Stock Price Prediction: Home Depot

Below is a brief bulleted description of the approach taken to predict the Home Depot stock price for December 31, 2021 which includes the thought process behind it, which methodologies to use, and what kind of model to use. Following that section is a formal summary.

Thought Process & Questions:

- Need to come up with a way to get the final end of the year stock price
  - Do we have to predict each day's value to get there?
  - Is it okay to do it by end of the month, year, or quarter?
- Should this only be an ARIMA forecast?
  - Can we even just look at an average change? Would that make sense?
- Can we use regression?
  - What are some relevant variables for stock pisces?
  - Need to come up with a way to select the best variables to include in the model
- Wait.. need to actually have end of the year values for variables
  - use ARIMA and/or ETS to predict possible variable outcomes
- Use predicted outcomes, use features selected to create models and plug and chug

Methodologies

1. Create regression Equation
   a. Feature Selection using correlations to variable of interest (Stock Price) along with logic, for example, we can't use Net EPS if we use PE ratio because you use Net EPS to calculate PE ratio.
   b. ETS and ARIMA to predict selected feature values for Dec 31st that we will plug into our regression equation
2. ARIMA and ETS using just Stock Price

Model(s):

1. Python + Alteryx: ARIMA (using auto ARIMA)
2. Python: ARIMA using best input based on MAPE
3. Alteryx: ETS
4. Alteryx: Linear Regression using ETS/ARIMA ensemble variable inputs

Summary:

The steps taken to predict a reasonable year-end stock price prediction for Home Depot included having to tackle the following issues:

1. Finding a good, relevant data of data
2. Feature selection
3. Predicting variable outcomes selected features for December 31st
4. Building a model that will predict the final stock price

Data

The original data includes the following variables: Date, High, Low, Open, Close , Volume, Adj Close. The data was then filtered so that there is only Date and Adj Close. Additionally, this original data is daily starting on January 2, 2004 until present day. The data is very granular so it was then subsetted into three categories, Year end stock prices, Month end stock prices, and Quarter end stock prices.

The second data set was found via www.macrotrends.net. I compiled together data with the following variables: Quarter, Price, Qtr_Revenue(B), Assets(B), Equity(B), TTM_Rev(B), TTM_GrossProfit(B), Gross_Margin, TTM_NetIncome(B), NetMargin, TTM_NetEPS, PE_Ratio, Current_Liabilities(B), LT_Investments_Debt(B). The (B) represents billions of dollars. From these variables, I logically picked out four variables to use. I reasoned that some of these variables are the result of equations using other variables. I stuck to using Equity, TTM_revenue, TTM_NetEPS, and Liabilities.

Process

The next step was to determine which of the variables in the data set would help accurately predict Stock Price. First, I did find that the variables I selected above were highly correlated with Stock Price. Then I ran a linear regression to see if the variables were in fact significant in predicting Stock Price. I found that Liabilities was in fact not significant. My final model only included Equity, TTM_revenue, and TTM_NetEPS.

In order to predict the Stock Price in my linear regression model, we need to also know what the variable inputs will be. In order to know those, we need to do a forecast. I forecasted using ARIMA and ETS models which I then created an ensemble for each variable to find the final forecasts for the inputs. These were input into the regression model to get the forecast for Linear Regression in the table on the last page. The model is as follows:

$$Price = -110.5 - 1.77Equity + 2.14TTMRev + 8.24TTMNetEPS$$

Finally, after running various models in Python and Alteryx, the predictions were combined/ensembled using the MAPE for each model. Additionally, there is one model that was done based on just the average change in Stock Price. There was a reasonable enough prediction so I decided to include it. For that, I based it's weight on trial and error to see how much impact it has on the change of the final result. The average model's prediction was actually very close to the ensemble model prediction with and without that model included. I decided to let it have a very small weight since it was almost the same value as the final prediction.

The final prediction is highlighted in the table below! The prediction shows a ~7.6% growth from current day (03/16)

Final Ensemble								
Python	**Model**	**Time**	**Notes**	**Final Forecast**	**MAPE**	**1 - MAPE**	**Weight**	**Final Prediction**
	ARIMA	Yearly	Auto ARIMA	338.55	8.65	-7.65	0.1756601607	**303.4476759**
	ARIMA	Monthly	Auto ARIMA	309.92	6.28	-5.28	0.1212399541	
	ARIMA	Quarterly	Auto ARIMA	236.8	15.15	-14.15	0.3249138921	
	ARIMA	Yearly	Configured based on forecast	377.34	11.92	-10.92	0.2507462687	
	ARIMA	Monthly	Configured based on forecast	272.25	2.96	-1.96	0.04500574053	
	ARIMA	Quarterly	Configured based on forecast	294.33	10.78	-9.78	0.2245694604	
Alteryx	ARIMA	Yearly	Auto ARIMA	338.55	0.138	0.862	-0.01979334099	
	ARIMA	Monthly	Auto ARIMA	303.81	0.0482	0.9518	-0.02185533869	
	ARIMA	Quarterly	Auto ARIMA	308.68	0.0901	0.9099	-0.02089322618	
	ETS	Yearly	-	282.08	0.1822	0.8178	-0.01877841561	
	ETS	Monthly	-	298.91	0.048	0.952	-0.02185993111	
	ETS	Quarterly	-	313.71	0.0835	0.9165	-0.02104477612	
	Linear Regression	Quarterly	Inputs Based on Avg. Change	277.894534	0.22	0.78	-0.01791044776	
Other	Average	Quarterly	Based on Avg Change	302.74	-	-	-0.001	