**Exploring the Bitcoin Cryptocurrency**

**1.1-Introduction**

**Python**, a high-level, interpreted scripting language was developed by Guido van Rossum in the late 1980s.

It is popular and also considered as one of the hottest skills to have and the most popular programming language in the world based on the Popularity of Programming Language Index. It has many built in packages which supports machine learning and Artificial Intelligence and helps in analysing the data using them.

**Machine learning** is a subfield of artificial intelligence (AI). The goal of machine learning generally is to understand the structure of data and fit that data into models that can be understood and utilized by people.

Although machine learning is a field within computer science, it differs from traditional computational approaches. In traditional computing, algorithms are sets of explicitly programmed instructions used by computers to calculate or problem solve. Machine learning algorithms instead allow for computers to train on data inputs and use statistical analysis in order to output values that fall within a specific range. Because of this, machine learning facilitates computers in building models from sample data in order to automate decision-making processes based on data inputs.

**Bitcoin** (₿) is a cryptocurrency, a form of electronic cash. It is a decentralized digital currency without a central bank or single administrator that can be sent from user to user on the peer-to-peer bitcoin network without the need for intermediaries.

Transactions are verified by network nodes through cryptography and recorded in a public distributed ledger called a blockchain. Bitcoin was invented by an unknown person or group of people using the name Satoshi Nakamoto and was released as open-source software in 2009.Bitcoins are created as a reward for a process known as mining. They can be exchanged for other currencies, products, and services. Research produced by University of Cambridge estimates that in 2017, there were 2.9 to 5.8 million unique users using a cryptocurrency wallet, most of them using bitcoin.

**1.2 Objectives of Research**

The main objective of this project is to describe

What are the causes of the sudden spikes and dips in cryptocurrency values? Are the markets for different altcoins inseparably linked or largely independent? How can we predict what will happen next?

**1.3 Problem Statement**

The projects main theme is to analyse the given cryptocurrency data and Fluctuations in market capitalization and total supply due to the other features.

To find the Input features that effect our market capitalization and total supply and to choose the best combination or model that effects the output features more.

**2. Review of literature**

In the existing model it would be difficult to predict what causes the fluctuations in market capitalization and total supply. The only why to predict them would be based on considering the previous inputs and estimating the required results manually or from any expert’s advice.

But Machine Learning through python made our work easy by predicting the estimated score and input features that effect the market capitalization and total supply through different models which exists in python as built-in functions in required packages.

In this model we have used regression algorithms (Linear regression and multiple linear regression for two, there, four, five features w.r.t market capitalization and total supply).

The result has predicted that percentage change in 1 hour, 24 hours and price of bitcoin effected the market capitalization as the accuracy of the input features w.r.t to output features is 99.99% and also 24 hours volume in usd, available supply, price usd effected the total supply, the accuracy value is 99.99%.

**3. Data Collection**

The data is collected from coinmarketcap.com which describes about 2018 bitcoin market capitalization values.

The input features are

1. **24 hours volume in usd**: The amount of the coin that has been traded in the last 24 hours. It is the total volume of buys or sells (orders that are filled whether it is buy order or sell order) and for bitcoin or altcoins since the market is 24/7 you add back to 24 hours.
2. **Available Supply:** Available supply or Circulating supply is the amount of coin available at the present time and circulating in the market.
3. **Last updated:** Last updated time duration of data.
4. **Percentage change in 1 hour:** Percentage change in 1 hour is a simple mathematical concept that represents the degree of change over time w.r.t dollar rate and market capitalization.
5. **Percentage change in 24 hours:** Percentage change in 24 hour is a simple mathematical concept that represents the degree of change over time w.r.t dollar rate and market capitalization.
6. **Percentage change in 7 days:** Percentage change in 7 days is a simple mathematical concept that represents the degree of change over time w.r.t dollar rate and market capitalization.
7. **Price of bitcoin:** Bitcoin's price is also quite dependent on the size of its mining network, since the larger the network is, the more difficult – and thus more costly – it is to produce new bitcoins.
8. **Price of usd:** Price of US dollar.

The output features are :

1. **Market Capitalization:** Current price of a coin times the total number of coins in the market, often referred to as circulating supply.
2. **Total supply:** Total supply is the amount of coin that is already in existence. However, not all of the coins may be circulating at the present time. Generally total supply is equal to or greater than the circulating supply.

**4. Methodology**

**4.1 Exploratory Data Analysis**

Exploratory Data Analysis (EDA) is an approach to analysing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modelling or hypothesis testing task.

**4.1.1 Figures and Tables**

There are many built-in functions in python to analyse the data. Some are mentioned below

**describe()** :

To view some basic statistical details like percentile, mean, std etc.

**Syntax:**

describe(percentiles = None, include = None, exclude = None)

**percentiles :** The percentile to include in the output. Should range should be 0 or 1.

**Include :** list of datatypes to be included while describing DataFrame. Default is None.

**Exclude :** list of datatypes to be excluded while describing DataFrame. Default is None.

Return type : Statistical summary of DataFrame.

**Count :** Total count of rows or observation.

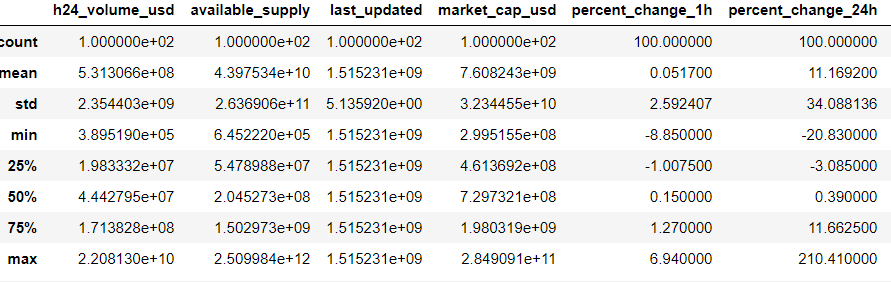
There are 100 rows in our dataset so in our case the count of each feature is 100

**Mean :** Ratio of collection of sum of values to the total number of values.

Each feature has a particular Mean which is “The total sum of the values of a feature to the total number of rows”

**Standard Deviation:** A quantity expressing by how much difference is in between members of group and mean value.

The value is always positive as it is squared up when calculated.



**Covariance:**

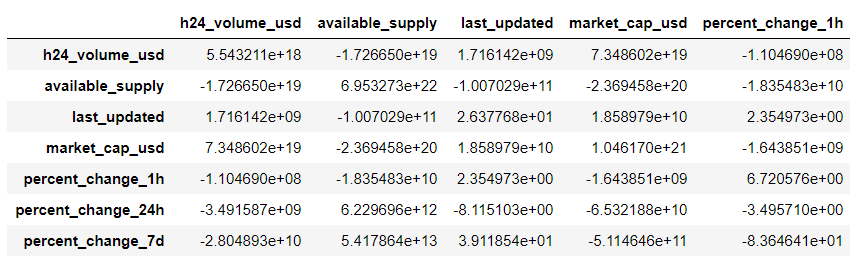
It is the measure of relationship between two random variables. The metric evaluates, how much, to what extent the value change together.

**Positive Covariance:**

It indicates that two variables tend to move in same direction.

**Negative Covariance:**

It indicates that two variables tend to move in inverse direction.

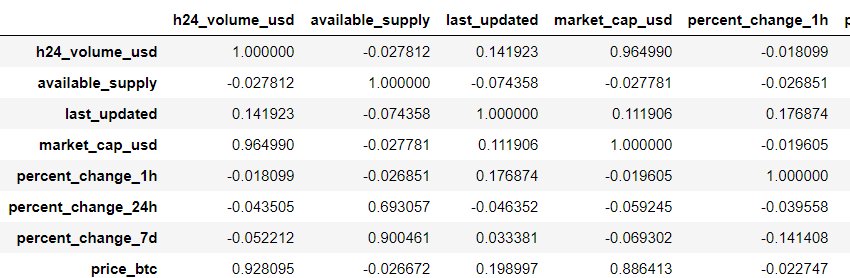


There exist a positive covariance between market capacity usd and total supply. The value is -9.531640e+20. It means they tend to move towards eachother

There exist a negative covariance between total\_supply and available supply 1.628830e+23. It means they tend to move in inverse direction.

**Correlation:**

It measures the strength of relationship between variables.



There exist a strong correlation between price of bitcoin and price of usd.

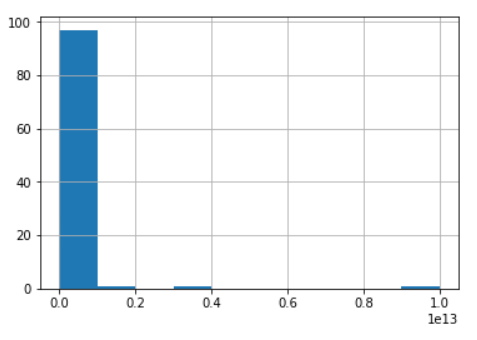
There is strong correlation between 24h\_volume\_supply, percentage change in 24 hours and percentage change in 7days and Market capitalization i.e., 0.964 , 0.69 ,0.9

There

**Hist()**

The Hist() automatically generates the histograms and returns the bit counts or probabilities.

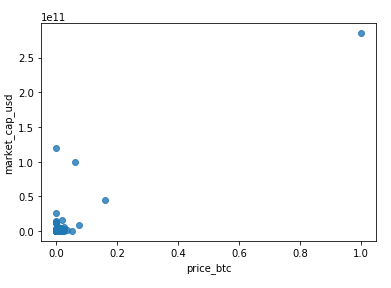
Here we take bins=10 and the market\_cap\_usd for first bin means for first 10 values is higher and for remaining it is decreased.



**Matplotlib scatterplot:**

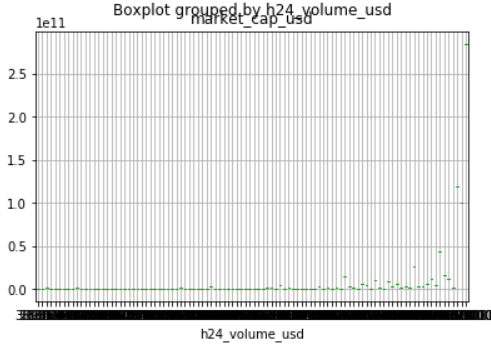
Matplot has a built-in function to create scatterplots called scatter(). A scatter plot is a type of plot that shows the data as a collection of points. The position of a point depends on its two-dimensional value, where each value is a position on either the horizontal or vertical dimension.

Here on X-axis we have taken price\_btc and market\_cap\_usd on y-axis.As the price-btc values are increasing the market\_cap\_usd is also increasing so we can say that price\_btc is directly proportional to market\_cap\_usd.



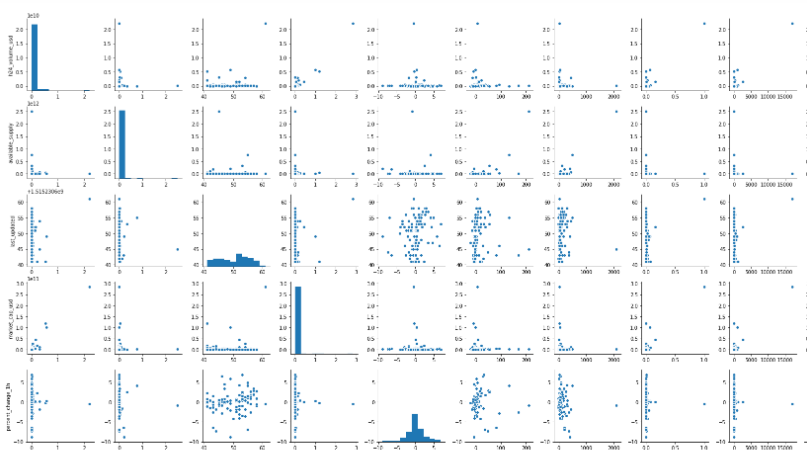
**Boxplot()**:  This method for graphically depicting groups of numerical data through their quartiles.

Here on X-axis we have taken h24\_volume\_usd and market\_cap\_usd on y-axis.As the h24\_volume\_usd values are increasing the market\_cap\_usd is also increasing so we can say that h24\_volume\_usd is directly proportional to market\_cap\_usd.



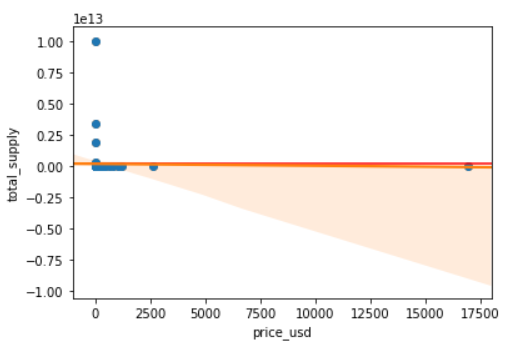
**Pairplot():** This function will create a grid of Axes such that each variable in data will by shared in the y-axis across a single row and in the x-axis across a single column.

This plot gives the graph of every individual with every other individual variables.



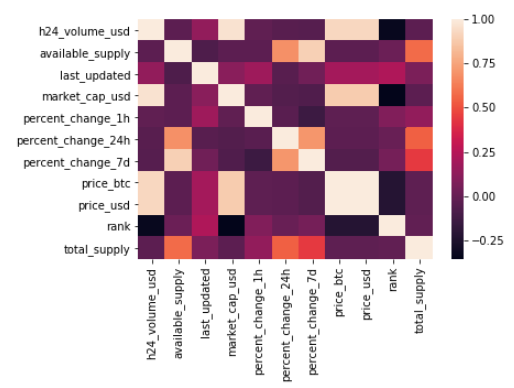
**Regplot():** This is a axis level function.

This is contains the scatter plot along with mean.Here we can say that whether the values are above the mean or below.



**Heatmap():** A heat map is a two-dimensional representation of data in which values are represented by colors.

Here in the plot every individual variable is mapped to every other variables.It gives the color representation of the plot.



**4.2 Data Modelling**

One of the best ways for modelling this data is by using regression algorithm as the output is in continuous values.

**Regression Algorithm :**

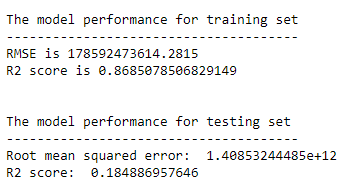
Regression searches for relationships among variables. Regression is classified into two types:

**Simple Linear Regression**

Simple linear regression is an approach for predicting a response using a single feature.

After applying the regression the following is the output

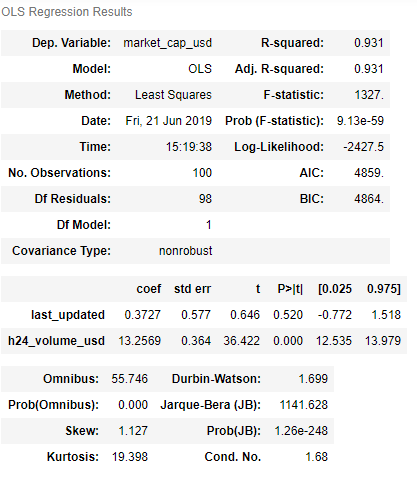
Accuracy of total\_supply and available supply is:86.65%



**Multiple Linear Regression**

Multiple linear regression attempts to model the relationship between two or more features and a response by fitting a linear equation to observed data.

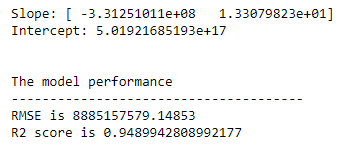
**Summary()**: It gives us the short document or section of a document of the data.



**Accuracy()**:It refers to the closeness of a measured value to a standard or Known value.

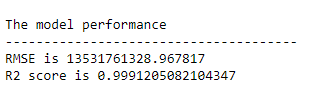
Accuracy of the last\_updated, h24\_volume\_usd and market\_cap\_usd: 94.89%

The RMSE and R2 score of the model is:



The accuracy of h24\_volume\_usd, available\_supply, price\_usd  
and total\_supply is :99.9%

The RMSE and R2 score values of the model is:



**5. Findings and Suggestions**

Acquired data from coinmarketcap.com

Refered the following about bitcoin :

<https://www.investopedia.com/terms/b/bitcoin.asp>

<https://www.investopedia.com/terms/p/percentage-change.asp>

<https://steemit.com/cryptocurrency/@devko100x/understanding-coin-supply-of-cryptocurrency>

The following model is analysed on Cryptocurrency rates of 2018.

**6. Conclusion**

The cryptocurrencies require only an internet connection and are not dependent on established institutions such as banks, they are ideally suited for society without a well-developed financial infrastructure.

Many individual markets skipped over landlines and went to straight for mobile phones, the same individuals may skip the overhead of traditional banking system and engage directly in mobile banking. For the reasons, we expect cryptocurrencies to become a major influence in emerging markets over next 3-5 years.

From our project analysis, we came to know about how various parameters are effecting the cryptocurrency market value.

We have got high accuracy on implementing multiple linear regression on input values 24h\_volume\_usd, available supply, price\_usd with the output variable total\_supply with 99.9% .

Our observation gave an overall view of our prediction and dependencies by which we can predict why currencies are differed every time. This result gave us a basic overview of what is cryptocurrency, why it differs which parameters are main reason for this change.