PREDICTED FORECAST FOR MICROSOFT'S STOCK PRICES

Presented By: Jaelyn Stewart

PROJECT INTRODUCTION

In this presentation, I aim to analyze the trend of high prices over the course of 365 days. By using historical data and regression analysis, I plan to provide an accurate description for traders who are considering buying. One of the key questions I will answer is: What is the highest possible price at which they can sell?

Furthermore, I will explore how these trends can inform trading strategies and decision-making steps. I hope to provide insights into the dynamics of the market.

he models used were regression models, univariate models, autocorrelation procedure, and moving average.

NULL & ALTERNATIVE HYPOTHESIS

NULL HYPOTHESIS

Ho: Microsoft’s open/close price has a direct correlation to the high/low price.

ALTERNATIVE HYPOTHESIS

Ha: Microsoft’s open/close price does not have a direct correlation to the high/low price.

MEANS PROCEDURE

A simple overview of the main variables that will be covered throughout the presentation showing the number of observations (N), mean, standard deviation (STD Dev), minimum, and maximum.

Means Procedure for All Variables except Volume and Date

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
Open	1511	107.3859762	56.6913335	40.3400000	245.0300000
High	1511	108.4374719	57.3822761	40.7400000	246.1300000
Low	1511	106.2945334	55.9771554	39.7200000	242.9200000
Close	1511	107.4220913	56.7022987	40.2900000	244.9900000

FINDING OUTLIERS

To find outliers, I used an SGPlot procedure and used a red line to follow my line of regression. Here is the result for the Close vs Low prices:

As you can see, there are no outliers present within the dataset which can allow me to confidently model the data.

FINDING OUTLIERS PT.2

To find outliers, I used an SGPlot procedure and used a red line to follow my line of regression. Here is the result for the Open vs High prices:

As you can see, there are no outliers present within the dataset which can allow me to confidently model the data.

AUTOCORRELATION PROCEDURE

As shown in the Pearson Correlation Coefficients, there are signs of strong correlations between the pairs of variables which are open, high, low, close. All of which show an almost perfect coefficient and extremely low p-value which indication strong, positive correlation and statistical significance.

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Open	1511	107.38598	56.69133	162260	40.34000	245.03000
High	1511	108.43747	57.38228	163849	40.74000	246.13000
Low	1511	106.29453	55.97716	160611	39.72000	242.92000
Close	1511	107.42209	56.70230	162315	40.29000	244.99000

Pearson Correlation Coefficients, N = 1511
Prob > |r| under H0: Rho=0

	Open	High	Low	Close
Open	1.00000	0.99977 <.0001	0.99976 <.0001	0.99954 <.0001
High	0.99977 <.0001	1.00000	0.99962 <.0001	0.99973 <.0001
Low	0.99976 <.0001	0.99962 <.0001	1.00000	0.99977 <.0001
Close	0.99954 <.0001	0.99973 <.0001	0.99977 <.0001	1.00000

UNIVARIATE PROCEDURE FOR OPEN

Central Tendency:

- **Mean:** 107.386
- **Median:** 93.990
- **Mode:** 44.480 which appeared twice

Variability

- **Standard Deviation:** 56.691
- **Range:** 204.690
- **Interquartile Range:** 81.590

Shape of Distribution

- **Skewness:** 0.8265 (more values on the lower end of price range)
- **Kurtosis:** -0.4724 (fewer extreme values)

Univariate Procedure for All Variables except Volume and Date

The UNIVARIATE Procedure
Variable: Open

Moments			
N	1511	Sum Weights	1511
Mean	107.385976	Sum Observations	162260.21
Std Deviation	56.6913335	Variance	3213.90729
Skewness	0.82653983	Kurtosis	-0.4724041
Uncorrected SS	22277471.1	Corrected SS	4853000.01
Coeff Variation	52.7921201	Std Error Mean	1.45842614

Basic Statistical Measures			
Location		Variability	
Mean	107.3860	Std Deviation	56.69133
Median	93.9900	Variance	3214
Mode	44.4800	Range	204.69000
		Interquartile Range	81.59000

Note: The mode displayed is the smallest of 102 modes with a count of 2.

UNIVARIATE PROCEDURE FOR HIGH

Central Tendency:

- **Mean:** 108.437
- **Median:** 95.100
- **Mode:** 47.770

Variability

- **Standard Deviation:** 57.382
- **Range:** 205.390
- **Interquartile Range:** 82.300

Shape of Distribution

- **Skewness:** 0.828 (more values on the lower end of price range)
- **Kurtosis:** -0.4762 (fewer extreme values)

Moments			
N	1511	Sum Weights	1511
Mean	108.437472	Sum Observations	163849.02
Std Deviation	57.3822761	Variance	3292.72561
Skewness	0.82826397	Kurtosis	-0.476279
Uncorrected SS	22739389.2	Corrected SS	4972015.68
Coeff Variation	52.9173865	Std Error Mean	1.47620115

Basic Statistical Measures			
Location		Variability	
Mean	108.4375	Std Deviation	57.38228
Median	95.1000	Variance	3293
Mode	47.7700	Range	205.39000
		Interquartile Range	82.30000

UNIVARIATE PROCEDURE FOR LOW

Central Tendency:

- **Mean:** 106.295
- **Median:** 92.920
- **Mode:** 54.00

Variability

- **Standard Deviation:** 55.977
- **Range:** 203.200
- **Interquartile Range:** 80.450

Shape of Distribution

- **Skewness:** 0.828 (more values on the lower end of price range)
- **Kurtosis:** -0.465 (fewer extreme values)

Moments			
N	1511	Sum Weights	1511
Mean	106.294533	Sum Observations	160611.04
Std Deviation	55.9771554	Variance	3133.44192
Skewness	0.82760677	Kurtosis	-0.4653697
Uncorrected SS	21803572.9	Corrected SS	4731497.3
Coeff Variation	52.6623087	Std Error Mean	1.44005339

Basic Statistical Measures			
Location		Variability	
Mean	106.2945	Std Deviation	55.97716
Median	92.9200	Variance	3133
Mode	54.0000	Range	203.20000
		Interquartile Range	80.45000

UNIVARIATE PROCEDURE FOR CLOSE

Central Tendency:

- **Mean:** 107.422
- **Median:** 93.860
- **Mode:** 62.300

Variability

- **Standard Deviation:** 56.702
- **Range:** 204.700
- **Interquartile Range:** 81.210

Shape of Distribution

- **Skewness:** 0.827 (more values on the lower end of price range)
- **Kurtosis:** -0.472 (fewer extreme values)

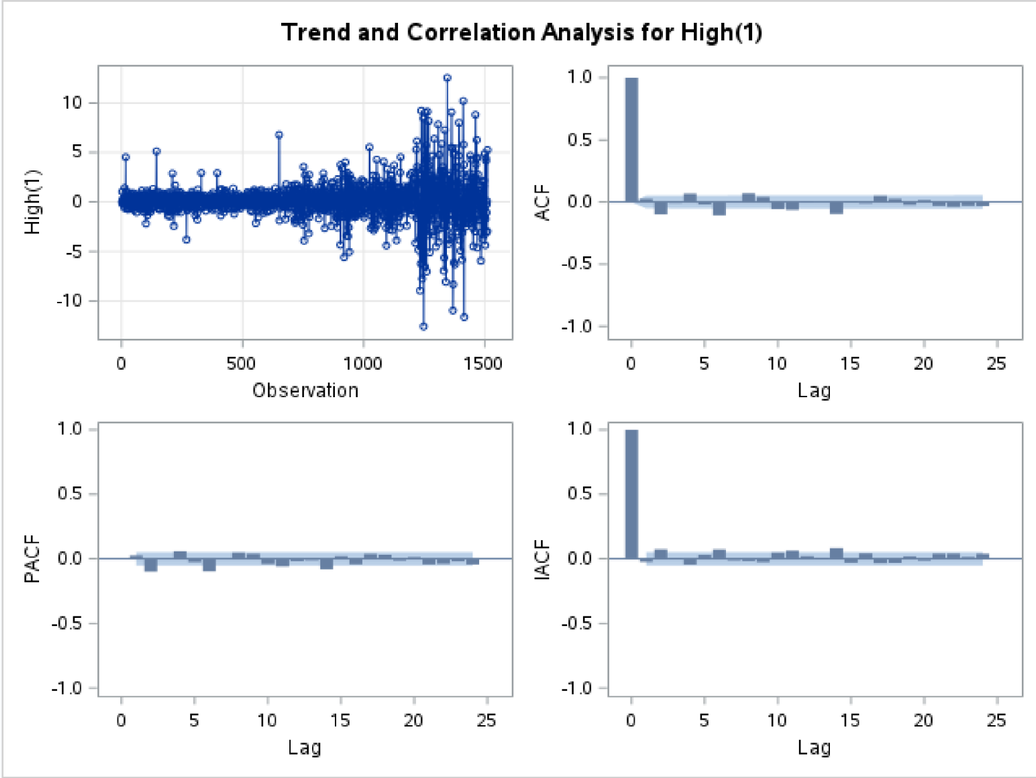
Moments			
N	1511	Sum Weights	1511
Mean	107.422091	Sum Observations	162314.78
Std Deviation	56.7022987	Variance	3215.15068
Skewness	0.8273771	Kurtosis	-0.4727047
Uncorrected SS	22291070.6	Corrected SS	4854877.53
Coeff Variation	52.784579	Std Error Mean	1.45870823

Basic Statistical Measures			
Location		Variability	
Mean	107.4221	Std Deviation	56.70230
Median	93.8600	Variance	3215
Mode	62.3000	Range	204.70000
		Interquartile Range	81.21000

ARIMA MODEL PT. 1

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	32.95	4	<.0001	-0.020	-0.067	-0.023	0.074	-0.023	-0.102
12	52.06	10	<.0001	-0.002	0.066	0.041	-0.056	-0.059	0.002
18	71.58	16	<.0001	0.006	-0.096	0.012	-0.015	0.051	0.025
24	78.60	22	<.0001	-0.022	0.018	-0.030	-0.036	-0.028	-0.029
30	97.91	28	<.0001	-0.070	-0.069	0.037	0.004	0.005	0.038
36	106.35	34	<.0001	0.029	-0.016	0.046	-0.038	-0.020	0.019
42	118.10	40	<.0001	0.047	0.016	0.034	-0.031	0.002	-0.055
48	136.78	46	<.0001	-0.036	-0.066	0.037	0.031	-0.060	0.019

Since the autocorrelation values are closer to zero, we can assume that the model is a good fit. Also, the p-value being at <0.0001 suggests that the model residuals have significant autocorrelation.



In the above scatterplot, it is shown that the data points are clustered around zero indicating some variability. Next is the ACF plot which shows us the correlation of the dataset at different lags. The bars represent the strength of correlation. Bars outside of the confidence interval suggest significant correlation

ARIMA MODEL PT. 2

Shown is statistical data from the ARIMA model with the high variable. The estimated mean is the average value of the variable after differencing, which is **0.1315.** The number of times the data has been differenced to make it stationary is **1.**

The autoregressive factors and moving average indicate the relationship between the current value and its previous values.

Model for variable High	
Estimated Mean	0.131494
Period(s) of Differencing	1

Autoregressive Factors	
Factor 1:	1 + 0.54773 B**(1)

Moving Average Factors	
Factor 1:	1 + 0.59957 B**(1)

MA(10) FOR HIGH VARIABLE (FIRST 20 OBS)

This could suggest that the stock price has increased significantly recently since the current price is much higher than the moving average

Moving Average of 10 for the High Price Variable

Obs	Date	Open	High	Low	Close	Volume	date_num	ObsNum	MA_10_ForHighVariable
1	4/1/15	40.60	40.76	40.31	40.72	36865322	04/01/2015	1	25.380
2	4/2/15	40.66	40.74	40.12	40.29	37487476	04/02/2015	2	25.370
3	4/6/15	40.34	41.78	40.18	41.55	39223692	04/06/2015	3	25.890
4	4/7/15	41.61	41.91	41.31	41.53	28809375	04/07/2015	4	25.955
5	4/8/15	41.48	41.69	41.04	41.42	24753438	04/08/2015	5	25.845
6	4/9/15	41.25	41.62	41.25	41.48	25723861	04/09/2015	6	25.810
7	4/10/15	41.63	41.95	41.41	41.72	28022002	04/10/2015	7	25.975
8	4/13/15	41.40	42.06	41.39	41.76	30276692	04/13/2015	8	26.030
9	4/14/15	41.80	42.03	41.39	41.65	24244382	04/14/2015	9	26.015
10	4/15/15	41.76	42.46	41.68	42.26	27343581	04/15/2015	10	26.230
11	4/16/15	41.95	42.34	41.82	42.16	22509652	04/16/2015	11	26.170
12	4/17/15	41.67	41.74	41.16	41.62	42387608	04/17/2015	12	25.870
13	4/20/15	41.73	43.17	41.68	42.91	46057733	04/20/2015	13	26.585
14	4/21/15	43.00	43.15	42.53	42.64	26013844	04/21/2015	14	26.575
15	4/22/15	42.67	43.13	42.55	42.99	25064273	04/22/2015	15	26.565
16	4/23/15	42.85	43.61	42.80	43.34	46309530	04/23/2015	16	26.805
17	4/24/15	45.66	48.14	45.65	47.87	130933665	04/24/2015	17	29.070
18	4/27/15	47.23	48.13	47.22	48.03	59248172	04/27/2015	18	29.065
19	4/28/15	47.78	49.21	47.70	49.16	60730778	04/28/2015	19	29.605
20	4/29/15	48.72	49.31	48.50	49.06	47804562	04/29/2015	20	29.655

Shown above are the moving averages of the past 10 data points at each point in time, which helps smooth out short-term fluctuations and highlight longer-term trends or cycles

MA(10) FOR LOW VARIABLE (FIRST 20 OBS)

This could suggest that the stock price has increased significantly recently since the current price is much higher than the moving average.

Shown below are the moving averages of the past 10 data points of the low variable at each point in time, which helps smooth out short-term fluctuations and highlight longer-term trends or cycles

Moving Average of 10 for the Low Price Variable

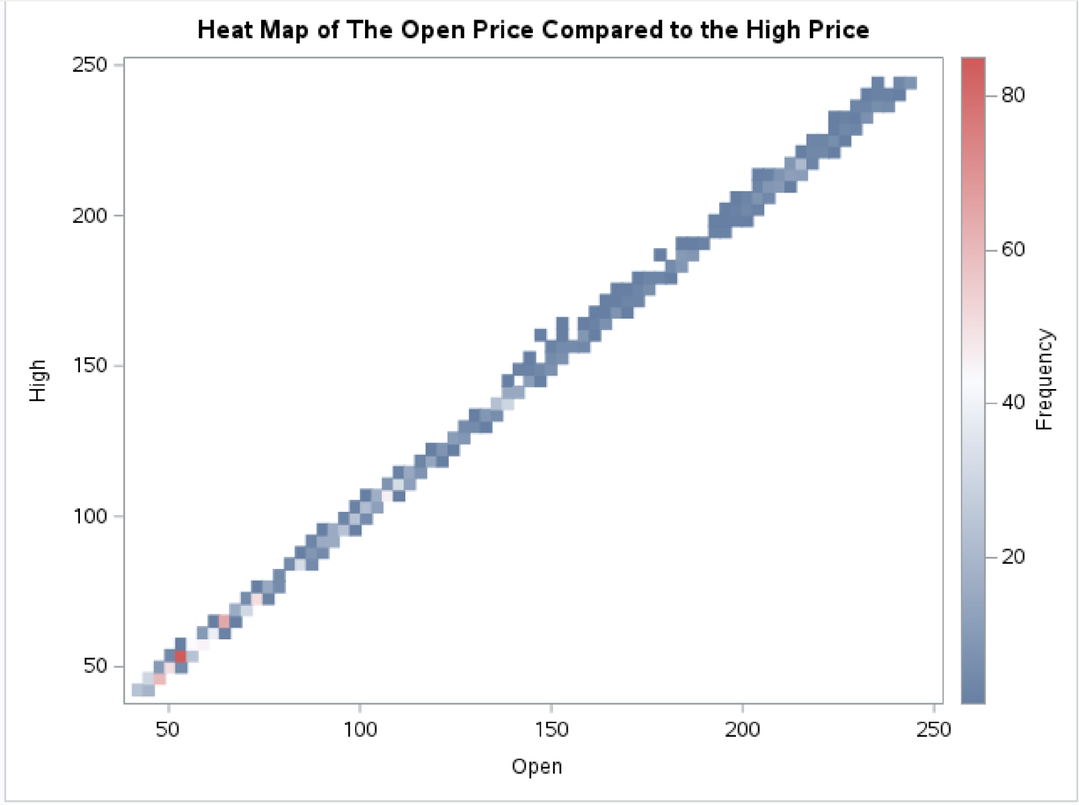
Obs	Date	Open	High	Low	Close	Volume	date_num	ObsNum	MA_10_ForLowVariable
1	4/1/15	40.60	40.76	40.31	40.72	36865322	04/01/2015	1	25.155
2	4/2/15	40.66	40.74	40.12	40.29	37487476	04/02/2015	2	25.060
3	4/6/15	40.34	41.78	40.18	41.55	39223692	04/06/2015	3	25.090
4	4/7/15	41.61	41.91	41.31	41.53	28809375	04/07/2015	4	25.655
5	4/8/15	41.48	41.69	41.04	41.42	24753438	04/08/2015	5	25.520
6	4/9/15	41.25	41.62	41.25	41.48	25723861	04/09/2015	6	25.625
7	4/10/15	41.63	41.95	41.41	41.72	28022002	04/10/2015	7	25.705
8	4/13/15	41.40	42.06	41.39	41.76	30276692	04/13/2015	8	25.695
9	4/14/15	41.80	42.03	41.39	41.65	24244382	04/14/2015	9	25.695
10	4/15/15	41.76	42.46	41.68	42.26	27343581	04/15/2015	10	25.840
11	4/16/15	41.95	42.34	41.82	42.16	22509652	04/16/2015	11	25.910
12	4/17/15	41.67	41.74	41.16	41.62	42387608	04/17/2015	12	25.580
13	4/20/15	41.73	43.17	41.68	42.91	46057733	04/20/2015	13	25.840
14	4/21/15	43.00	43.15	42.53	42.64	26013844	04/21/2015	14	26.265
15	4/22/15	42.67	43.13	42.55	42.99	25064273	04/22/2015	15	26.275
16	4/23/15	42.85	43.61	42.80	43.34	46309530	04/23/2015	16	26.400
17	4/24/15	45.66	48.14	45.65	47.87	130933665	04/24/2015	17	27.825
18	4/27/15	47.23	48.13	47.22	48.03	59248172	04/27/2015	18	28.610
19	4/28/15	47.78	49.21	47.70	49.16	60730778	04/28/2015	19	28.850
20	4/29/15	48.72	49.31	48.50	49.06	47804562	04/29/2015	20	29.250

HEAT MAP OF THE CLOSE PRICE COMPARED TO THE LOW PRICE

This heatmap can help traders identify patterns and frequencies of close prices relative to low prices, which is useful for making informed trading decisions. The diagonal pattern suggests that it's not uncommon for the close price to be the same as the low price for the day.

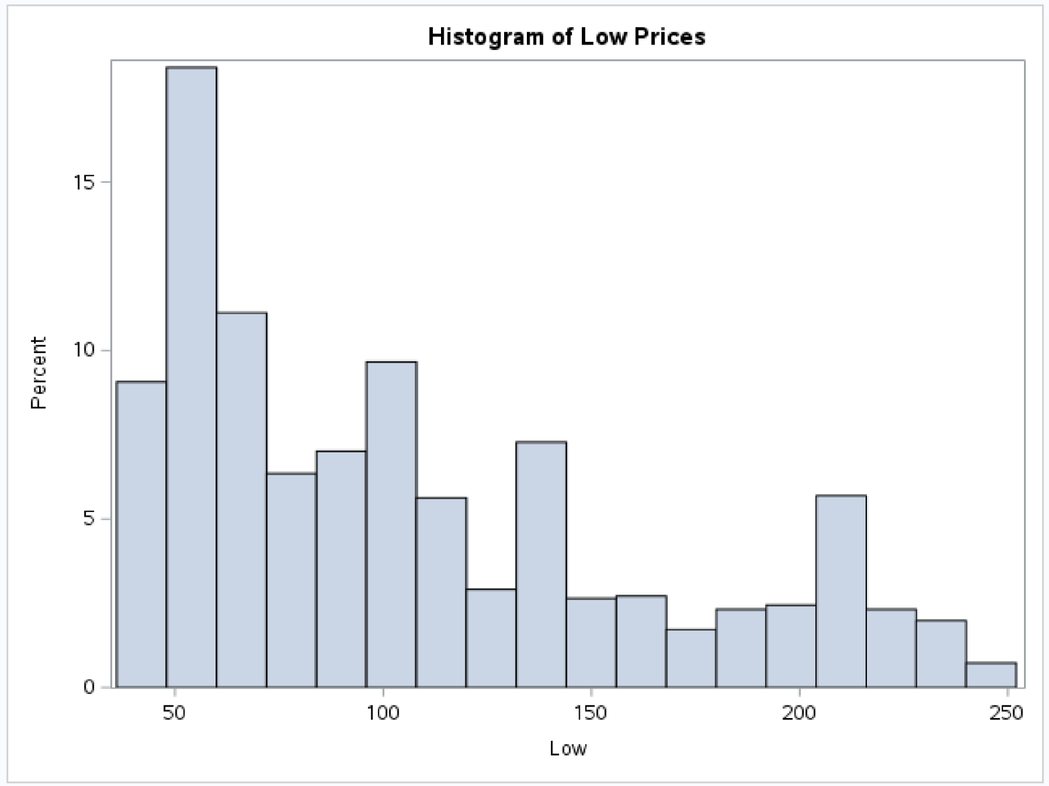
HEAT MAP OF THE OPEN PRICE COMPARED TO THE HIGH PRICE

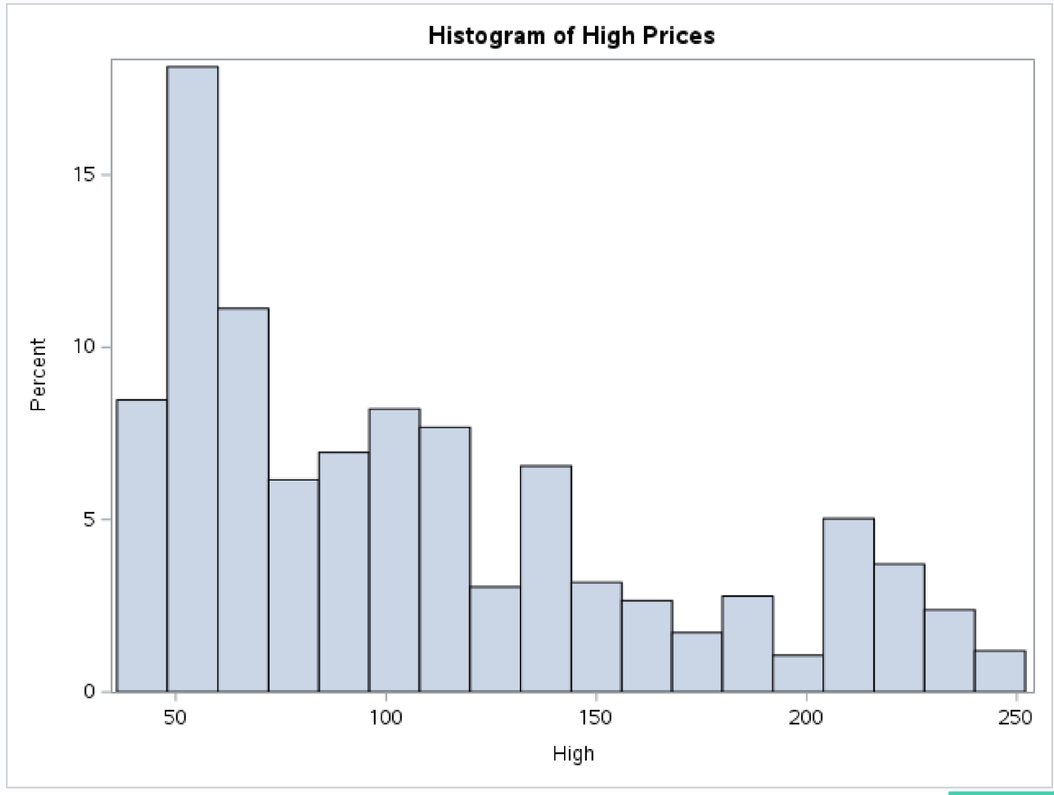
This heatmap can help traders identify patterns and frequencies of open prices relative to high prices, which is useful for making informed trading decisions. The diagonal pattern suggests that it's not uncommon for the open price to be the same as the high price for the day.



HISTOGRAM OF LOW AND HIGH PRICES

Here we can see the frequencies of stock prices between the high and low variables. The change is subtle, but noticeable enough to see the ranges compared to the >170 range. Also, as mentioned before, while seeing the univariate procedure, the skewness of the graph shows mmore values on the lower end of the spectrum





LINEAR REGRESSION PROCEDURE PT.1

As shown by the regression procedure, the data provided shows evidence to fail to reject the null hypothesis (<0.0001). An R-square of 0.9995 tells us that 99.95% of the variability in the high price can be explained by the open price.

The formula would be y = −0.233 + 1.012*Open

Linear Regression with the High Price being the Dependent Variable and the Open Price being Independent

The REG Procedure
Model: MODEL1
Dependent Variable: High

Number of Observations Read	1511
Number of Observations Used	1511

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4969777	4969777	3349204	<.0001
Error	1509	2239.15661	1.48387		
Corrected Total	1510	4972016			

Root MSE	1.21814	R-Square	0.9995
Dependent Mean	108.43747	Adj R-Sq	0.9995
Coeff Var	1.12336		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.23282	0.06714	-3.47	0.0005
Open	1	1.01196	0.00055296	1830.08	<.0001

LINEAR REGRESSION PROCEDURE PT.2

As shown by the regression procedure, the data provided shows evidence to reject the null hypothesis (<0.0001). An R-square of 0.9995 tells us that 99.95% of the variability in the low price can be explained by the close price.

The formula would be y = 0.271 + 0.987*Close

Linear Regression with the Low Price being the Dependent Variable and the Close Price being Independent

The REG Procedure
Model: MODEL1
Dependent Variable: Low

Number of Observations Read	1511
Number of Observations Used	1511

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4729296	4729296	3241670	<.0001
Error	1509	2201.49088	1.45891		
Corrected Total	1510	4731497			

Root MSE	1.20785	R-Square	0.9995
Dependent Mean	106.29453	Adj R-Sq	0.9995
Coeff Var	1.13633		

Parameter Estimates

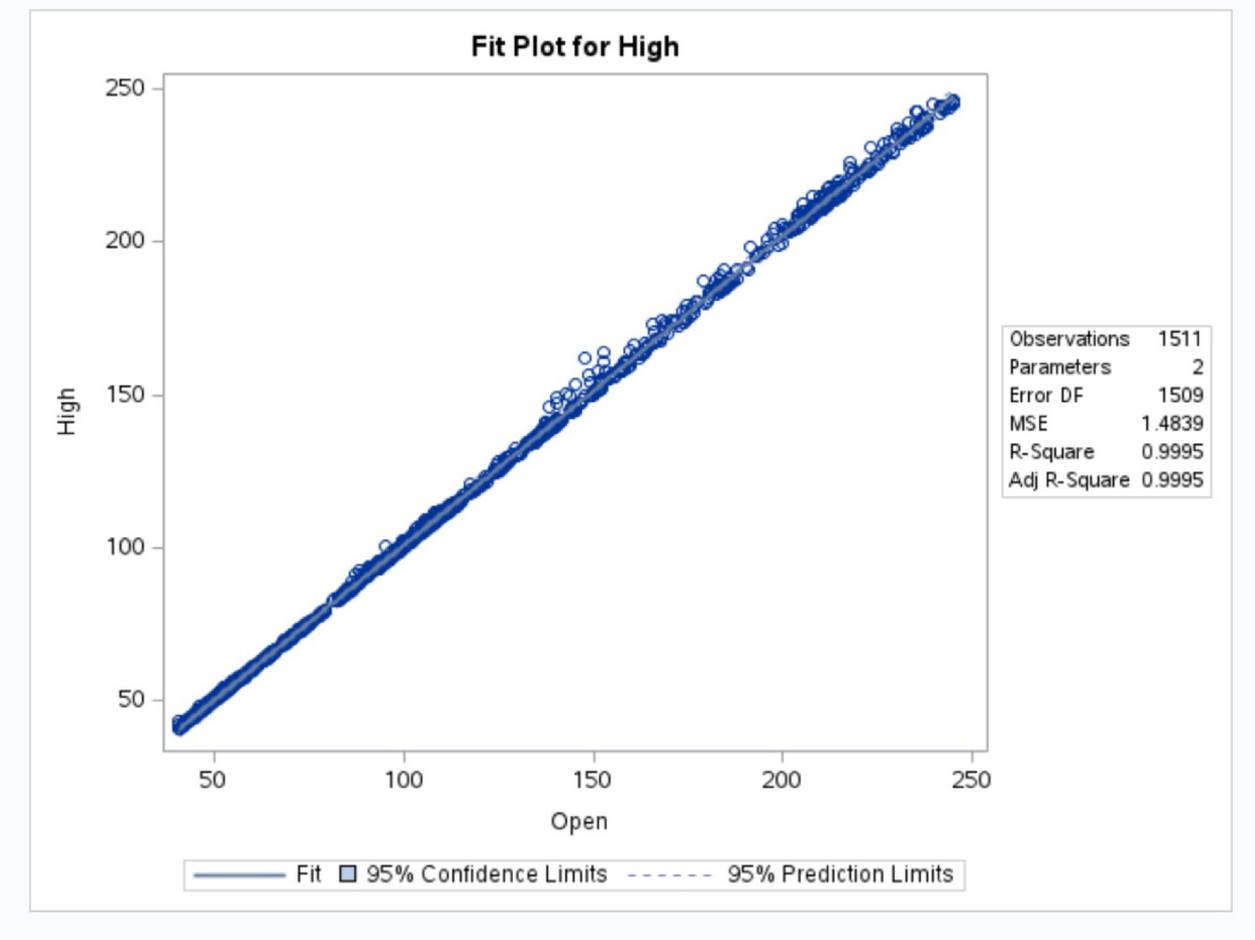
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.27090	0.06658	4.07	<.0001
Close	1	0.98698	0.00054818	1800.46	<.0001

GOODNESS OF FIT GRAPH

The fit plot provides a visual representation of the dataset that shows the relationship between the regression model and the data.

The goodness of fit graph demonstrates accuracy of the regression model. It also shows how closely the model's predictions align with the actual dataset.

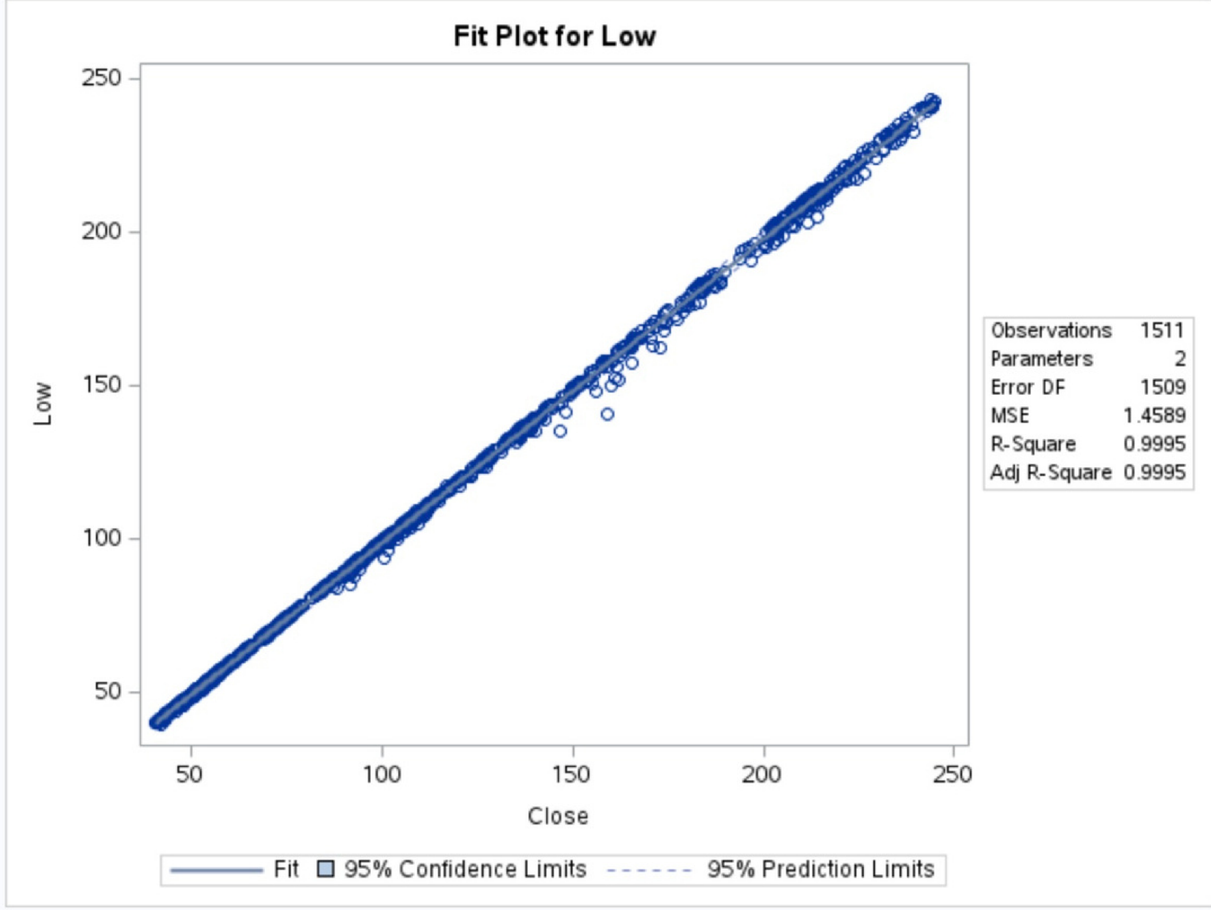
So, by viewing the fit plot, you can see that the data's predictions align with the actual dataset and are very accurate



GOODNESS OF FIT GRAPH

The fit plot provides a visual representation of the dataset that shows the relationship between the regression model and the data.

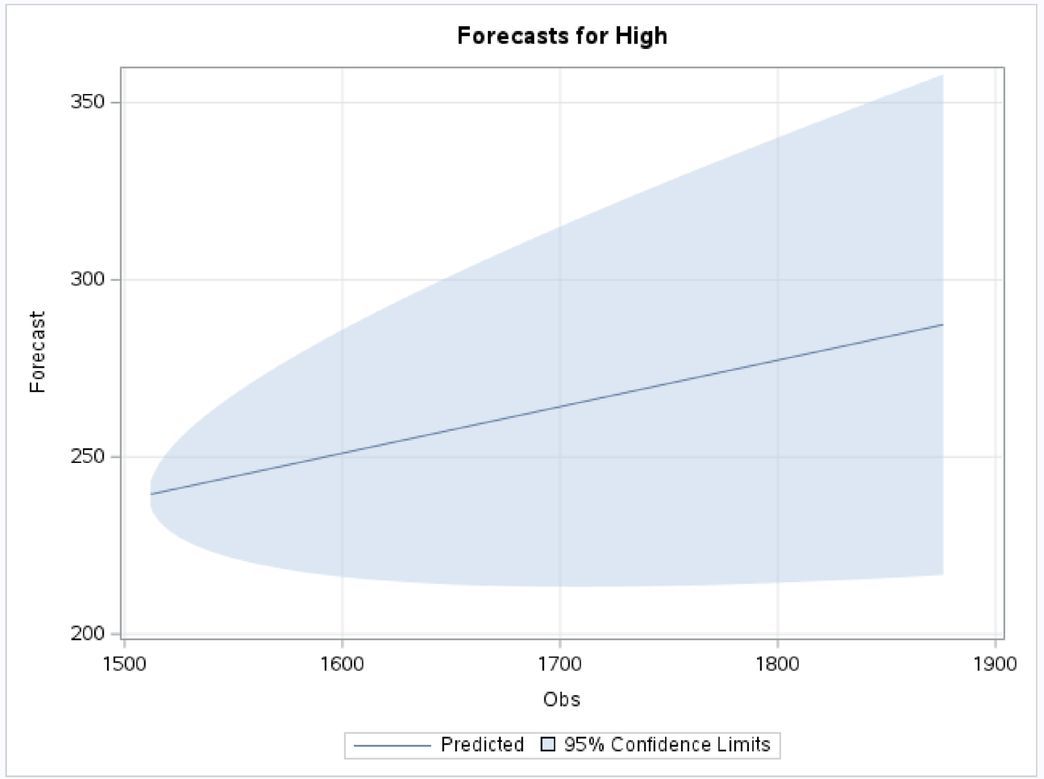
The goodness of fit graph demonstrates accuracy of the regression model. It also shows how closely the model's predictions align with the actual dataset.

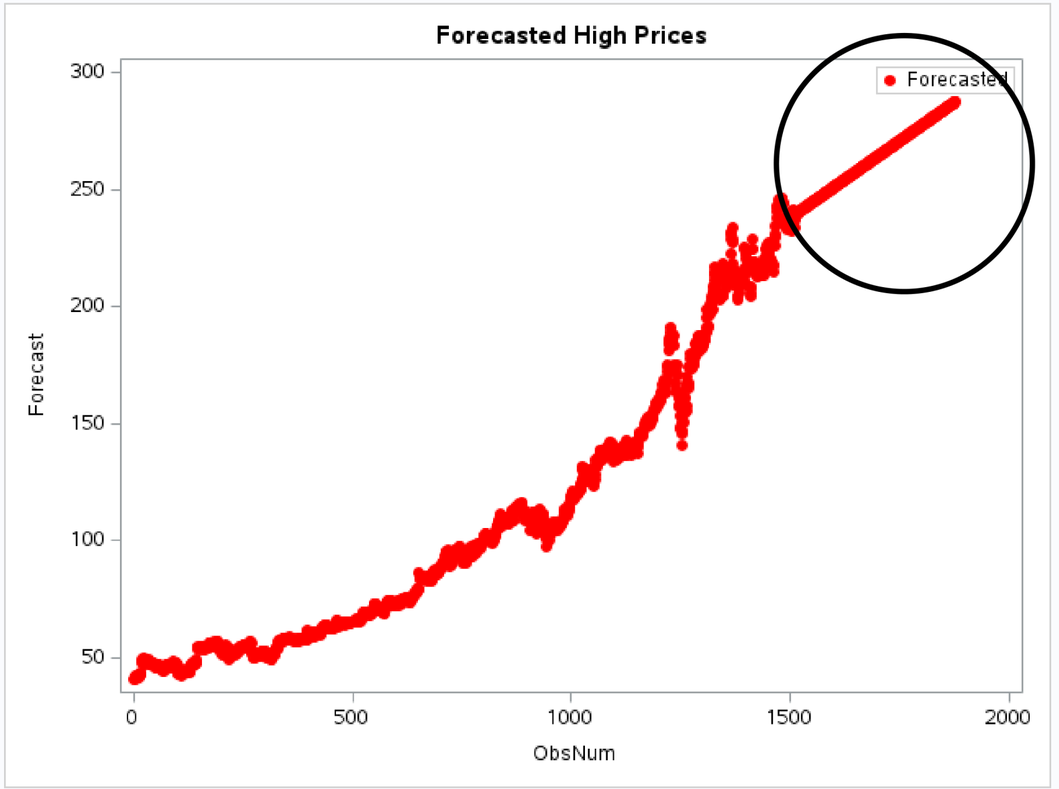


PREDICTED FORECAST

PREDICTED FORECAST

Below, we have the well-awaited forecasted values of the stock prices (high). Based on historical data, we can estimate that the stock prices are going to continue going in an upward trend. We also have a view of the confidence limits which tells us that the wider the confidence interval, the greater the uncertainty, which is clear here. Also, a 95% confidence interval means that we can be 95% confident that the actual values will fall within this range.





CONCLUSION

While there are many methods to predict stock prices, pinpointing the highest price of the day or the optimal time to sell a stock isn't an exact science. We can, however, make educated estimates based on historical data and statistical models. This presentation aimed to provide traders with insights into potential stock price movements, assisting them in making informed decisions on whether to sell or hold their stocks.