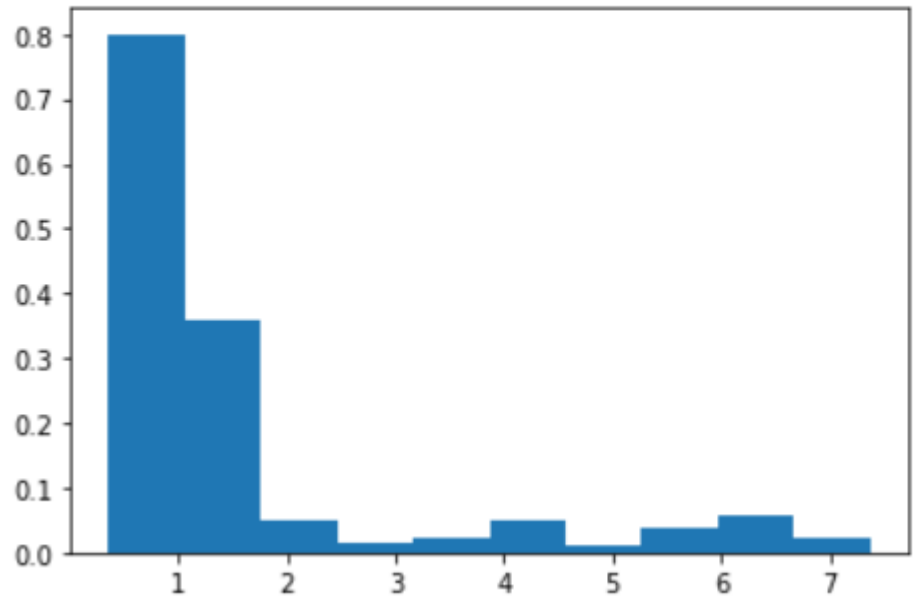
The predicted variance and standard deviation for k=1000 , N=10000 is fairly close to the actual value which is

USD 7734536.902256386

Then I started looking at the other random variable Return Ratio and plotted its actual histogram distribution and calculated its true mean and true variance

```
rr=df['Return Ratio']

plt.hist(rr , density=True);
```




```
print('True mean' ,np.mean(rr))
```

```
True mean 1.5963172995371757
```

```
print('True Variance' ,np.var(rr ,ddof=1))
```

```
True Variance 2.7256830183119907
```

Then I calculated the sampling distribution using k samples and obtaining N values

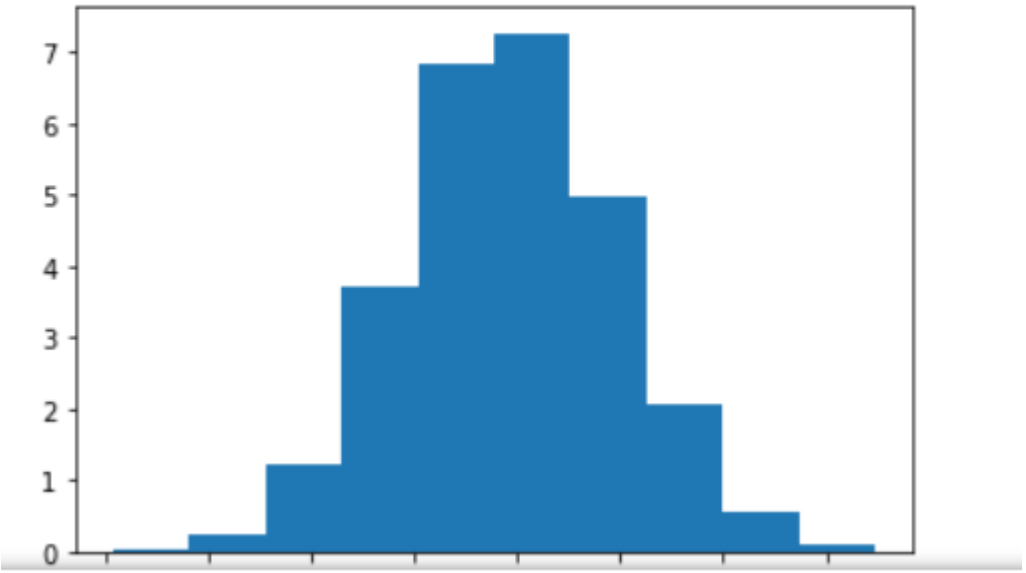
```
def s_distr2(N=10000 , k=1000):

    mean=[]
    for x in range(N):
        s_v=np.random.choice(rr , size=k)
        mean.append(np.mean(s_v))
    print('Predicted Mean ' ,np.mean(mean))
    plt.hist(mean , density =True);


interact(s_distr2 , N=(10,100000) , k=(2,100000))
```

I predicted the mean of Return Ratio using this distribution

```
Predicted Mean  1.596404041589826
```



The histogram resembled a normal distribution as expected
The predicted mean came close to the actual mean which is 1.5963172995371757

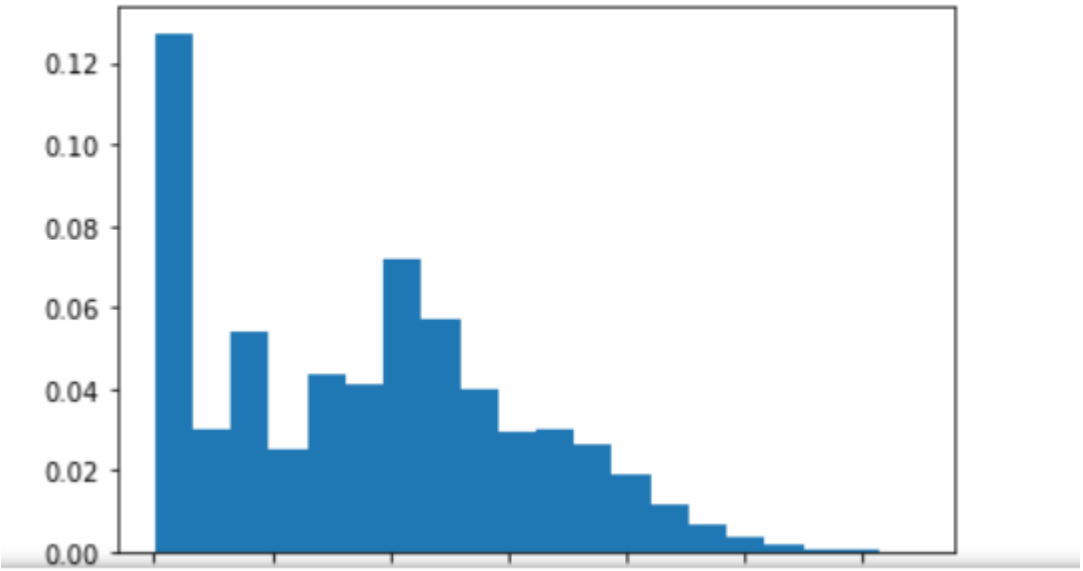
Then I calculated the sampling distribution of the variance using k samples and N experiments

```
def var_distr2(N=10000 , k=1000):
    v2=[]
    true_v=np.var(rr ,ddof=1)
    variance=[]
    for x in range(N):
        s_v=np.random.choice(rr , size=k)
        variance.append(np.var(s_v , ddof=1))
        v2.append(np.var(s_v , ddof=1)*(k-1)/true_v)
    print('Predicted Variance ' , np.mean(variance))
    plt.hist(v2 , density =True ,bins=20);

interact(var_distr2 , N=(10,100000) , k=(2,30))
```

And predicted the value of variance based on this distribution

```
Predicted Variance  2.7351327434745403
```



The predicted value came close to actual variance of return ratio came which is 2.725683018311990

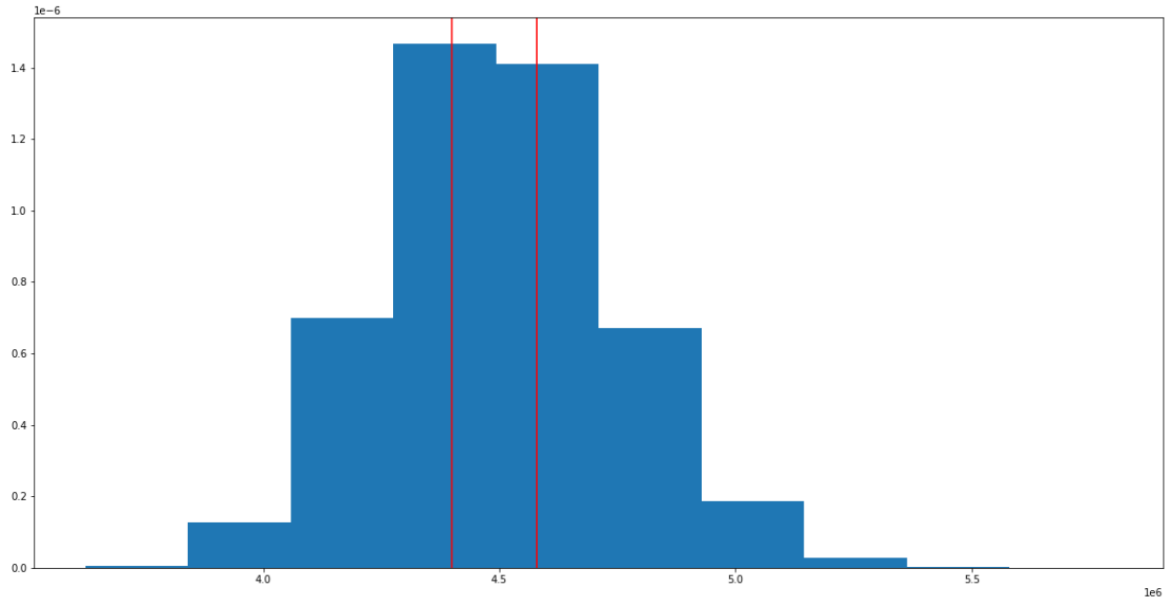
After that I calculated 95% and 99% confidence intervals for the true mean of Volume USD and Return Ratio.

Using (for volume USD) :

```
#confidence intervals  of mean
def fun(n,alpha):
    nrv=norm(0,1)
    sample=np.random.choice(volume , size=n)
    m=np.mean(sample)
    std_volume=np.std(volume , ddof=1)
    print((1-alpha)*100 , '% Confidence Interval for true mean of Volume USD is')
    print(m-nrv.ppf(1-alpha/2)*std_volume/np.sqrt(n) ,',' ,m  + nrv.ppf(1-alpha/2)*std_volume/np.sqrt(n))
    mean=[]
    variance=[]
    N=10000
    k=1000
    for x in range(N):
        s_v=np.random.choice(volume , size=k)
        mean.append(np.mean(s_v))
    plt.figure(figsize=(20,10))
    plt.hist(mean , density =True);
    plt.axvline(m-nrv.ppf(1-alpha/2)*std_volume/np.sqrt(n) , 0 ,1 ,color='r')
    plt.axvline(m+nrv.ppf(1-alpha/2)*std_volume/np.sqrt(n) , 0 ,1 ,color='r')

interact( fun , n=(2,100000) , alpha =[0.01 ,0.05])
```

```
99.0 % Confidence Interval for true mean of Volume USD is
4400255.433642612 , 4578449.010735995
```

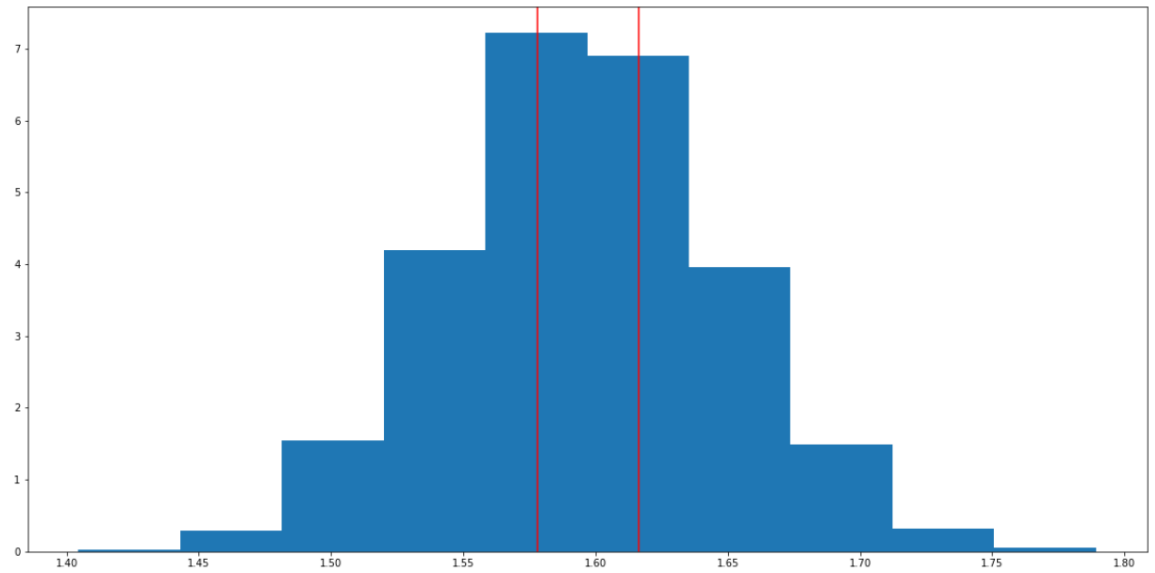


For Return Ratio:

```
#confidence intervals  of mean  return ratio
def fun(n,alpha):
    nrv=norm(0,1)
    sample=np.random.choice(rr , size=n)
    m=np.mean(sample)
    std_rr=np.std(rr , ddof=1)
    print((1-alpha)*100 , '% Confidence Interval for true mean of Return Ratio is')
    print(m-nrv.ppf(1-alpha/2)*std_rr/np.sqrt(n)  ,m  + nrv.ppf(1-alpha/2)*std_rr/np.sqrt(n))
    mean=[]
    N=10000
    k=1000
    for x in range(N):
        s_v=np.random.choice(rr , size=k)
        mean.append(np.mean(s_v))
    plt.figure(figsize=(20,10))
    plt.hist(mean , density =True);
    plt.axvline(m-nrv.ppf(1-alpha/2)*std_rr/np.sqrt(n) , 0 ,1 ,color='r')
    plt.axvline(m+nrv.ppf(1-alpha/2)*std_rr/np.sqrt(n) , 0 ,1 ,color='r')

interact( fun, n=(2,100000) , alpha =[0.01 ,0.05])
```

```
99.0 % Confidence Interval for true mean of Return Ratio is
1.5781967190005561 1.6162327685413729
```



I repeated the same process for XRP USD, ETH USD , LTC USD and obtained corresponding results

	BTCUSD	XRP USD	ETH USD	LTC USD
TRUE MEAN (VOL USD)	USD 4499277.337603658	USD 1010129.6886238761	USD 1372963.1550725193	USD 257190.11035441552
TRUE VARIANCE (VOL USD)	USD 59823061092365.82	USD 4892736773846.435	USD 14146863106984.918	USD 379373816681.37897
TRUE MEAN (Return Ratio)	1.596317299537 1757	0.515819581171507 7	0.694699083182 1285	0.5655806613977282
TRUE VARIANCE (Return Ratio)	2.725683018311 9907	0.116671 33715373 454	0.883068787746 2138	0.16001696409369734
PREDICTED MEAN (VOL USD)	USD 4497805.665642482	USD 1010396.7280836992	USD 1371334.6088460006	USD 257352.93997723516
PREDICTED VARIANCE (VOL USD)	USD 59986403074255.38	USD 4892655821118.655	USD 14153777767081.617	USD 379355541584.3565
PREDICTED MEAN (RETURN RATIO)	1.596404041589826	0.515968 28906079 75	0.694053897558 6292	0.5653927232611764
PREDICTED VARIANCE (RETURN RATIO)	2.735132743474 5	0.116794 58758685 946	0.884980013257 7198	0.15988323897354026

95% CONFIDENCE INTERVAL OF MEAN (VOL USD)	[USD 4401139.015213068 , USD 4536727.578049201]	USD 983566.209704268, USD 1022342.3579810716	USD 1332586.6708817403 ,USD 1398522.096987818	USD 251951.33050315868 ,USD 262748.8107574999
99% CONFIDENCE INTERVAL OF MEAN (VOL USD	USD 4430321.560945826 ,USD 4608515.138039209	USD 990152.057521995 9, USD 1041112.5546525524	USD 1355191.0950145589, USD 1441844.933577758	USD 248704.57526507252 ,USD 262894.8697107889
95% CONFIDENCE INTERVAL OF MEAN (RETURN RATIO)	1.579321778820 6009 1.60826 3637316 9967	0.514993 03008110 19 , 0.520980 87893455 78	0.68996 0628750 8471 , 0.70643 4133635 6652	0.5608832876337232 0.5700992510784225
99% CONFIDENCE INTERVAL OF MEAN (RETURN RATIO	1.57479 2492365 7886 1.61282 8541906 6053	0.509684 39043307 91 , 0.517553 75776003 7	0.68429 0183430 6718 , 0.70594 0038708 2055	0.5608832876337232 ,0.5700992510784225

HYPOTHESIS TESTING I (BTC AND ETH)

Bitcoin and Ethereum have been among the earliest coins traded.What makes these popular among the investors?

The Decentralized Blockchain Technology:

One of the reasons Bitcoin and Ethereum remain successful is the blockchain technology. The blockchain network was essentially built to create an environment that allows BTC and ETH transactions to be facilitated between users on the network without the interference of a third party.

Hence, this is a fully decentralized system which means some attributes like the transparency of the data, the anonymity of the users are only contributing to the popularity of the technology and currencies.

Online Trading Sites:

Another reason why Bitcoin and Ethereum are very popular is online exchange sites. Because these are among the first cryptocurrencies that have been launched, they are accepted on a majority of trading sites.

Given these basic attributes,what should be our ideal choice as a naive investor?Well,the answer to this is not that straight forward.Among the plethora of factors that could influence our choice,here are a few that have been worked out by the usual method of hypothesis testing.

Hypothesis 1:

The average return ratios of both the coins are compared here to get an idea about which coin can fetch more returns over a certain period.

Known variances can be assumed here.

* Null Hypothesis : The average return ratio, on an hourly basis, of Bitcoin is equal to that of Ethereum.
* Alternate Hypothesis : The average return ratio, on an hourly basis, of Bitcoin is greater than that of Ethereum.

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim \mathcal{N}(0, 1)$$

The level of significance assumed here is 1% and the approach of the critical region is used here.

The null hypothesis is rejected and we find that bitcoin gets higher returns indicating its popularity.

Hypothesis 2:

Looking at the data,one can make an initial judgement that the trade volume of Bitcoin in USD is at least twice that of Ethereum on an average.

The null hypothesis was considered for a factor of three.

Unknown variances were assumed here since the large values made the computation of true variances a little cumbersome.The approach used was through p-value calculation.

* Null Hypothesis : The average volume of trade(in USD), on an hourly basis, of Bitcoin is thrice that of Ethereum.
* Alternate Hypothesis : The average volume of trade, on an hourly basis, of Bitcoin is greater than thrice that of Ethereum.

The level of significance is assumed to be 5%.The null hypothesis was eventually rejected which indicates that the volume of trade for Bitcoin is more than three times that of Ethereum!

Hypothesis 3:

In order to get a measure for the stability of Bitcoin and Ethereum on an hourly basis,a comparison was made between fluctuations in the prices of the two currencies.

A one sided test on the left of the distribution was performed using the critical region approach.The known variances case was applied.

* Null Hypothesis : The average fluctuation, on an hourly basis, of Bitcoin is equal to that of Ethereum.
* Alternate Hypothesis : The average fluctuation, on an hourly basis, of Bitcoin is lesser than that of Ethereum.

The level of significance was fixed at 10% which gives the liberty to make a type 1 error 10% of the times.

The null hypothesis was rejected.this gives us some insight on the stability of Bitcoin compared to Ethereum.

Points to Take Away:

1.Bitcoin is the more traded currency as evident from its volume of trade.

2.It is also the more preferred currency to invest due to its greater return ratio,stability and predictability.These features allow investors to have more confidence in their investments.

3.The only negative from an investor's point of view about Bitcoin is high price per coin that discourages small investors.This is due to the large gap between demand and supply.Such investors feel they are better off investing in Ethereum,Litecoin or XRP.This is the main factor that can govern the rise of these relatively new currencies in the cryptocurrency market as explained in the next section of hypothesis testing.

HYPOTHESIS TESTING II(LTC AND XRP):

So we had a look at the Return Ratio and Trade Volumes for Bitcoin and Ethereum. And by now you might have seen that on an average Bitcoin and Ethereum have a higher return ratio as compared to other two. So from the point of view of an investor you have the answer to the question we first started with: “Can I make profits by investing and which Cryptocurrency should I invest in?”

You now know that Bitcoin and Ethereum are the currencies to look forward to. So should you just go and buy some of these two? Wait before you click Buy. You see Cryptocurrency transactions have a dedicated process of verification to prevent wrong data. This transaction has its own cost and time for each Cryptocurrency. For Bitcoin and Ethereum, the transaction time and cost are much higher than LTC and XRP. So if you are looking to start out and don’t want to start with very big amounts, you can start with these to prevent high payments in transaction charges and if you have to make high frequency deals in future, you can rely on these coins.

In hypothesis testing we will be analysing the Return Ratio and the Volume USD for both of these coins.

Hypothesis Testing for Return Ratio:

In this hypothesis we will be analysing the Return Ratios of LTC and XRP and try to find out which one has a better return ratio. Seeing the brief statistical description of LTC and XRP using the describe() function of Dataframe in Pandas, I put forward the following Null and Alternate Hypothesis:

H_0: $X_{LTC}=X_{XRP}$

H_1: $X_{LTC}>X_{XRP}$

Where X_{LTC} is the Average Return Ratio for LTC and X_{XRP} is the Average Return Ratio for XRP

Using a 95% Confidence Interval i.e alpha=0.05, I was able to reject the Null Hypothesis using a one-sided p value test and I was able to conclude with 95% confidence that LTC has a higher return ratio as compared to XRP.

Hypothesis Testing for Trade Volume:

In this hypothesis we will be analysing the Trade Volumes of LTC and XRP and try to find out which one has a higher Trade Volume. Seeing the brief statistical description of LTC and XRP using the describe() function of Dataframe in Pandas, I put forward the following Null and Alternate Hypothesis:

H_0: $V_{LTC}=V_{XRP}$

H_1: $V_{XRP}>V_{LTC}$

Where V is the Volume USD.

Using a 95% Confidence Interval i.e alpha=0.05, I was able to reject the Null Hypothesis using a one-sided p value test and I was able to conclude with 95% confidence that XRP has a higher trade volume as compared to LTC.

```
In [25]: if (pval>alpha):
             print("Null Hypothesis Accepted")
         else:
             print("Null Hypothesis Rejected")

         Null Hypothesis Rejected
```

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \sim \mathcal{N}(0,1)$$

REGRESSION ON BTC

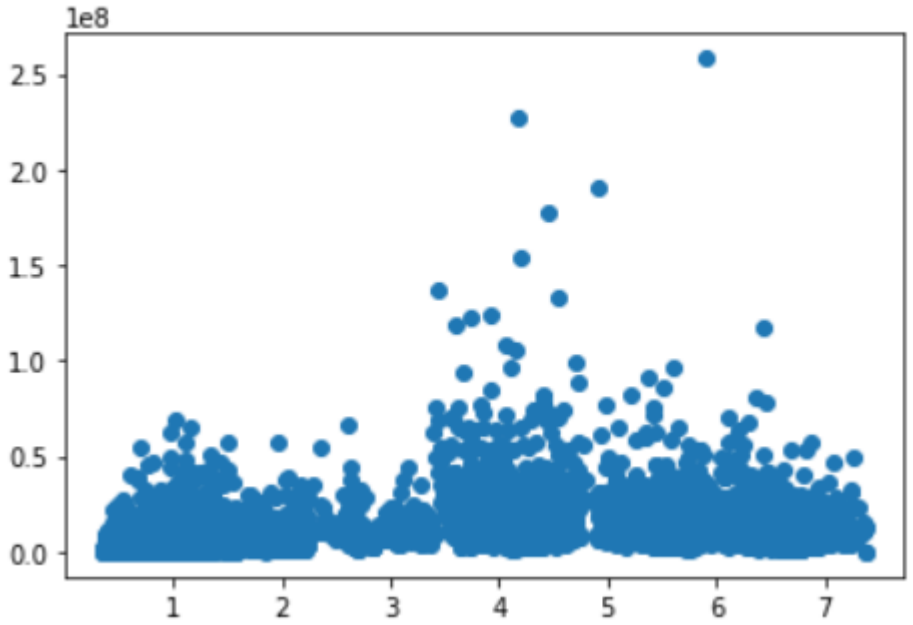
Random Variables : Volume USD and Return Ratio

I imported the .csv file to a dataframe and then separated Volume USD and Return Ratio columns , into separate series and then converted these series into numpy array

```
x=rr=df['Return Ratio']
y=vol=df['Volume USD']
x=np.array(x)
y=np.array(y)
```

The scatter plot of the data looked like this:

```
plt.scatter(rr ,vol);
```



Then I calculated b1 and b0 using S_{xx} , S_{xy}

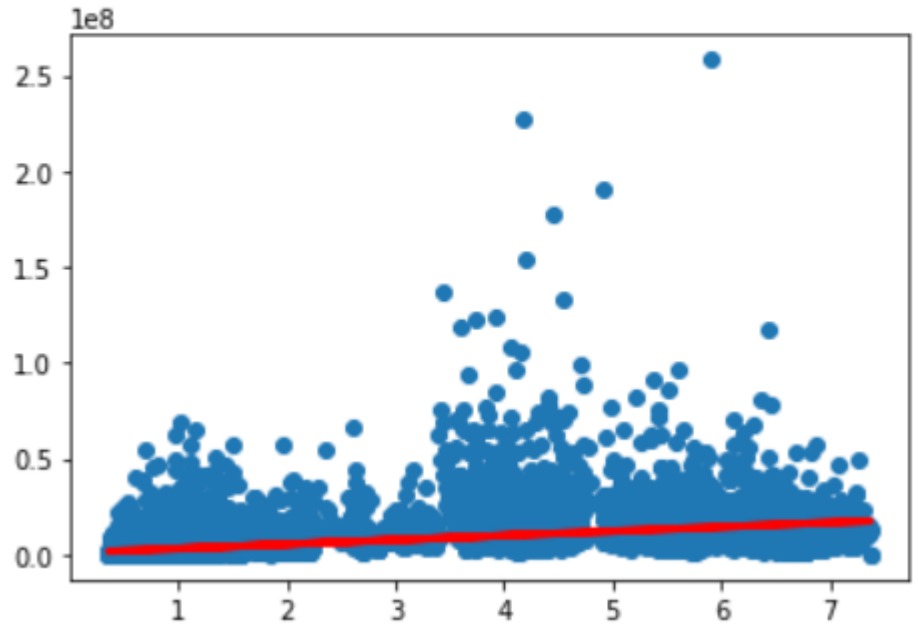
```
x_mean=np.mean(x)
y_mean=np.mean(y)
sxx=np.sum((x-x_mean)**2)
sxy=np.sum((x-x_mean)*(y-y_mean))
b1 = sxy/sxx
b0 = y_mean - b1*x_mean
```

```
print('The Linear Regression Coefficients are ', b0 , ' and ',  b1)
```

```
The Linear Regression Coefficients are  891341.9698658073  and  2260161.791633721
```

Assuming a linear fit I plotted the best fit line and calculated the R^2 value

```
yp=b0 + b1*x
plt.plot(x, yp, 'r', lw=3)
plt.scatter(x,y);
```




```
sse=np.sum((yp-y)**2)
sst=np.sum((y-y_mean)**2)
ssr=sst-sse
rsq=ssr/sst

print('R Sqaured Value is  :' ,rsq)
```

```
R Sqaured Value is  : 0.23274790170338844
```

CONCLUSION:

The R^2 value came to be 0.23274 , which is closer to zero than to one , and we don't have any prior information of the underlying model being linear , thus I conclude that there is no linear relation between return ratio and volume transacted in USD.

REGRESSION ON ETH

Assuming a simple linear model and applying supply and demand principles,we expect a rise in trade volume with rise in prices.This is both due to increase in supply as well as shift in demand curve due to change in sentiments of investors as they grow in confidence about the market.

A linear least squares fit does not capture the complete essence of the data which is indicated by a low R^2 value of 0.344.

We also see that a higher degree polynomial can do a better fit for our given data.Here,we see how a third degree polynomial can be a better fit although there is no conclusive evidence or theoretical backing to a higher polynomial fit.

Also shown is the perfect fit for return ratio and average prices which is a consequence of the definition of return ratio.

Conclusion:

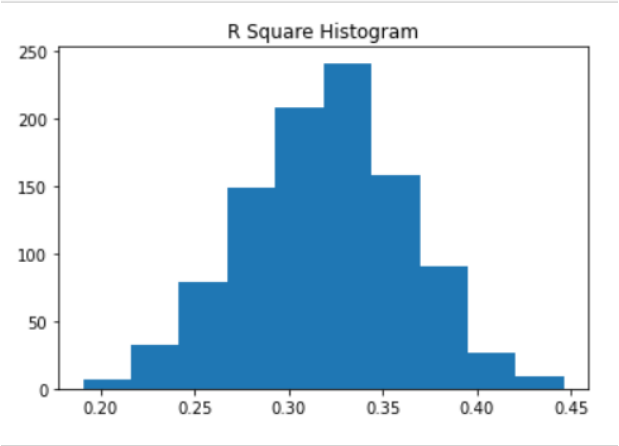
Hence,we can conclude that there are a lot of other factors that need to be taken into account to come to a fair conclusion about the relation between the two.These could be market sentiments(bullish or bearish),market structure,ease of transactions,types of markets,presence of middlemen and arbitrageurs etc.

REGRESSION ON LTC

In this, I tried to find if there exists a linear regression between the Return Ratio and Volume in USD for LTC. Using all the 26852 points, I tried to find the Least Squares fit between the two.

I got a R^2 value of about 0.3125 using all the points and this is not a good sign for a linear fit.

Now taking random samples of size 1000, I tried to get a rough distribution of R^2 and see the histogram of these values. The histogram which I got was roughly centred around 0.32 and it was spread between 0.20 and 0.45. Hence, we see that the R^2 values we get either taking all points together or random samples is pretty low to be called a good fit. Hence, we conclude that there is no linear relation between Return Ratio and Volume USD.



In terms of reasoning, think of it as this: Volume USD is the total Volume of Trade in that particular hour that is how many dollars worth of trades was made in that particular hour and this involves both buying and selling. So an hour with a high Return Ratio will have high Volume USD because of people selling more whereas an hour with a low Return Ratio can also have high Volume USD due to people who are buying.

REGRESSION ON XRP

Assuming a simple linear model between Return Ratio and Trade Volume backed by intuition we expect higher Volume USD corresponding to high Return Ratio.

In this, I tried to find if there exists a linear regression between the Return Ratio and Volume in USD for LTC. Using all the 26852 points, I tried to find the Least Squares fit between the two.

Using the R^2 value as an indicator of how good the linear fit is, I decided to back my conclusion of whether there exists a linear fit or not on this statistic and doing the computation I got a R^2 value of 0.213

With this low R^2 value, I hereby conclude that there is no linear relation between the two quantities and this was against the intuition on which I tried to find the linear fit between these quantities.

I feel this has happened due to the various non numerical factors coming into play like Market sentiment or something as simple as Elon Musk Tweets :p

CONCLUSION

Through this project, we have made an attempt to characterize the financial markets through data analysis and through the process have tried to provide insights into the domain of Cryptocurrencies.We chose this as a topic due to the easy availability of historical data, the random nature of markets and a zeal to make an attempt to find some order into the highly random and fluctuating world of Cryptocurrencies.

With financial markets of course things ascend much beyond simple Data Analysis. We need to consider the impact sentiments, emotions and perceptions have in determining the statistics which we analyzed using pure numbers. Due to the non characterization of non numerical factors like public sentiment, there is always a chance that mathematical models can go wrong but still trying to get valuable conclusions from this data has been a very enticing process for us.

This project was just the beginning of our understanding of finance and we learnt quite a significant bit about finance through this project.The learning will not stop and hopefully we will get to know more both in the field of Data Science and Finance.

We hope that you liked our project and it helped you get some insight into the domain of Cryptocurrencies. We had a great time doing this project and we are thankful for this opportunity.

Signing Off,
Team Crypto Trackers

CONTRIBUTIONS

- DOCUMENT FORMATTING: DEEPANSHU VERMA
- INTRODUCTION : MAYANK KUMAR PANDA
- HISTORY OF CRYPTOCURRENCIES : MAYANK KUMAR PANDA
- SOME DATA VISUALISATION: DEEPANSHU VERMA
- MARGINAL AND JOINT DISTRIBUTIONS: DHWANIL PATEL
- SAMPLING AND CONFIDENCE INTERVALS: MOHD. TAHA ABBAS
- HYPOTHESIS TESTING I(BTC AND ETH) : ROHIT THIRUPASUR
- HYPOTHESIS TESTING II(LTC AND XRP) : MAYANK KUMAR PANDA
- REGRESSION (BTC) : MOHD. TAHA ABBAS
- REGRESSION (ETH) : ROHIT THIRUPASUR
- REGRESSION (LTC) : MAYANK KUMAR PANDA
- REGRESSION (XRP) : ROHIT THIRUPASUR(JUPYTER NOTEBOOK)
  MAYANK KUMAR PANDA(REPORT)

➢ PROJECT VIDEO LINK :
https://drive.google.com/file/d/1LIL0uuptI9Xgo1HIjCbZKVq85G2Jr1rO/view?usp=sharing

➢ JUPYTER NOTEBOOK LINK :
https://drive.google.com/file/d/10lPoCXKrNdzIsw0cd9qJ2ziwKVSG1_9Y/view?usp=sharing

THANK YOU