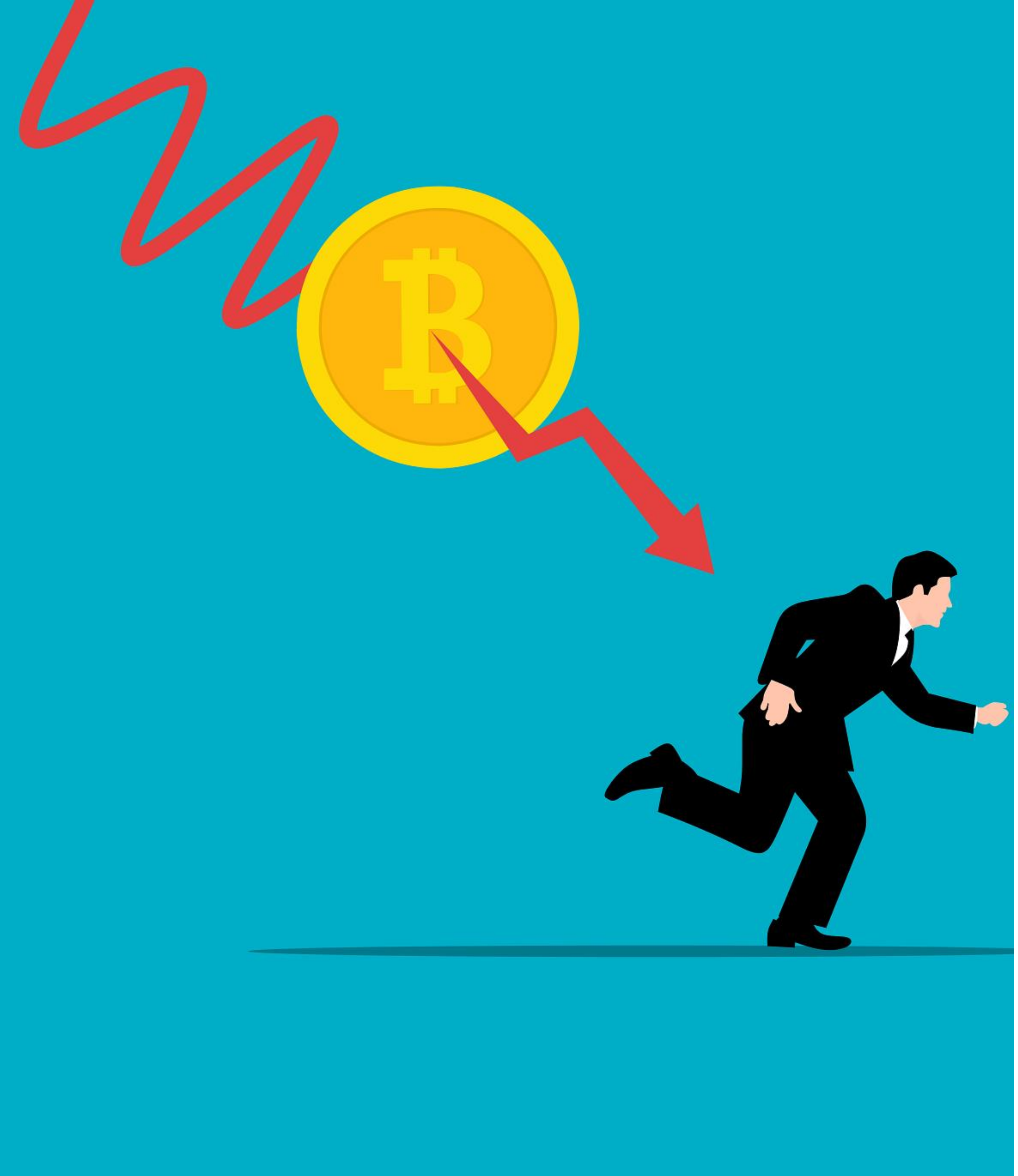
Predictive Analysis of Market Movements in Stock Markets: Prediction of Trade Volumes using Stock and Inflation Data for the Stock Market in the US

Akshaya Reddy Ragi
MSDS Data Science
Practicum -1

Introduction/Background

- Presently, the business environment is fast paced and dynamic; a consequence of advances in technology, which have improved the pace at which business activities are conducted across the world (Zhang et al., 2023; Bindeeba et al., 2025).

- For investors in stock markets, the dynamic business environment implies more dynamic movements in the stock markets (Pancić et al., 2023). Market movements being representative of market reaction hence become among the best placed for consideration for investment in the stock markets.



Trade Volumes

- Trade volumes are indicators of market movements in terms of confidence and reaction of the markets (Yamani, 2023; Care & Cumming, 2024).

- Thus, predictive analytics for trade volumes would be important in giving direction in market reaction and market confidence in the present dynamic markets.

Research Methodology

- Data for the project was collected from through the following approaches:
  - API Data Access
  - Web scrapping

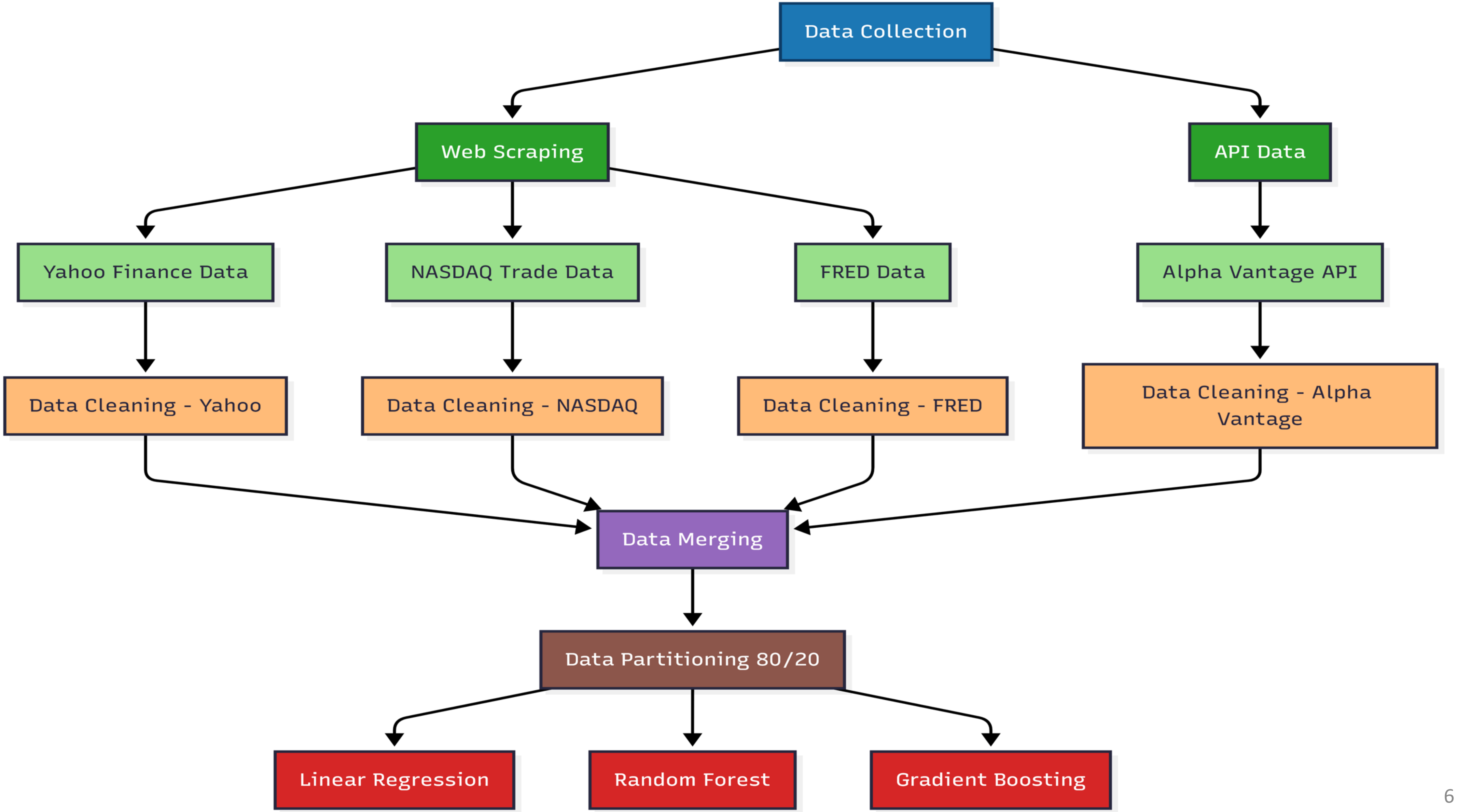
- Under API Data access, data was extracted from the Alpha Vantage API while web scraping was employed for the following sources:
  - Yahoo Finance
  - Nasdaq Trader
  - FRED

- Predictive analytics offers an approach for modelling where a know combination of output and inputs trains a model for the determination of the output given a completely new set of inputs (Jamarani et al., 2024; Kaur et al., 2024).

- For the predictive analytics of trade volumes, the following three dynamic models will be employed:
  - Linear Regression model
  - Random Forest model
  - Gradient Boosting model

Project Workflow - Logic

Project Workflow – Data Pipeline



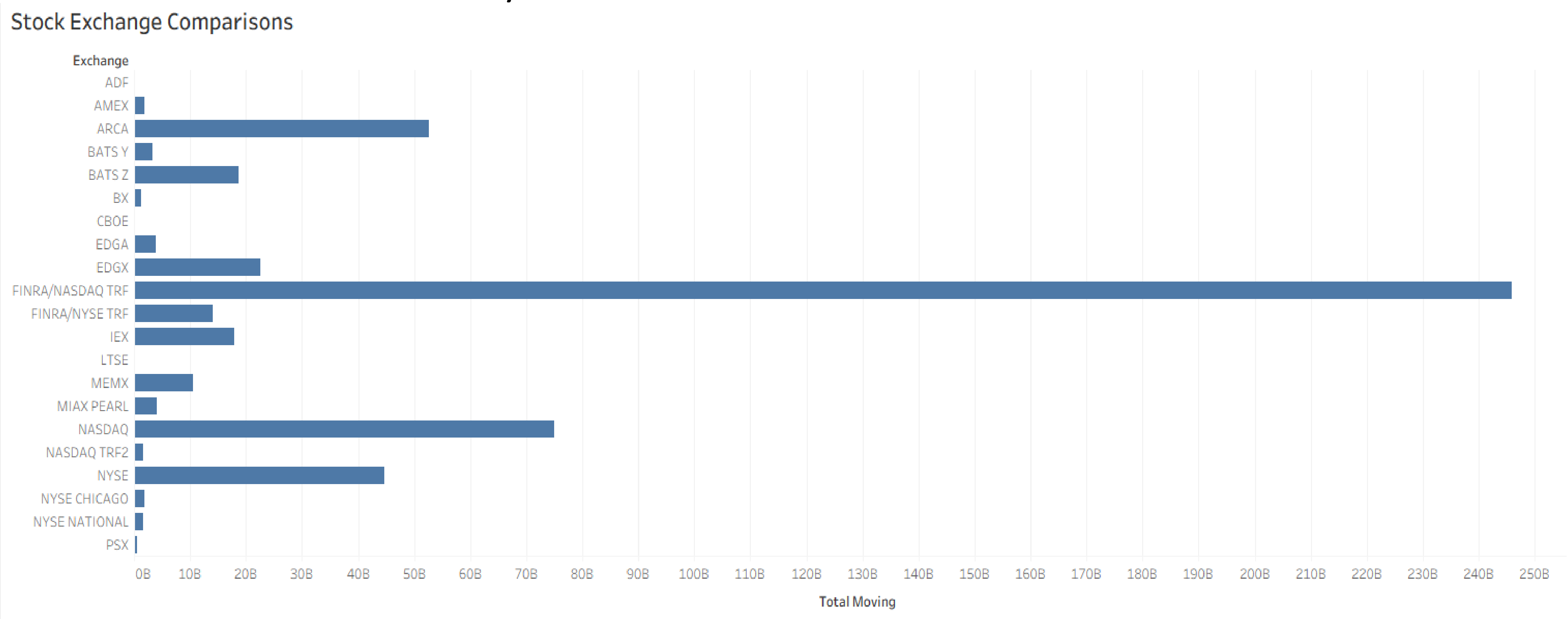
Predictive Analytics Models

- Azure (2024) notes that an equation based approach to predictive analytics is the linear regression; which describes the equation as the output being a linear function of the inputs.

- The random forest amplifies the classification tree through employing several decision based algorithms that split conditions for inputs in a tree like manner until all inputs are exhausted and the output predicted at the base (Ojo & Ayodele, 2025).

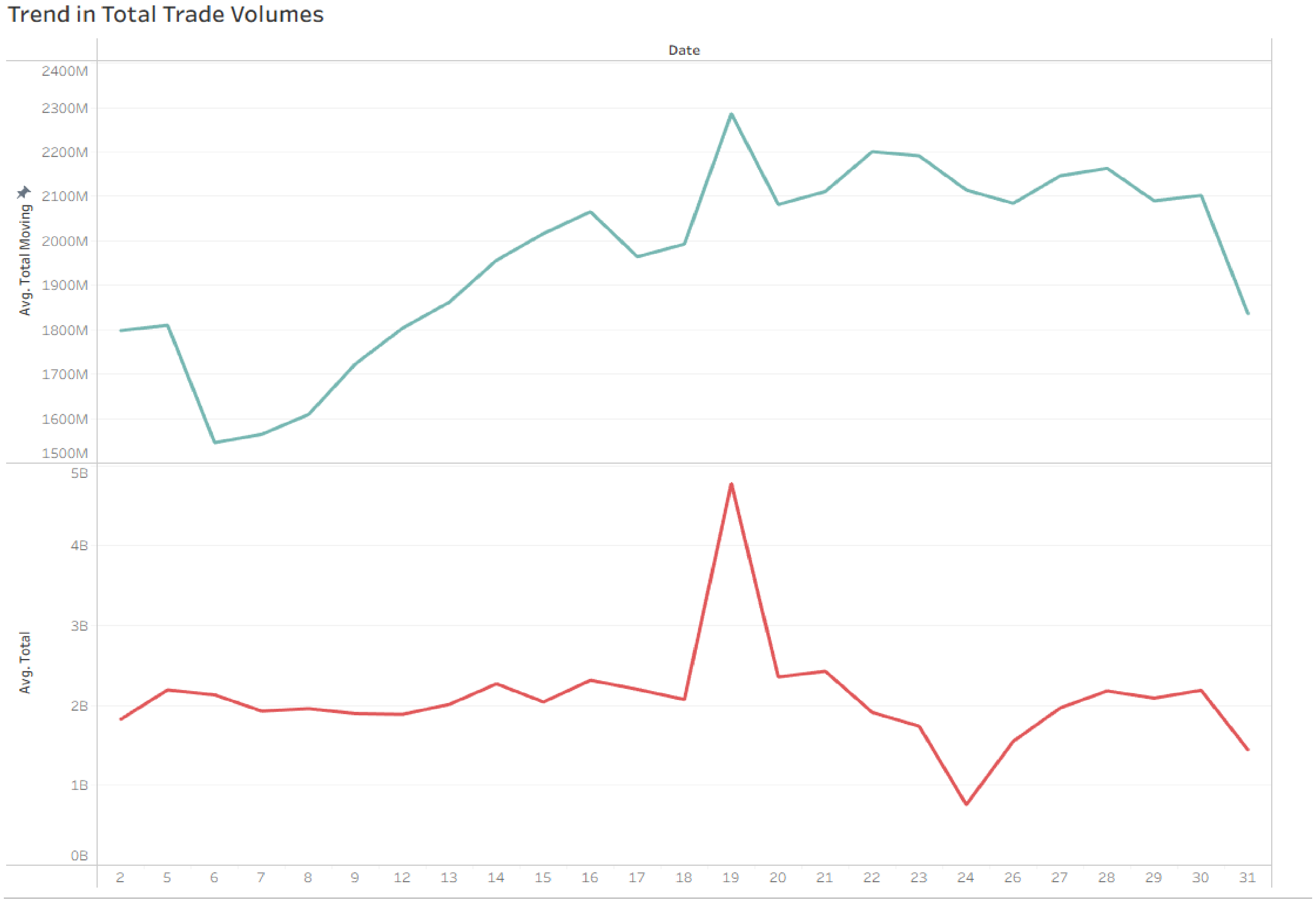
- Gradient boosting is explained in Rizkallah (2025) as an ensemble tool that trains weak predictive models sequentially with each new model learning from the previous model’s residuals; and combines these models to form a predictive tool.

Trade Volumes Analysis



- The comparison in the volumes traded by different stock exchanges over the past 30 days is presented in the bar chart above.
- I note that majority of the volumes traded were in FINRA/NASDAQ TRF Exchange, followed by the NASDAQ, ARCA and NYSE Exchanges.

Trade Volumes Analysis



- The plot on the left shows the 30 day trend for the traded volumes across the exchanges in the US.
- The average moving volumes show a general rising trend as the month progresses while the average total volumes generally fluctuates about the 2 billion shares mark with a notable spike (day 19) and drop (24) as the month progresses.

Modelling Approach

- The final data after cleaning contained the following variables:
  - Trade Volumes
  - Adjusted Close
  - Stock Exchange Name (converted to dummies)
  - FF (Representing Interest Rates)
  - Regular Market Volumes
  - Regular Market Open
  - Regular Market Change Percentage
  - Market Capital

- Data preparation for the modelling involved partitioning the data into train and test with 80% going to the former and 20% to the latter.

- Two models were trained and evaluated for the linear regression, random forest and gradient boosting models:
  - Baseline Model
  - Tuned Model

Linear Regression Model

I note from the baseline model, that the mean squared error was equal to 1402172245462.9373 while the R-Squared was equal to 0.8632. Hence the model was able to explain 86.32% of the variation in the trade volumes.

On the other hand, from the tuned model, that the mean squared error was equal to 1427702554271.3047 while the R-Squared was equal to 0.8606.Hence the tuned model was able to explain 86.06% of the variation in the trade volumes.

Random Forest Model

I note from the baseline model that the mean squared error was equal to 1135919899657.5408 while the R-Squared was equal to 0.8892. Hence the model was able to explain 88.92% of the variation in the trade volumes.

For the tuned model, the mean squared error was equal to 745364017404.2378 while the R-Squared was equal to 0.9273. Hence the tuned model was able to explain 92.73% of the variation in the trade volumes.

Gradient Boosting Model

I note for the baseline model that the mean squared error was equal to 1039289046245.1575 while the R-Squared was equal to 0.8986. Hence the model was able to explain 89.86% of the variation in the trade volumes.

The tuned model showed that the mean squared error was equal to 779681848900.6945 while the R-Squared was equal to 0.9239. Hence the tuned model was able to explain 92.39% of the variation in the trade volumes.

Key Insights

Linear Regression

Model	MSE	R-Squared
Baseline Model	1402172245462.9373	0.8632
Tuned Model	1427702554271.3047	0.8606

Random Forest

Model	MSE	R-Squared
Baseline Model	1135919899657.5408	0.8892
Tuned Model	745364017404.2378	0.9273

Gradient Boosting

Model	MSE	R-Squared
Baseline Model	1039289046245.1575	0.8986
Tuned Model	779681848900.6945	0.9239

Conclusion

- Observing the R-Squared across the baseline and the tuned models across all three models, I note that the predictive performance exceed 85%; hence Adjusted Close, Stock Exchange, FF (Representing Interest Rates), Regular Market Volumes, Regular Market Open, Regular Market Change Percentage and Market Capital can sufficiently predict the trade volumes for the market movements.

- The best model for the prediction of market movements would be the tuned Random Forest with a prediction performance of 92.73%.

References

Azure, I. (2024). *Predictive modeling for industrial productivity: Evaluating linear regression and decision tree regressor approaches*. Journal of AppliedMath, 2(4), Article 1435. https://doi.org/10.59400/jam.v2i4.1435

Bindeeba, D. S., Tukamushaba, E. K., & Bakashaba, R. (2025). *Digital transformation and its multidimensional impact on sustainable business performance: evidence from a meta-analytic review*. Future Business Journal, 11, Article 90. https://doi.org/10.1186/s43093-025-00511-z

Carè, R., & Cumming, D. J. (2024). *Technology and automation in financial trading: A bibliometric review*. Research in International Business and Finance, 71, Article 102471. https://doi.org/10.1016/j.ribaf.2024.102471

Jamarani, A., Haddadi, S., Sarvizadeh, R., Haghi Kashani, M., Akbari, M., & Moradi, S. (2024). *Big data and predictive analytics: A systematic review of applications*. Artificial Intelligence Review, 57, Article 176. https://doi.org/10.1007/s10462-024-10811-5

Kaur, B., Sidhu, B. K., & Bhathal, G. S. (2024). *Predictive analytics models for commodity market: A literature review*. International Journal of Intelligent Systems and Applications in Engineering, 12(3), 4525–4535. https://doi.org/10.17762/ijisae.v12i3.7958

Ojo, G. O., & Ayodele, E. (2025). *Evaluating the predictive power of random forest regression on economic growth*. Federal Polytechnic Ilaro Journal of Pure and Applied Sciences, 7(1). https://fepi-jopas.federalpolyilaro.edu.ng/index.php/journal/article/view/165

Pancić, M., Ćućić, D., & Serdarušić, H. (2023). Business Intelligence (BI) in Firm Performance: Role of Big Data Analytics and Blockchain Technology. Economies, 11(3), 99. https://doi.org/10.3390/economies11030099.

Rizkallah, L. W. (2025). *Enhancing the performance of gradient boosting trees on regression problems*. Journal of Big Data, 12, Article 35. https://doi.org/10.1186/s40537-025-01071-3

Yamani, E. (2023). *Return–volume nexus in financial markets: A survey of research*. Research in International Business and Finance, 65, Article 101910. https://doi.org/10.1016/j.ribaf.2023.101910

Zhang, F., Yang, B., & Zhu, L. (2023). Digital technology usage, strategic flexibility, and business model innovation in traditional manufacturing firms: The moderating role of the institutional environment. Technological Forecasting and Social Change, 194, Article 122726. https://doi.org/10.1016/j.techfore.2023.122726