

**Module 1 R Practice**

Dia Khosla

College of Professional Studies, Northeastern University  
ALY6010: Probability Theory and Introductory Statistics

Instructor - Dr. Selcuk Baran

May 30, 2024

Abstract:

This project involved analyzing the Netflix Stock Price Dataset using R programming. The dataset provided a comprehensive record of Netflix's stock price changes over time, including essential columns such as date, opening price, highest price of the day, lowest price of the day, closing price, adjusted closing price, and trading volume.

Guided by project instructions, we utilized R's built-in functions for Frequency Tables and Cross-Tabulations to explore the data's distributions and relationships. We employed various visualization techniques such as histograms, line plots, and time series analysis to uncover insights into the stock price dynamics.

Through this project, we demonstrated proficiency in statistical analysis and visualization techniques, highlighting the application of statistical methods in real-world scenarios. The findings contribute to a better understanding of Netflix's stock price behavior and the factors influencing it.

Introduction:

In this Module 1 project, we explore and analyze the Netflix Stock Price Dataset, focusing on understanding the stock price dynamics over time. The dataset provides a comprehensive record of Netflix's stock price changes.

Throughout Module 1, Week 1, we follow a structured approach to perform data analysis and visualization tasks using R. We begin by importing necessary libraries such as dplyr, ggplot2, plotly, lubridate, and readr, which are essential for data manipulation, visualization, and handling date complexities.

Next, we read the Netflix dataset from a CSV file and apply appropriate data structures and type conversions to ensure data integrity. We conduct a summary analysis of the dataset to gain insights into its structure and distribution.

To understand the temporal distribution of the data, we generate frequency tables for the year variable and perform cross-tabulations to explore the relationship between year and month.

We then identify high-volume trading days by calculating a volume threshold and adding a binary indicator column to the dataset.

Subsequently, we calculate daily returns to analyze the stock price movements over time. Utilizing the mutate function from the dplyr library, we compute the percentage change in closing prices from one day to the next.

Finally, we visualize the Netflix stock price data using various plots, including time series plots for opening, closing, adjusted closing prices, and trading volume. Additionally, we plot the total volume of stocks traded in the first quarter of 2024 to illustrate temporal trends.

**Key Findings:**

**Variables of Interest:** The variables of interest in the dataset (Netflix Dataset) include:

1. Date: The date of the trading day.
2. Open: The opening price of Netflix stock on the given day.
3. High: The highest price of Netflix stock during the trading day.
4. Low: The lowest price of Netflix stock during the trading day.
5. Close: The closing price of Netflix stock on the given day.
6. Adjusted Close: The adjusted closing price of Netflix stock.
7. Volume: The trading volume of Netflix stock on the given day.

**Date Range**: The dataset spans from May 24, 2002, to May 24, 2024, indicating a historical record of Netflix stock prices. The median date falls on May 24, 2013, suggesting a balanced distribution of dates in the dataset.

**Opening Price (Open):** The opening prices range from a minimum of 0.38 to a maximum of 692.35. The median opening price is 35.97, indicating that half of the opening prices fall below this value.

**Highest Price of the Day (High):** The highest prices during the trading day range from 0.41 to 700.99. The median highest price is 36.78, reflecting the central tendency of highest prices in the dataset.

**Lowest Price of the Day (Low):** The lowest prices during the trading day range from 0.3464 to 686.09. The median lowest price is 35.38, indicating the typical low price during a trading day.

**Closing Price (Close):** The closing prices range from a minimum of 0.373 to a maximum of 691.69. The median closing price is 36.2014, representing the central tendency of closing prices in the dataset.

**Adjusted Closing Price (Adj.Close):** The adjusted closing prices range from 0.373 to 691.69. The median adjusted closing price is 36.2014, consistent with the median closing price.

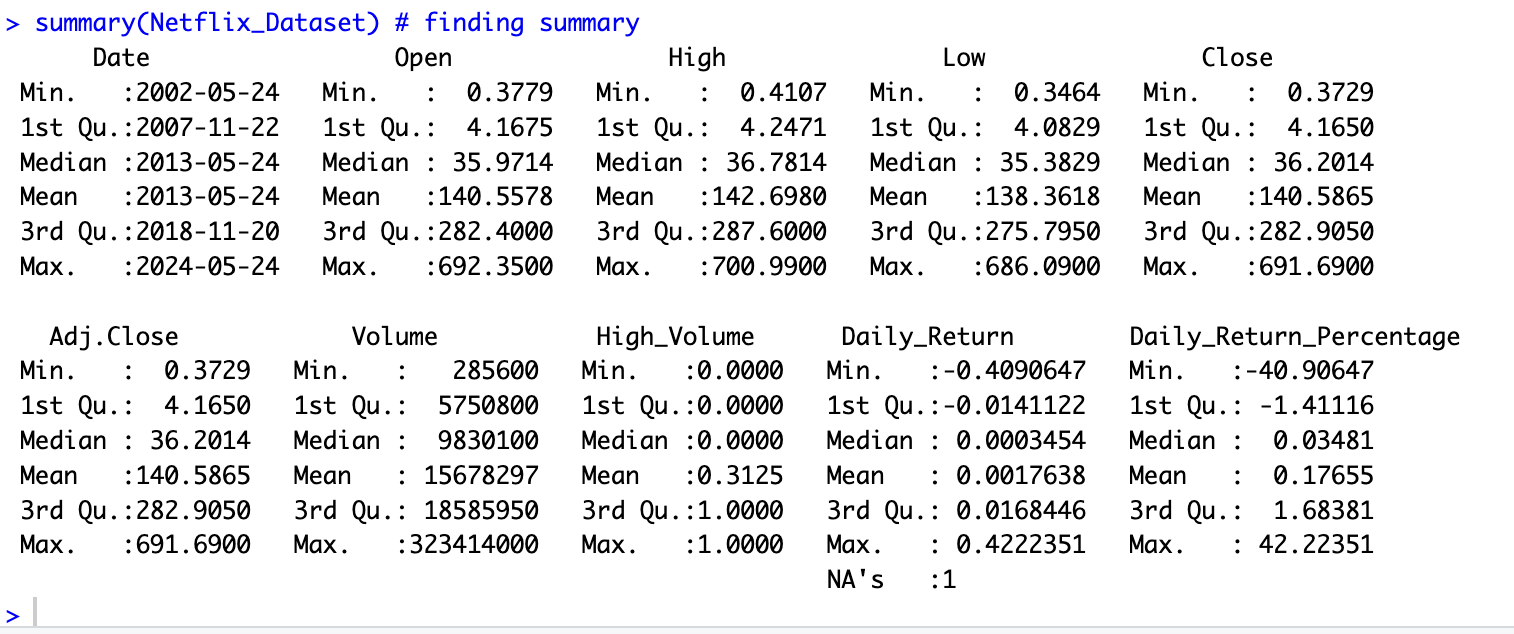
**Trading Volume (Volume):** The trading volumes range from a minimum of 285,600 to a maximum of 323,414,000 shares. The median trading volume is 9,830,100 shares, indicating the typical level of trading activity.

**High Volume Indicator (High\_Volume):** The high volume indicator takes values of 0 or 1, with a mean of 0.3125, suggesting that approximately 31.25% of trading days have high volume.

**Daily Returns:** Daily returns range from a minimum of -0.41 to a maximum of 0.42. The median daily return is 0.0003454, indicating that half of the daily returns are positive.

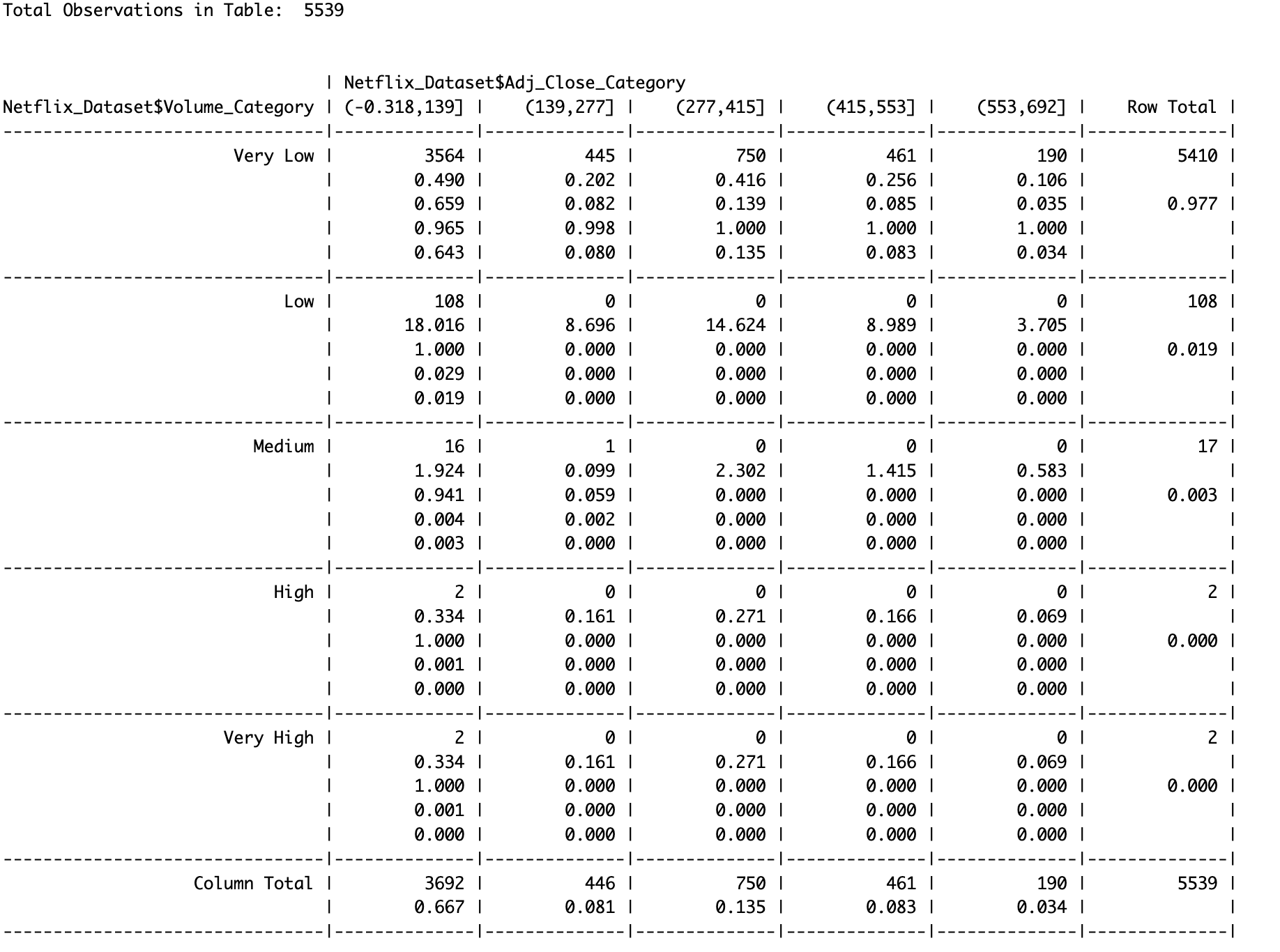
**Daily Return Percentage:** Daily return percentages range from a minimum of -40.90% to a maximum of 42.21%.

Here using the summary function, I have derived summary of the Netflix Dataset below, conducting statistical analysis.



### **Cross-Tabulations with gmodels**

To further analyze the distribution of data across different time periods, I utilized the gmodels library to examine the relationship between two categorical variables by displaying (Adjusted Close). Also, I labelled them “Very Low", "Low", "Medium", "High", "Very High" in order to find the frequency relationship between these two variables. Here, is the result :-



Observations :-

**Very Low Volume Category:**

* Contains the majority of observations (5410 out of 5539).
* The proportion of the grand total (5539) for this category is 0.977.
* The majority of the observations fall into the "Very Low" volume category, indicating that Netflix stock is most often traded in very low volumes. This category alone accounts for 97.7% of the total observations.
* Within the "Very Low" volume category, the adjusted closing prices are fairly spread out, with the highest counts in the lowest price interval (-0.318, 139] and the interval (277, 415].

**Low, Medium, High, and Very High Volume Categories:**

* Contain very few observations (108, 17, 2, and 2, respectively).
* Each of these categories represents a very small proportion of the grand total.
* The "Low" volume category has 108 observations, all in the (0.318, 139] price interval.
* The "Medium" volume category has 17 observations, with 16 in the (0.318, 139] price interval and 1 in the (139, 277] price interval.
* Both "High" and "Very High" volume categories have only 2 observations each, all in the (0.318, 139] price interval.

Conclusion from this key observation :- The data suggests that the majority of trading volume for Netflix stock is very low, regardless of the adjusted closing price category. This could indicate low overall trading activity or a significant portion of days with low trading volumes.

High adjusted closing prices do not appear frequently with higher trading volumes. Most of the higher price intervals have much fewer observations.

Summary

**Predominantly Low Volumes:**

* Netflix stock is mostly traded in very low volumes, indicating infrequent large institutional trades.
* High volume trades are rare.

**Price Distribution:**

* Most trades occur in the lowest price interval across all volume categories.
* This suggests periods when Netflix stock was predominantly trading at lower prices.

**Investment Strategy Insights:**

* Investors should consider the impact of low liquidity on price movements.
* The rarity of high volume trades may imply less risk of significant price swings, which is beneficial for long-term stability.

Next, I converted the Date column to the Date type to conduct date-specific analyses and conducted Volume Threshold Calculation where I worked on calculating The average trading volume to understand the threshold for categorizing high and low volume days and created a High Volume Indicator to indicate whether the trading volume for a day exceeds the average volume.

The code here is :- volume\_threshold <- mean(Netflix\_Dataset$Volume)  
Netflix\_Dataset$High\_Volume <- ifelse(Netflix\_Dataset$Volume > volume\_threshold, 1, 0)

#### **Observations and Analysis:-**

The dataset shows that the majority of trading days have volumes below the average threshold, indicated by High\_Volume being 0 for most rows. High trading volume days are infrequent, suggesting that large institutional trades are uncommon.

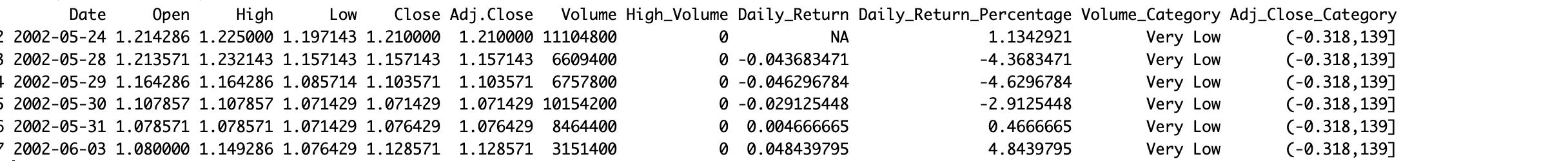
**Use of Library(dplyr) in Calculating daily returns:-**

Here, I worked on calculating daily returns (Daily\_Return) netflix where I calculated the percentage change from the previous day's closing price:

Code :- Netflix\_Dataset <- mutate(Netflix\_Dataset, Daily\_Return = (Close - lag(Close))/lag(Close))

head(Netflix\_Dataset)

Here are the calculated daily returns for the sample data provided:

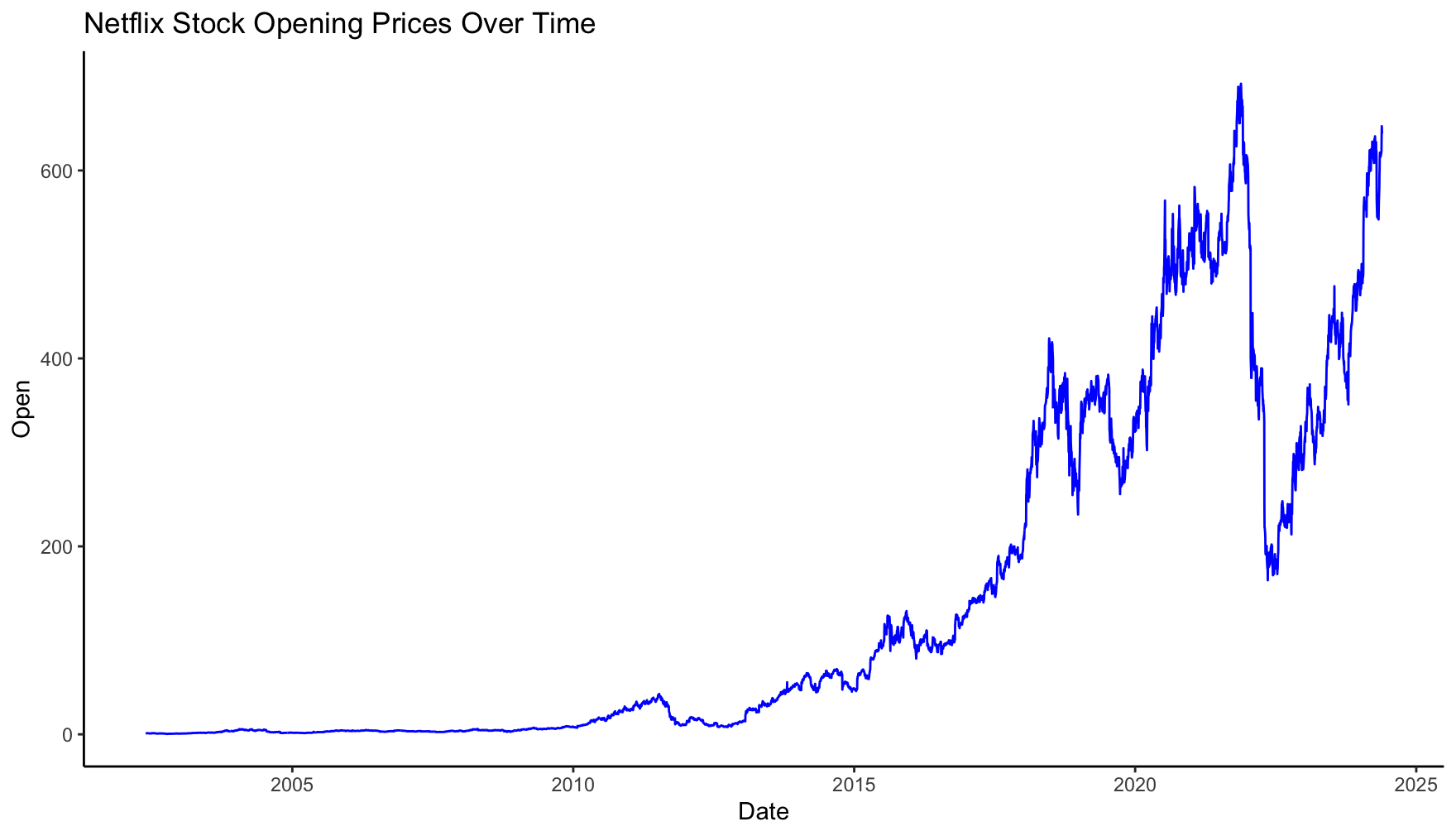
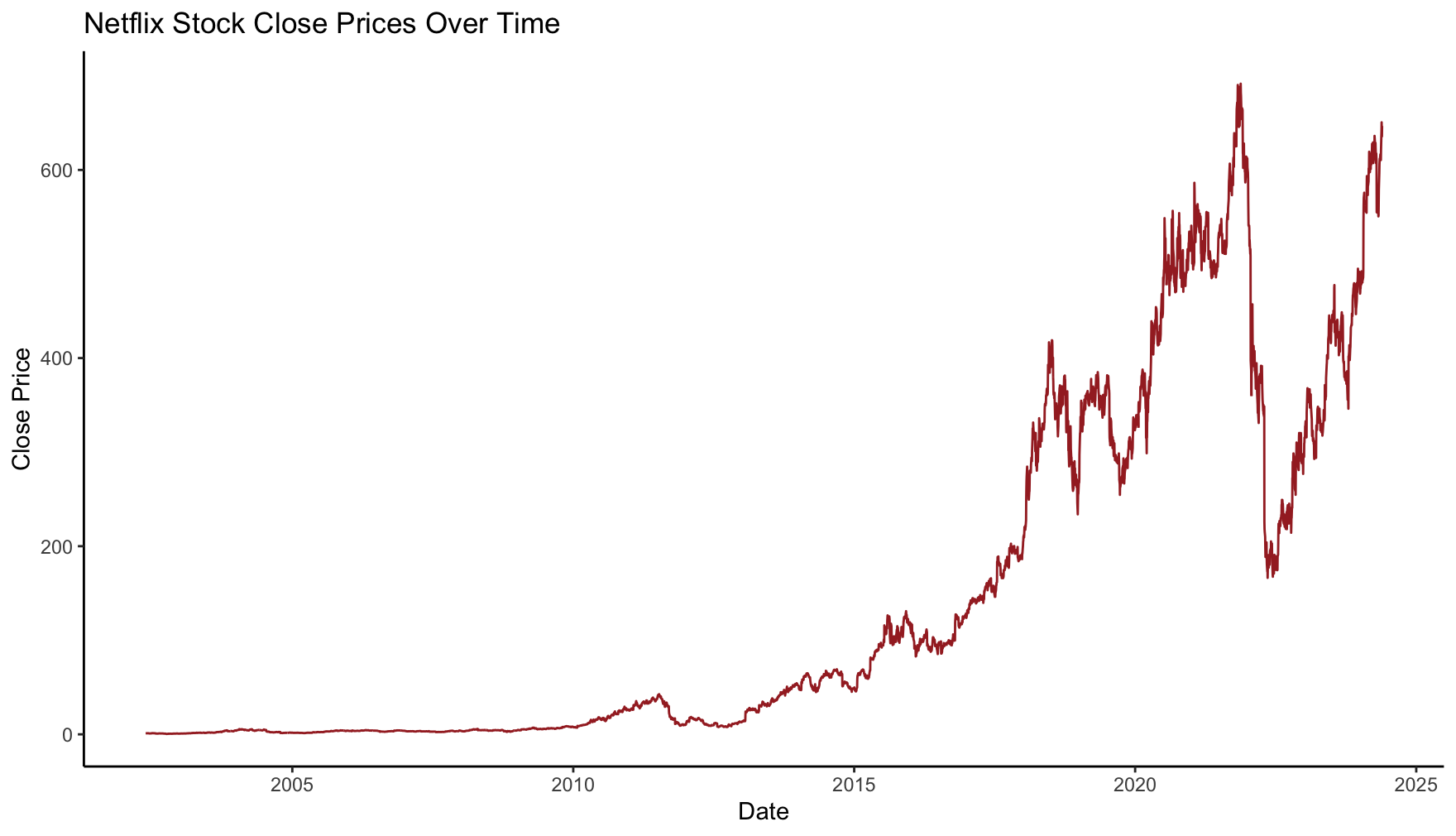


**Interpretation of Daily Returns:**

* **Negative Returns:**
  + **2002-05-28:** The stock price fell by approximately 4.37% compared to the previous trading day.
  + **2002-05-29:** The stock price dropped further by approximately 4.63%.
  + **2002-05-30:** Another decline of about 2.91%.
* **Positive Returns:**
  + **2002-05-31:** A slight increase in stock price by about 0.47%.
  + **2002-06-03:** A more significant increase of about 4.84%.

These returns indicate periods of both decline and recovery within a short timeframe.

**Time series plot for Date :- Netflix Stock Opening and Closing Prices Over Time**

**Overview**

This analysis focuses on the time series trends of Netflix's stock prices from 2015 to 2025, with specific emphasis on the following observations:

* Stable open and close prices from 2015 until mid-2020.
* Increased volatility and price fluctuations from mid-2020 onwards.
* Significant price movements in 2022-2024, with prices rising above 650 and then dropping below 200.
* Expected price rise towards the end of 2024 or beginning of 2025.

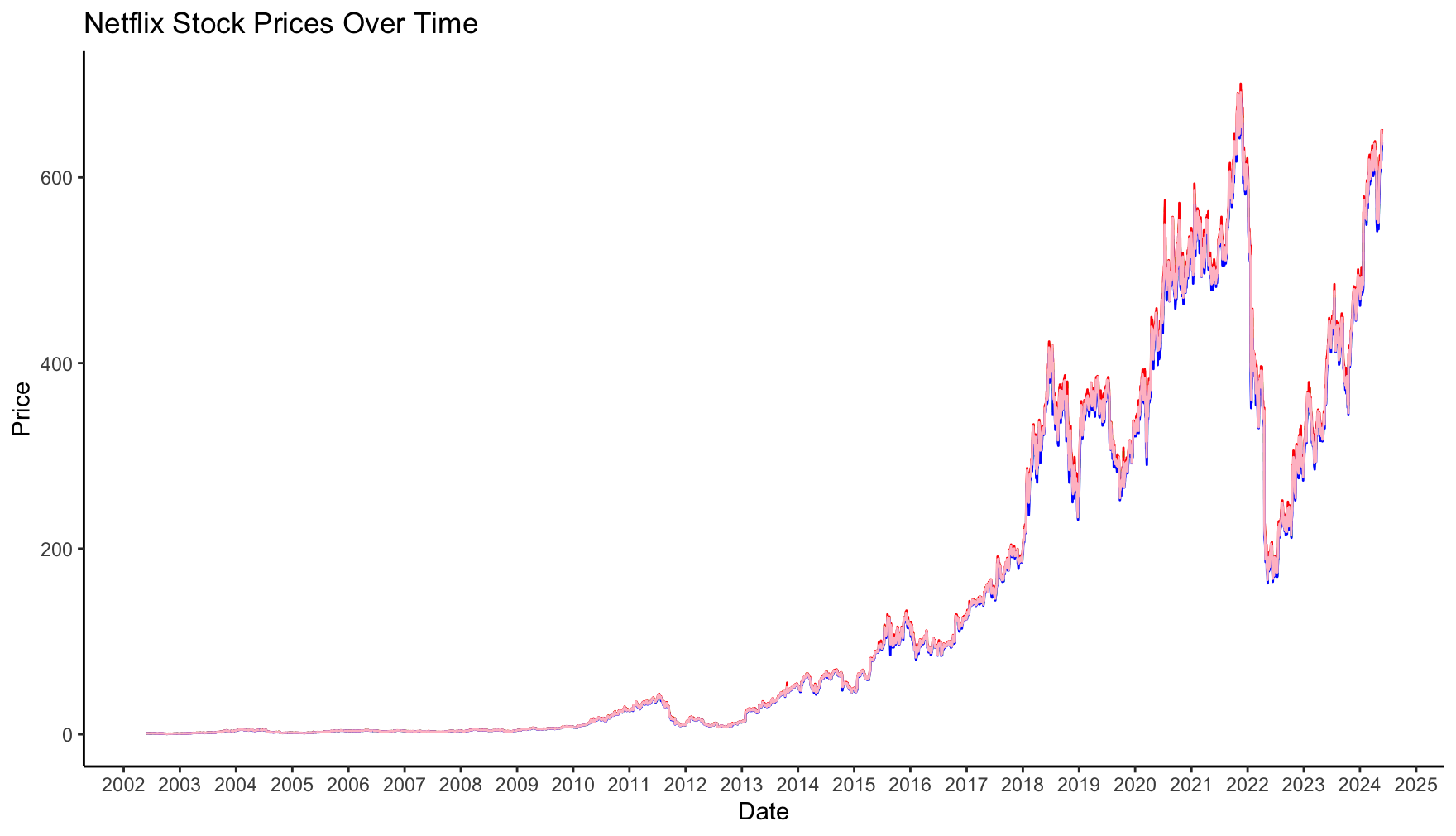
**Analysis: -** Despite recent volatility, the long-term growth potential of Netflix remains strong, particularly if the company continues to innovate and expand its market presence.

The expected rise in prices towards the end of 2024 or beginning of 2025 suggests that periods of decline may be followed by recovery phases.F or long-term investors, periods of price decline could present buying opportunities, assuming confidence in Netflix's future performance.

**Line plots for Netflix Stock Prices Over Time**

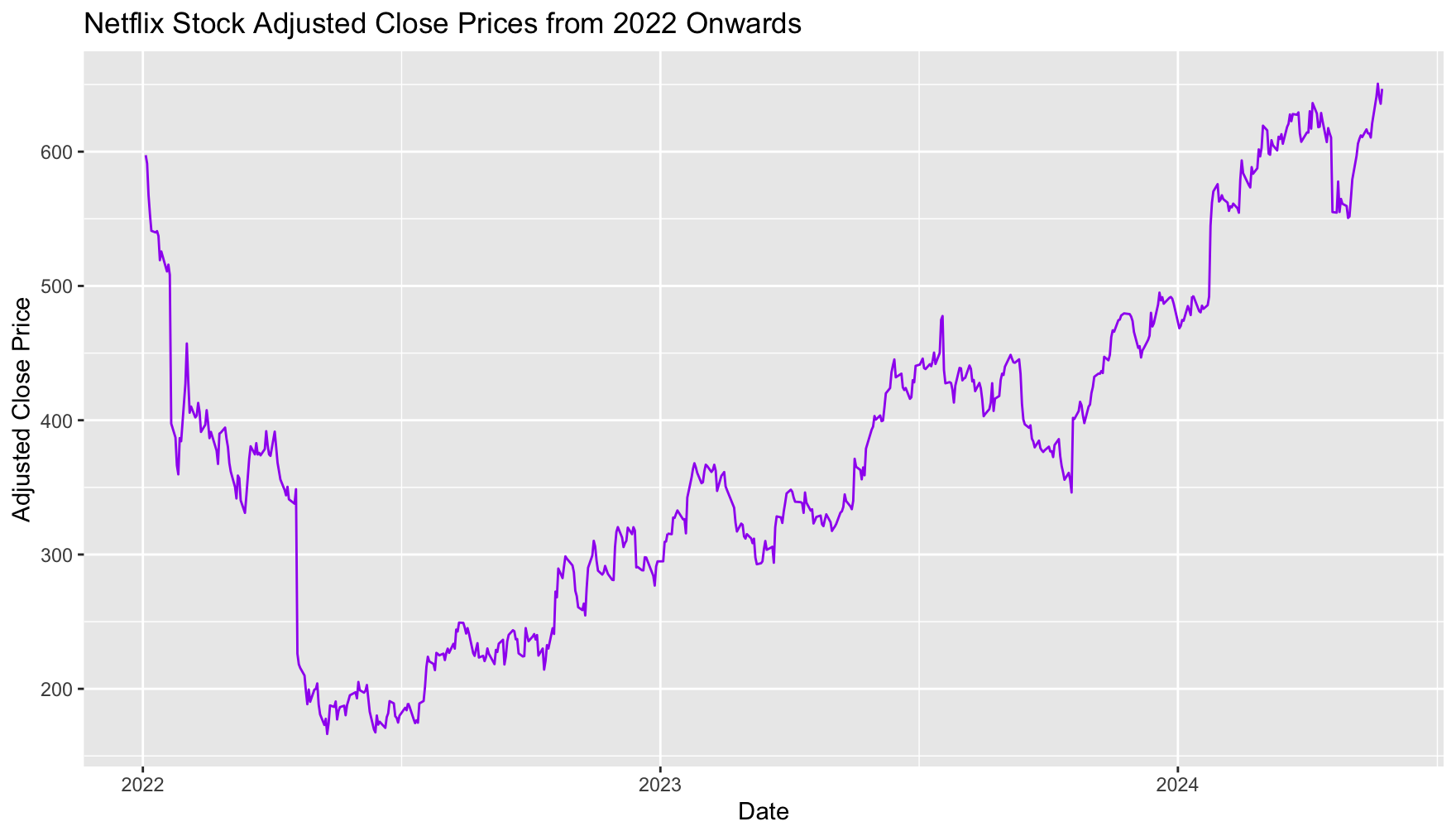
I utilized ggplot to create a line plot depicting four distinct stock prices over time. Each line represents a different metric:

* Open, indicated in green;
* High, represented in red; Low, shown in blue;
* and Close, depicted in pink .



Analysis :-

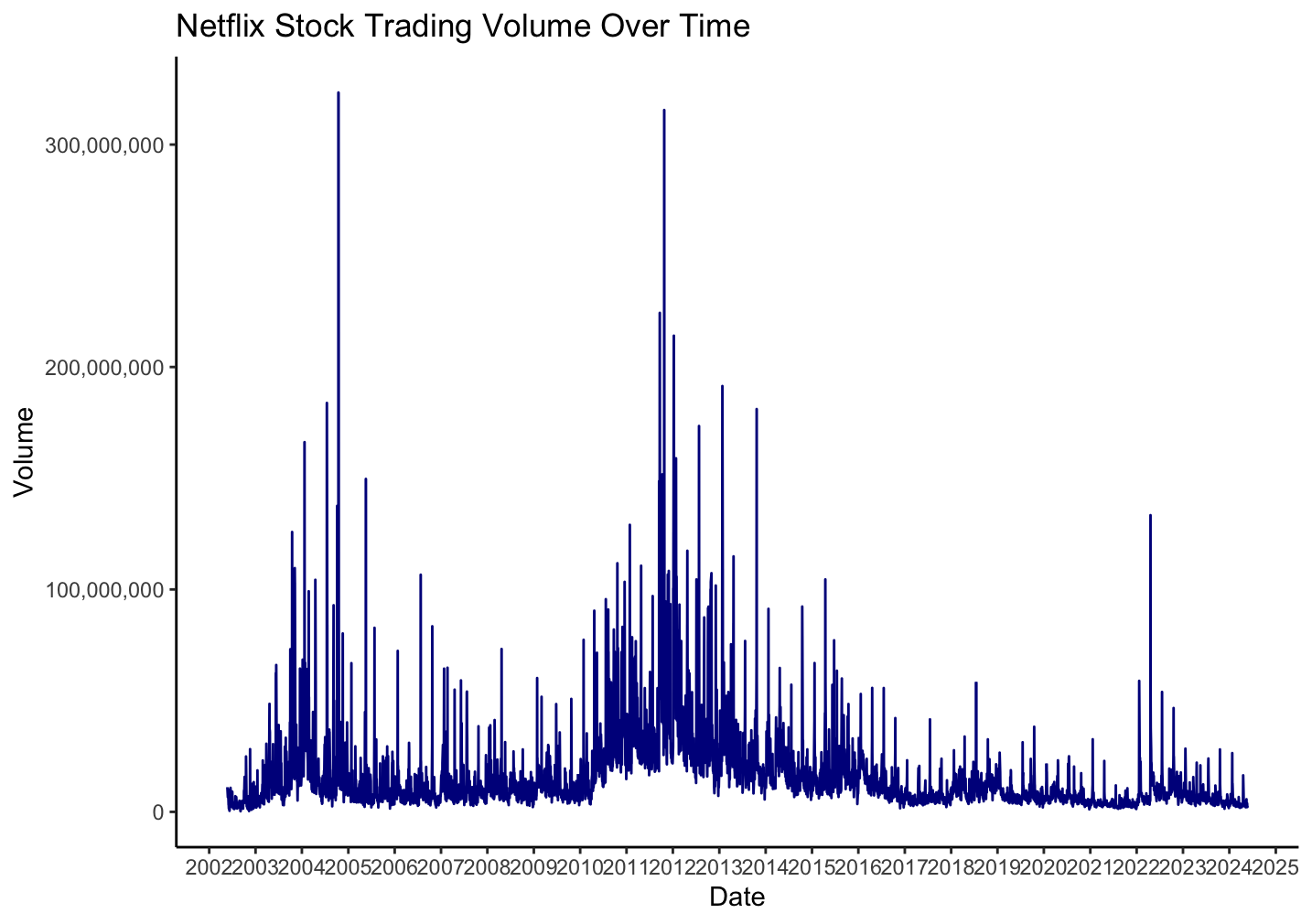
* Stability (2002 - Mid-2021): Closing prices showed stability with minimal fluctuations.
* Adverse Decline (Mid-2021 - 2022): A significant downturn occurred, suggesting negative market sentiment.
* Rise (End of 2022 - 2025): Gradual increase in closing prices indicated recovery and positive market sentiment



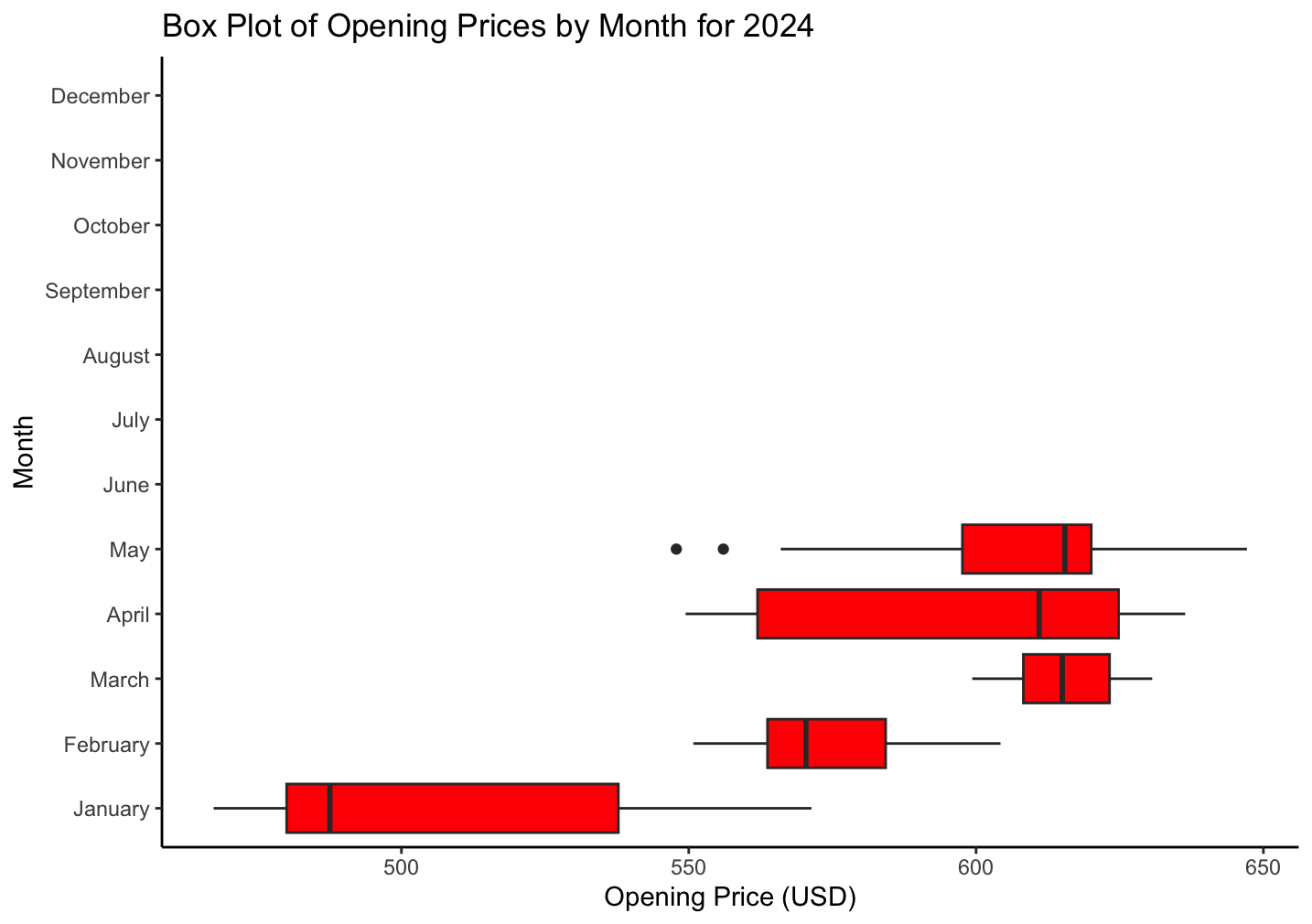
From 2022, Netflix's adjusted close prices initially declined but gradually increased toward the end of the year. This upward trend continued into 2023 and 2024, suggesting growing investor confidence. The pattern implies potential for improved performance in 2025, reflecting positive market sentiment and company outlook.

**Line plots for Netflix Stock Trading Volume Over Time**

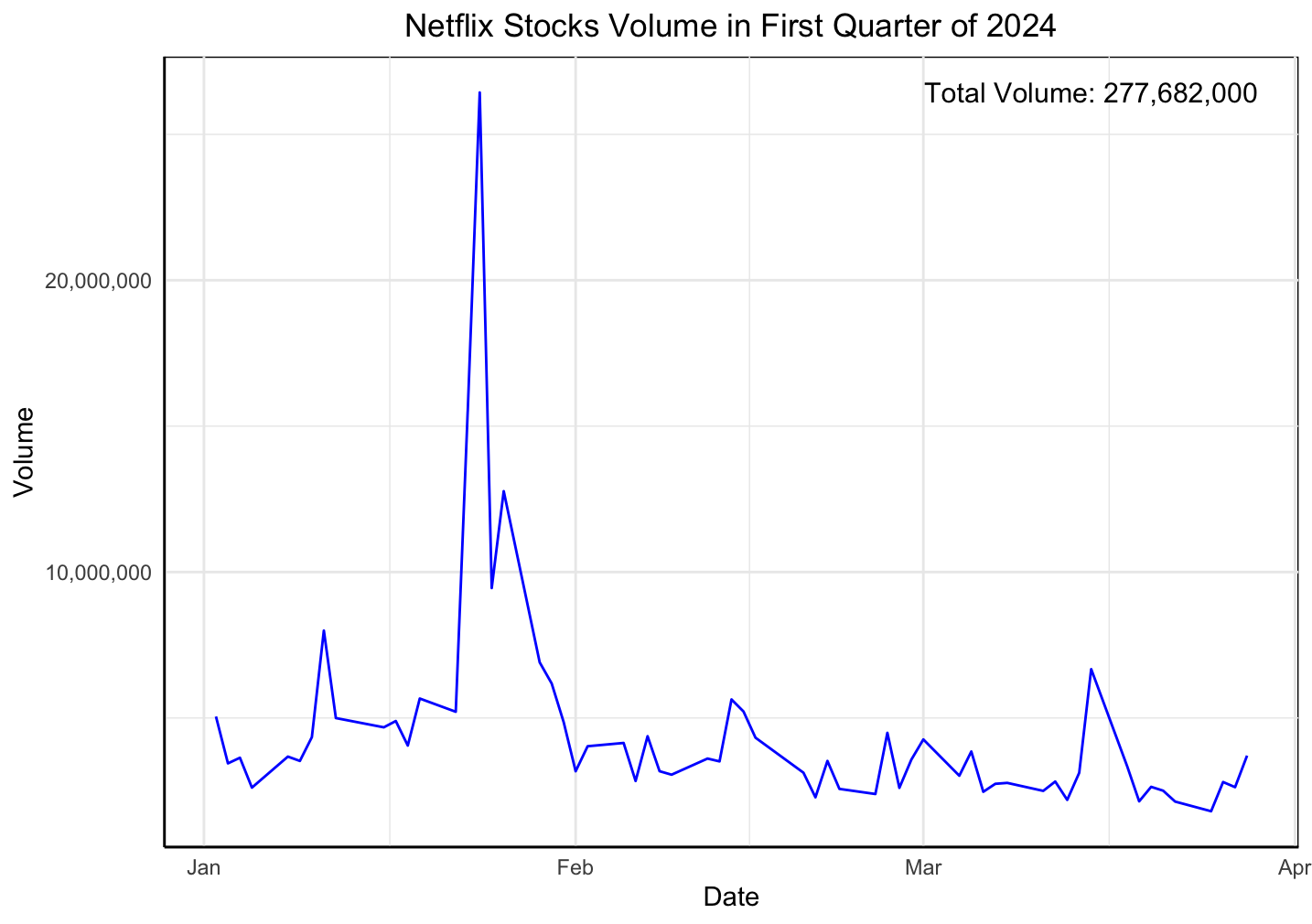
* Highest trading volumes occurred in 2024 and 2012, indicating significant investor interest and market activity.
* The second-highest volumes were observed in 2004, 2005, 2013, 2014, and the end of 2022, reflecting substantial market participation.
* Conversely, low volumes were noted in 2002 and from 2023 onwards to 2024, suggesting decreased investor interest or market uncertainty.



**Plotting Box plot of opening prices by month for the year 2024**



**Analyzing Netflix Stocks Volume in First Quarter of 2024**



During January and February, the volume of Netflix stocks traded is high, followed by a downward trend afterward. This pattern suggests heightened market activity initially, potentially driven by subsequent decline in volume indicates reduced trading activity or decreased investor interest.

Conclusion:-

In this Module 1 project, the focus was on understanding Netflix's stock price trends over time. The process began with checking and adjusting the data to make sure it was accurate. Then, we explored how the prices changed over the years and months, identifying patterns in trading activity. Additionally, we calculated the daily changes in stock prices to better understand market movements. By visualizing the data using charts, we were able to communicate these trends clearly, providing insights for investors and decision-makers in the financial market.

Reference :

* Redirecting. (n.d.). <https://northeastern.instructure.com/courses/174180/modules>
* https://www.kaggle.com/datasets/mayankanand2701/netflix-stock-price-dataset/data