KEITHSTON FASHION

STOCK PRICE PREDITION OF NVIDIA

As we provide our yearly report on digital marketing for 2024, we go into the exciting fashion world. This report showcases our brand's journey in the digital sphere and captures a year of innovation, creativity, and strategic skill.

TABLE OF *Contents*

01 INTRODUCTION
02 GOALS & OBJECTIVES
03 STEP 1: DATA PREPROCESSING
04 STEP 2: EDA
07 STEP 3: MODEL DEVELOPMENT
10 STEP 4: MODEL EVALUATION
11 STEP 5: ENTIRE DATASET
14 FORECAST OF THE BEST MODELS & CONCLUSION

01

Introduction

In the ever-evolving landscape of the financial markets, understanding and predicting stock price movements have become pivotal for investors, traders, and analysts. The stock market is influenced by a myriad of factors including economic indicators, company performance, industry trends, and global events, making the prediction of stock prices a challenging yet vital task. Among the multitude of companies traded publicly, Nvidia Corporation stands out due to its significant impact on the technology sector, especially in areas like gaming, automotive, data centers, and artificial intelligence.

***Nvidia's** stock, represented by the ticker **NVDA**, offers a rich dataset for analysis due to its active trading volume and the company's innovative edge in the market. This report aims to dissect Nvidia's stock data to uncover underlying patterns, trends, and potential future movements. Through a comprehensive exploratory data analysis and the application of various forecasting models, we aim to provide insights into Nvidia's stock performance and predict its future price trajectory.*

02

Goals & Objectives

The primary goal of this report is to analyze historical stock price data for Nvidia (NVDA) and develop predictive models to forecast future stock prices. The objectives to achieve this goal include:

- **Data Preprocessing:** To prepare the stock price and volume data for analysis by ensuring it is clean, sorted, and formatted correctly, particularly focusing on the 'Date' column for time series analysis.
- **Exploratory Data Analysis (EDA):** To conduct a thorough analysis of Nvidia's stock, including trends in closing prices, trading volumes, and other relevant metrics such as moving averages. This step aims to uncover any underlying patterns or insights that could inform predictive modeling.
- **Model Development:** To employ various statistical and machine learning models, including linear regression, quadratic models, ARIMA (AutoRegressive Integrated Moving Average), and ETS (Error, Trend, Seasonality) models, for forecasting Nvidia's stock prices. This will involve splitting the data into training and testing sets to validate the models' performance.
- **Model Comparison and Evaluation:** To compare the performance of the developed models using appropriate metrics, such as the Mean Absolute Percentage Error (MAPE), to determine the most accurate and reliable method for predicting future stock prices.
- **Insights and Recommendations:** To provide actionable insights based on the analysis and model forecasts, including potential investment strategies and considerations for investors interested in Nvidia's stock.

03

DATA PREPROCESSING

1: Load and Read Data:
We start by loading the necessary libraries and reading the dataset from a CSV file into an R data frame named nvda_data. This step involves using read_csv() from the readr package, which is known for its efficiency in handling CSV files.

2: Date Column Conversion and Sorting
After loading the data, we convert the 'Date' column from a string format to a Date object using as.Date(). This conversion is essential for any time series analysis as it enables R to recognize and appropriately handle the dates for plotting, filtering, and modeling.

3: Handling Trading Volume Data
We preprocess the 'Vol.' column by removing the 'M' (which stands for million) and converting the values to numeric, multiplying by 1 million to reflect the actual trading volumes in numeric form. This conversion is crucial for any analysis involving volume, as it allows for quantitative operations and comparisons.

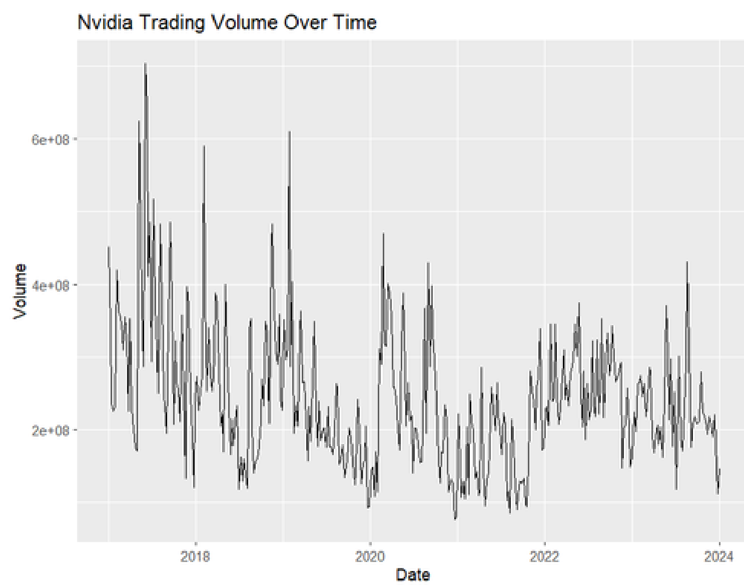
4. Additional Consideration :

- Checked for Missing Values
- Checked for Outliers
- Ensured Data Integrity

04

EDA

Performance



NVIDIA STOCK CLOSING PRICES OVER TIME

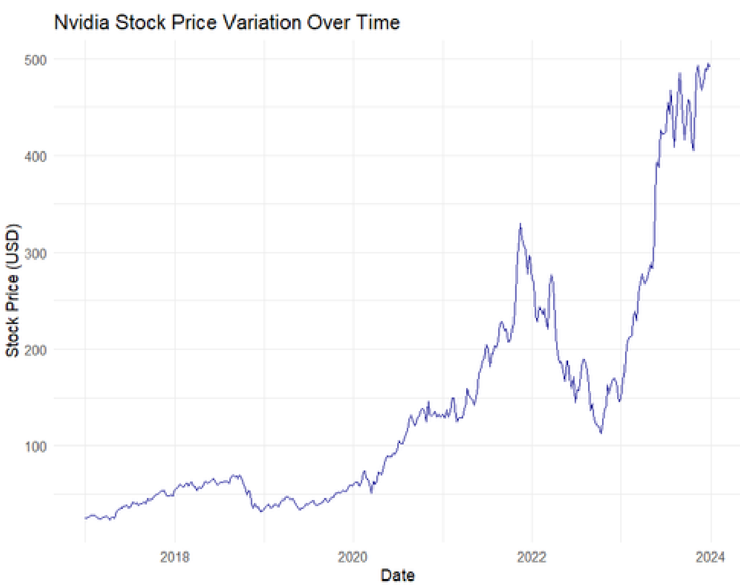
Analyzing the closing prices chart, we notice a pronounced surge in the stock price beginning around 2022. This spike can be technically attributed to Nvidia's substantial role in the AI industry. As AI technologies have advanced, there has been a growing demand for high-performance computing and GPUs, which are Nvidia's core products. Their chips are essential for deep learning and AI applications, which are increasingly used across various sectors such as automotive, healthcare, finance, and more. This period likely correlates with Nvidia's successful penetration into AI markets and possibly the release of new, high-demand GPU models or AI-driven technologies.

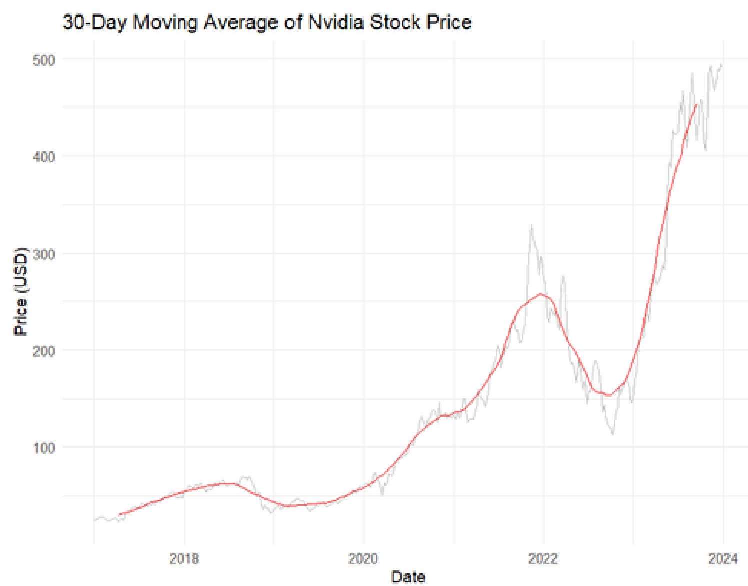
NVIDIA TRADING VOLUME OVER TIME

The trading volume chart shows spikes that could correspond with key events or product launches by Nvidia. A sudden increase in volume, especially if aligned with a price surge, might indicate strong investor confidence and public interest, potentially boosted by Nvidia's advancements in AI and their adoption in emerging technologies. Such peaks in trading volume often coincide with market optimism about Nvidia's growth prospects due to AI advancements.

EDA

Performance





NVIDIA STOCK PRICE VARIATION OVER TIME

This chart further emphasizes the dramatic uptick in Nvidia's stock price. Given Nvidia's strategic positioning in the AI industry, any news or predictions about breakthroughs in AI technology or lucrative partnerships could have contributed to this price variation. The company's commitment to AI and machine learning, combined with strong financial performance, could lead to heightened investor optimism, which is reflected in the stock's price trajectory.

30-DAY MOVING AVERAGE OF NVIDIA STOCK PRICE

The 30-day moving average chart during the spike period shows a clear upward trend, smoothing out the short-term volatilities and providing a more generalized view of the stock's momentum. Such a consistent upward trend in the moving average could be a sign that the company is steadily gaining ground, likely fueled by its successes and innovations in the AI space.

06

EDA

Performance

CANDLESTICK CHART OF NVIDIA STOCK

The candlestick chart provide the granular details of daily trading during the spike, revealing days with significant price ranges (high volatility) and strong closings (bullish sentiment). If this period reflects Nvidia's announcements of new AI-driven products or services, such bullish candlestick patterns is possibly a technical confirmation of positive market reception to Nvidia's AI initiatives.

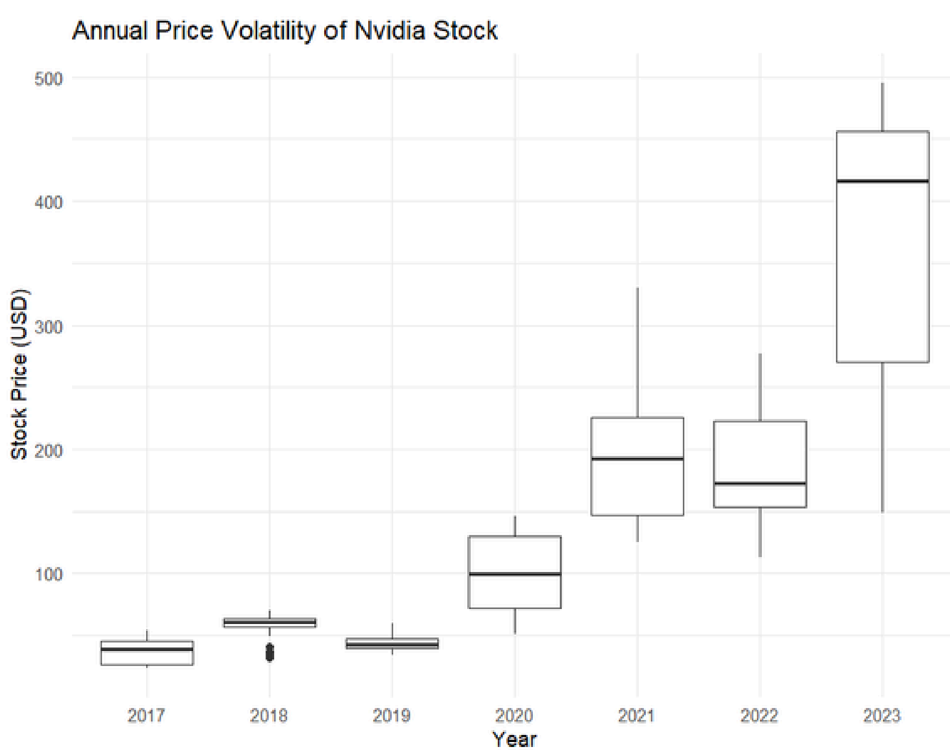
07

EDA

Performance

ANNUAL PRICE VOLATILITY OF NVIDIA STOCK

The box plot showing annual price volatility might reveal a broader range in 2022, suggesting increased variability in stock prices. This could be due to the fast-paced nature of the AI industry, where Nvidia is a key player. The upper quartile and whiskers extending upwards indicate that the stock hit particularly high prices during that year, which could be attributed to investors reacting to Nvidia's potential in the booming AI market.



In summary, Nvidia's prominence in the AI sector, primarily through its GPU and AI computing solutions, is likely a significant factor contributing to the observed spike in its stock price. The company's involvement in AI has not only diversified its product offerings but also positioned it as a leader in a high-growth industry. This strategic positioning likely bolstered investor confidence, as reflected in the price and volume charts, and helped drive the stock price to new heights around 2022.

Model Development

```
tslm(formula = train.ts ~ trend)

Residuals:
    Min      1Q  Median      3Q     Max 
-64.409 -34.201  -1.412  25.950 149.531 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -8.6459     4.6991   -1.84   0.0668 .  
trend         0.7410     0.0279   26.56   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 39.98 on 289 degrees of freedom
Multiple R-squared:  0.7094,	Adjusted R-squared:  0.7084 
F-statistic: 705.6 on 1 and 289 DF,  p-value: < 2.2e-16
```

LINEAR REG

The model summary shows a multiple r squared and an adjusted r squared at 70.94% and 70.84% respectively. The p value is less than 0.05. Overall, the model is a good fit and statistically significant and can be used for prediction.

```
tslm(formula = train.ts ~ trend + I(trend^2))

Residuals:
     Min       1Q   Median       3Q      Max 
-108.600  -18.530   -2.706   19.059  130.798 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 46.7246572  5.5908130   8.357 2.76e-15 ***
trend       -0.3928303  0.0884156  -4.443 1.27e-05 ***
I(trend^2)   0.0038831  0.0002932  13.242  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 31.57 on 288 degrees of freedom
Multiple R-squared:  0.8194,	Adjusted R-squared:  0.8181 
F-statistic: 653.3 on 2 and 288 DF,  p-value: < 2.2e-16
```

QUADRATIC REG

The quadratic model shows a multiple r squared and an adjusted r squared at 81.94% and 81.81% respectively. The p value is less than 0.05. The model is a good fit and statistically significant and can be used for forecasting.

```
ETS(M,A,N) 

Call:
 ets(y = train.ts, model = "ZZZ") 

  Smoothing parameters:
    alpha = 0.9999 
    beta  = 1e-04 

  Initial states:
    l = 27.1637 
    b = 0.4936 

  sigma:  0.0616

     AIC     AICc      BIC 
2567.495 2567.705 2585.861 
```

HOLT-WINTER'S

The HW model has the following options: multiplicative error, additive trend and no seasonality. The smoothing parameters are as follows: exponential smoothing constant alpha is 0.9999 and the smoothing constant for trend is 0.0001.

```
tslm(formula = train.ts ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-68.166 -32.473  -0.937  26.449 131.635 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.03036   18.06403   0.112    0.911    
trend        0.74449    0.03026  24.600   <2e-16 ***

Residual standard error: 43.17 on 238 degrees of freedom
Multiple R-squared:  0.721,	Adjusted R-squared:   0.66 
F-statistic: 11.83 on 52 and 238 DF,  p-value: < 2.2e-16
```

LIN W/ TREND + SEAS

The model has a multiple r squared and an adjusted r squared of 72.1% and 66% respectively. The model can still be a good fit. The p value is below 0.05 which shows that the model is statistically significant. This model can be used for prediction.

09

Model Development

```
tslm(formula = train.ts ~ trend + I(trend^2) + season)

Residuals:
    Min      1Q  Median      3Q     Max
-98.713 -15.045  -1.592  15.762 104.649

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.729e+01  1.413e+01   4.056 6.78e-05 ***
trend       -4.776e-01  9.222e-02  -5.179 4.77e-07 ***
I(trend^2)   4.185e-03  3.061e-04  13.671  < 2e-16 ***
season2     -2.660e+00  1.868e+01  -0.142    0.887
season3     -3.630e+00  1.868e+01  -0.194    0.846


Residual standard error: 32.35 on 237 degrees of freedom
Multiple R-squared:  0.844,	Adjusted R-squared:  0.8091
F-statistic: 24.19 on 53 and 237 DF,  p-value: < 2.2e-16
```

```
Series: train.ts
ARIMA(0,1,1)

Coefficients:
         ma1
      0.1348
s.e.  0.0570

sigma^2 = 66.15:  log likelihood = -1018.83
AIC=2041.65   AICc=2041.7   BIC=2048.99
```

QUAD W/ TREND + SEAS

The summary shows a multiple r squared and adjusted r squared of 84.4% and 80.91%. This means the model is a good fit. The p value is also below 0.05 which indicates a statistically significant model. This can be used for forecasting.

ARIMA

The auto arima model shows an arima with non seasonal components of (0,1,1). The coefficient of this model is only the ma1 which is 0.1348.

2 LEVEL FORECAST

We also created a 2 level forecast combining the quadratic model with trend and seasonality with the trailing ma forecast of its residuals.

Model Evaluation

	Model	MAE	RMSE	ME	MPE	MAPE	MASE	ACF1
1	Linear Regression	65.249	134.864	111.602	4.521	36.725	0.970	6.050
2	Quad Regression	-37.698	98.224	89.678	-33.184	44.725	0.967	8.648
3	Holt-Winter's	99.733	158.271	124.951	18.028	36.353	0.971	5.978
4	Lin Trend + Seas	62.418	134.804	114.059	2.905	38.548	0.972	6.323
5	Quad Trend + Seas	-50.803	104.122	89.774	-38.876	47.595	0.969	9.596
6	ARIMA	117.186	177.066	138.576	23.413	39.196	0.971	6.511
7	2 Level Forecast	8.428	90.815	81.245	-13.750	34.069	0.969	6.298

ACCURACY MEASURES

The models are evaluated using their forecasts in the validation set. The measures that are used during evaluation are the mean absolute percentage error (MAPE) and the root mean squared error (RMSE). From the accuracy measures, the model with the lowest MAPE and RMSE is the 2 level forecast of the quadratic model with trend and seasonality and its residuals. The model's MAPE is 34.07 and RMSE is 90.82.

Using the Models in the Whole Dataset

```
tslm(formula = nvda_ts ~ trend)

Residuals:
    Min      1Q  Median      3Q     Max
-141.62  -48.45  -11.45   41.40  187.24

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -34.73594    6.85608  -5.066 6.47e-07 ***
trend         0.95570    0.03256  29.355  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 65.27 on 362 degrees of freedom
Multiple R-squared:  0.7042,	Adjusted R-squared:  0.7034
F-statistic: 861.7 on 1 and 362 DF,  p-value: < 2.2e-16
```

LINEAR REG

The model summary shows a multiple r squared and an adjusted r squared at 70.42% and 70.34% respectively. The p value is less than 0.05. Overall, the model is a good fit and statistically significant and can be used for prediction.

```
tslm(formula = nvda_ts ~ trend + I(trend^2))

Residuals:
     Min       1Q   Median       3Q      Max
-154.842  -17.810   -0.575   23.442  144.506

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 56.1844489  8.1114552   6.927 1.99e-11 ***
trend       -0.5347985  0.1026269  -5.211 3.16e-07 ***
I(trend^2)   0.0040836  0.0002723  14.997  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

Residual standard error: 51.3 on 361 degrees of freedom
Multiple R-squared:  0.8177,	Adjusted R-squared:  0.8167
F-statistic: 809.8 on 2 and 361 DF,  p-value: < 2.2e-16
```

QUADRATIC REG

The quadratic model shows a multiple r squared and an adjusted r squared at 81.77% and 81.67% respectively. The p value is less than 0.05 as well. This model is a good fit and statistically significant and can be used for forecasting the stock price.

```
ETS(M,A,N)

Call:
 ets(y = nvda_ts, model = "ZZZ")

  Smoothing parameters:
    alpha = 0.9999
    beta  = 1e-04

  Initial states:
    l = 26.0492
    b = 0.5261

  sigma:  0.0629

     AIC     AICc      BIC
3486.518 3486.685 3506.004
```

HOLT-WINTER'S

The HW model has the following options: multiplicative error, additive trend and no seasonality. The smoothing parameters are as follows: exponential smoothing constant alpha is 0.9999 and the smoothing constant for trend is 0.0001.

```
tslm(formula = nvda_ts ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max
-139.98  -46.07  -12.04   41.60  176.27

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -42.16143   26.86793  -1.569    0.118
trend         0.94506    0.03507  26.951   <2e-16 ***
season2      -2.34077   37.19078  -0.063    0.950
season3      -0.66155   37.19083  -0.018    0.986

Residual standard error: 69.58 on 311 degrees of freedom
Multiple R-squared:  0.7112,	Adjusted R-squared:  0.6629
F-statistic: 14.73 on 52 and 311 DF,  p-value: < 2.2e-16
```

LIN W/ TREND + SEAS

The model has a multiple r squared and an adjusted r squared of 72.12% and 66.29% respectively. The model is still a good fit. The p value is below 0.05 which shows that the model is statistically significant. This model can be used to forecast the dataset.

Using the Models in the Whole Dataset

```
tslm(formula = nvda_ts ~ trend + I(trend^2) + season)

Residuals:
     Min       1Q   Median       3Q      Max
-153.154  -19.495    1.115   26.530  126.427

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept) 47.0813545 21.8816216   2.152   0.0322 *
trend       -0.5464068  0.1086580  -5.029 8.37e-07 ***
I(trend^2)   0.0040862  0.0002881  14.183  < 2e-16 ***
season2     -2.1364630 29.0094015  -0.074   0.9413
season3     -0.2610985 29.0094503  -0.009   0.9928

Residual standard error: 54.27 on 310 degrees of freedom
Multiple R-squared:  0.8248,	Adjusted R-squared:  0.7949
F-statistic: 27.54 on 53 and 310 DF,  p-value: < 2.2e-16
```

```
Series: nvda_ts
ARIMA(0,1,1) with drift

Coefficients:
         ma1   drift
      0.2021  1.2722
s.e.  0.0546  0.7004

sigma^2 = 124:  log likelihood = -1389.04
AIC=2784.08   AICc=2784.15   BIC=2795.77
```

QUAD W/ TREND + SEAS

The summary shows a multiple r squared and adjusted r squared of 82.48% and 79.49%. This means the model is a good fit. The p value is also below 0.05 which indicates a statistically significant model. This model can be used for forecasting.

ARIMA

The auto arima model shows an arima with non seasonal components of (0,1,1) and a drift component. The coefficient of this model is the ma1 which is 0.1348 and the drift coefficient which is 1.272.

2 LEVEL FORECAST

The 2 level forecast of the quadratic model with trend and seasonality and its residuals are used as well in the whole data set.

Final Model Evaluation

	Model	MAE	RMSE	ME	MPE	MAPE	MASE	ACF1
1	Linear Regression	0.000	65.089	52.381	-3.885	61.481	0.974	12.410
2	Quad Regression	0.000	51.091	35.978	-9.711	30.456	0.971	6.217
3	Holt-Winter's	0.742	11.311	6.736	-0.106	4.713	0.167	0.990
4	Lin Trend + Seas	0.000	64.313	51.648	-3.401	61.882	0.975	13.121
5	Quad Trend + Seas	0.000	50.084	35.199	-9.196	30.125	0.974	6.181
6	ARIMA	0.000	11.092	6.683	-0.917	4.833	-0.013	1.002
7	2 Level Forecast	5.420	17.507	13.875	2.796	18.569	0.823	3.973
8	SNaive	53.767	113.190	78.536	19.496	47.379	0.976	7.570

ACCURACY MEASURES

We used the models in the whole data set and calculated their accuracy measures. We included the seasonal naive forecast in the comparison so that we will know the baseline of our comparison. Given the accuracy measures, the best model to use for forecasting the NVDA stock price is the Holt-Winter's Model. This model has a MAPE of 4.713 and an RMSE of 11.311. Close second to this is the ARIMA model with a MAPE of 4.833 and RMSE of 11.092.

Forecasts of Best Models

HOLT-WINTER'S

```
Time Series:
Start = c(2024, 1)
End = c(2025, 52)
Frequency = 52
  [1] 488.8532 489.4063 489.9594 490.5125 491.0656 491.6187 492.1718 492.7249 493.2780 493.8311
 [11] 494.3842 494.9373 495.4904 496.0435 496.5966 497.1497 497.7028 498.2559 498.8090 499.3621
 [21] 499.9152 500.4683 501.0214 501.5745 502.1276 502.6807 503.2338 503.7869 504.3400 504.8931
 [31] 505.4462 505.9993 506.5524 507.1055 507.6585 508.2116 508.7647 509.3178 509.8709 510.4240
 [41] 510.9771 511.5302 512.0833 512.6364 513.1895 513.7426 514.2957 514.8488 515.4019 515.9550
 [51] 516.5081 517.0612 517.6143 518.1674 518.7205 519.2736 519.8267 520.3798 520.9329 521.4860
 [61] 522.0391 522.5922 523.1453 523.6984 524.2515 524.8046 525.3577 525.9108 526.4639 527.0170
 [71] 527.5701 528.1232 528.6763 529.2294 529.7825 530.3356 530.8887 531.4418 531.9949 532.5480
 [81] 533.1011 533.6542 534.2073 534.7604 535.3135 535.8666 536.4197 536.9727 537.5258 538.0789
 [91] 538.6320 539.1851 539.7382 540.2913 540.8444 541.3975 541.9506 542.5037 543.0568 543.6099
[101] 544.1630 544.7161 545.2692 545.8223
```

ARIMA

```
Time Series:
Start = c(2024, 1)
End = c(2025, 52)
Frequency = 52
  [1] 488.7446 490.0167 491.2889 492.5611 493.8333 495.1055 496.3777 497.6499 498.9221 500.1943
 [11] 501.4665 502.7387 504.0108 505.2830 506.5552 507.8274 509.0996 510.3718 511.6440 512.9162
 [21] 514.1884 515.4606 516.7328 518.0049 519.2771 520.5493 521.8215 523.0937 524.3659 525.6381
 [31] 526.9103 528.1825 529.4547 530.7269 531.9990 533.2712 534.5434 535.8156 537.0878 538.3600
 [41] 539.6322 540.9044 542.1766 543.4488 544.7210 545.9931 547.2653 548.5375 549.8097 551.0819
 [51] 552.3541 553.6263 554.8985 556.1707 557.4429 558.7151 559.9872 561.2594 562.5316 563.8038
 [61] 565.0760 566.3482 567.6204 568.8926 570.1648 571.4370 572.7092 573.9813 575.2535 576.5257
 [71] 577.7979 579.0701 580.3423 581.6145 582.8867 584.1589 585.4311 586.7033 587.9754 589.2476
 [81] 590.5198 591.7920 593.0642 594.3364 595.6086 596.8808 598.1530 599.4252 600.6974 601.9695
 [91] 603.2417 604.5139 605.7861 607.0583 608.3305 609.6027 610.8749 612.1471 613.4193 614.6915
[101] 615.9636 617.2358 618.5080 619.7802
```

CONCLUSION

Non-linear Trends: The superior performance of the Quadratic Regression model over the Linear Regression model indicates that NVIDIA's stock prices do not follow a purely linear trend. Instead, there are non-linear dynamics at play, which the quadratic term helps to capture more effectively.

Seasonality and Trend Components: The inclusion of seasonal adjustments in the models has generally improved forecasting accuracy. This finding underscores the presence of seasonal patterns in NVIDIA’s stock price movements, which are crucial for accurate forecasting. Models that can capture both trend and seasonality, like Holt-Winters and seasonally adjusted regressions, are more adept.

Forecasting with Holt-Winters and ARIMA: The Holt-Winters and ARIMA models have shown promising results, suggesting a gradual increase in NVIDIA's stock prices over the forecast period. The consistency in the upward trend prediction by these models highlights an optimistic outlook for NVIDIA's stock.

Model Selection for Future Forecasting: Between the two, the choice of model may depend on the specific requirements of the forecast (e.g., the importance of capturing seasonal patterns or the preference for a simpler model with fewer parameters like ARIMA). Each model has its strengths and can serve as a valuable tool for investors and analysts looking to forecast NVIDIA’s stock prices.

Handling Trading Volume Data Importance of External Factors: While the statistical models provide a methodological way of forecasting based on historical data, it is essential to acknowledge the impact of external factors. These include market conditions, technological advancements, regulatory changes, and NVIDIA’s own financial health and innovation pipeline. Such factors can significantly influence stock prices.

Recommendation for Investors and Analysts: Investors and analysts are advised to use these forecasting models as part of a broader investment strategy. The models can provide a statistical basis for understanding potential future trends but integrating this information with market analysis, financial performance indicators, and sector-specific news will yield the most robust investment decisions.

Thank You

Shriya Arora

Julian Honrado

Pallavi Nair

Madhu