**PAWAN BHADANA**

2002pawanbhadana@gmail.com

**CLASSIFICATION NOTEBOOK PREDICTING PRICE MOVEMENTS( NEXT HOUR PRICE INCREASE OR DECREASE** [**https://colab.research.google.com/drive/1tDpHQuYo7n1n4PggimbOvrXP\_hISV95f?usp=sharing**](https://colab.research.google.com/drive/1tDpHQuYo7n1n4PggimbOvrXP_hISV95f?usp=sharing)

**REGRESSION NOTEBOOK PREDICTING THE CLOSING PRICE OF NEXT 30 HOURS USINGLSTM:**[**https://colab.research.google.com/drive/1SUey6c9MSfe\_Cl0\_QrzRB6CbaSA7wjz8?usp=sharing**](https://colab.research.google.com/drive/1SUey6c9MSfe_Cl0_QrzRB6CbaSA7wjz8?usp=sharing)

**1. Introduction**

**Purpose of the Analysis**

This report presents an analysis of the BTCINR price movements using historical Kline (candlestick) data. The primary goal is to uncover trends, compute technical indicators, and develop a predictive model for future price forecasting. Additionally, the backtesting strategy is implemented to validate the predictions.

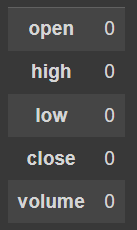
**Dataset Used**

The data consists of open, high, low, close prices, and volume for Bitcoin (BTC) against the Indian Rupee (INR) obtained from the Pi42 exchange API. The dataset spans from July 2024 to September 2024, with hourly intervals.

**2. Exploratory Data Analysis (EDA):**

**Checking Null Values**

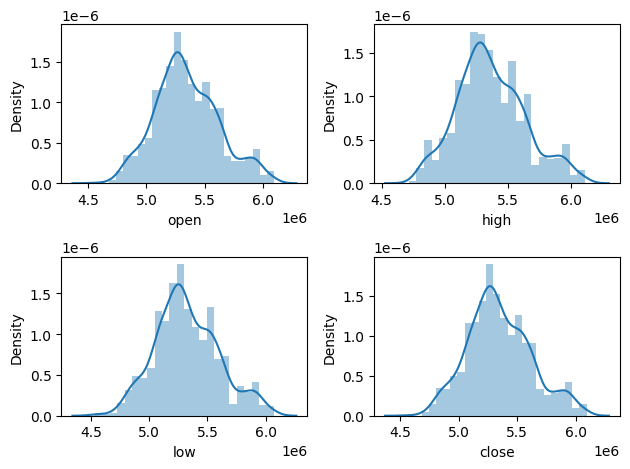
No null values found in data

.

**Checking the distribution of variables:**

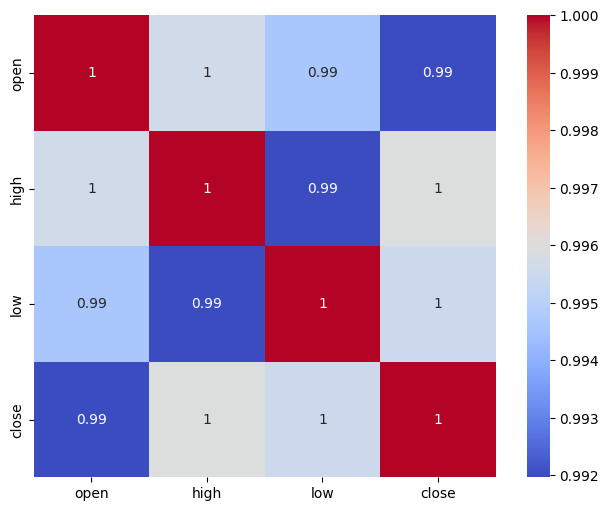
I made a histplot of each variable to see they are distributed.

After observing the graphs, It looks like they follow **the gaussian distribution**.



**Checking the Correlation Among Price variables:**

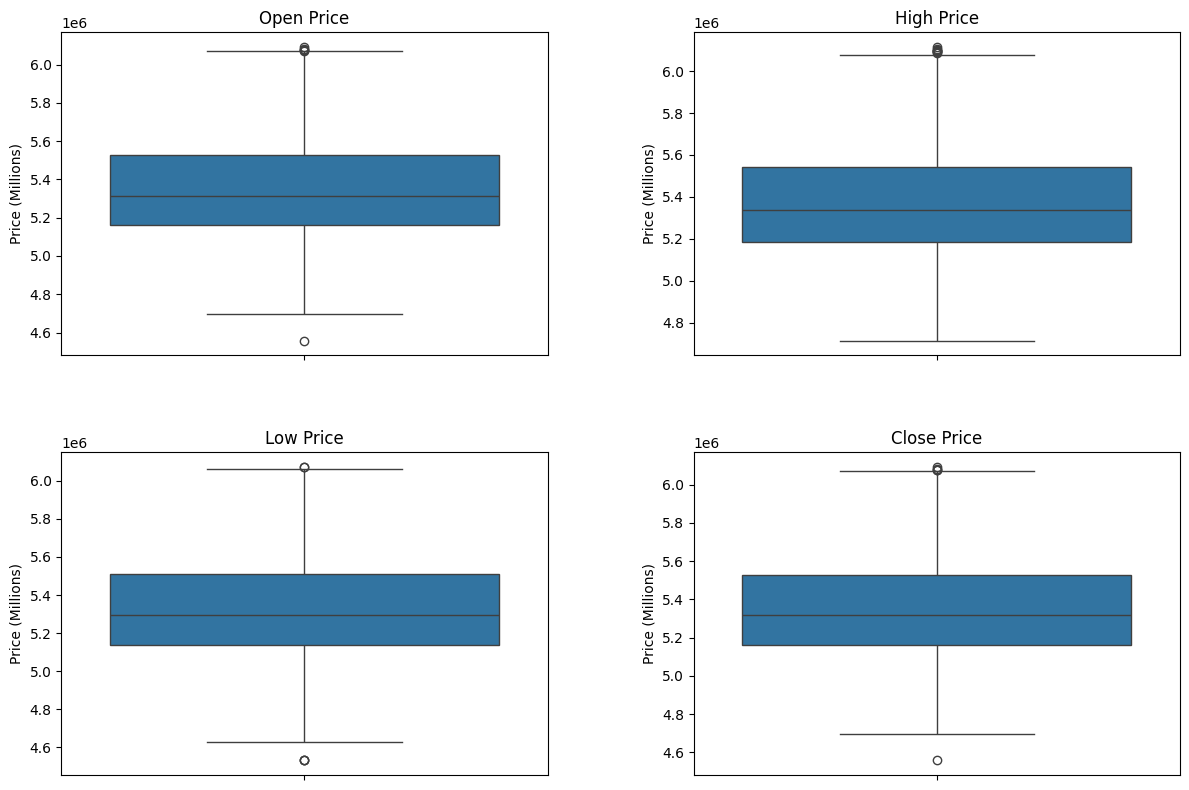
After seeing the distribution curves of all prices, they all look similar so checking if the variables are correlated or not. Because they seem correlated.



After observing the heatmap , I made the decision that all these 4 price variables are highly correlated , so for predicting one price , other prices would not do any good in prediction .SO decided not to include others in predictors of that price.

**Checking Outliers**

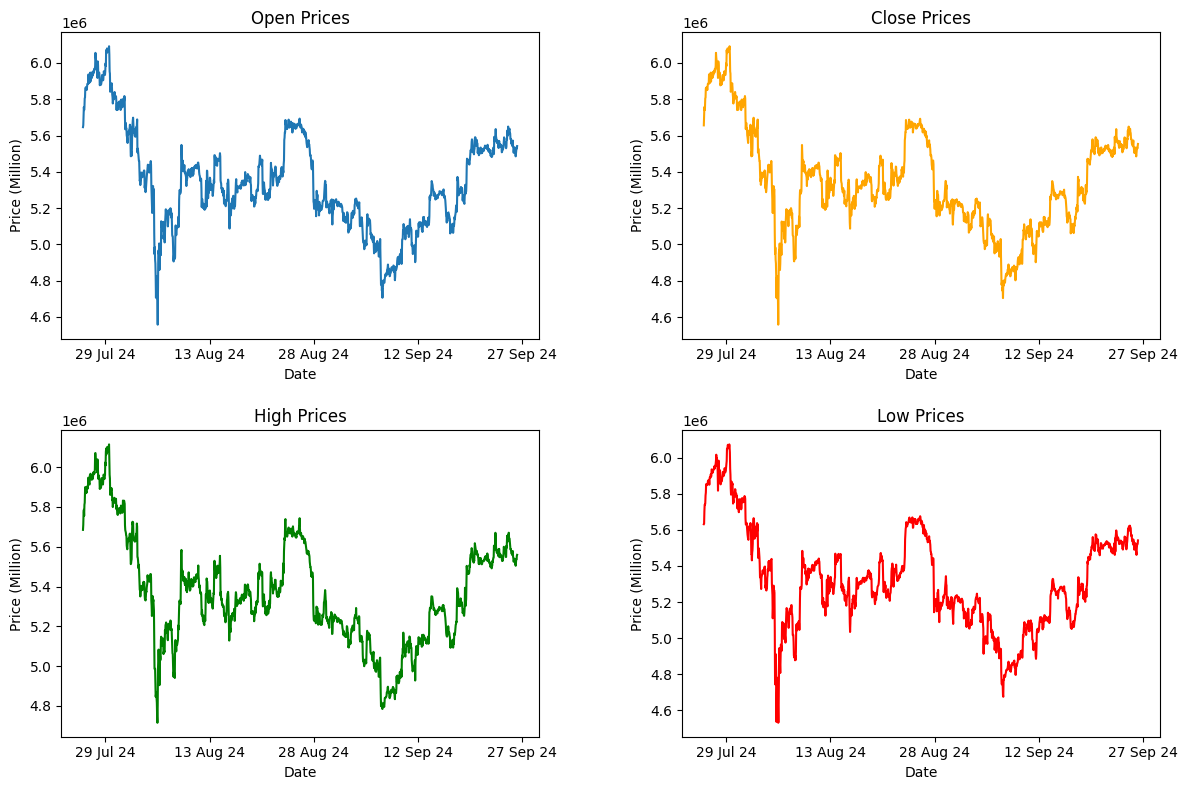
Plotted the boxplot to see if there are outliers



After seeing these graphs we can see there are few outliers in the dataset, but **removing these is not a good idea as this is a time series data , we can do one thing that reducing the impact of outliers and can set the larger value of outliers equal to 0.95 quantile and smaller values to 0.05.**

Price Movements

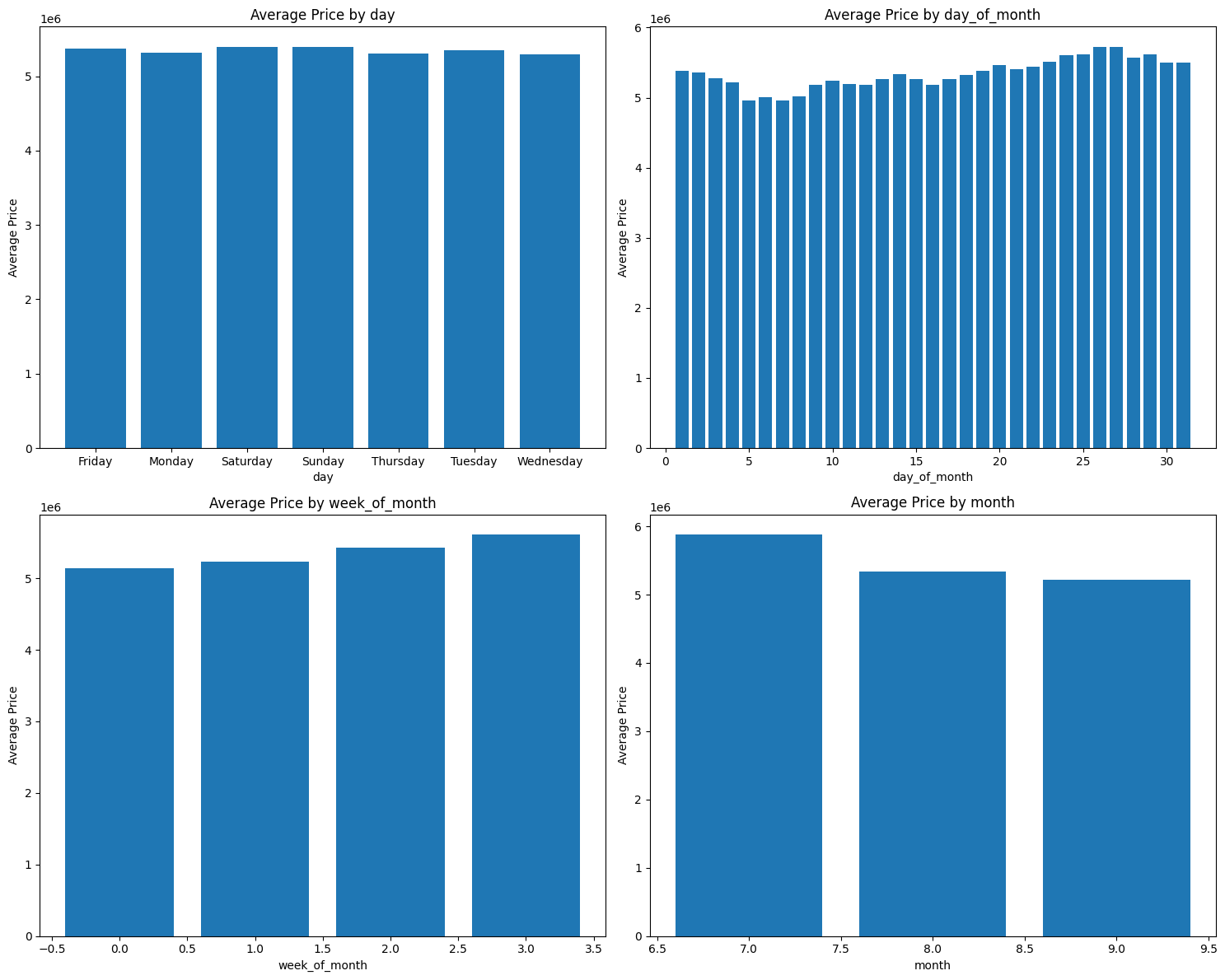
I visualised the BTCINR price movements using open, close, high, and low prices. Here's the graph illustrating these movements:



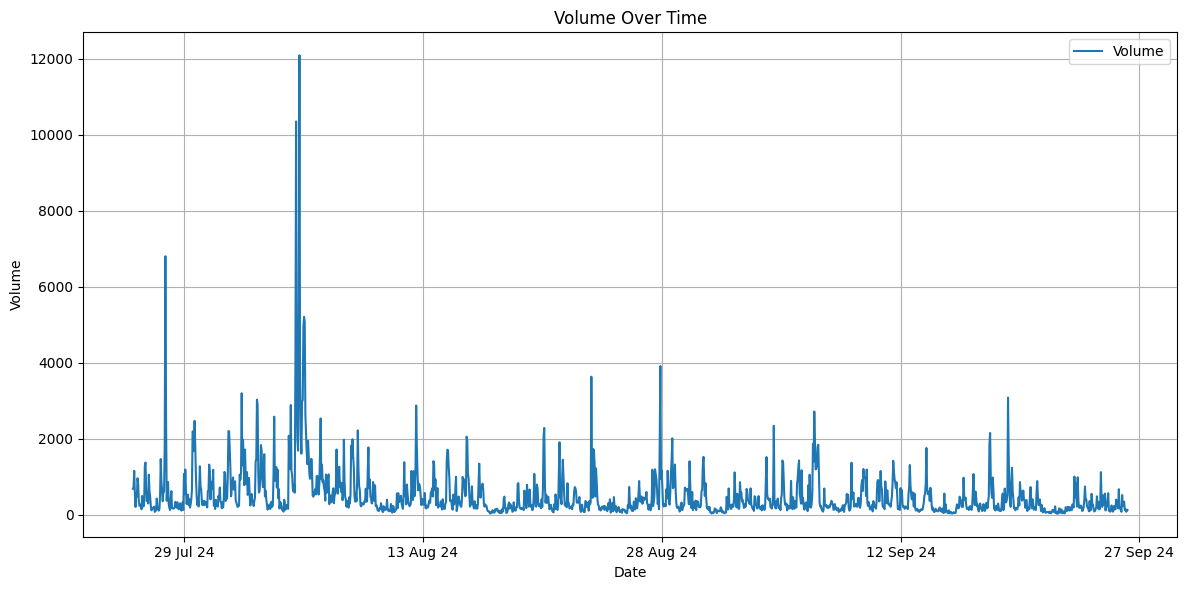
After observing these maps it looks that **1st Week of Aug 2024 was a very bad day for trading, prices went down drastically and had not reached the same level as at the end of July till the end september. Maybe some international issues or some bad events caused that fall.**

**ALSO VISUALISED THE HOW AVERAGE OPENING PRICE DAY , WEEK , MONTH WISE**

After analysing these trends one observation can be made that as week progresses the opening price is increasing in this period of time (graphs below)

****

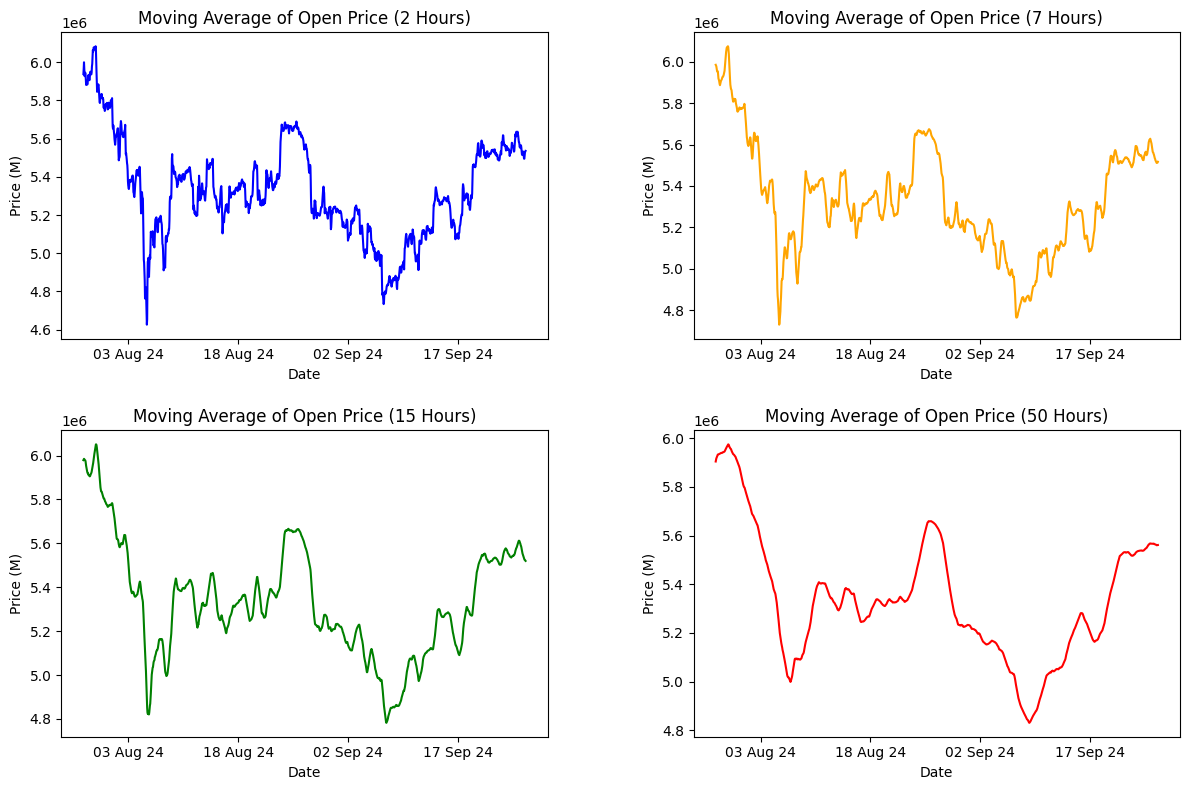
**Volume Analysis:**

I also analysed the trading volume.

From seeing this curve , we can see that the **market went up high after the 1st Week of Aug 2024.(around 7-8 Aug)**

**Moving Averages (MA):**

Calculated short-terms(2,7,15 hours) and long-term(50 hours) simple moving averages to smooth out price fluctuations and identify trends.

****

**4. Predictive Modeling:**

Made 2 kinds of models, one for predicting the price movements,other for predicting the future prices.

Classification Task: Predicting the future direction of price movements (increase or decrease).

Regression Task: Predicting future price levels for opening and closing prices of next 30-hours using LSTM.

**Notebook 1: Price Movement Classification**

[**https://colab.research.google.com/drive/1tDpHQuYo7n1n4PggimbOvrXP\_hISV95f?usp=sharing**](https://colab.research.google.com/drive/1tDpHQuYo7n1n4PggimbOvrXP_hISV95f?usp=sharing)

**Approach**

In the first notebook, I treated the problem as a classification task, where I predicted whether the prices (open, close, high, low) would increase or decrease in future intervals. The models used were:

* RandomForestClassifier
* XGBoostClassifier

**SCALER:** STANDARD **SCALER (AS THE DIS**TRIBUTION FOLLOWS THE GAUSSIAN)

**Models and Hyperparameter Tuning:**

Both models were optimised using GridSearchCV, testing a variety of parameters:

* RandomForestClassifier: n\_estimators, max\_depth, min\_samples\_split, etc.
* XGBoostClassifier: learning\_rate, max\_depth, n\_estimators, etc.

**EVALUATION METRIC: PRECISION SCORE**

**Used backtesting for further evaluation**

GOT THE BACKTESTING SCORE FOR LOW PRICE : 0.6052631578947368

**Notebook 2 : Price Forecasting Using LSTM:**

[**https://colab.research.google.com/drive/1SUey6c9MSfe\_Cl0\_QrzRB6CbaSA7wjz8?usp=sharing**](https://colab.research.google.com/drive/1SUey6c9MSfe_Cl0_QrzRB6CbaSA7wjz8?usp=sharing)

Here I treated the problem as a regression task by predicting the exact future prices (open and close) for the next 30 hours using an LSTM (Long Short-Term Memory) model.

**Approach**

Here I take the last 100 hours of prices as the dependent variables and based on this predict the price of the 101th hour.

**SCALER**: STANDARD **SCALER**

MODEL: LSTM

EVALUATION METRIC: PRECISION SCORE

Used backtesting for further evaluation

#### **Improvements** for future work**:**

Additional Features: Incorporating external market data (e.g., market sentiment analysis, social media trends, or volume-related features) could help the model better understand price movements.