International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Leveraging R for Analyzing and Visualizing Stock Market Data

PARTHIBAN.S[1], RAMACHANDRAN.K[2]

[1]*AVUDAI SK INFOTECH SOLUTIONS, AMBATTUR, CHENNAI, https://skinfotechsolutions.com/*
[2]*AVUDAI SK INFOTECH SOLUTIONS, AMBATTUR, CHENNAI, https://skinfotechsolutions.com/*

---***---

Abstract - R is a powerful tool for analyzing and visualizing share market data, offering capabilities for data acquisition, cleaning, and advanced analytics like time-series forecasting and sentiment analysis. This article highlights R's strengths in identifying market trends, predicting price movements, and creating dynamic visualizations that support informed decision-making. By showcasing case studies and practical applications, it demonstrates how financial professionals can leverage R to gain actionable insights and optimize strategies in the ever-changing landscape of the financial markets.

Key Words: Data Analytics, marketing, dynamic visualizations, trends and time-series.

1.INTRODUCTION

In this report, R is applied to examine stock market returns of the specific set of firms as well in the short term, which is 30 days and in the long term, which is 365 days. The goal is to investigate the fluctuation of stock prices, correlation between volume in stock exchange and price changes and finally identify good bargains for investments. The analytical tools used in the methodology include cleaning of data, use of basic statistical measures, correlation measures, and charts. A hypothetical portfolio is also constructed as a guide in the application of the presented findings in arriving at sound investment decisions.

1.1 Objectives

This script is an R script that analyses historical stock market data, mainly stock price, stock fluctuation, and their returns over a period of some years for some companies. It measures relevant movement and stock price fluctuations focusing on the most recent 30 days as well as the past 365 days. The general picture of the script is portfolio construction using 30-day changes in prices and activities such as data preparation, descriptive statistics, volatility checks, correlation matrix, and visualization (histograms, box and whisker plots, line graphs, scatter plots, heat plots). The objective is to gain knowledge and understand Stock trends and the relation between volume and price to be used in the investment process.

2. LITERATURE REVIEW

Bhandari et al., (2022). This work focuses on proposing a machine learning approach using Long Short Term Memory (LSTM) to predict the closing price of S&P 500 index in the forthcoming day. The study utilizes nine predictors including technical, macroeconomic and market characteristics single-layer and multilayer LSTM models. The performance of the presented models is evaluated using the coefficient of determination, root mean square error and mean absolute percentage error. Therefore, the results reveal that for the volatility and modelling of nonlinear stock market fluctuations, the single-layer LSTM model is superior to the multi-layer LSTM model.

3. METHODOLOGY

To achieve this research objective, this study employs an observational, quantitative research approach to evaluate trends and performances in the stock exchange based on data retrieved from the NIFTY 50 CSV file of the Market Watch known as MW-NIFTY-50-24-Dec-2024. This data set covers such values as prices, volume and percentage change in 30- and 365-days averages. The data analysis uses R programming language with the help of dplyr, ggplot2, tidyr, and reshape!2 for data operation, graphical display and statistical assessment within the context of market behavior and trends.

- ***Data Preparation:*** The dataset was also read and prepared on the raw format, column names were also made universal and the character type features were made into numerical for uniformity.
- ***Descriptive Statistics:*** Descriptive statistics included use of mean, median, and quartiles done with the summary() function on R.
- ***Volatility Analysis:*** To measure volatility, the high-low price difference method was used to determine the.User's Generate most volatile stocks in the data set.
- ***Correlation Analysis:*** Pearson correlation coefficients quantified partnership between them such as trade volume, differences in prices, percentage differences for certain periods.
- ***Statistical Calculations:*** Descriptive statistics of 30-day and 365-day changes were calculated from statistical functions in the statistical software R.
- ***Data Visualization:*** Trends of the data were deducted from histograms, boxplots and lines, scatter plot and heat map were generated using ggplot.
- ***Ranking and Filtering:*** The stocks were then sorted and selected based on some factors including percent changes and nearness to the 52 vars lows or highs.
- ***Portfolio Simulation:*** To further analyze and arrange the data, a hypothetical portfolio with 30-day price changes was assumed and simple returns and data organization were performed by dplyr functions.

4. RESULTS

4.1 Histogram

The histograms represent the 30-day and 365-day stock changes on percentage basis, respectively. This is the nature of fluctuations for a 30-day period ranging from approximately -40 percent to 10 percent, although the majority of stocks are within a range that shows little movement at +/- 0 percent. Smaller bars define a plethora of gains and losses, while large negative ones are observed only near -40%. The histogram that covers the area of 365 days which is from -100% to 120% is also oriented around 0% but with greater fluctuation. There are stock prices which have recorded one hundred percent drop and at the same time recorded one hundred and twenty percent gain. This suggests of more fluctuations in the stock performance over the year than in the 30 days period.

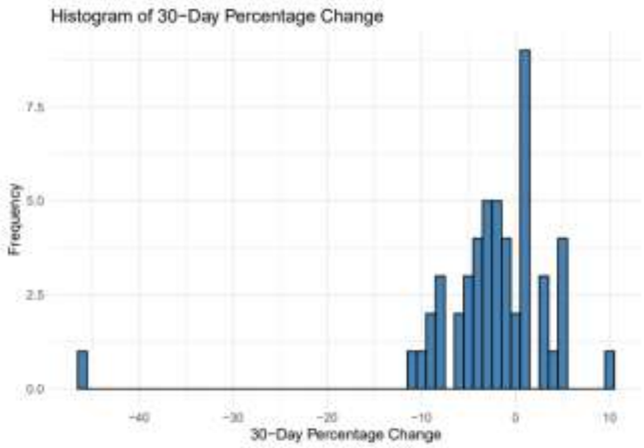

Figure 1: Histogram of 30-day percentage change

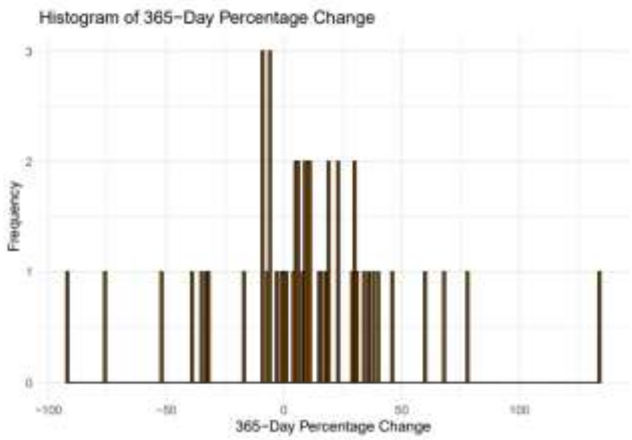

Figure 2: Histogram of 365- day percentage change

4.2 Boxplot

The first figures below the scattered boxplots illustrate the distribution of the daily 30-day and 365-day % changes of stocks. The boxplot for the absolute changes for the 30-day cohort are depicted below By definition, the chronological middle is slightly below 0%, and 75% of all fluctuation are between -5% as well as + 5%. The above table also presents a variety of outliers with one value, a little over 10% and another around -45 %. The 365-day boxplot of the changes approximate the zero percent and majority of the changes range between -20% and +30%. It is more erratic than the compounded return and has outlier figures that depict an operating loss of about -80 percent, and an operating profit of about 120 percent. This is true because, the variability observable in period 365 clearly shows more expansion than the variability of the 30 percent changes which illustrate fluctuations in stock performance within one year.

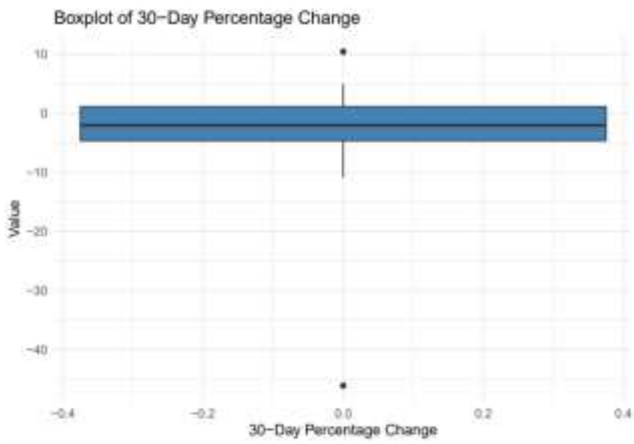

Figure 3: Boxplot of 30-day percentage change

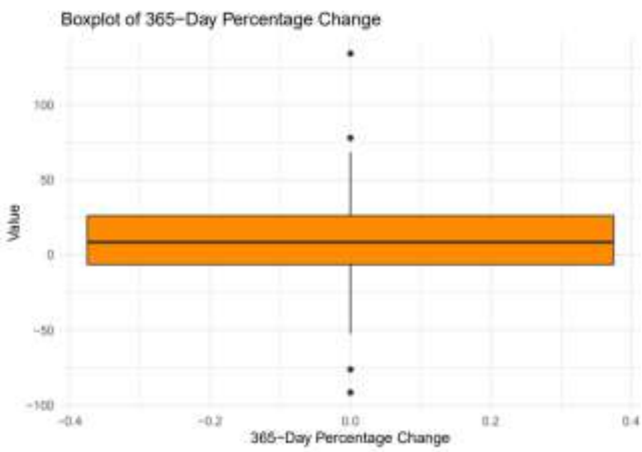

Figure 4: Boxplot of 365-day percentage change

4.3 Time series plot with a trend line

This is a time series plot of daily stock price by percentage of change over the last 30 days starting from January ending mid-February. Increased volatility is colour blue and fluctuates between -10% and + 10% being at -45% in middle of February. The black dashed line means a linear trend, for which the line turns slightly downwards, and the grey area is the confidence interval. Altogether the picture corresponds to fluctuations by days and shows a gradual deteriorating tendency with a significant contribution made by the period of the middle of February.

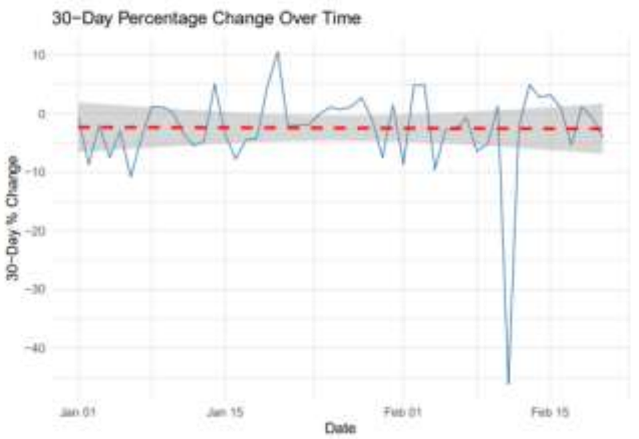

Figure 5: Time series plot with trend line (30-day percentage change over time)

The displayed time series graph represents daily relative changes in stock by 365 after January till the middle of February. The blue line is greatly oscillating, nearing -80% and exceeding 120%, with a sudden rise and fall in mid of January. There is a very slight positive slope shown by the indicate by the red dashed line that is heavily masked by significant variance as depicted by the wide confidence interval. This has the effect of flattening the actual trend due to such Figures below depict that the 365-day changes are way more abrupt than the 30-day differences.

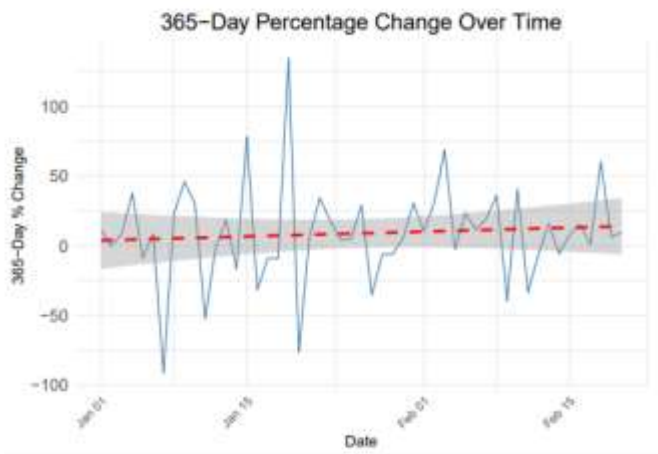

Figure 6: Time series plot with trend line (365-day percentage change over time)

4.4 Volatility analysis

This bar chart shows the percentage change of 30 days for different stock types based on their changes from the highest to the lowest. The y-axis provides stock symbols whereas the x-axis portrays the correspondent 30-day percent change. An increase is shown as bars pointing to the right while a decrease pointing to the left. The chart shows that the positive trend has changed steeply from reaching a maximum of around +10% (DRREDDY) to going down to nearly -40% (WIPRO) with rare stocks lying just around 0%. Indeed, it presents a recent movement of stock prices, inclusive both the rise and the drop, within the last 30 days.

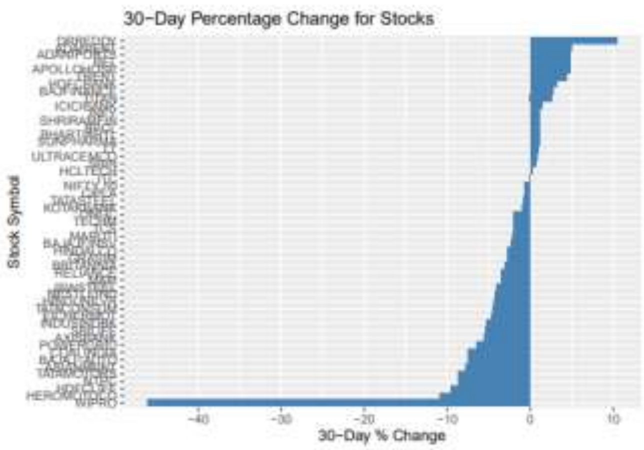

Figure 7: 30 - day percentage change for stocks

4.5 Scatter plots

This scatterplot analyzes the nature of trading volume (measured on the logarithmic scale) and price change in stocks. Variability of orange dots in the low volume area: price changes vary from +30 to -60 and are located in quantity at all four trading zones. A case of an outlier at high volume level resulted to very minimal price change (-0.25) thus demonstrating that high crude oil trading does not necessarily mean large price differentials. All in all, this scatterplot indicated that there is little sign of linearity to the plot, which means that the correlation of trading volume and price change is low.

Figure 8: Volume Vs price change

4.6 Heatmap

This section presents percentage changes for stocks in a table with 30 days and 365 days periods, where intensity colors show positive (blue) values, close to zero (light) and negative (purple). The heatmap plot presents low 30-day changes in percentage (as represented by the light colors) with limited variations for most stocks. However, the relative changes for each day within the individual year scatter more and include large positive (blue) as well as negative (purple) percentages. This points to low intra daily fluctuations and high inter daily fluctuations of stocks necessary to assess variations in performance over long term.

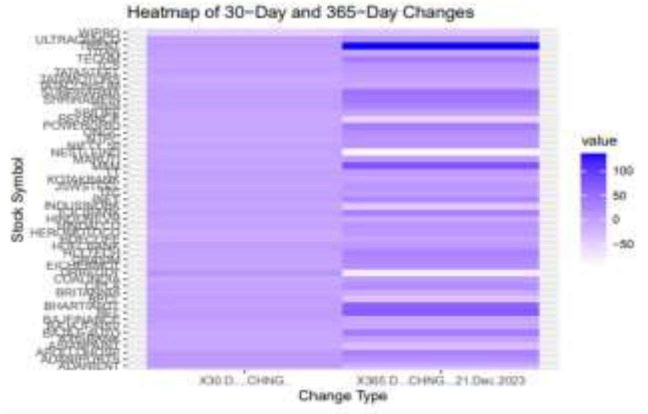

Figure 9: Heatmap of 30-day and 365-day changes

5. DISCUSSIONS

This analysis of the stock market shows the peculiarities of the market dynamics using various quantitative and graphical approaches. Some of the main analytical insights revealed gross variations between the 30 day and 365-day trends. The data range obtained in last 30 days shows relatively less fluctuation and a slight up and down movement while the range obtained in last 365 days shows higher fluctuation and high movement up and down. Financial ratio analysis focuses on change in prices and individual securities, and likewise, time series and bar charts are used to present the changes in the stock prices. The next heatmap shows the daily/ monthly fluctuations in the short run and annual fluctuations in the long run; it shows less fluctuation in the short run. In More detail, the consideration of short long and intermediate horizons proves relevant coupled with the use of quantitative measures alongside qualitative analysis with regard to stocks.

6. CONCLUSION

From the stock market data analysis of stocks, the present work finds out that there are little changes within 30 days – these are almost insignificant- while changes within 365 days are very high –these are highly significant-. The measures of dispersion for short-term periods were characterized by narrow ranges, while the long-term variations were modeled using wider ranges. Yearly data were depicted as versus daily oscillations presenting moderate daily changes as opposed to highly increased or decreased 365-day values. Imprecise cyclical associations were depicted by bar charts and scatterplots; however, heatmaps demonstrated a strong regularity of short-term stability and significant variability of long-term performance. In all, the study compares short- and long-term stocks behaviours that can be useful in investments through statistical analysis and charts.

REFERENCES

1. Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, R. K. (2022). Predicting stock market index using LSTM. Machine Learning with Applications, 9, 100320. https://doi.org/10.1016/j.mlwa.2022.100320
2. Gandhmal, D. P., & Kumar, K. (2019). Systematic analysis and review of stock market prediction techniques. Computer Science Review, 34, 100190. https://doi.org/10.1016/j.cosrev.2019.08.001
3. Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2019). A systematic review of fundamental and technical analysis of stock market predictions. Artificial Intelligence Review, 53(4), 3007-3057. https://doi.org/10.1007/s10462-019-09754-z
4. Bhowmik, R., & Wang, S. (2020). Stock market volatility and return analysis: A systematic literature review. Entropy, 22(5), 522. https://doi.org/10.3390/e22050522
5. Mai, Z., Nawaz Saleem, H. M., & Kamran, M. (2023). The relationship between political instability and stock market performance: An analysis of the MSCI index in the case of Pakistan. PLOS ONE, 18(10), e0292284. https://doi.org/10.1371/journal.pone.0292284
6. Budiharto, W. (2021). Data science approach to stock prices forecasting in Indonesia during COVID-19 using long short-term memory (LSTM). Journal of Big Data, 8(1). https://doi.org/10.1186/s40537-021-00430-0
7. Takács, A., Ulbert, J., & Fodor, A. (2019). Have investors learned from the crisis? An analysis of post-crisis pricing errors and market corrections in US stock markets based on the reverse DCF model. Applied Economics, 52(20), 2208-2218. https://doi.org/10.1080/00036846.2019.1686114
8. Li, Y., Wang, S., & Zhao, Q. (2021). When does the stock market recover from a crisis? Finance Research Letters, 39, 101642. https://doi.org/10.1016/j.frl.2020.101642
9. Marbun, W., Suparti, & Maruddani, D. A. (2020). Modeling of composite stock price index (CSPI) using semiparametric regression truncated spline based on GUI R. Journal of Physics: Conference Series, 1524(1), 012096. https://doi.org/10.1088/1742-6596/1524/1/012096
10. Pang, X., Zhou, Y., Wang, P., Lin, W., & Chang, V. (2018). An innovative neural network approach for stock market prediction. The Journal of Supercomputing, 76(3), 2098-2118. https://doi.org/10.1007/s11227-017-2228-y