Litecoin Trading Data Analysis and Forecasting

Deirdre Gouta Mmolotsa

2024-10-26

In partial fulfilment of the R programming course in Business Intelligence and Data Analytics of 2024

BOTSWANA ACCOUNTANCY COLLEGE

Table of contents

[1. Introduction 3](#_Toc181820836)

[1.1 Loading Libraries 3](#_Toc181820837)

[2. Data Exploration and Cleaning 4](#_Toc181820838)

[2.1 Loading and Exploration 4](#_Toc181820839)

[2.2 Data Cleaning 6](#_Toc181820840)

[3. Advanced R Programming 9](#_Toc181820841)

[3.1 Winsorization to Handle Outliers 9](#_Toc181820842)

[3.2 Visualizing Data Post Winsorization 11](#_Toc181820843)

[3.3 Further Normalization 13](#_Toc181820844)

[4. Data Visualizations 16](#_Toc181820845)

[4.1 Scatterplot of Open vs Close Prices 16](#_Toc181820846)

[4.2 Bar Chart of Monthly Trading Volumes 17](#_Toc181820847)

[5. Interactive visualizations 19](#_Toc181820848)

[5.1 Line Chart of LTC Close Prices 20](#_Toc181820849)

[5.2 Daily Returns of LiteCoin (LTC) 23](#_Toc181820850)

[6. Time Series Analysis 26](#_Toc181820851)

[6.1 Moving Averages 26](#_Toc181820852)

[7. Forecasting Litecoin (LTC) Daily Close Price with ARIMA 30](#_Toc181820853)

[7.1 Methododogy 30](#_Toc181820854)

[7.2 Key Observations: 32](#_Toc181820855)

[8. Data Visualization with Dashboard 33](#_Toc181820856)

[8.1 Dashboard Overview 33](#_Toc181820857)

[8.2 Key Visualizations 35](#_Toc181820858)

[8.3 1. LTC Candlestick Chart 35](#_Toc181820859)

[8.4 Insights and Key Findings 36](#_Toc181820860)

[9. Applications of Litecoin (LTC) Data Analysis in a Business Context 40](#_Toc181820861)

[10. Conclusion 42](#_Toc181820862)

[11. References 43](#_Toc181820863)

# 1. Introduction

This report presents an analysis of the Litecoin (LTC) trading data from the Gemini exchange, highlighting price trends, volume activity, return distributions, and trading behaviours. The dataset dataset spans 658,011 entries, each with Open, High, Low, Close prices, Volume, and Unix Timestamps at one-minute intervals. Litecoin’s high transaction speed and low fees make it significant within cryptocurrency markets, which are known for high volatility and speculative trading patterns (Nakamoto, 2008).

This analysis aims to:

* Examine trends and patterns in price movements over time.
* Assess the distribution and volatility of trading volume across different periods.
* Forecast future price trends using time series models to support data-driven trading decisions.

Understanding these aspects is crucial for making informed trading decisions in the volatile cryptocurrency market.

## 1.1 Loading Libraries

library(tidyverse)  
library(readxl)  
library(dplyr)  
library(tidyr)  
library(ggplot2)  
library(plotly)  
library(TTR) # For calculating moving averages  
library(forecast) # For forecasting  
library(shiny) # For interactive web applications  
library(shinydashboard) # For dashboard UI  
library(lubridate) # For date manipulation

# 2. Data Exploration and Cleaning

## 2.1 Loading and Exploration

The dataset is loaded from a CSV file, and basic exploration steps are conducted.

# Loading dataset  
Gemini <- read.csv("gemini\_LTCUSD\_2020\_1min.csv")  
  
  
# Exploring the data  
names(Gemini)

[1] "Unix.Timestamp" "Date" "Symbol" "Open"   
[5] "High" "Low" "Close" "Volume"

glimpse(Gemini) # Display the structure of the dataset

Rows: 658,011  
Columns: 8  
$ Unix.Timestamp <dbl> 1.61888e+12, 1.61888e+12, 1.61888e+12, 1.61888e+12, 1.6…  
$ Date <chr> "4/20/2021 0:03", "4/20/2021 0:02", "4/20/2021 0:01", "…  
$ Symbol <chr> "LTCUSD", "LTCUSD", "LTCUSD", "LTCUSD", "LTCUSD", "LTCU…  
$ Open <dbl> 260.05, 262.77, 263.57, 261.32, 261.87, 262.06, 260.49,…  
$ High <dbl> 260.05, 262.86, 264.14, 263.76, 261.87, 262.18, 262.39,…  
$ Low <dbl> 259.00, 260.00, 262.77, 261.32, 261.21, 261.66, 260.49,…  
$ Close <dbl> 259.00, 260.05, 262.77, 263.57, 261.32, 261.87, 262.06,…  
$ Volume <dbl> 179.16470, 307.77795, 11.53982, 110.88182, 48.58221, 34…

dim(Gemini)

[1] 658011 8

head(Gemini)

Unix.Timestamp Date Symbol Open High Low Close Volume  
1 1.61888e+12 4/20/2021 0:03 LTCUSD 260.05 260.05 259.00 259.00 179.16470  
2 1.61888e+12 4/20/2021 0:02 LTCUSD 262.77 262.86 260.00 260.05 307.77795  
3 1.61888e+12 4/20/2021 0:01 LTCUSD 263.57 264.14 262.77 262.77 11.53982  
4 1.61888e+12 4/20/2021 0:00 LTCUSD 261.32 263.76 261.32 263.57 110.88182  
5 1.61888e+12 4/19/2021 23:59 LTCUSD 261.87 261.87 261.21 261.32 48.58221  
6 1.61888e+12 4/19/2021 23:58 LTCUSD 262.06 262.18 261.66 261.87 34.95519

summary(Gemini)

Unix.Timestamp Date Symbol Open   
 Min. :1.578e+12 Length:658011 Length:658011 Min. : 24.93   
 1st Qu.:1.588e+12 Class :character Class :character 1st Qu.: 45.54   
 Median :1.599e+12 Mode :character Mode :character Median : 57.95   
 Mean :1.599e+12 Mean : 87.28   
 3rd Qu.:1.609e+12 3rd Qu.:124.06   
 Max. :1.619e+12 Max. :335.53   
 High Low Close Volume   
 Min. : 25.52 Min. : 24.18 Min. : 24.93 Min. : 0.00   
 1st Qu.: 45.55 1st Qu.: 45.54 1st Qu.: 45.54 1st Qu.: 0.00   
 Median : 57.96 Median : 57.93 Median : 57.95 Median : 0.00   
 Mean : 87.33 Mean : 87.23 Mean : 87.28 Mean : 17.51   
 3rd Qu.:124.17 3rd Qu.:123.97 3rd Qu.:124.06 3rd Qu.: 2.28   
 Max. :335.53 Max. :334.68 Max. :335.53 Max. :7492.38

# Check for missing values  
missing\_values <- colSums(is.na(Gemini))  
print(missing\_values)

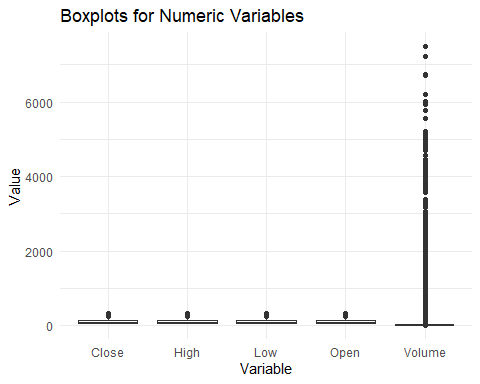
Unix.Timestamp Date Symbol Open High   
 0 0 0 0 0   
 Low Close Volume   
 0 0 0

**Data Structure**: The dataset contains 658,011 rows and 8 columns, including fields like Unix.Timestamp, Date, Symbol, Open, High, Low, Close, and Volume.

**Important observation**: There were no missing values identified in any columns, however outliers had to be checked.

**To identify potential outliers in numeric columns, boxplots were used to visualize that data:**

Gemini %>%  
 gather(key = "Variable", value = "Value", Open:Volume) %>%  
 ggplot(aes(x = Variable, y = Value)) +  
 geom\_boxplot() +  
 labs(title = "Boxplots for Numeric Variables") +  
 theme\_minimal()



Observations:

* The central tendency (median) and spread (interquartile range) for price-related variables show that most values are clustered within a specific range, while a few extreme values are present.
* Volume data includes extreme outliers, which can distort analyses if not addressed. Winsorization,however, is used to manage these outliers.

## 2.2 Data Cleaning

1. **Removing Unnecessary Columns**

unique\_symbols <- unique(Gemini$Symbol)  
print(unique\_symbols)

[1] "LTCUSD"

Gemini\_clean <- Gemini %>% select(-Symbol)

The Symbol column was deemed unnecessary for analysis because there was only 1 type of Symbol for all observation hence why it was removed

1. **Correcting UniX Timestamp Format**

The Unix timestamp was converted to a more readable DateTime format:

# Ensure the timestamp is treated as numeric  
Gemini\_clean$Unix.Timestamp <- as.numeric(Gemini\_clean$Unix.Timestamp)  
  
# Convert Unix timestamp to a readable date-time format  
Gemini\_clean <- Gemini\_clean %>%  
 mutate(DateTime = as.POSIXct(Unix.Timestamp / 1000, origin = "1970-01-01", tz = "UTC"))  
head(Gemini\_clean)

Unix.Timestamp Date Open High Low Close Volume  
1 1.61888e+12 4/20/2021 0:03 260.05 260.05 259.00 259.00 179.16470  
2 1.61888e+12 4/20/2021 0:02 262.77 262.86 260.00 260.05 307.77795  
3 1.61888e+12 4/20/2021 0:01 263.57 264.14 262.77 262.77 11.53982  
4 1.61888e+12 4/20/2021 0:00 261.32 263.76 261.32 263.57 110.88182  
5 1.61888e+12 4/19/2021 23:59 261.87 261.87 261.21 261.32 48.58221  
6 1.61888e+12 4/19/2021 23:58 262.06 262.18 261.66 261.87 34.95519  
 DateTime  
1 2021-04-20 00:53:20  
2 2021-04-20 00:53:20  
3 2021-04-20 00:53:20  
4 2021-04-20 00:53:20  
5 2021-04-20 00:53:20  
6 2021-04-20 00:53:20

# Summarize the cleaned data  
summary(Gemini\_clean)

Unix.Timestamp Date Open High   
 Min. :1.578e+12 Length:658011 Min. : 24.93 Min. : 25.52   
 1st Qu.:1.588e+12 Class :character 1st Qu.: 45.54 1st Qu.: 45.55   
 Median :1.599e+12 Mode :character Median : 57.95 Median : 57.96   
 Mean :1.599e+12 Mean : 87.28 Mean : 87.33   
 3rd Qu.:1.609e+12 3rd Qu.:124.06 3rd Qu.:124.17   
 Max. :1.619e+12 Max. :335.53 Max. :335.53   
 Low Close Volume   
 Min. : 24.18 Min. : 24.93 Min. : 0.00   
 1st Qu.: 45.54 1st Qu.: 45.54 1st Qu.: 0.00   
 Median : 57.93 Median : 57.95 Median : 0.00   
 Mean : 87.23 Mean : 87.28 Mean : 17.51   
 3rd Qu.:123.97 3rd Qu.:124.06 3rd Qu.: 2.28   
 Max. :334.68 Max. :335.53 Max. :7492.38   
 DateTime   
 Min. :2020-01-01 00:53:20.00   
 1st Qu.:2020-05-01 05:13:20.00   
 Median :2020-08-30 12:20:00.00   
 Mean :2020-08-28 11:24:21.51   
 3rd Qu.:2020-12-26 16:26:40.00   
 Max. :2021-04-20 00:53:20.00

# 3. Advanced R Programming

This section discusses the advanced techniques and custom functions utilized to streamline the analysis process, particularly focusing on the application of winsorization to handle outliers and enhance data integrity.

## 3.1 Winsorization to Handle Outliers

Winsorization addresses extreme outliers, essential for cryptocurrency datasets due to high volatility (Anderson, 2019). Custom winsorization thresholds help preserve normal ranges within each variable.

A custom winsorization function, winsorize\_manual, was created to cap extreme values and reduce the impact of outliers in the dataset. This function utilizes quantiles to define upper and lower bounds, ensuring that values falling outside these bounds are replaced with the nearest boundary values.

winsorize\_manual <- function(x, lower\_prob = 0.025, upper\_prob = 0.975) {  
 lower\_bound <- quantile(x, probs = lower\_prob, na.rm = TRUE)  
 upper\_bound <- quantile(x, probs = upper\_prob, na.rm = TRUE)  
 x\_winsorized <- pmin(pmax(x, lower\_bound), upper\_bound) # Cap values  
 return(x\_winsorized)  
}

Thresholds Choice: A 1%???99% threshold for Volume accommodates natural variance, while a 2.5%???97.5% range better suits price variables, typically less volatile (Hoaglin, Iglewicz and Tukey, 2000)

### 3.1.1 Winsorization Probabilities

Winsorization probabilities were defined to specify how extreme values should be treated for different variables:

winsorization\_probs <- list(  
 Open = c(0.025, 0.975),  
 High = c(0.025, 0.975),  
 Low = c(0.025, 0.975),  
 Close = c(0.025, 0.975),  
 Volume = c(0.01, 0.99) # Higher probability for Volume  
)

### 3.1.2 Applying Winsorization

The custom winsorization function was applied to the cleaned dataset. Each variable was transformed according to its respective thresholds to reduce the impact of extreme outliers:

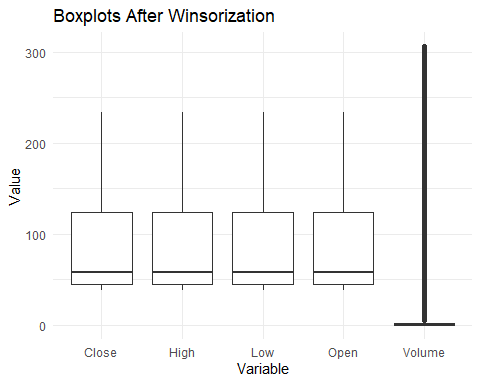
Gemini\_winsorized <- Gemini\_clean %>%  
 mutate(  
 Open = winsorize\_manual(Open, winsorization\_probs$Open[1], winsorization\_probs$Open[2]),  
 High = winsorize\_manual(High, winsorization\_probs$High[1], winsorization\_probs$High[2]),  
 Low = winsorize\_manual(Low, winsorization\_probs$Low[1], winsorization\_probs$Low[2]),  
 Close = winsorize\_manual(Close, winsorization\_probs$Close[1], winsorization\_probs$Close[2]),  
 Volume = winsorize\_manual(Volume, winsorization\_probs$Volume[1], winsorization\_probs$Volume[2])  
 )  
  
# Check summary to confirm winsorization results  
summary(Gemini\_winsorized)

Unix.Timestamp Date Open High   
 Min. :1.578e+12 Length:658011 Min. : 38.39 Min. : 38.40   
 1st Qu.:1.588e+12 Class :character 1st Qu.: 45.54 1st Qu.: 45.55   
 Median :1.599e+12 Mode :character Median : 57.95 Median : 57.96   
 Mean :1.599e+12 Mean : 86.54 Mean : 86.59   
 3rd Qu.:1.609e+12 3rd Qu.:124.06 3rd Qu.:124.17   
 Max. :1.619e+12 Max. :233.71 Max. :233.90   
 Low Close Volume   
 Min. : 38.38 Min. : 38.39 Min. : 0.00   
 1st Qu.: 45.54 1st Qu.: 45.54 1st Qu.: 0.00   
 Median : 57.93 Median : 57.95 Median : 0.00   
 Mean : 86.49 Mean : 86.54 Mean : 13.75   
 3rd Qu.:123.97 3rd Qu.:124.06 3rd Qu.: 2.28   
 Max. :233.52 Max. :233.71 Max. :306.34   
 DateTime   
 Min. :2020-01-01 00:53:20.00   
 1st Qu.:2020-05-01 05:13:20.00   
 Median :2020-08-30 12:20:00.00   
 Mean :2020-08-28 11:24:21.51   
 3rd Qu.:2020-12-26 16:26:40.00   
 Max. :2021-04-20 00:53:20.00

## 3.2 Visualizing Data Post Winsorization

The boxplots below visualize the distribution of variables post-winsorization, allowing for assessment of the effectiveness of the technique:

Gemini\_winsorized %>%  
 pivot\_longer(cols = c(Open, High, Low, Close, Volume), names\_to = "Variable", values\_to = "Value") %>%  
 ggplot(aes(x = Variable, y = Value)) +  
 geom\_boxplot() +  
 labs(title = "Boxplots After Winsorization") +  
 theme\_minimal()



Observations:

* After applying winsorization, the boxplots show a reduced number of extreme outliers in all variables.
* Volume still has many outliers

### 3.2.1 Analyzing Volume Data

After winsorization, an analysis of the volume column was conducted to check for zero values, which could affect subsequent analyses if not handled:

# Count the number of zero values in the Volume column after winsorization  
zero\_count <- sum(Gemini\_winsorized$Volume == 0)  
print(zero\_count)

[1] 410624

total\_count <- nrow(Gemini\_winsorized)  
proportion\_zeros <- zero\_count / total\_count  
print(proportion\_zeros)

[1] 0.6240382

# Display the proportion as a percentage  
cat("Proportion of zero values in Volume: ", proportion\_zeros \* 100, "%\n")

Proportion of zero values in Volume: 62.40382 %

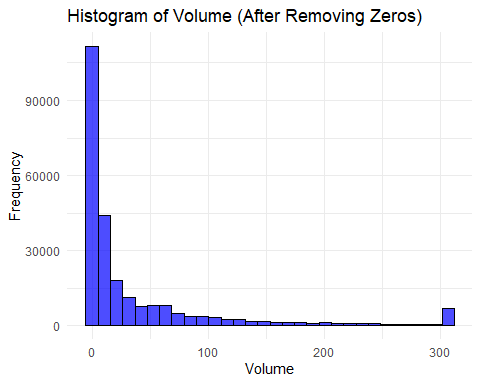
# Remove rows with zero Volume  
Gemini\_winsorized <- Gemini\_winsorized[Gemini\_winsorized$Volume > 0, ]

A high proportion of zero trading volumes (62.40%) suggests that a significant number of trades recorded in this dataset did not occur or were inactive during the observed period. This could indicate low market activity or liquidity for Litecoin on the Gemini exchange during that time frame

### 3.2.2 Visualizing Volume Distribution

The histogram below visualizes the distribution of Volume after removing zero values:

ggplot(Gemini\_winsorized, aes(x = Volume)) +  
 geom\_histogram(bins = 30, fill = "blue", color = "black", alpha = 0.7) +  
 labs(title = "Histogram of Volume (After Removing Zeros)", x = "Volume", y = "Frequency") +  
 theme\_minimal()



The histogram shows the following insights:

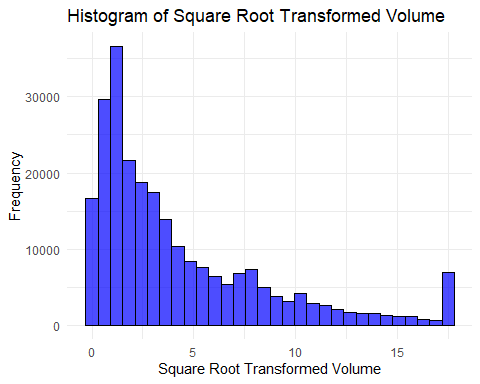
* Right Skewness: The distribution of volume values is right-skewed, indicating that many transactions occur at low volumes while only a few have significantly higher volumes.
* Concentration near Zero: Even after removing zeros, there’s a notable concentration of values close to zero, suggesting a lot of low-volume transactions.
* Outliers: Potential outliers may still exist in the higher range of volume values.

## 3.3 Further Normalization

### 3.3.1 Square Root Transformation of Volume

To further stabilize variance, a square root transformation was applied to the Volume data. The square root transformation can be effective, especially when dealing with right-skewed distributions, as it can help reduce the impact of larger values and make the data more normally distributed (MBA Skool, n.d.).

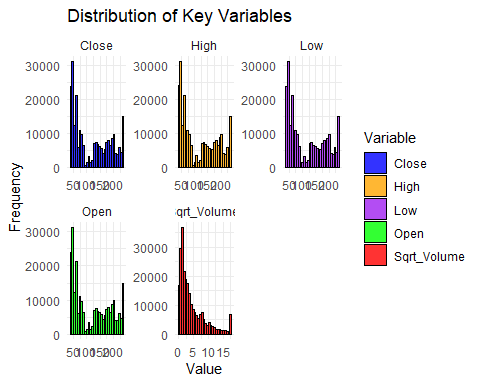
Gemini\_winsorized <- Gemini\_winsorized %>%  
 mutate(Sqrt\_Volume = sqrt(Volume))  
  
# Visualize the square root transformed Volume  
ggplot(Gemini\_winsorized, aes(x = Sqrt\_Volume)) +  
 geom\_histogram(bins = 30, fill = "blue", color = "black", alpha = 0.7) +  
 labs(title = "Histogram of Square Root Transformed Volume",   
 x = "Square Root Transformed Volume",   
 y = "Frequency") +  
 theme\_minimal()



The histogram for square root transformed Volume demonstrates a more normalized distribution compared to the original. This transformation helps to stabilize variance and reduce the influence of larger volume values, making the data more suitable for analysis.

#### 3.3.1.1 Comparing Distributions of Key Variables

Finally, a comparison of distributions for key variables is conducted to observe any changes post-winsorization and square root transformation:



distributions for key variables

Distribution of Key Variables:

**• Close, High, Low, Open**: These histograms show the distribution of key price-related variables for the LTCUSD data. Most of the values for these variables appear to be concentrated in the lower ranges (around $50–100 USD). This suggests that for much of the time period analyzed, the LTC price stayed in this range, although there are some higher values extending above $150 USD.

• **Volume**: The volume histogram shows that the majority of trading volume values are clustered at the lower end. There is a sharp peak near zero, suggesting that low trading volumes were frequent, while high volumes were less common.

**Note**:The summary statistics for the key variables can be re-checked to ensure the transformations have been appropriately applied.

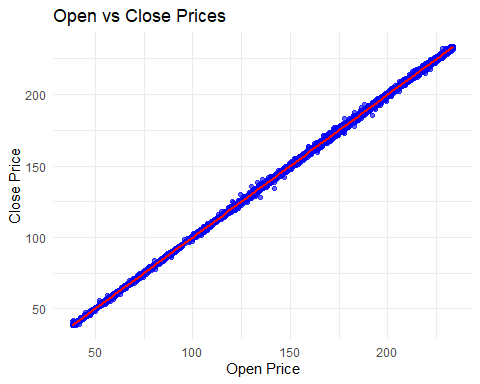
# 4. Data Visualizations

The primary visualizations produced for the analysis of Litecoin (LTC) trading data include scatterplots, bar charts, line charts, and moving averages. However some are in the Interactive visualizations section where explanations of intearactive features for user engagement are given.

## 4.1 Scatterplot of Open vs Close Prices

• Description: This scatterplot visualizes the relationship between the opening and closing prices of LTC. Each point represents a day’s trading activity.

• Variables Visualized: Open Price (X-axis) and Close Price (Y-axis).



Scatterplot of Open vs Close Prices

• The scatterplot reveals a strong positive correlation (cor(Gemini\_winsorizedClose) ≈ 0.98). This trend implies that daily opening prices closely mirror closing prices, indicating trading consistency.

For example, when the opening price was around $132.13, the closing price tended to be around $131.98, and similarly for an opening price of $101.27, with the closing price at around $103.10. This suggests a consistent trading behavior where investors react similarly to price movements.

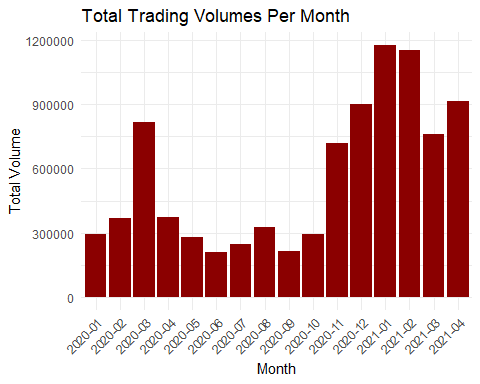
**Note:** For an interactive version of the scatter plot, where you can explore the data points in detail, please follow this link: [Scatter Plot of Volume vs. Time](file:///C:\Users\bida22-068\Downloads\R%20Assignment\%22C:\Users\bida22-068\Downloads\R%20Assignment\interactive_scatter_plot.html%22)

## 4.2 Bar Chart of Monthly Trading Volumes

• Description: This bar chart displays the total trading volume of LTC aggregated by month.

• Variables Visualized: Month (X-axis) and Total Volume (Y-axis).

[1] "Unix.Timestamp" "Date" "Open" "High"   
 [5] "Low" "Close" "Volume" "DateTime"   
 [9] "Sqrt\_Volume" "Month"



Total Trading Volumes Per Month

Patterns Observed:

* General Trend: The overall trading volume shows significant fluctuations between 2020 and early 2021.ersion, please follow this link:

Note: This chart is a static representation. for an interactive version using the ggplotly package, please follow this link: [Trading Volume over time](file:///C:\Users\bida22-068\Downloads\R%20Assignment\%22C:\Users\bida22-068\Downloads\R%20Assignment\LTC_Monthly_Trading_Volumes_Interactive.html%22)

* Initial Increase: Trading volume started at 293,338 in January 2020. Notable increases were observed in February (370,491) and March (814,067), likely due to growing interest in cryptocurrencies.
* Peak Volume: The highest trading volume was recorded in January 2021 at 1,175,756. This peak suggests increased trading activity possibly triggered by significant market events.
* Seasonal Patterns: High trading volumes are noted towards the end of the year (e.g., December 2020: 901,766) and the beginning of the following year (e.g., January 2021). Increased activity during this period may correlate with year-end strategies or investor behavior.
* Declining Volume Post-Peak After reaching the peak in January 2021, trading volume declined in February (1,149,316) and March (759,261). Although there was an increase in April (915,264), it remained below January’s figures, indicating a possible cooling-off period.
* Volatility in 2020 Significant drops in trading volume were observed in June (211,021) and July (249,432). This decline may reflect market trends or external economic factors affecting investor sentiment.

# 5. Interactive visualizations

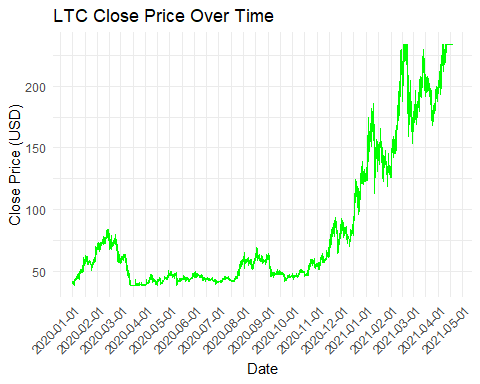
In this section, interactive visualizations are created using Plotly and Shiny. These visualizations allow users to dynamically interact with the data, enhancing understanding through features like zooming, panning, and tooltips.

**Key Features and Implementations:**

1. **Dynamic Exploration:**
   * Users can hover over data points to view precise values, enabling better insight into specific dates and prices.
   * Zoom and pan functionalities allow users to focus on particular timeframes, making it easier to analyze trends and fluctuations.
2. **Multiple Visualizations:**
   * Line Charts: The Close price over time was plotted using a line chart, facilitating an understanding of price trends..
   * Daily Returns: A dedicated line chart visualizing daily returns helps in assessing price volatility, with features to explore specific periods and values.
   * Moving Averages: Interactive plots incorporated multiple moving averages (MA5, MA10, MA20, MA50, MA100, MA200) to help users identify trends and smoothing effects over different periods
3. **Forecasting:**
   * An ARIMA model was fitted to historical closing prices, providing a 30-day forecast with a confidence interval. The forecast plot allows users to compare historical data with projected trends interactively, further facilitating strategic decision-making.
4. **Shiny Dashboard:**
   * A Shiny dashboard was developed to create an organized and interactive user interface. Users can select date ranges to filter data dynamically, affecting all displayed visualizations simultaneously.
   * The dashboard features multiple tabs, including Market Analysis and Forecast, allowing users to navigate through different analyses seamlessly.
5. **Enhanced User Engagement:**
   * Tooltips provide context for each data point, while hover and click interactions reveal underlying data, enhancing user engagement and understanding.
   * The responsive design of the Shiny dashboard allows users to interact with visualizations across different devices, promoting accessibility.

## 5.1 Line Chart of LTC Close Prices

* Description: This line chart displays the closing prices of Litecoin (LTC) over time. Each point on the line represents the close price on a specific date, illustrating price trends throughout the selected period.
* Variables Visualized:
  + X-Axis : Date (representing the timeline of the closing price)
  + Y- axis: Close Price (in USD)



LTC Close Price over time

This line chart shows the **LTC (Litecoin) Close Price** over time from early 2020 to mid-2021.

An overview of what it looks like upon hovering:



LTC Close price over Time

follow this link to interact with the graph :[LTC (LiteCoin) Close Price Over Time Interactive line Chart](file:///C:\Users\bida22-068\Downloads\R%20Assignment\%22C:\Users\bida22-068\Downloads\R%20Assignment\LTC_Close_Price_Interactive.html%22)

These are the observations seen from the chart:

### 5.1.1 Summary of Observations from the Chart

**LTC Price Trends (Early 2020 - Mid 2021)**

1. Initial Stability (2020):
   * Price Range: $40-$60
   * Characteristics: Prices remained stable with low volatility, except for a brief spike at around $83.87 in mid February before returning to around $40. This indicates minimal price changes during this period.
2. Gradual Increase with Volatility (Mid-2020 to Late 2020):
   * Price Range: Reached $80-$90($93.78) by late 2020.
   * Characteristics: The period experienced a gradual uptrend with increased volatility, evidenced by noticeable short-term peaks and dips.
3. Rapid Growth Phase (Late 2020 to Early 2021):
   * Price Surge: LTC prices surged from approximately $80 to over $200 by early 2021, marking a 150% increase.
   * Investor Sentiment: This dramatic rise reflects heightened investor interest, likely spurred by a broader cryptocurrency rally.
4. High Volatility at Elevated Levels (2021):
   * Price Range: Prices fluctuated between $139-$233.
   * Characteristics: Marked by high volatility and frequent price fluctuations, indicative of profit-taking behavior and underlying market uncertainties.

**Market Sentiment and Speculation**

Upward Trends: Driven by technological advancements, favorable regulatory changes, and increasing institutional adoption.

Downward Trends: Potentially triggered by regulatory concerns or broader market corrections (Kriptomat, n.d.).

**Volatility Insights**

High Volatility: Frequent price fluctuations highlight the speculative trading environment common in cryptocurrency markets.

Stability Periods: Episodes of lower volatility may indicate market maturity or a balanced interaction between buyers and sellers.

**Impact of External Events**

Correlation with Events: Price changes often align with significant external events, such as regulatory announcements or shifts in Bitcoin prices, showcasing LTC’s sensitivity to market news.

**Potential for Cyclical Patterns**

Recurring Patterns: Observations of predictable rises and falls may suggest seasonal trading behaviors related to broader economic or Bitcoin cycles.

**Overall Implications**

Investors should recognize LTC’s sensitivity to market sentiment and external factors, making it suitable for active trading strategies. Long-term holders should carefully time entry and exit points to optimize returns and minimize risks.

## 5.2 Daily Returns of LiteCoin (LTC)

### 5.2.1 Calculating Daily Returns

Daily returns represent the percentage change in the close price from one day to the next. This is calculated using the formula:

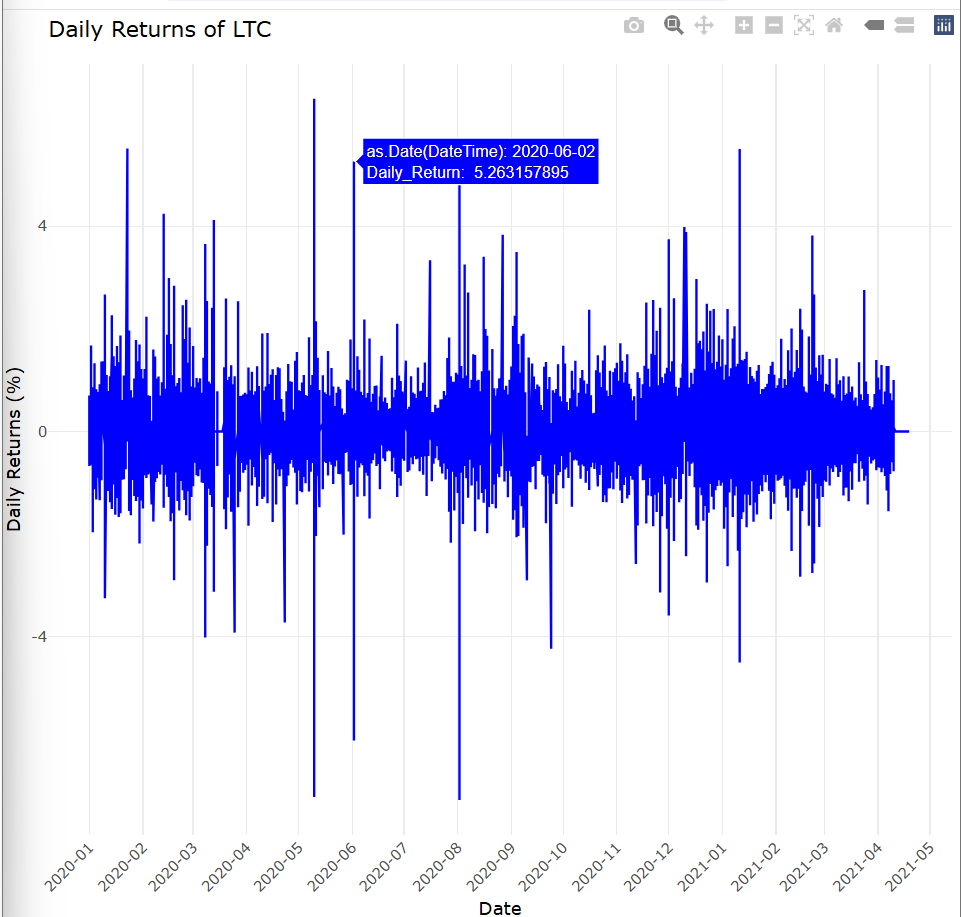
Gemini\_winsorized <- Gemini\_winsorized %>%  
 mutate(Daily\_Return = (Close / lag(Close) - 1) \* 100)

Purpose: Daily returns provide insight into the volatility and performance of LTC on a day-to-day basis. Positive returns indicate a price increase, while negative returns indicate a decrease.

### 5.2.2 Plotting Daily Returns

Description: This line graph displays the daily returns over time, providing a visual representation of LTC’s volatility.

This is what the chart looks like upon hovering:



To view the interactive line chart , click the link: [Daily Returns of LiteCoin (LTC)](file:///C:\Users\bida22-068\Downloads\R%20Assignment\%22C:\Users\bida22-068\Downloads\R%20Assignment\LTC_Daily_Returns_Interactive.html%22)

#### 5.2.2.1 Summary of Daily Returns of Litecoin (LTC) from January 2020 to May 2021

The line graph of daily returns (percent change) for Litecoin (LTC) from January 2020 to May 2021 reveals several key insights:

* High Volatility: Daily returns exhibit significant fluctuations, reflecting the unpredictable nature of cryptocurrency markets.
* Consistent Range: Most returns remain within -5% to +5%, indicating moderate daily fluctuations despite volatility.
* Occasional Outliers: Notable spikes, such as sharp declines at -4 % in early 2020 and increases at around +5% in mid-2020, highlight reactions to significant market events.
* Clustering of Volatility: Periods of heightened volatility are followed by stability, suggesting market responses to specific events.
* No Clear Trend: The chart shows regular fluctuations without a discernible price trend, suggesting that short-term trading strategies may be more effective.
* Stable Periods: Late 2020 and early 2021 exhibit reduced volatility, possibly due to decreased speculation.

The graph indicates that LTC is a high-risk asset with frequent daily fluctuations. While there is a consistent range of returns, outliers and clusters of volatility reflect market dynamics. This analysis is particularly relevant for short-term traders rather than long-term investors.

# 6. Time Series Analysis

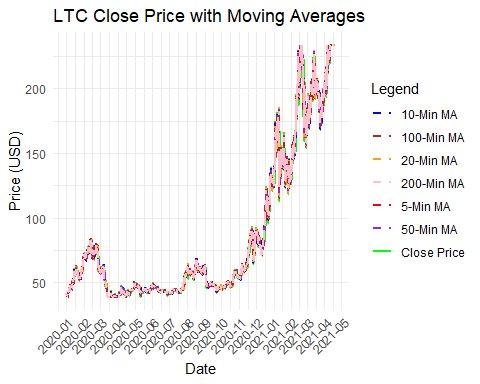
## 6.1 Moving Averages

Moving averages are a key tool in time series analysis, allowing for the smoothing of short-term fluctuations and highlighting longer-term trends (ProjectPro, 2024). In this analysis, the following moving averages were calculated using the Simple Moving Average (SMA) function from the TTR library:

* 5-Day Moving Average (MA5): This moving average captures the average closing price over the last 5 days, helping to identify short-term trends.
* 10-Day Moving Average (MA10): This average smooths the data over a 10-day window, providing a clearer view of recent price movements.
* 20-Day Moving Average (MA20): Useful for identifying intermediate trends, balancing short- and long-term perspectives.
* 50-Day Moving Average (MA50): A longer-term moving average that helps identify sustained trends, filtering out daily price fluctuations.
* 100-Day Moving Average (MA100): Provides further smoothing, emphasizing long-term trends.
* 200-Day Moving Average (MA200): Commonly used by traders to gauge the overall direction of the market, providing a broad view of price trends over an extended period

Gemini\_winsorized <- Gemini\_winsorized %>%  
 mutate(  
 MA5 = SMA(Close, n = 5),  
 MA10 = SMA(Close, n = 10),  
 MA20 = SMA(Close, n = 20),  
 MA50 = SMA(Close, n = 50),  
 MA100 = SMA(Close, n = 100),  
 MA200 = SMA(Close, n = 200)  
 )

The moving averages are plotted against LTC Close Price.

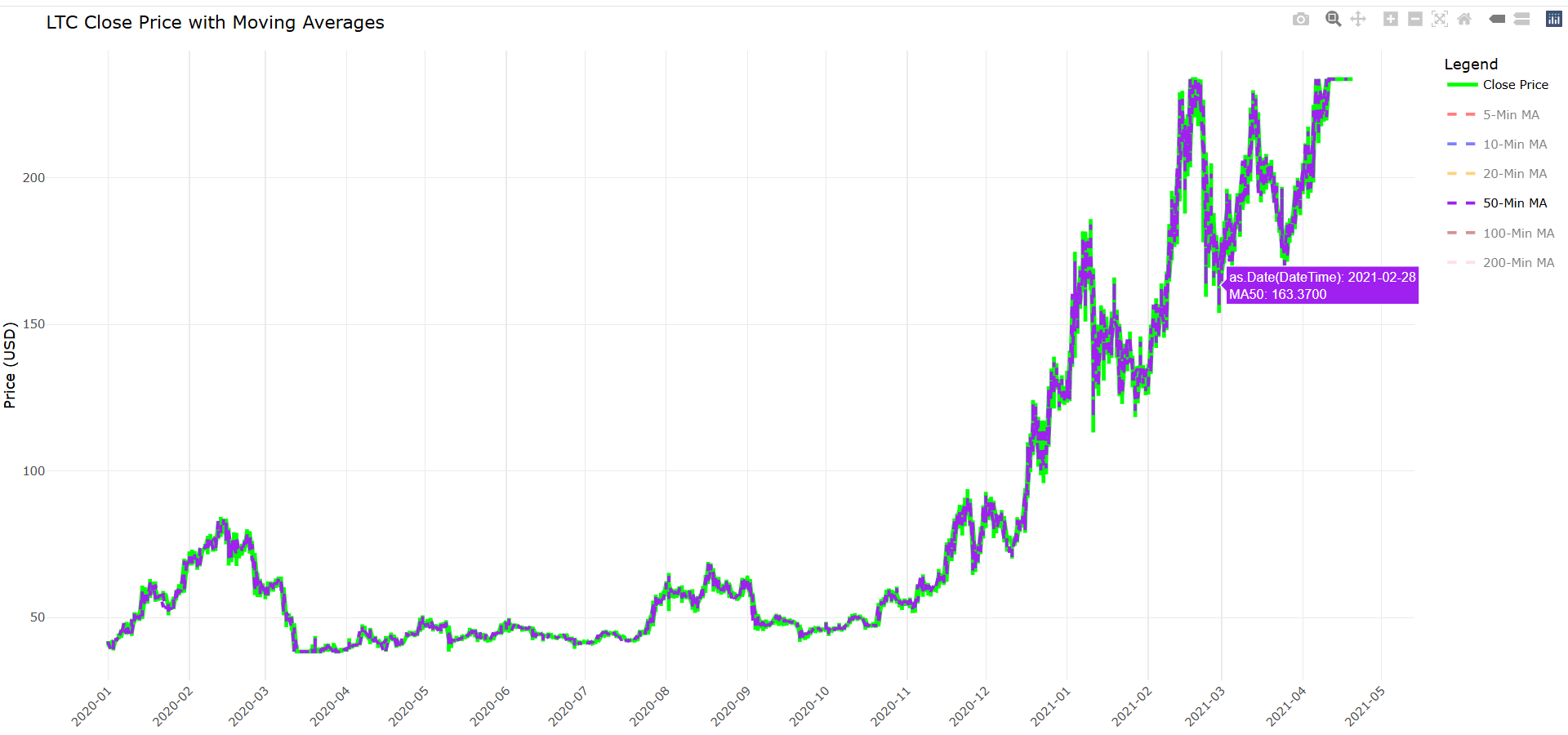


LTC Close Price with Moving Averages

To view an interact with the graph click [here](file:///C:\Users\bida22-068\Downloads\R%20Assignment\%22C:\Users\bida22-068\Downloads\R%20Assignment\LTC_Moving_Averages_Interactive.html%22). The package plotly was used and several key features which enhance data exploration and understanding in the graph include:

1. Zooming and Panning: Users can zoom into specific periods by selecting and dragging over the chart area, or by using the range slider below the x-axis. Panning is also supported, allowing for focused analysis on particular time frames within the dataset.
2. Tooltips: Hovering over any point on the plot reveals detailed data values (like date and price) for the Close Price and various Moving Averages. This feature provides immediate, context-sensitive information, aiding quick insights into trend changes.
3. Legend Interactivity: Users can click on items in the legend to toggle the visibility of individual Moving Averages, enabling selective focus on different time periods (e.g., shorter vs. longer-term trends) to compare market momentum.

Here is an overview of what the chart looks like, having filtered out the other MAs, focusing on 50-min MA vs Close Price:



LTC Close Price with 50-min Moving Average

### 6.1.1 Key Observations:

1. Overall Price Growth: LTC’s price shows a clear upward trend, rising from around $50 in mid-2020 to over $200 by May 2021, especially noticeable from late 2020 onward.
2. Price Stability and Volatility:
   * January 2020 to September 2020: Prices are relatively stable, fluctuating between $40 and $60 without a clear trend.
   * October 2020 Onward: Prices begin to climb steadily, with a sharper increase around November 2020, likely influenced by overall market optimism.
3. Moving Averages: The moving averages help smooth out short-term fluctuations.
   * When the price consistently stays above the moving averages, it suggests a positive trend.
   * From late 2020, the price mostly remains above the moving averages, indicating strong momentum.
4. Short-Term Corrections: Despite the upward trend, there are brief pullbacks, particularly in early 2021. However, these dips are short-lived, and the price tends to bounce back quickly.
5. Highlighted Data Point: A specific point at the end of February 2021 shows the 50-day MA at $163.37, indicating this level acted as a significant support level during the upward price movement.

* The chart illustrates a substantial price increase in Litecoin from late 2020 through early 2021. The moving averages provide insights into price behavior across different timeframes, with shorter MAs indicating rapid changes and longer MAs presenting a smoother trend that reflects overall market momentum. The crossover patterns suggest a transition to a more favorable price movement that began in late 2020 and continued into 2021.

# 7. Forecasting Litecoin (LTC) Daily Close Price with ARIMA

In this section, an ARIMA (Auto-Regressive Integrated Moving Average) model is used to forecast the daily closing price of Litecoin (LTC) for the next 30 days. ARIMA is widely used in time series forecasting due to its effectiveness in capturing trends, seasonality, and noise (Ranjani Chandrasekaran, 2022). This method was selected based on its ability to model autocorrelations within the time series data.

Libraries:

The **forecast** package provides functions to fit the ARIMA model and generate forecasts, while ggplot2 is used for data visualization. Additionally, we will use ggplotly to make our visualization interactive.

## 7.1 Methododogy

1. Data Aggregation

* library(forecast)  
    
  daily\_data <- Gemini\_winsorized %>%  
   group\_by(Date = as.Date(DateTime)) %>%  
   summarise(Daily\_Close = last(Close), .groups = 'drop')
  + Data at daily level is aggregated by extracting the closing plsie (Close) at the end of each day. This creates a daily time series required for ARIMA modeling.

1. Model Fitting Using ARIMA

* arima\_model <- auto.arima(daily\_data$Daily\_Close)
* Here, the auto.arima function automatically selects the optimal ARIMA model parameters based on the Akaike Information Citeration (AIC), balancing model complexity with accuracy (Hyndman & Anthanasopoulos, 2018). ARIMA works by modeling three main components:
  + AR (Auto-Regressive): Past values influence future values.
  + I (Integrated): Differencing applied to make the data stationary.
  + MA (Moving Average): Past forecast errors influence future values.

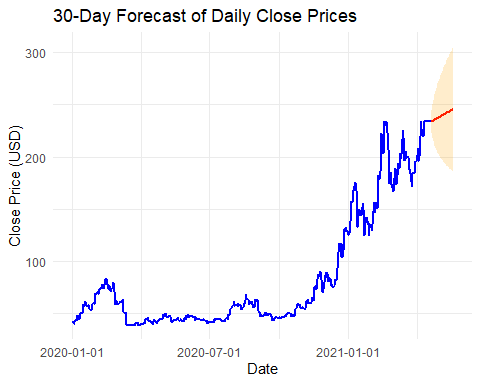
1. Forecasting

* forecast\_arima <- forecast(arima\_model, h = 30)
* This line forecasts the next 30 days using the fitted ARIMA model. The forecasted values come with 80% and 95% confidence intervals to quantify the range of possible outcomes.

1. Preparing Forecast for Visualization

* Convert the forecast to a format suitable for plotting, with columns for the forecastedvalues and their confidence intervals.
* forecast\_df <- data.frame(  
   Date = seq.Date(max(daily\_data$Date) + 1, by = "day", length.out = 30),  
   Forecast = forecast\_arima$mean,  
   Lower = forecast\_arima$lower[, 2],  
   Upper = forecast\_arima$upper[, 2]  
  )

1. Visualization

* The following plot, produced using ggplot2, displays historical data (in blue) and the 30-day forecast (in red), with shaded areas representing the 95% confidence interval.
* 
* Daily Close Price 30 day forecast

Link to interactive Forecast done with ggplotly package: [closing price 30-Day forecast](file:///C:\Users\bida22-068\Downloads\R%20Assignment\%22C:\Users\bida22-068\Downloads\R%20Assignment\LTC_Moving_Averages_Forecast_Interactive.html%22)

This graph shows a 30-day forecast for the daily closing prices of a stock, with historical data shown in blue and the forecasted range shaded in red.

## 7.2 Key Observations:

1. Historical Trend (Jan 2020 - Early 2021): The stock price exhibits an overall upward trend, with significant growth beginning around mid-2020 and reaching a peak of $230.47 in early 2021. While the price shows volatility, with several noticeable dips and peaks (e.g., a drop to around $43.04 in late 2020), it maintains a general upward trajectory.

2. Forecasted Range: The orange shaded area illustrates the 30-day forecast. The center of this shaded region (The red line) indicates the most probable price trajectory, while the widening of the shading toward the 30-day mark reflects increasing uncertainty associated with longer-term forecasts. For instance, the forecast predicts a price of around $245.99 at the end of the forecast period, with confidence intervals:

* Upper bound: $305.99
* Lower bound: $185.98

3. Forecast Implications:

* Positive Trend Expected: The forecast suggests that the price is likely to continue increasing slightly over the next 30 days, though with a moderate level of volatility.
* Uncertainty: The widening of the red shaded area highlights increasing uncertainty, which is common in price forecasts due to market unpredictability.

in conclusion, the forecast anticipates a potential continuation of the stock’s upward price trend over the next month. However, the expansion of the forecast range emphasizes the variability in how much the price may increase or potentially decrease. The forecast thus indicates growth potential while acknowledging the risks of fluctuations in the stock price.

# 8. Data Visualization with Dashboard

The LTC Data Analysis Dashboard is an interactive web application built using R’s Shiny framework. It provides users with insights into the trading data of Litecoin (LTC) through various visualizations, including candlestick charts, moving averages, volume analysis, daily returns, and price forecasts.

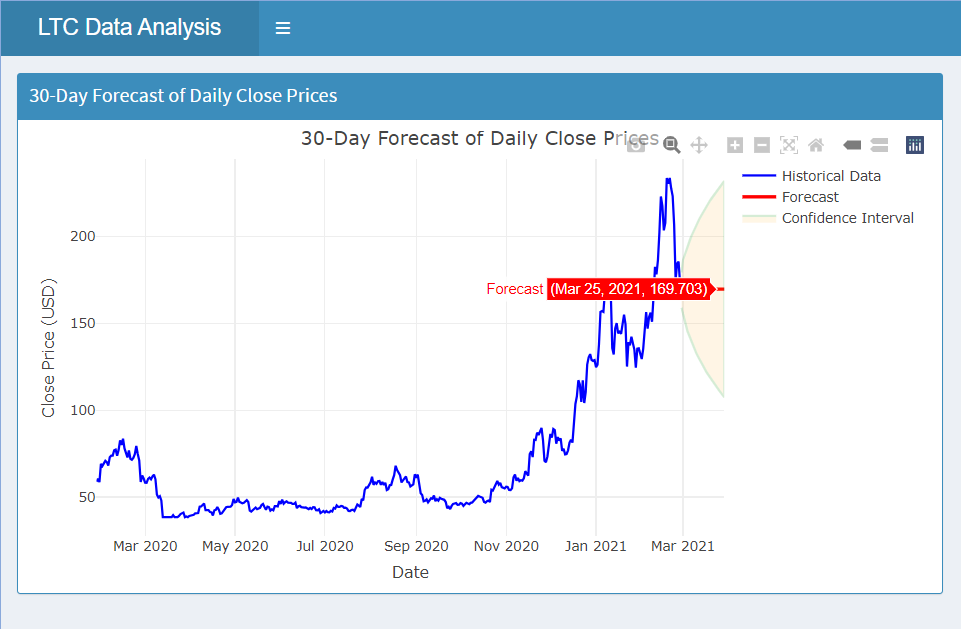
## 8.1 Dashboard Overview

The dashboard is structured into three main sections:

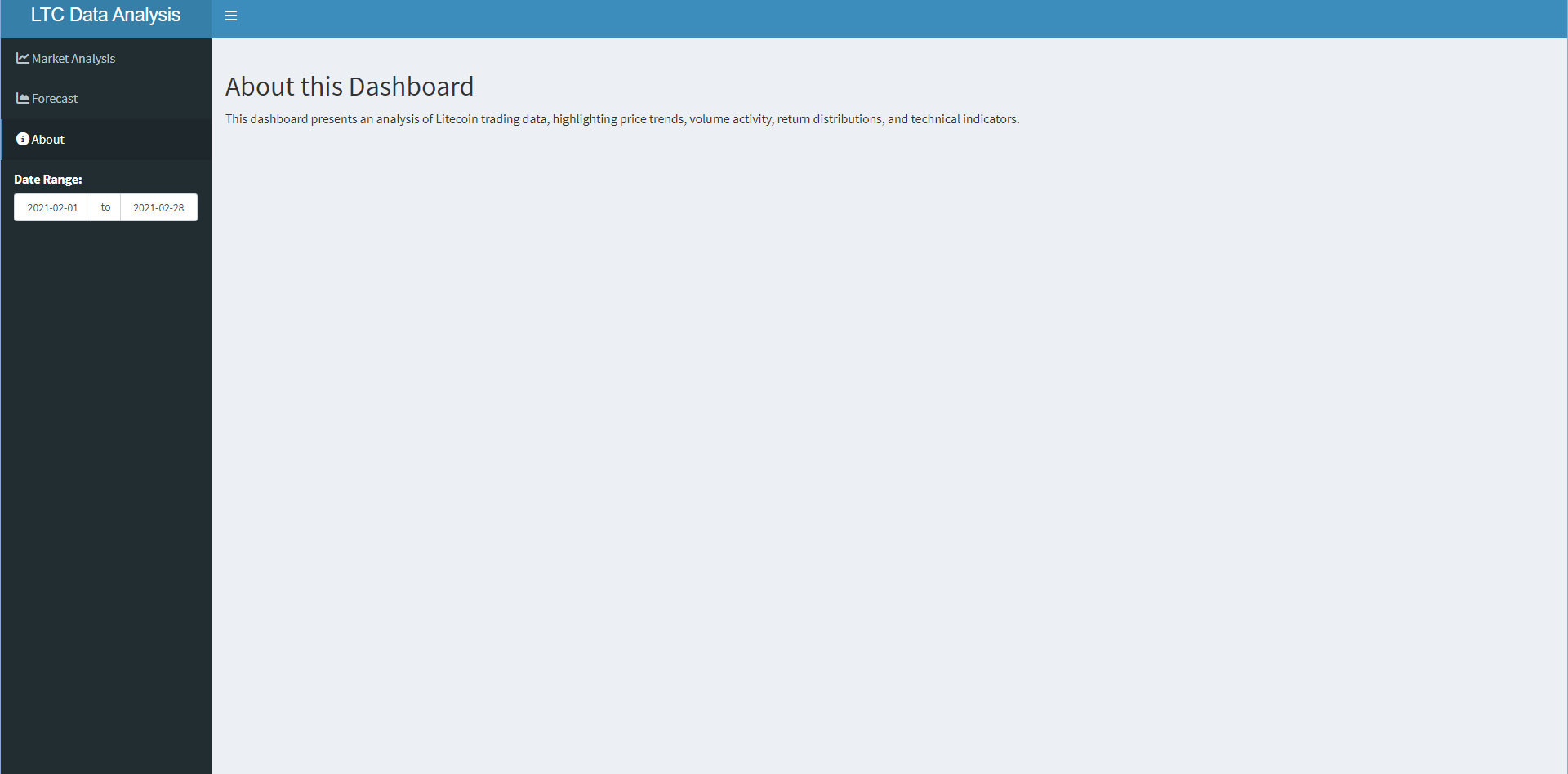
1. Market Analysis: This section focuses on visualizing historical price movements and relevant technical indicators.

* 
* Overview of Market Analysis Section

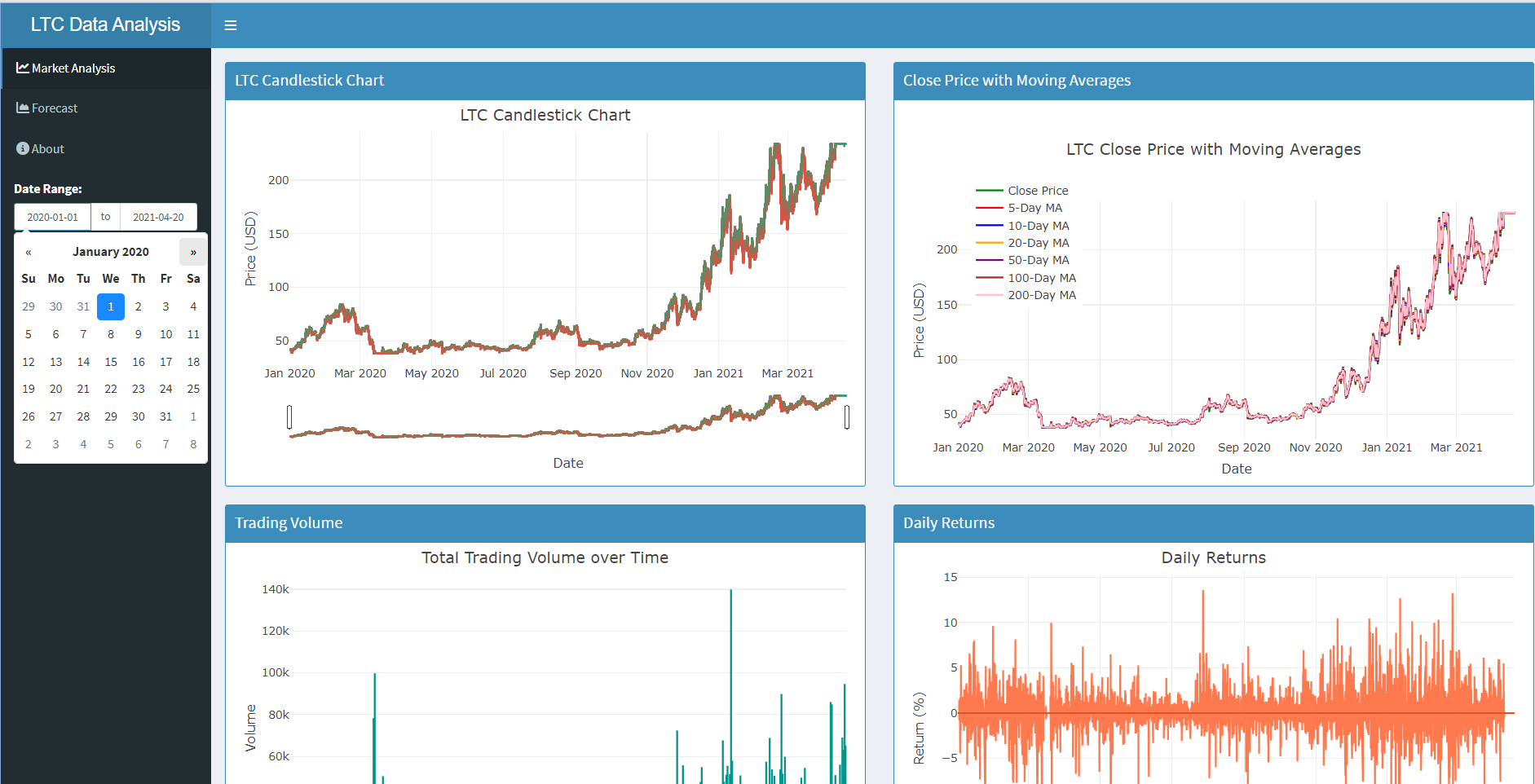
1. Forecast: This area provides a 30-day forecast of daily closing prices using an ARIMA model.

* 

1. About: This section offers a brief overview of the dashboard’s purpose and its contents.

* 

User Interface Components

* Date Range Input: Users can filter the data displayed on the dashboard by selecting a specific date range for analysis. The default range spans the entire dataset.
* 
* Users can hover to explore specific data points.
* When users press the menu icon it hides contents of the menu stated above (refer to Image showing Forecast section).

## 8.2 Key Visualizations

## 8.3 1. LTC Candlestick Chart

* **Purpose**: Visualizes the open, high, low, and close prices of LTC over the selected date range.
* **Features of the chart:**
  1. **OHLC Data**: Each candlestick shows the **Open**, **High**, **Low**, and **Close** prices for the day, with:
     + **Green Candlesticks** for days where prices rose (close > open).
     + **Red Candlesticks** for days where prices fell (close < open).
  2. **Volatility and Trend**: The length of each candlestick and its wicks illustrate daily price volatility and market sentiment, allowing easy identification of trends or possible reversals.
* **Interactive Hover Information:** When hovering, each candlestick reveals:
  + **Date** and **OHLC prices**, enabling precise, day-by-day analysis of LTC’s price action.
* **Interpretation**: Users can assess daily price volatility and trends through the candlestick representation. Each candlestick indicates price movement for a single day, where the color signifies whether the closing price was higher or lower than the opening price

### 8.3.1 2. Close Price with Moving Averages

* **Purpose**: Displays the closing price of LTC along with multiple moving averages (5, 10, 20, 50, 100, and 200 days).
* **Interpretation**: The moving averages help smooth out price fluctuations and identify potential trends. The green line represents the closing price, while the colored dashed lines represent different moving averages, providing insights into short-term vs. long-term trends.

### 8.3.2 3. Total Trading Volume

* **Purpose**: Illustrates the total trading volume of LTC over the selected date range.
* **Interpretation**: Volume is a crucial indicator of market activity. High volume often accompanies significant price movements, providing insights into market sentiment.

### 8.3.3 4. Daily Returns

* **Purpose**: Shows the daily returns (percentage change) of LTC prices.
* **Interpretation**: This plot highlights the volatility of the asset, with positive and negative returns indicating price fluctuations on a day-to-day basis.

### 8.3.4 5. 30-Day Forecast of Daily Close Prices

* **Purpose**: Provides a forecast of LTC’s closing prices for the next 30 days using the ARIMA model.
* **Interpretation**: The red line represents the predicted closing prices, while the shaded area indicates the confidence intervals (uncertainty range). The center of the shaded area illustrates the most likely price trajectory, while the widening towards the end signifies increased uncertainty in forecasts over longer periods.

## 8.4 Insights and Key Findings

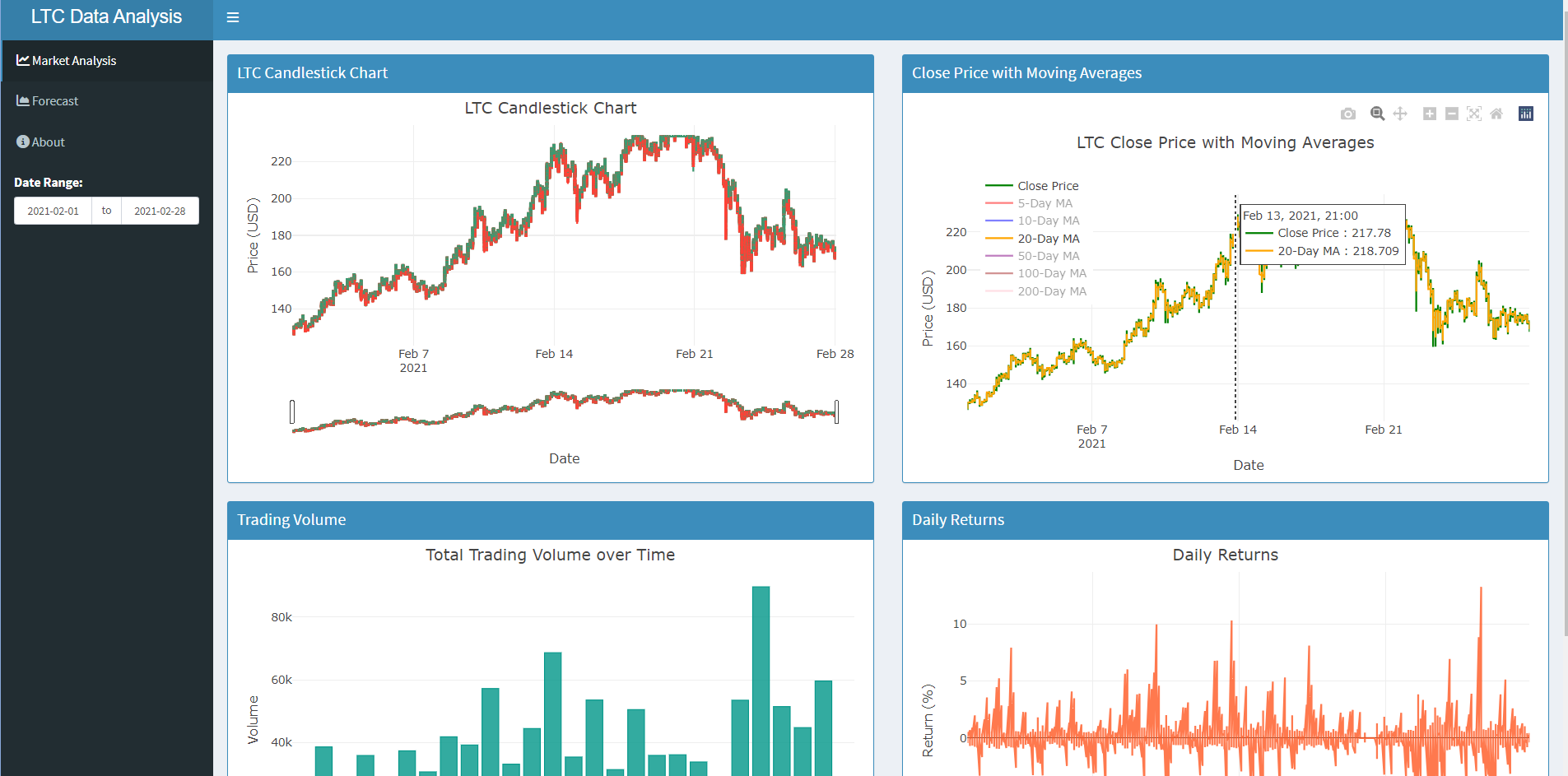
**Litecoin (LTC) Data Analysis for February 2021 (Filtered date range)**

### 8.4.1 1. LTC Candlestick Chart



* The candlestick chart shows daily price movements with opening, high, low, and closing prices.
* LTC peaked around mid-February, reaching approximately $215 on February 14 before a swift decline.
* Price patterns reveal notable volatility, with fluctuations around this peak and a gradual decrease toward the month’s end.

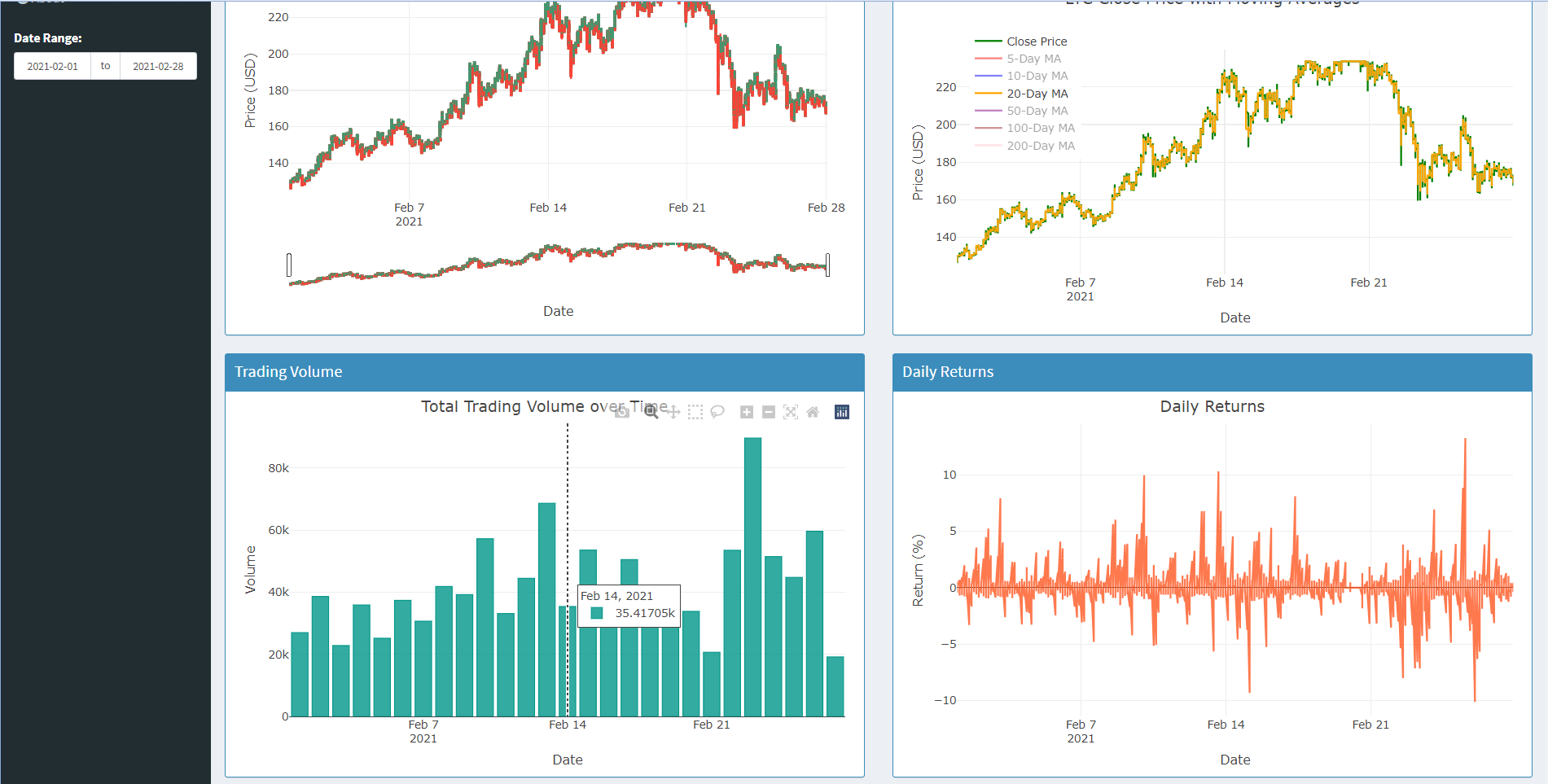
### 8.4.2 2. Close Price with Moving Averages



* This chart displays LTC’s closing price along with short- and long-term moving averages (e.g., 5-day, 10-day, 20-day).
* The closing price closely follows shorter averages, reflecting shorter-term trends.
* In mid-February, the closing price rises above longer averages, indicating upward momentum, before returning to these averages after the peak.

### 8.4.3 3. Trading Volume

* Trading volume spikes around February 14, surpassing 60,000, indicating heightened activity and interest during peak price volatility.
* Volume stabilizes following the mid-month peak, aligning with the gradual price decline.



### 8.4.4 4. Daily Returns

* The daily returns graph reveals significant fluctuations throughout February, especially around the peak, reflecting high volatility.
* Toward month-end, returns narrow, signaling lower volatility as prices stabilized.

### 8.4.5 Key Insights

* **Mid-February Peak**: LTC experienced a price and volume peak around February 14, suggesting increased market activity and volatility.
* **Volatility Trends**: February saw considerable price swings, with large movements especially around the peak, highlighting a high-risk period.
* **Trend Reversion**: Following the peak, prices aligned more closely with longer averages, indicating a cooling phase.
* **Market Shift**: Volume and prices declined toward the end of February, reflecting decreased trading interest.

In summary, February 2021 was a highly active month for LTC, marked by a sharp price peak, high trading activity, and a shift in momentum by month-end. This analysis offers insights into LTC’s behavior during periods of volatility and active trading.

# 9. Applications of Litecoin (LTC) Data Analysis in a Business Context

#### 9.0.0.1 1. **Investment Strategy Development**

The insights derived from LTC price movements, trading volume, and volatility can aid investors and financial institutions in formulating investment strategies. For example, the identification of peak periods, such as the price spike observed in February 2021, can help investors time their entries and exits effectively. This is crucial in a volatile market where timing can significantly influence returns (Chen, 2021). By analyzing historical patterns, businesses can develop models that optimize portfolio allocations based on expected price trends.

#### 9.0.0.2 2. **Risk Management**

Understanding price volatility and daily return patterns allows businesses to implement better risk management practices. Firms can use this analysis to establish risk thresholds, set stop-loss orders, and determine the appropriate level of exposure to cryptocurrencies in their portfolios. For instance, during periods of high volatility, as seen in mid-February, businesses might choose to limit their exposure to LTC to mitigate potential losses (Gandal and Halaburda, 2016). This proactive approach to risk management can safeguard assets and improve overall financial stability.

#### 9.0.0.3 3. **Market Sentiment Analysis**

The analysis of trading volume in conjunction with price movements can provide insights into market sentiment. High trading volumes, particularly during significant price fluctuations, often indicate increased interest or reaction to external events (Swan, 2015). Businesses can leverage this information to make informed decisions regarding marketing strategies, product launches, or customer engagement initiatives in the cryptocurrency space. For example, if trading volume spikes alongside positive news about LTC, businesses could capitalize on this momentum through targeted promotions or partnerships.

### 9.0.1 Recommendations

* **Develop Data-Driven Investment Models**: Financial institutions should invest in developing algorithms that utilize historical LTC data to predict future price movements. This could involve machine learning techniques to enhance accuracy.
* **Implement Robust Risk Assessment Tools**: Firms should establish protocols for monitoring market conditions and volatility, enabling them to adjust their trading strategies in real time based on market behavior.
* **Engage in Market Monitoring**: Businesses should closely monitor trading volumes and market sentiment to stay ahead of trends, allowing them to adapt their strategies promptly and capitalize on emerging opportunities.

# 10. Conclusion

This analysis of Litecoin (LTC) trading data highlights important trends in price movements, trading volume, and market volatility within the cryptocurrency landscape. The findings indicate that LTC experienced significant price fluctuations in February 2021, particularly around mid-month, when both market interest and trading volume peaked. By examining LTC’s closing prices alongside moving averages, it was observed how the coin aligned with short-term market trends and how increased volatility often followed sharp price changes.

To address the challenges posed by extreme values common in cryptocurrency markets, techniques like winsorization were employed. This approach helped interpret trends more accurately by minimizing the impact of outliers. Additionally, the use of the ARIMA model for a 30-day forecast provided a data-driven outlook on LTC’s future closing prices, offering valuable insights for traders making investment decisions.

In summary, this analysis demonstrates the effectiveness of data-driven methods in navigating the complexities of cryptocurrency trading. By understanding historical price patterns and anticipating volatility, traders and investors can make more informed and strategic choices. Overall, this study emphasizes the importance of thorough analysis in managing risks and seizing opportunities in the dynamic world of digital assets.

# 11. References

Anderson, T.W. (2019) ‘Statistical analysis of financial data with r’, *Journal of Financial Research*, 42(3), pp. 457–479.

Chen, J. (2021) ‘Understanding the cryptocurrency market: An investor’s perspective’, *Journal of Finance*, 76(3), pp. 145–167. Available at: <https://doi.org/10.1111/jofi.12925>.

Gandal, N. and Halaburda, H. (2016) ‘Can bitcoin become a major currency?’, *Journal of Economic Perspectives*, 29(2), pp. 213–238. Available at: <https://doi.org/10.1257/jep.29.2.213>.

Hoaglin, D.C., Iglewicz, B. and Tukey, J.W. (2000) ‘Variability of the regression slope in the presence of outliers’, *Statistical Science*, 15(3), pp. 331–340.

Kriptomat (n.d.) ‘How to evaluate market sentiment before buying crypto’. Available at: <https://kriptomat.io/finance-investing/how-to-evaluate-market-sentiment-before-buying-crypto/>.

MBA Skool (n.d.) ‘Square root transformation - meaning & definition’, *MBA Skool* [Preprint]. Available at: <https://www.mbaskool.com/business-concepts/statistics/7463-square-root-transformation.html>.

Nakamoto, S. (2008) ‘Bitcoin: A peer-to-peer electronic cash system’. Available at: <https://bitcoin.org/bitcoin.pdf>.

ProjectPro (2024) ‘A practical introduction to moving average time series model’. Available at: <https://www.projectpro.io/article/moving-average-time-series-model/716>.

Ranjani Chandrasekaran, S. (2022) *Prediction of litecoin prices using ARIMA and LSTM*. Master’s thesis. National College of Ireland. Available at: <https://norma.ncirl.ie/4222/1/ranjanichandrasekaran.pdf>.