

Applying Multiple Regression to predict ULTA's closing stock price using adjacent stocks in Consumer Discretionary Components in the S&P 500

John I. Anderies, April 2024.

Abstract:

This report aims to use multiple regression to predict Ulta Corporation's closing stock price using the closing stock price of other companies in the consumer discretionary components list from the S&P 500. For the development of the model, SAS was used supplemented with Tableau for visualizations. The first step of this analysis was to use stepwise regression on different stocks from companies that were considered to have a similar consumer base as Ulta and therefore might help predict the stock price. These stocks are as follows: Amazon, Bath & Body Works, Etsy, Expedia Group, Lululemon, Marriott International, MGM Resorts, Nike, Pool Corporation, Ralph Lauren Corporation, Starbucks, Tapestry, Inc., Wynn Resorts, Ross stores, Royal Caribbean and Yum! Brands. Stepwise regression was then applied to these variables and returned the first-order model with the highest R-squared. The model included these variables; Bath and Body Works, Amazon, Etsy, Lululemon, Marriott, Nike, Ralph Lauren, Tapestry, Wynn Resorts, and Yum Brands.

The next step in the regression process was to check for multicollinearity, this was done using the Variance Inflation Factor (VIF) of all independent variables. Models were tested by removing high VIF variables and then re-running the model to see if the removed variable was the cause of the inflation. The final model that had VIF values all below the desired cut-off of 10 was: Bath and Body Works, Etsy, Lululemon, MGM, Nike, Ralph Lauren, Wynn Resorts, and Yum Brands. The next step was to add second-order terms, with all 10 original variables being tested as second-order terms, this test was done through an F-test of the second-order terms. The F-test came back significant meaning that at least 1 of the second-order coefficients was significant, all t-tests for second-order terms were examined and 3 were found to be insignificant those being: Ralph Lauren,

Lululemon, and Yum! These terms were removed and the model's performance did not change. The next step was to look at interaction terms with Lululemon and Nike, as well as, MGM and Wynn resorts being selected as 2 interaction terms given they cover very similar markets. These terms were tested with interaction plots, when added to the model they increased the R-squared adjusted and both had significant t-tests. These terms were added to the final model giving the following:

$$E(\text{Ulta}) = -992.65411 + 26.25259(\text{BBWI}) + 3.44117(\text{LULU}) + 28.82128(\text{MGM}) - 5.79296(\text{NKE}) \\ + 1.60998(\text{RL}) - 9.22255(\text{WYNN}) + 0.58940(\text{YUM}) - 0.332676(\text{BBWI}^2) + 0.58108(\text{MGM}^2) \\ + 0.09340(\text{NKE}^2) + 0.22493(\text{WYNN}^2) - 0.03203(\text{LULU} * \text{NKE}) - 0.80242(\text{MGM} * \text{WYNN})$$

Residual analysis was then done to ensure that the assumptions of equal variances and normality hold true. The residual versus predicted, fit-mean, cooks D, and histogram of the residuals all showed the expected behaviors that support these assumptions which are necessary for regressions validity. The overall R-squared adjusted was .91 which means that 91% of the variation in Ulta's stock price was accounted for by the model. All metrics of performance were met meaning that this model is performing well and is trustworthy.

The next steps for this model would be to explore all of the relevant interaction terms and incorporate them into the model to see what effect they might have on improving the overall performance. Once this is checked more stocks that relate to Ulta's consumer base outside of the Consumer Discretionary Components could be brought in to see if they help with predictions.

Exploratory Data Analysis:

There are 52 stocks in this dataset from the Consumer Discretionary Components in the S&P 500, I selected Ulta which is a cosmetic retail chain that specializes in beauty products. I thought it would be interesting to explore what other stocks are linked to the performance of Ulta stock as it seems like a fairly stable market that doesn't have much competition outside of large retailers that don't necessarily specialize in beauty products. While there are 51 other variables that I could have chosen for this analysis I chose to look at stocks of companies that are related to either retail or lifestyle and fashion as it would seem that these cover the same market and therefore would be useful in predicting Ulta's stock price. I also threw in Starbucks as every time I have been to an Ulta I have seen a

vast majority of people in the store drinking Starbucks so maybe there is a connection through the type of consumers for both companies.

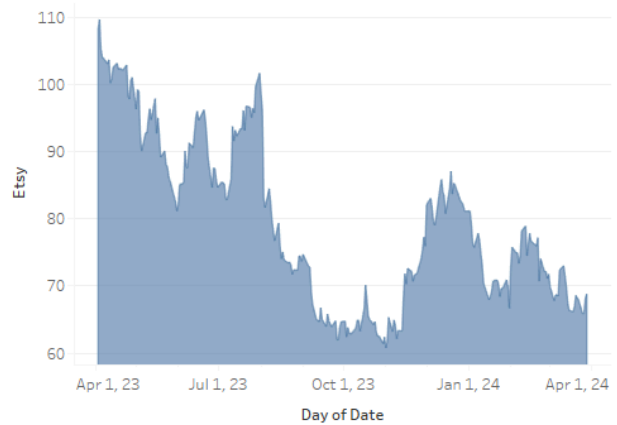
The stocks that I selected are as follows: Amazon, Bath & Body Works, Inc., Etsy, Expedia Group, Lululemon Athletica, Marriott International, MGM Resorts, Nike, Inc., Pool Corporation, Ralph Lauren Corporation, Starbucks, Tapestry, Inc., Wynn Resorts, Ross Stores, Royal Caribbean and Yum! Brands. All of these companies fall into either entertainment, fashion, luxury, or lifestyle which are all adjacent to the products that Ulta sells.

Firstly let's look at the trends of all of the stocks over the given timeframe. We can see in Fig 1.1 the closing prices of each day from April 2023 to March 2024, the stocks that are in Fig 1.1 are Ulta, Bath and Body Works, Etsy, and Amazon, we can see that for all stocks except Amazon, there is a distinct drop around early November. This drop is seen in Amazon but it is much less significant when compared to the other 3 stocks.

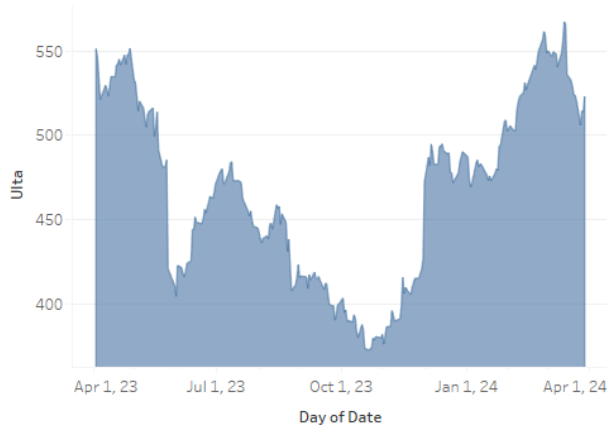
Bath and Body works Closing price from Arpil 23' to March 24'



Etsy Closing price from Arpil 23' to March 24'



Ultas Closing price from Arpil 23' to March 24'



Amazons Closing price from Arpil 23' to March 24'



Fig 1.1

The next set of stocks are Expedia, Lululemon, Marriott, and MGM Resorts. We see a similar pattern with distinct drops around early November in Expedia and MGM, while in Lululemon and Marriott we see pretty consistent growth across the year. There is an interesting spike around February in Expedia's stock where the price quickly drops and then stabilizes, this could be due to a stock split but more investigation would be needed to confirm that.

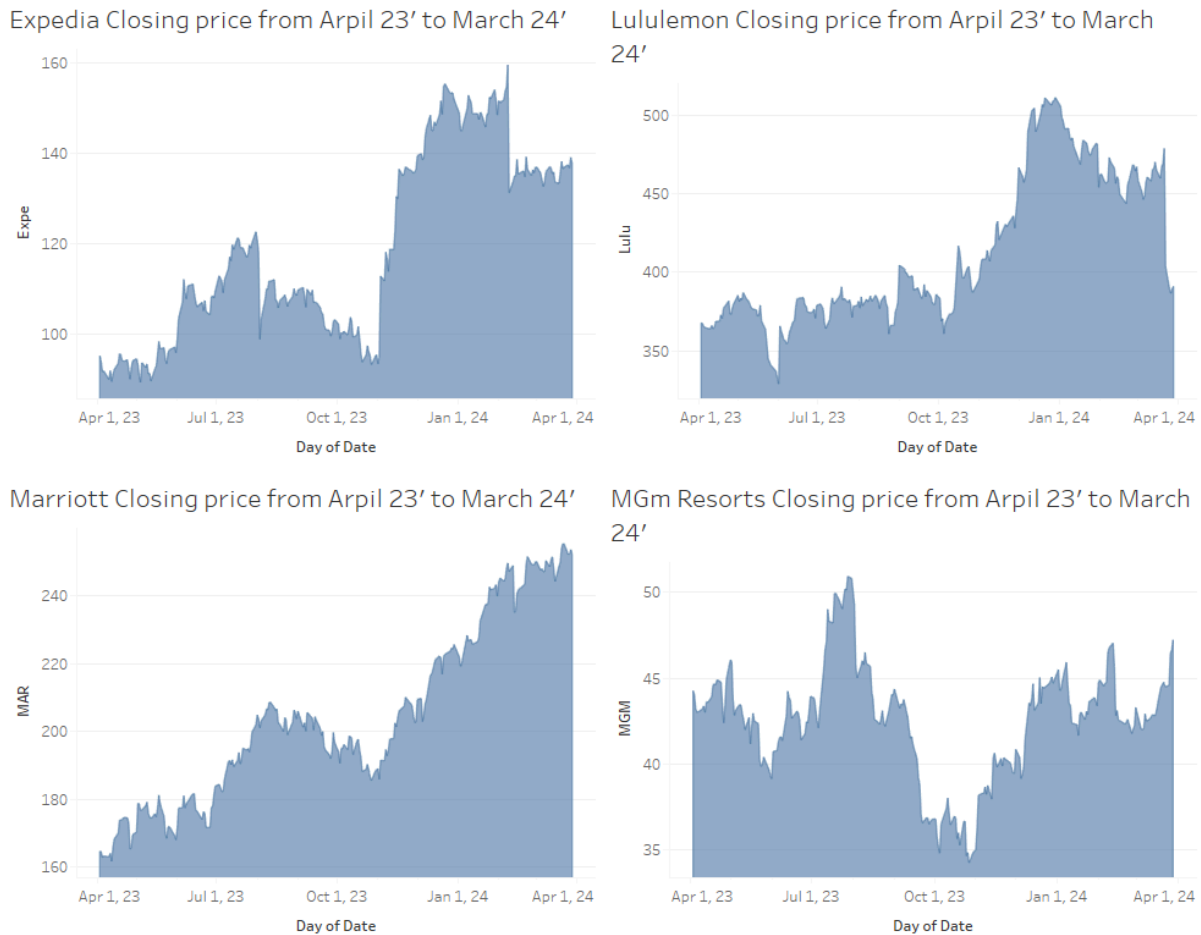
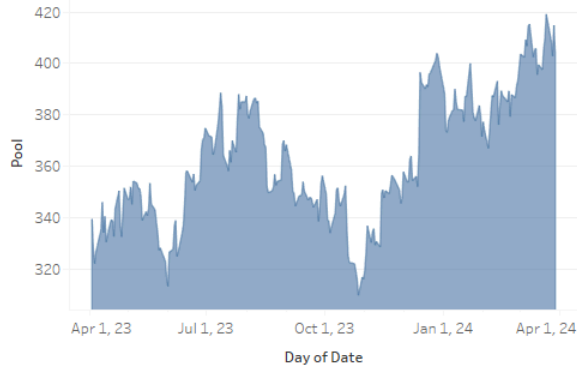


Fig 1.2

The next set of stocks we have in Fig 1.3 are the Pool Corporation, Ralph Lauren, Ross Stores, and Nike. There aren't any crazy patterns that would stand out, we see the drop in prices around November in the Pool corporation, and early October in Nike but Nike's stock quickly shoots back up during December which makes sense given they are a very popular brand for fashion as well as a lifestyle over the holidays. Apart from that Ralph Lauren has had steady growth from their lowest point in late October with a large jump around March of 2024. Ross stores has a very similar pattern with the stock price steadily growing since October. Overall no unexpected patterns or extreme outliers.

Pool Corporation Closing price from April 23' to March 24'



Ralph Lauren Closing price from April 23' to March 24'



Ross Stores Closing price from April 23' to March 24'



Nike Closing price from April 23' to March 24'

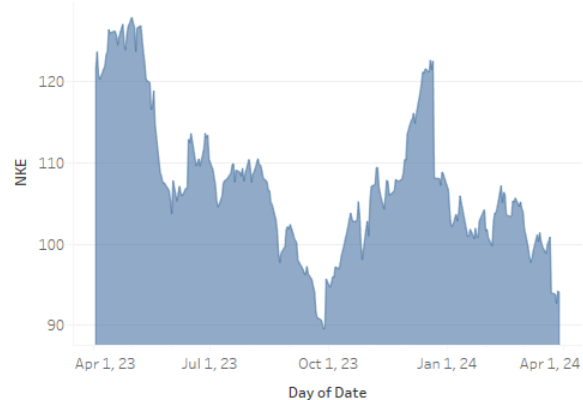
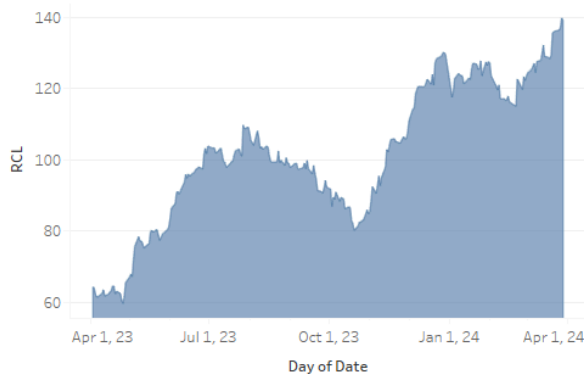


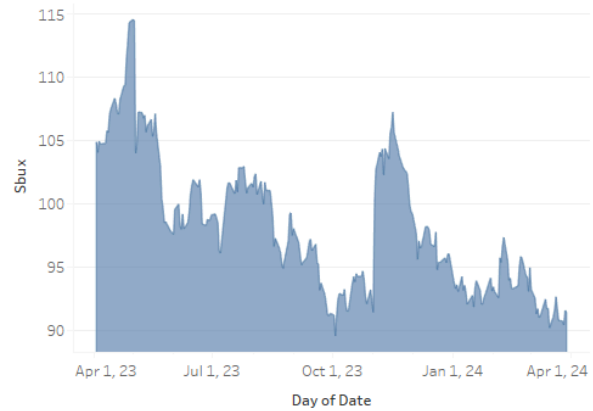
Fig 1.3

In Fig 1.4 we have Royal Caribbean, Starbucks, Tapestry Inc., and Wynn Resorts. For Wynn Resorts we see a pattern of very high stock prices during the summer months a steady decline leading into the holidays and then a resurgence at the beginning of the new year. For Starbucks, it goes against the trend of being at its lowest in November and has a large spike just at the beginning of November (probably a byproduct of the legendary pumpkin spice latte becoming available at this time). There is an interesting trend with Tapestry Inc. as there is a huge drop starting around August which hits a low around November, I am not sure what could have caused this but it might be an outlier during that time stretch.

Royal Caribbean Closing price from April 23' to March 24'



Starbucks Closing price from April 23' to March 24'



Tapestry, inc Closing price from April 23' to March 24'



Wynn Resorts price from April 23' to March 24'



Fig 1.4

The last stock in our initial variable pool is Yum! Which is a fast-food corporation that owns many large restaurant brands. As seen in Fig 1.5 its performance also has that distinct drop around November but overall is fairly steady with no strange patterns.



Fig 1.5

Overall the stocks selected share common behaviors and don't seem to have any wild stock fluctuations apart from the fairly consistent drop in November. Overall the sample of stocks we have covers a lot of industries that relate to luxury lifestyle, travel, and fashion which I think are all related to Ulta's primary market and therefore should be able to predict its stock price given a set combination of stocks.

Methodology:

For the development of this model, SAS was used, primarily Proc reg for analyzing multiple regression as well as for the variable selecting methods for R-squared, R-squared adjusted, and Cp. Proc GLMselect was also used for variable selection regarding the PRESS statistic. Proc GLM and PLM were used to identify interaction terms through interacting plots and all hypothesis tests for second-order terms were done through the Proc reg test statement. For the visualization of the correlation between different variables and the time series visualizations of the different stocks, Tableau was used.

Correlation analysis:

To see what variable might contribute more information than others we can look at the correlation between the dependent variable and the possible independent variables to see which ones are highly correlated with each other. This will help give a better picture as to which variables might be better for predicting the dependent variable. This can be done using the Proc Corr function in SAS which returns the following output. In Fig 2.1 we can see that 7 of the variables have a correlation coefficient above 0.5 and the rest have less than 0.5 with Starbucks having the lowest at 0.15.

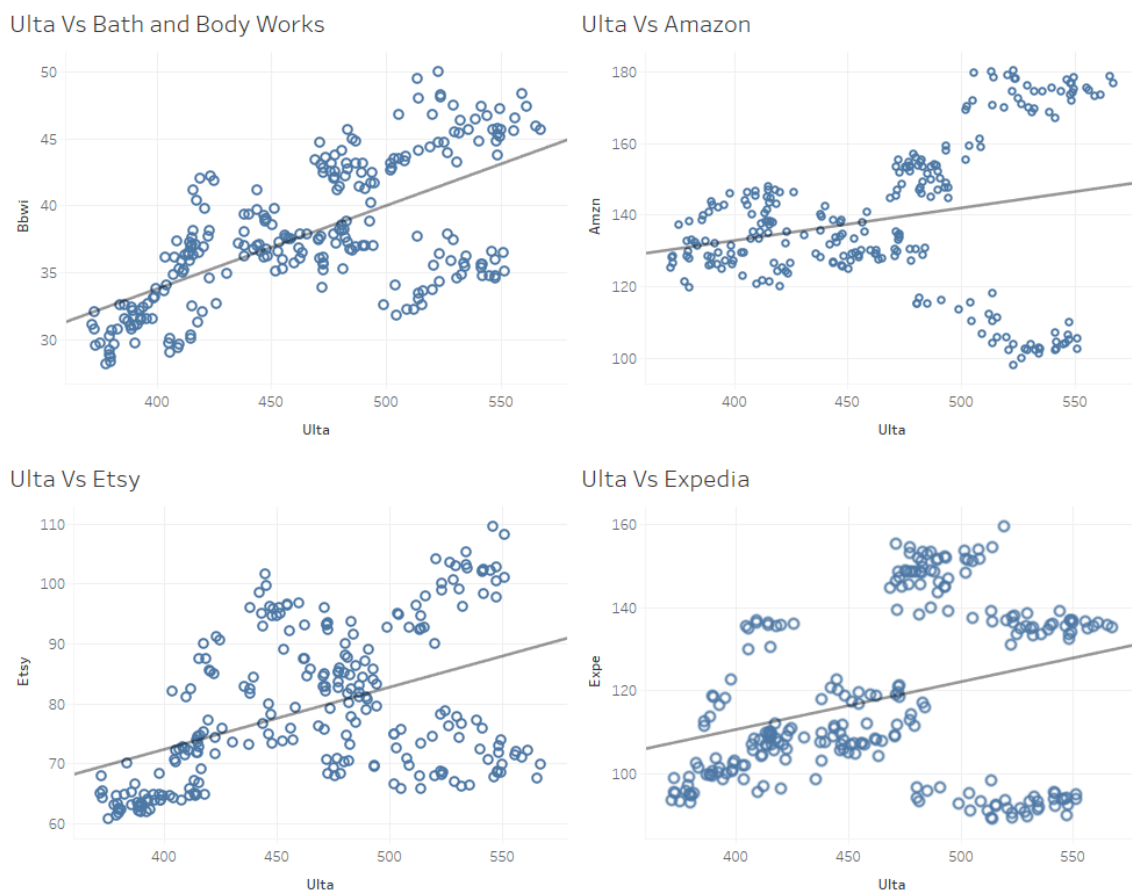
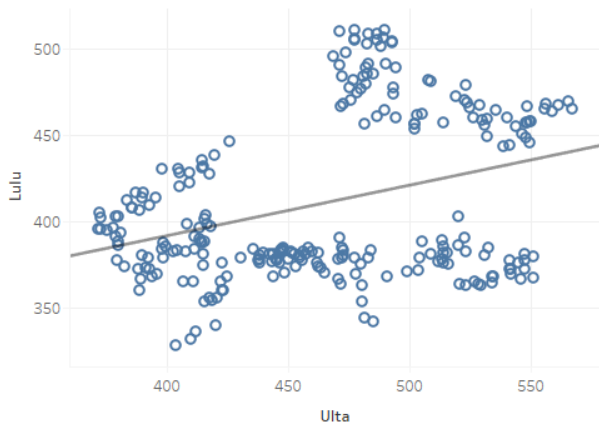


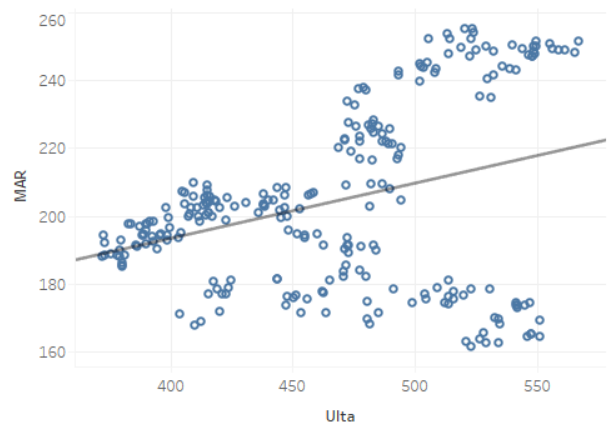
Fig 2.2

We can see that the trend line shows us a positive correlation for all of the variables with the 2 strongest correlations being that of Bath and Body Works and Etsy. Amazon has an interesting pattern where after Ulta hits 500 per share we see a fork on either side of the trend line which is a strange behavior that might cause it to not be a good variable for predicting the higher values of Ulta stock price.

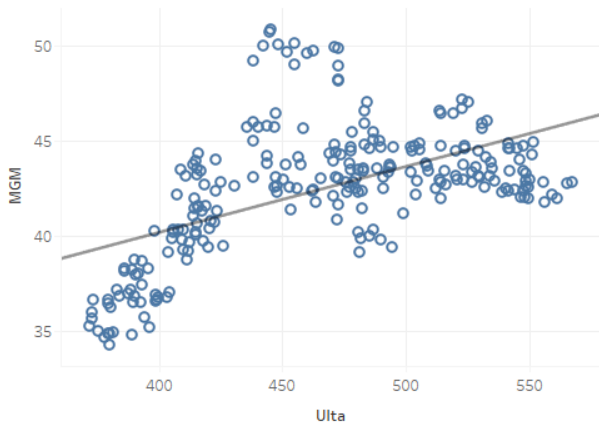
Ulta Vs Bath and Lululemon



Ulta Vs Marriott International



Ulta Vs MGM Resorts



Ulta Vs Nike

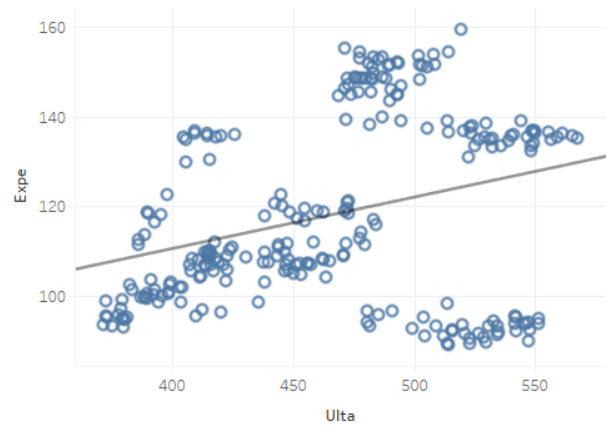
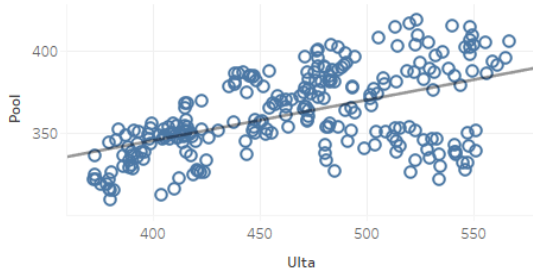


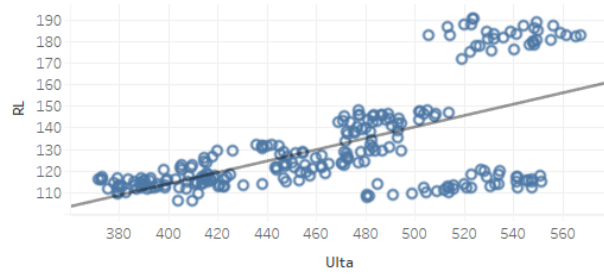
Fig 2.3

In Fig 2.3 we can see that the trend in the correlation between Ulta and other lifestyle brands continues. We see the same forking pattern in Marriott that we see in Amazon meaning that we might not want to use it as a predictor of Ulta. Another strange trend we see is that of MGM resorts where there is a quadratic shape to the data, this indicates that using a squared term for MGM might yield better results overall in the modeling process. Lululemon has an interesting pattern where it seems to spread out on either side of the trend line as Ulta stock price increases.

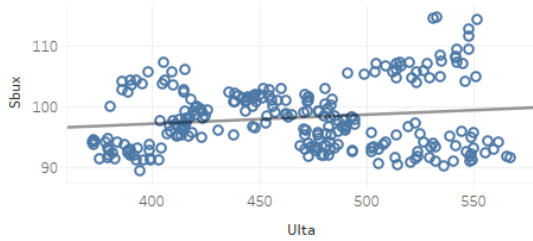
Ulta Vs Pool Corporation



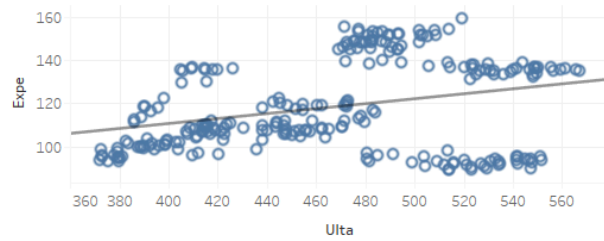
Ulta Vs Ralph Lauren



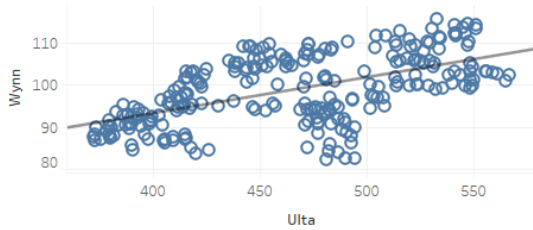
Ulta Vs Starbucks



Ulta Vs Tapestry, Inc.



Ulta Vs Wynn Resorts



Ulta Vs Yum!

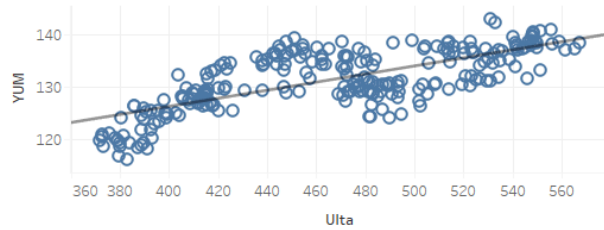


Fig 2.4

In Fig 2.4 we see more of the fork pattern in Tapestry Inc. and Ralph Lauren. Overall the rest follow a decent correlation with Ulta and don't have any strange patterns. In Fig 2.5 below we can see that for Ross stores there is a dramatic fork in as Ulta stock price gets into the higher ranges.

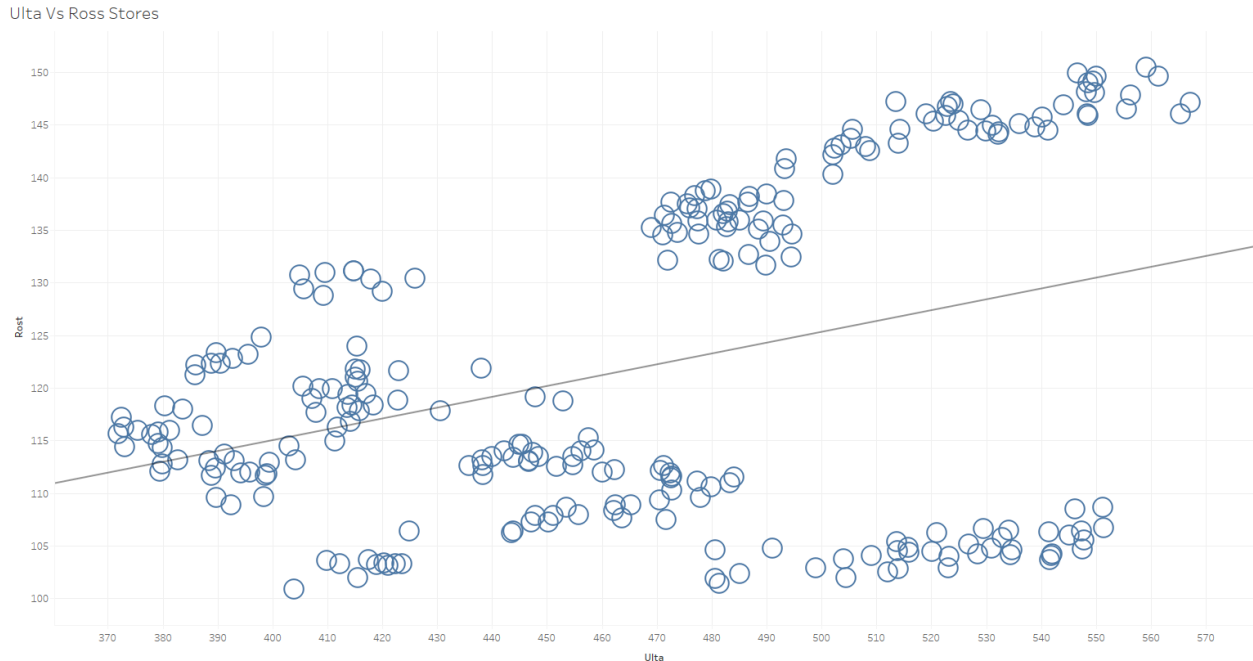


Fig 2.5

Overall the fork pattern is most likely due to the lurking variable that is time given the data is time series so over a given period fluctuations at different times could lead to data points on either side of the trend line. There is a positive correlation in all of the variables with Ulta, but this could be just due to market conditions so the correlation alone isn't enough meaning that further regression analysis is required.

Variable Selection:

Now that we have looked at the raw data and made sure there aren't any outliers or strange behaviors we can start to use more advanced techniques to select variables for the regression model. The first is stepwise regression, which goes through all possible combinations of the possible variables and selects the model with the best performance overall. For the initial 14 independent variables when stepwise regression is applied, the resulting model from the stepwise regression is Bath and Body works, Amazon, Etsy, Lululemon, Marriott, Nike, Ralph Lauren, Tapestry, Wynn resorts, and Yum Brands. This initial step has eliminated 4 variables from our initial pool. The next step is to check for multicollinearity,

which means that two variables are contributing redundant information which we can assess using the Variance Inflation Factor (VIF). We want the VIF of all variables to be less than 10. In the proc reg command, we can have SAS compute the VIF of all independent variables.

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	Intercept	1	-372.86212	59.86756	-6.23	<.0001	0
BBWI	BBWI	1	1.40282	0.61232	2.29	0.0228	6.77911
AMZN	AMZN	1	-1.47659	0.28325	-5.21	<.0001	23.85815
ETSY	ETSY	1	0.71957	0.43402	1.66	0.0987	20.08273
LULU	LULU	1	0.13995	0.07124	1.96	0.0506	7.57598
MAR	MAR	1	1.66253	0.24129	6.89	<.0001	26.90812
MGM	MGM	1	-4.61876	0.90338	-5.11	<.0001	6.47729
NKE	NKE	1	2.08492	0.33457	6.23	<.0001	6.03492
RL	RL	1	0.59773	0.21833	2.74	0.0067	17.66328
TPR	TPR	1	1.42228	0.86623	1.64	0.1019	20.10502
WYNN	WYNN	1	1.67559	0.42773	3.92	0.0001	8.17037
YUM	YUM	1	1.60805	0.53483	3.01	0.0029	6.31178

We can see that the highest VIF is 26.9 for Marriott so to fix this we need to remove this variable and re-run the model again.

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	Intercept	1	-320.50554	64.92271	-4.94	<.0001	0
BBWI	BBWI	1	2.09121	0.66047	3.17	0.0017	6.59863
AMZN	AMZN	1	-0.91387	0.29653	-3.08	0.0023	21.87467
ETSY	ETSY	1	-0.34422	0.44346	-0.78	0.4384	17.54129
LULU	LULU	1	0.32743	0.07198	4.55	<.0001	6.47063
MGM	MGM	1	-1.91102	0.88930	-2.15	0.0327	5.25148
NKE	NKE	1	1.93189	0.36497	5.29	<.0001	6.00832
RL	RL	1	1.26985	0.21354	5.95	<.0001	14.13707
TPR	TPR	1	0.50133	0.93569	0.54	0.5926	19.62632
WYNN	WYNN	1	2.03820	0.46408	4.39	<.0001	8.04669
YUM	YUM	1	1.62529	0.58472	2.78	0.0059	6.31164

Removing Marriott had a very small effect on the overall VIFs of the model so we are going to add it back and remove the second-highest VIF to see if that is the variable causing the Inflation. The second-highest of the original was Amazon.

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	Intercept	1	-503.34364	57.29646	-8.78	<.0001	0
BBWI	BBWI	1	1.75909	0.64108	2.74	0.0065	6.69466
ETSY	ETSY	1	1.55417	0.42501	3.66	0.0003	17.34994
LULU	LULU	1	0.10193	0.07466	1.37	0.1735	7.49656
MAR	MAR	1	1.29985	0.24342	5.34	<.0001	24.67108
MGM	MGM	1	-6.29293	0.88959	-7.07	<.0001	5.65867
NKE	NKE	1	2.21769	0.35147	6.31	<.0001	5.99995
RL	RL	1	0.19280	0.21497	0.90	0.3707	15.42741
TPR	TPR	1	0.11584	0.87359	0.13	0.8946	18.42216
WYNN	WYNN	1	2.83240	0.38525	7.35	<.0001	5.97122
YUM	YUM	1	1.46384	0.56272	2.60	0.0099	6.29489

We see that just removing Amazon didn't have much of an effect therefore the next step is to remove both of the variables and re-run the model.

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	Intercept	1	-416.29345	58.00419	-7.18	<.0001	0
BBWI	BBWI	1	2.23032	0.67053	3.33	0.0010	6.56781
ETSY	ETSY	1	0.37580	0.38357	0.98	0.3282	12.67272
LULU	LULU	1	0.27415	0.07110	3.88	0.0001	6.09743
MGM	MGM	1	-3.43985	0.75109	-4.58	<.0001	3.61743
NKE	NKE	1	2.04404	0.36955	5.53	<.0001	5.94859
RL	RL	1	0.89754	0.17919	5.01	<.0001	9.61270
TPR	TPR	1	-0.24493	0.91975	-0.27	0.7902	18.31198
WYNN	WYNN	1	2.76568	0.40661	6.80	<.0001	5.96494
YUM	YUM	1	1.52541	0.59411	2.57	0.0109	6.29225

Removing both variables has reduced the VIF of most of the original high values but there are still a few that are above the boundary with Tapestry Inc. being the highest. The next step is to remove it and re-run the model.

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	Intercept	1	-407.86918	48.52418	-8.41	<.0001	0
BBWI	BBWI	1	2.14258	0.58288	3.68	0.0003	4.98193
ETSY	ETSY	1	0.31646	0.31158	1.02	0.3108	8.39463
LULU	LULU	1	0.28010	0.06737	4.16	<.0001	5.49541
MGM	MGM	1	-3.37789	0.71276	-4.74	<.0001	3.27028
NKE	NKE	1	2.06617	0.35940	5.75	<.0001	5.64795
RL	RL	1	0.86693	0.13719	6.32	<.0001	5.65690
WYNN	WYNN	1	2.74372	0.39739	6.90	<.0001	5.71959
YUM	YUM	1	1.44195	0.50373	2.86	0.0046	4.54105

That did the trick! we now have the best-performing first-order model that has VIFs all below the threshold of 10 meaning that we do not have severe multicollinearity.

Regression Analysis:

Now to pick the final first-order model that will be regressed 4 different metrics will be used to determine the best overall model. These are R-squared, R-squared adjusted, predicted residual error sum of squares (PRESS), and Cp statistic.

In the table below all 4 selection criteria were used to find the most parsimonious model with the best predictive power.

Metric used	Model	R ²	R ² -adj	CV	MSE
R ²	$E(\text{Ultra}) = -428.14066 + 1.457(\text{Yum}) + 0.26232(\text{Lulu}) + 2.80312(\text{Wynn}) - 3.03522(\text{MGM}) + 2.33971(\text{NKE}) + 0.82933(\text{RL}) + 2.29631(\text{BBWI})$.8502	.8458	4.54	465.82
PRESS	$E(\text{Ultra}) = -407.86918 + 2.14258(\text{BBWI}) + 0.31646(\text{ETSY}) + 0.28010(\text{LULU}) - 3.37789(\text{MGM}) + 2.06617(\text{NKE}) + 0.86693(\text{RL}) + 2.74372(\text{WYNN}) + 1.44195(\text{Yum})$.8508	.8458	4.54	447.71
CP	$E(\text{Ultra}) = -428.14066 + 2.29631(\text{BBWI}) + 0.26232(\text{Lulu}) - 3.03522(\text{MGM}) + 2.33971(\text{NKE}) + 0.82933(\text{RL}) +$.8502	.8458	4.54	447.76

	2.80312(Wynn) + 1.457(Yum)				
R ² -Adj	E(Ulta) = -407.86918 + 2.14258(BBWI) + 0.31646(ETSY) + 0.28010(LULU) - 3.37789(MGM) + 2.06617(NKE) + 0.86693(RL) + 2.74372(WYNN) + 1.44195(Yum)	.8508	.8458	4.54	447.71

Overall the selection criteria arrive at very similar models with R-squared and Cp returning the same model, then Press and R-squared adjusted returning the same model. The R-squared adjusted for all of the models is the same and the CV is the same the only thing that varies between the models is the MSE and R-squared but the difference is so small it is negligible. Therefore the model that I will be selecting is the model with the fewest variables meaning it is the most parsimonious model. The model that will be moving forward is the one selected by Cp and R-squared as it has 1 fewer variable than the other model but has the same performance.

The next step in this analysis is to add some second-order terms to see if they add predictive power to the model. We can do this by using an F-test for all of the second-order term coefficients and testing whether at least 1 of them is non-zero.

$$H_0: \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = \beta_{15} = 0$$

$$H_a: \text{at least 1 } \beta \text{ is non-zero.}$$

When we perform this test using SAS we get the following output.

Source	DF	Mean Square	F Value	Pr > F
Numerator	6	5038.74992	15.43	<.0001
Denominator	234	328.38900		

The p-value is significant which means that there is sufficient evidence to reject the null hypothesis and say that at least 1 of the coefficient of the square terms is non-zero. When we run the regression with all of the square terms we get the following output.

Given that 3 of the second-order terms are non-significant(as denoted by the p-value in the far right column) we

Variable	DF	Mean Square	F Value	Pr > F
BBWI_sq	1	0.07492	0.23	0.627
LULU_sq	1	0.00089025	0.0027	0.958
mgm_sq	1	0.11311	0.35	0.555
nke_sq	1	0.01387	0.04	0.831
rl_sq	1	0.00478	0.01	0.917
wynn_sq	1	0.03198	0.10	0.750
yum_sq	1	0.04395	0.14	0.711

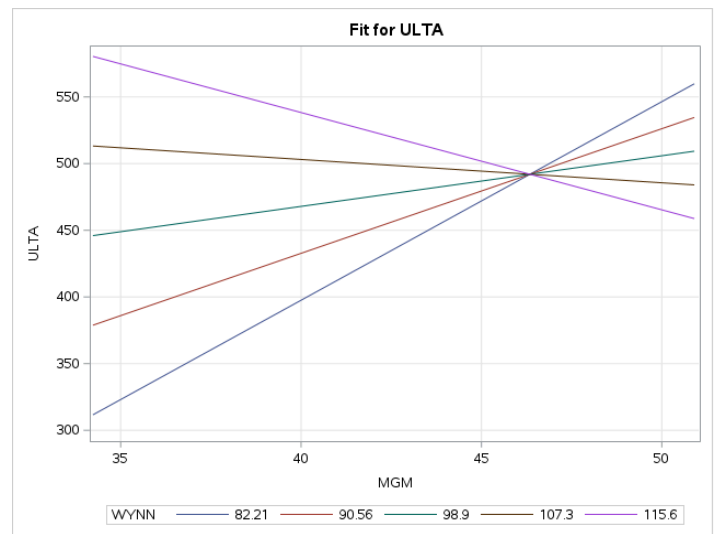
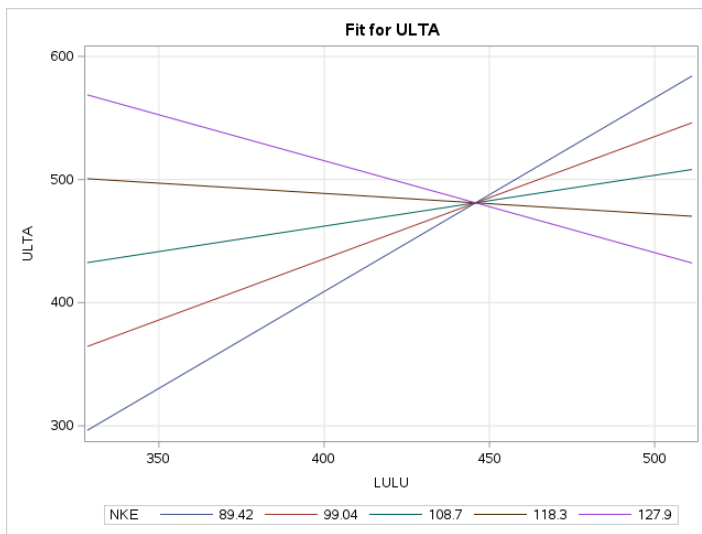
can remove them and re-run the model.

After removing the insignificant variables the model R-squared is .8882 which is a

BBWI_sq	1	-0.27963	0.06682	-4.18	<.0001
mgm_sq	1	-0.25451	0.09441	-2.70	0.0075
nke_sq	1	0.10677	0.01295	8.25	<.0001
wynn_sq	1	0.06832	0.02945	2.32	0.0212

jump of almost 4% from the original first-order model. The next thing that could add some predictive power to the model is interaction terms, due to

the scope of this project testing all of the interaction terms is not possible but I will test 2 interaction terms and see how they affect the model's performance. The 2 interaction terms I am testing are Nike and Lululemon as they are closely related in the type of products they produce i.e. sportswear, the second pair will be MGM Resorts and Wynn Resorts as again they cover the same customer base and potentially interact. To test whether or not there is an interaction we can use the built-in interaction plot in SAS. The interaction plot shows how one independent variable changes the prediction of the dependent variable as another independent variable also varies. If there is no interaction then the line produced should all have the same slope and directions. The output for both interaction terms is shown below.



We can see that in both plots there is a clear change in prediction from one variable given a change in the other showing that there is an interaction between the two variables when predicting the dependent variable. This indicates that adding these

interaction terms to the model might help improve the performance as there is an interaction-type relationship.

LULU_NKE	1	-0.03203	0.00422	-7.60	<.0001
MGM_WYNN	1	-0.80242	0.18811	-4.27	<.0001

As we can see both of the interaction terms are statistically significant and therefore are kept in the model. Once we add the 2 interaction terms to the model we get the following complete model.

$$E(\text{Ulta}) = -992.65411 + 26.25259(\text{BBWI}) + 3.44117(\text{LULU}) + 28.82128(\text{MGM}) - 5.79296(\text{NKE}) + 1.60998(\text{RL}) - 9.22255(\text{WYNN}) + 0.58940(\text{YUM}) - 0.332676(\text{BBWI}^2) + 0.58108(\text{MGM}^2) + 0.09340(\text{NKE}^2) + 0.22493(\text{WYNN}^2) - 0.03203(\text{LULU} * \text{NKE}) - 0.80242(\text{MGM} * \text{WYNN})$$

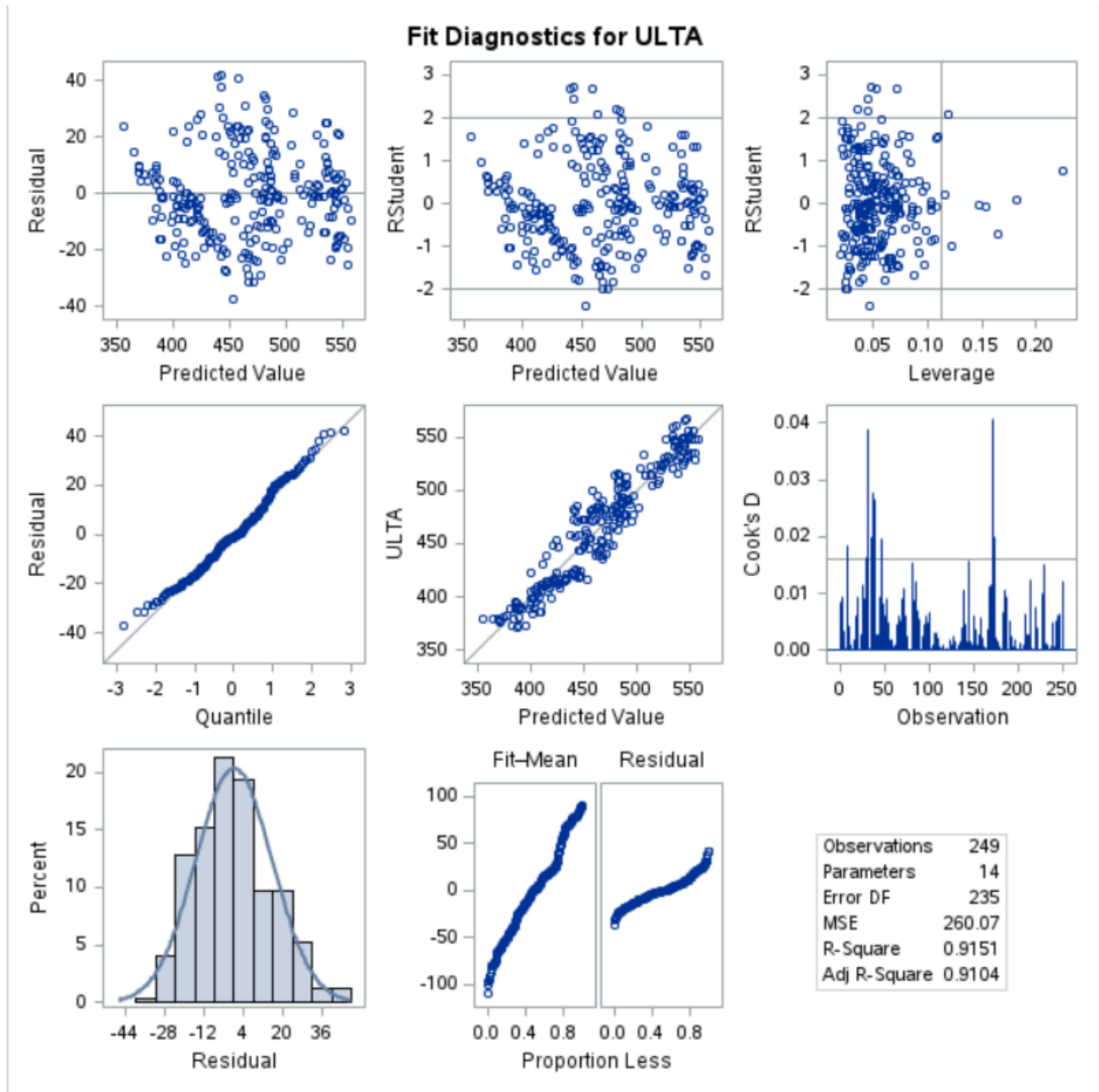
Root MSE	16.12677	R-Square	0.9151
Dependent Mean	465.82920	Adj R-Sq	0.9104
Coeff Var	3.46195		

For all independent variables including all of the interaction and second-order terms, they are statistically significant. The summary statistics for the model are also all satisfactory with an adjusted R-squared of 0.9104 meaning that the model accounts for 91% of the variation in Ulta stock price. We can also see that the coefficient of variation is below the threshold of 10 which is desired.

Now that we have our final model we need to check the residuals to ensure that the assumption of equal variances and normality holds true as it is required for the regression model to be valid.

Residual Analysis

For the residual analysis, we need to ensure that the residuals have equal variances as this is one of the key assumptions of linear regression, and if it is not true the model cannot be used. To check for these we can use the residual vs predicted plot and check to see if there are any patterns present.



Starting with the residual vs predicted values (top left), there don't seem to be any patterns and the points are scattered which is what we would expect given equal variances so the assumption holds. Moving to the Q-Q plot (middle left), we can check for the normality of the residuals. We expect the points to lie along a straight line given that the variances are normal, we do indeed see this pattern with some variance at the extremes which is to be expected. The pattern indicates that the variances are normally distributed which can be seen in the residual histogram (bottom left) which is normally distributed. The Fit-Mean plot shows that the spread of the residuals is small when compared to that of the fitted values. Looking

at the Cook's D plot we see that while some values are of high influence on the beta predictions, overall most singular data points don't influence the model at large. Overall the residuals all look normal supporting the assumption of equal variance meaning that we can approve the model for further use as it meets all of the requirements needed to be considered a valid model.

Conclusion and Next Steps.

Now that we have ensured equal variances for the model, we can use the model below.

$$E(\text{Ulta}) = -992.65411 + 26.25259(\text{BBWI}) + 3.44117(\text{LULU}) + 28.82128(\text{MGM}) - 5.79296(\text{NKE}) + 1.60998(\text{RL}) - 9.22255(\text{WYNN}) + 0.58940(\text{YUM}) - 0.332676(\text{BBWI}^2) + 0.58108(\text{MGM}^2) + 0.09340(\text{NKE}^2) + 0.22493(\text{WYNN}^2) - 0.03203(\text{LULU} * \text{NKE}) - 0.80242(\text{MGM} * \text{WYNN})$$

For all of the coefficients in the model, the practical interpretations in the context of the data are as follows, for all of the terms it is assumed that all other variables are held constant. The intercept represents when all other stocks are 0 the expected value of ULTA stock price is - 992.65. For every increase of one dollar in the stock price for Bath and Body Works Ulta's expected stock price increases by 26.25 dollars. For every increase of one dollar in the stock price of Lululemon Ulta's expected stock price increases by 3.44 dollars. For every increase of one dollar in the stock price of MGM Resorts Ulta's expected stock price increases by 28.82 dollars. For every increase of one dollar in the stock price of Nike, Ulta's expected stock price decreases by 5.79 dollars. For every increase of one dollar in the stock price of Ralph Lauren Ulta's expected stock price increases by 1.61 dollars. For every increase of one dollar in the stock price of Wynn Resorts Ulta's expected stock price decreases by 9.22 dollars. For every increase of one dollar in the stock price of Yum! Corporation Ulta's expected stock price increases by 0.59 dollars. Every increase of one dollar in the stock price of Bath and Body Works squared Ulta's expected stock price decreases by 0.33 dollars. Every increase of one dollar in the stock price of MGM Resorts squared Ulta's expected stock price increases by 0.58 dollars. For every increase of one dollar in the stock price of Nike squared Ulta's expected stock price increases by 0.09 dollars. Every increase of one dollar in the stock price of Wynn Resorts squared Ulta's expected stock price increases by 0.22 dollars. Every increase of one dollar in the stock price of Nike and Lululemon

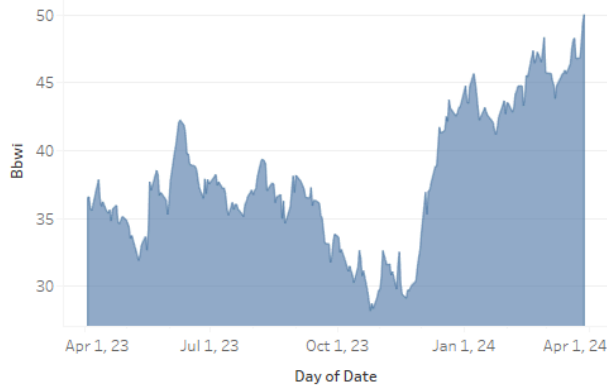
has a combined effect on Ulta's expected stock price to decrease by 0.03 dollars. Every increase of one dollar in the stock price of MGM Resorts and Wynn Resorts has a combined effect on Ulta's expected stock price to decrease by 0.8 dollars.

Overall the model performs well with an adjusted R-squared of 0.91 meaning that 91% of the variability of Ulta's stock price is accounted for by the model. The model's residuals support the assumption of equal variance as well as the normality of the residuals.

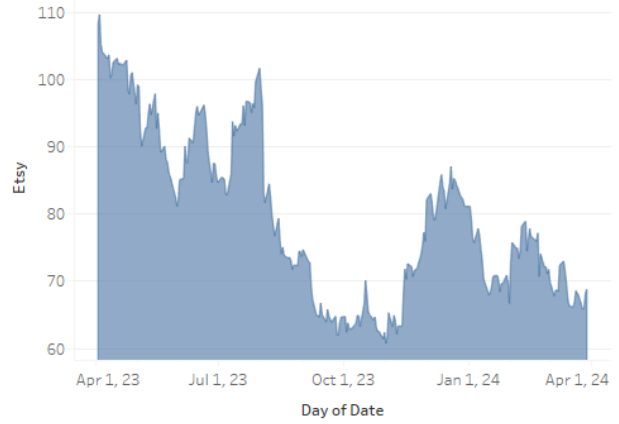
The next steps for the model would be to investigate all of the remaining interaction terms to see if adding more would make the model more effective in predicting Ulta's stock price. Along with adding all of the interaction terms starting with more potential independent variables in the initial variable selection might lead to better models that emerge from the raw data rather than the selection based on a shared customer base as in this analysis. This approach would take more time than this report was allotted but is fully possible if development is continued. The current state of the model is very accurate with all of the key statistics performing well and all pitfalls checked for in the modeling process. To test the effectiveness of the model it should be used on an out-of-sample data set to see how it performs on new data as this would validate the model given it performs well.

Appendix

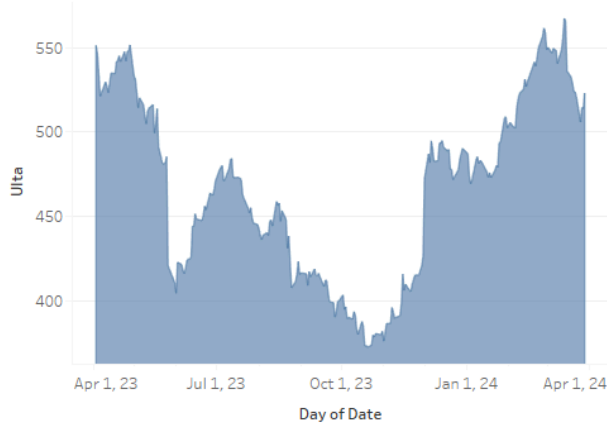
Bath and Body works Closing price from Arpil 23' to March 24'



Etsy Closing price from Arpil 23' to March 24'



Ultas Closing price from Arpil 23' to March 24'



Amazons Closing price from Arpil 23' to March 24'

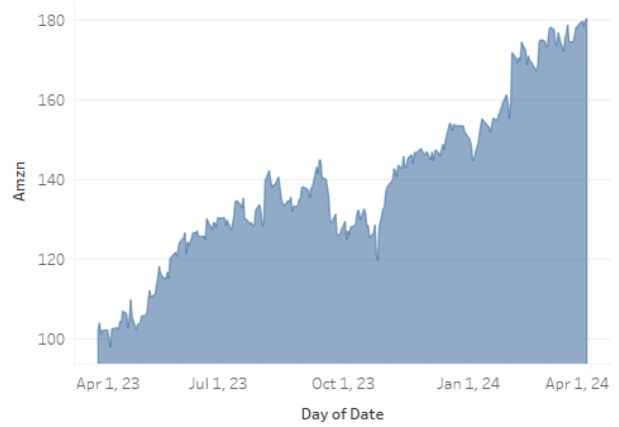
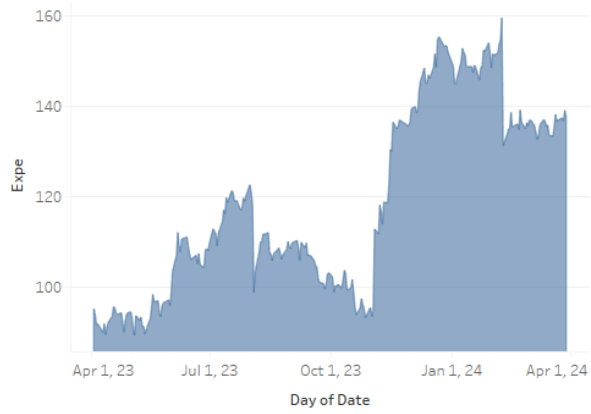


Fig 1.1

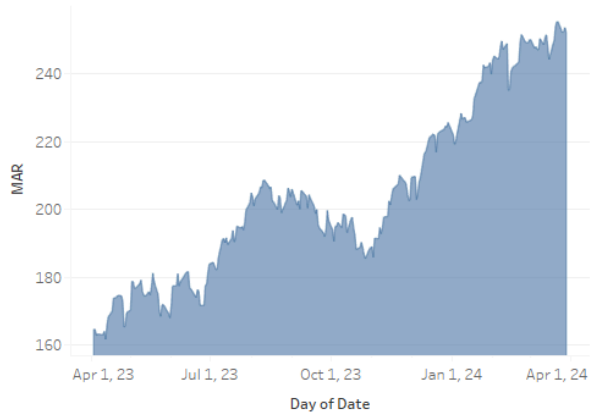
Expedia Closing price from Arpil 23' to March 24'



Lululemon Closing price from Arpil 23' to March 24'



Marriott Closing price from Arpil 23' to March 24'



MGM Resorts Closing price from Arpil 23' to March 24'

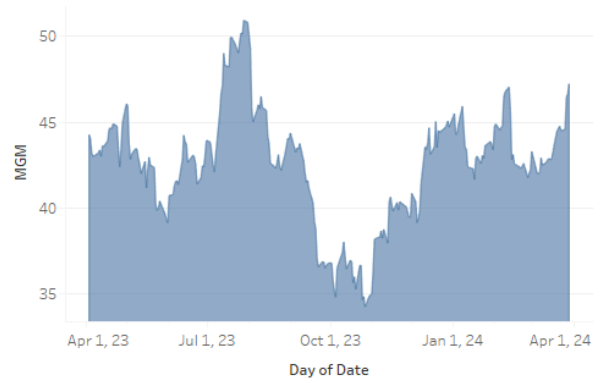
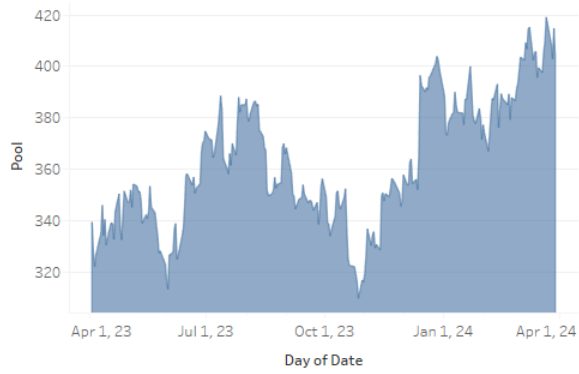


Fig 1.2

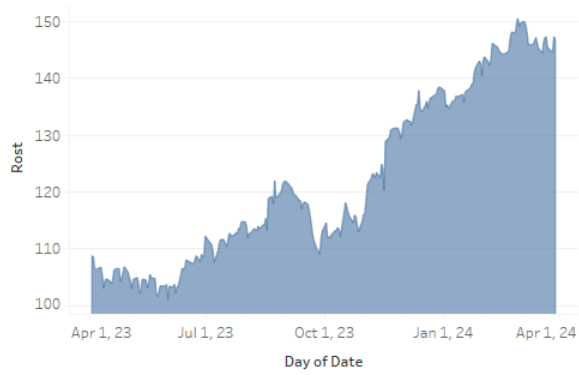
Pool Corporation Closing price from April 23' to March 24'



Ralph Lauren Closing price from April 23' to March 24'



Ross Stores Closing price from April 23' to March 24'



Nike Closing price from April 23' to March 24'

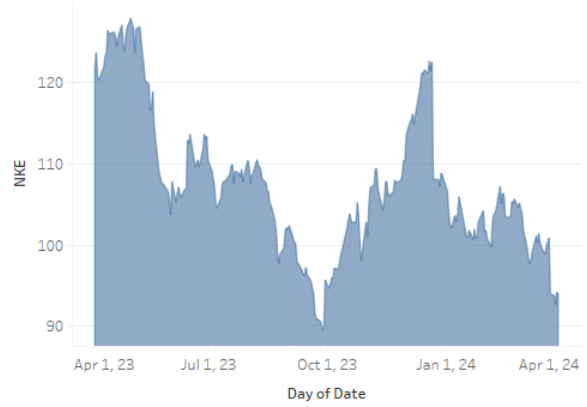
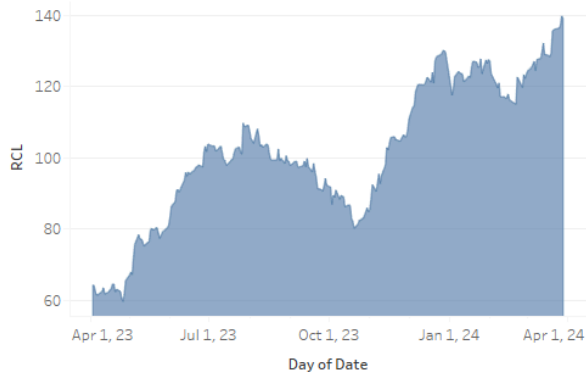
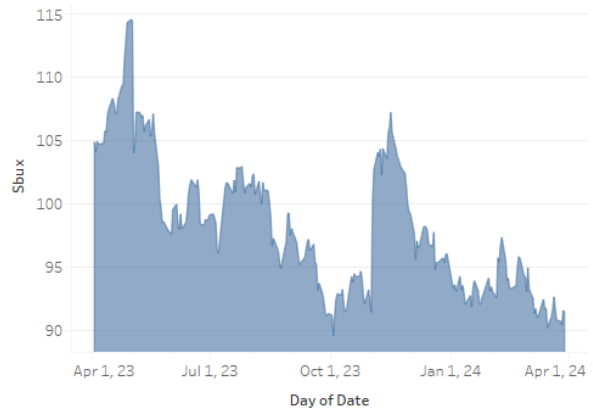


Fig 1.3

Royal Caribbean Closing price from April 23' to March 24'



Starbucks Closing price from April 23' to March 24'



Tapestry, inc Closing price from April 23' to March 24'



Wynn Resorts price from April 23' to March 24'



Fig 1.4

Yum! Closing price from April 23' to March 24'

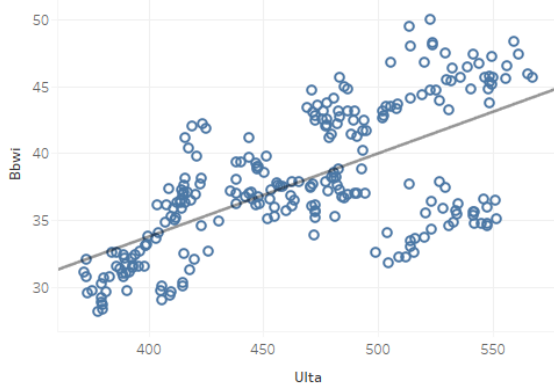


Fig 1.5

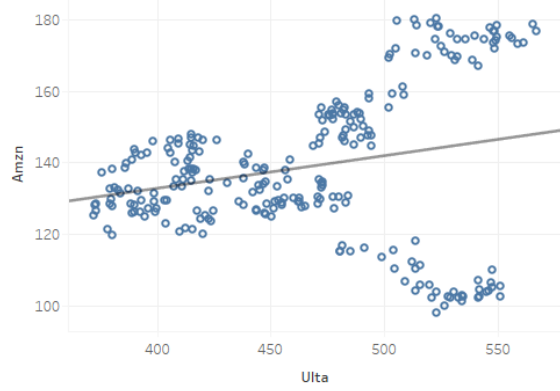
	ULTA
ULTA ULTA	1.00000
BBWI BBWI	0.65097 <.0001
AMZN AMZN	0.23479 0.0002
ETSY ETSY	0.44808 <.0001
EXPE EXPE	0.30231 <.0001
LULU LULU	0.33693 <.0001
MAR MAR	0.33591 <.0001
MGM MGM	0.54838 <.0001
NKE NKE	0.43581 <.0001
POOL POOL	0.52555 <.0001
RL RL	0.60747 <.0001
ROST ROST	0.36734 <.0001
RCL RCL	0.20700 0.0010
SBUX SBUX	0.14978 0.0180
TPR TPR	0.78659 <.0001
WYNN WYNN	0.56731 <.0001
YUM YUM	0.72573 <.0001

Fig 2.1

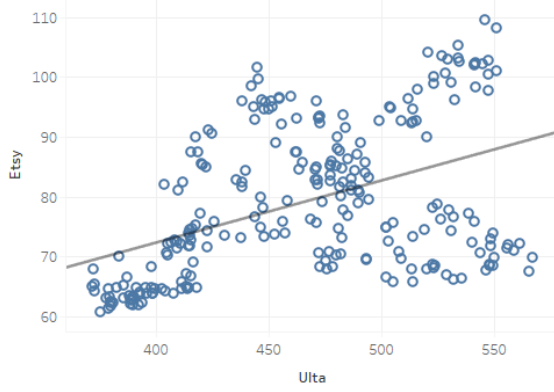
Ulta Vs Bath and Body Works



Ulta Vs Amazon



Ulta Vs Etsy



Ulta Vs Expedia

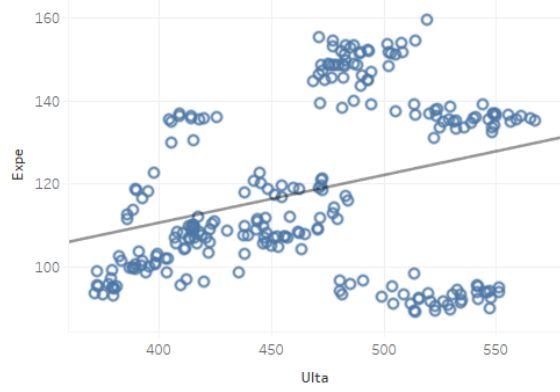
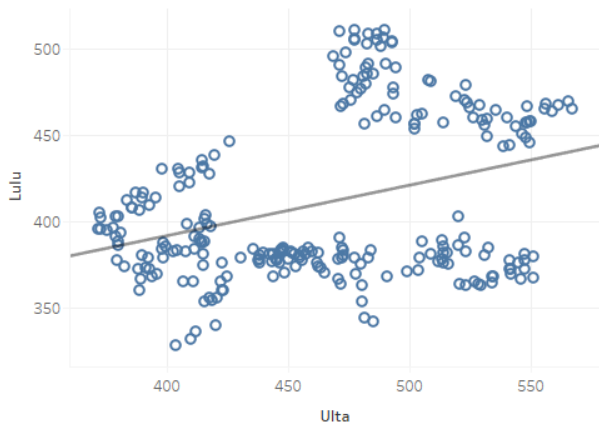
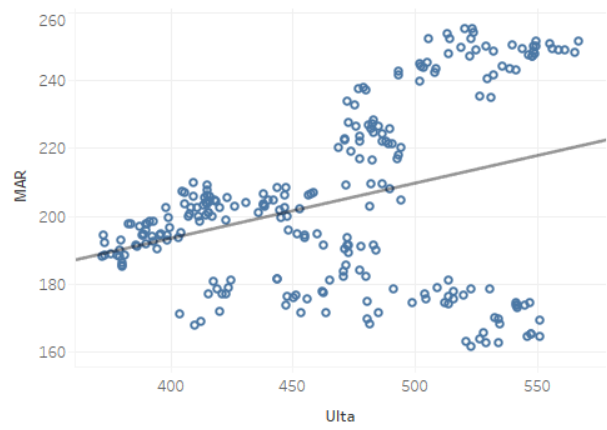


Fig 2.2

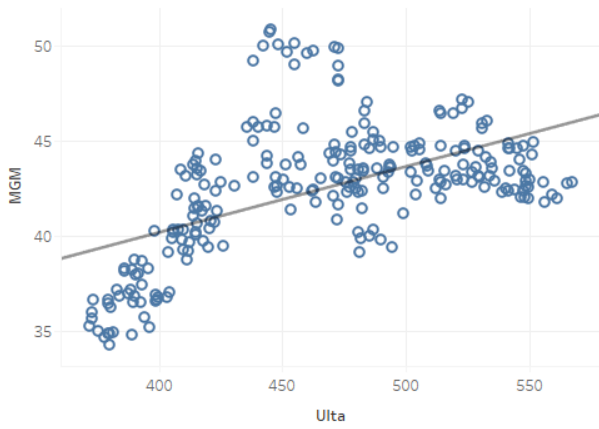
Ulta Vs Bath and Lululemon



Ulta Vs Marriott International



Ulta Vs MGM Resorts



Ulta Vs Nike

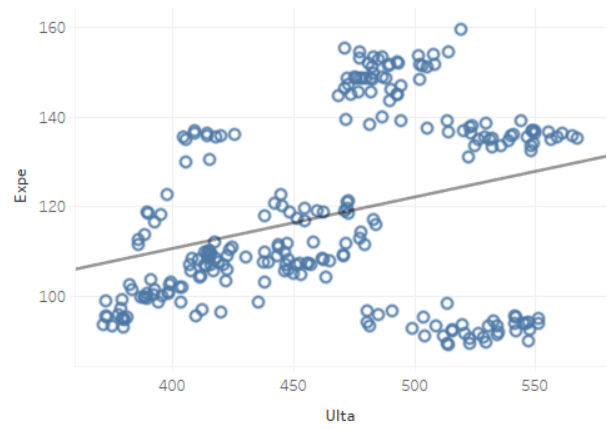
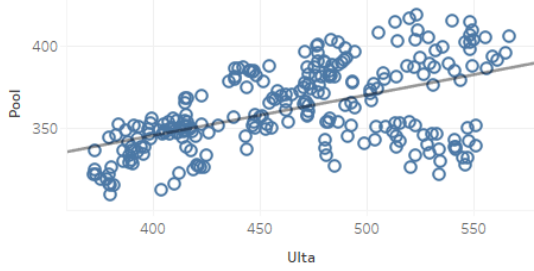
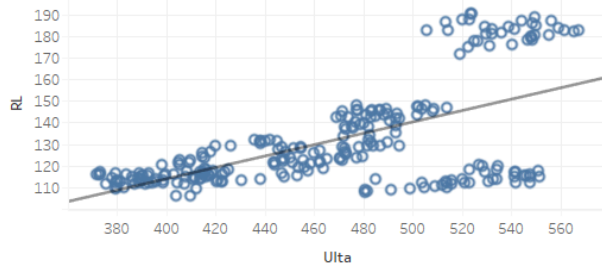


Fig 2.3

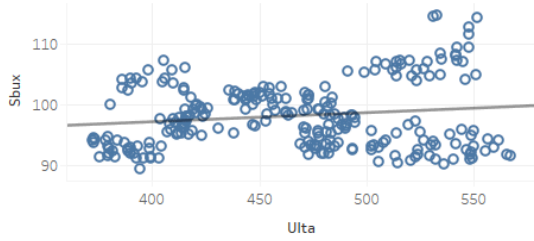
Ulta Vs Pool Corporation



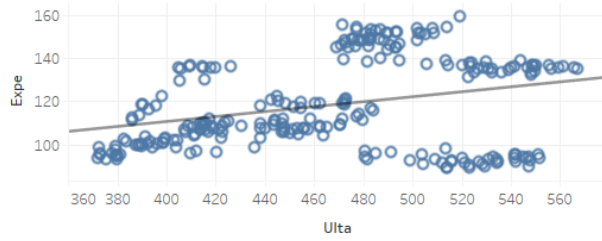
Ulta Vs Ralph Lauren



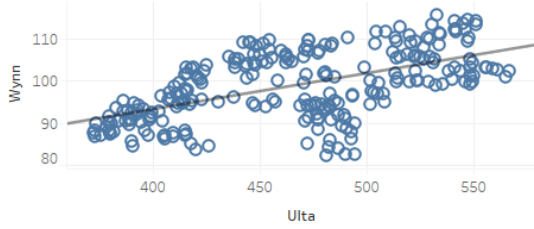
Ulta Vs Starbucks



Ulta Vs Tapestry, Inc.



Ulta Vs Wynn Resorts



Ulta Vs Yum!

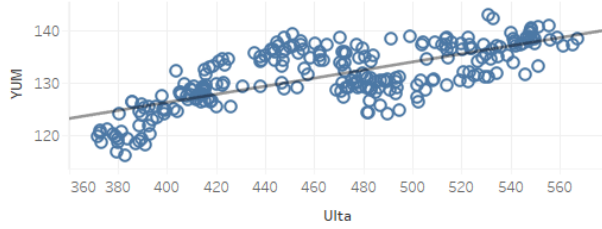


Fig 2.4

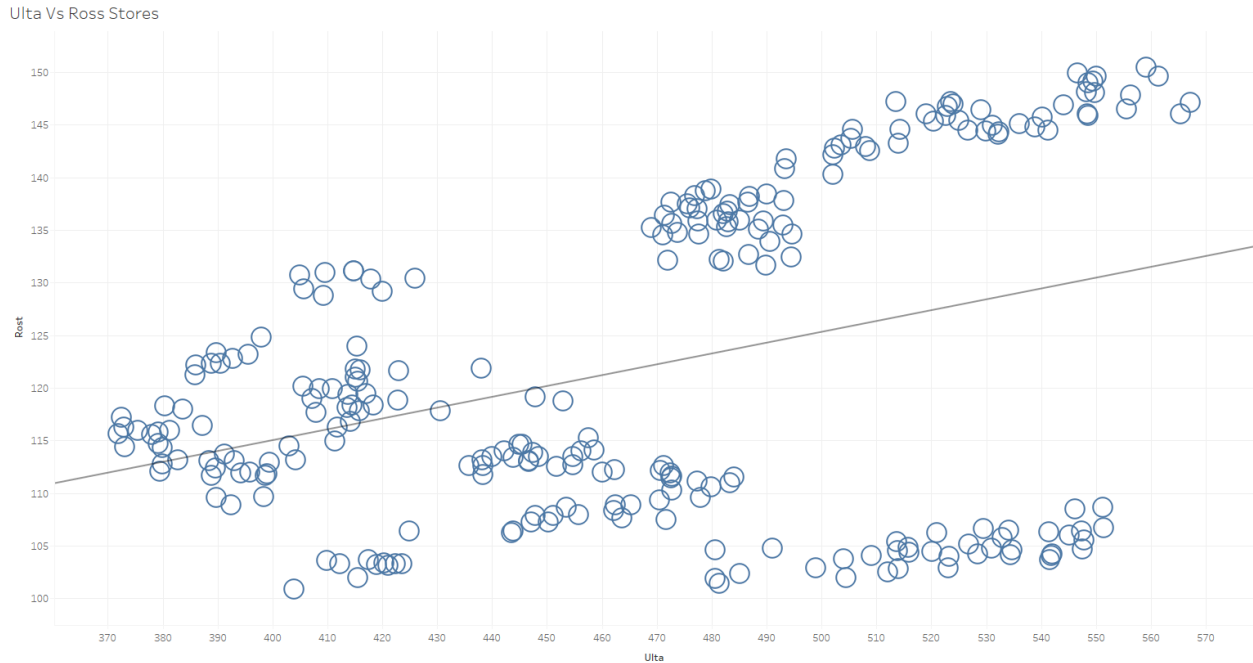


Fig 2.5

SAS CODE:

```
/* Generated Code (IMPORT) */
/* Source File: Spring 2024 Lab #3 Closing Prices.xlsx */
/* Source Path: /home/u63620296/sasuser.v94 */
/* Code generated on: 4/21/24, 12:53 PM */
```

```
%web_drop_table(WORK.IMPORT);
```

```
FILENAME REFFILE '/home/u63620296/sasuser.v94/Spring 2024 Lab #3
Closing Prices.xlsx';
```

```
PROC IMPORT DATAFILE=REFFILE
  DBMS=XLSX
  OUT=WORK.price;
  GETNAMES=YES;
```

```
RUN;
```

```
PROC CONTENTS DATA=WORK.price; RUN;
```

```
%web_open_table(WORK.IMPORT);
```

```
proc corr data=price;  
var ULTA BBWI AMZN ETSY EXPE LULU MAR MGM NKE POOL RL  
ROST RCL SBUX TPR WYNN YUM;  
run;
```

```
proc reg data=price;  
model ULTA = BBWI AMZN ETSY EXPE LULU MAR MGM NKE POOL  
RL SBUX TPR WYNN YUM/selection=stepwise;  
run;
```

```
proc reg data=price;  
model ULTA = BBWI AMZN ETSY LULU MAR MGM NKE RL TPR  
WYNN YUM/VIF;  
run;
```

```
proc reg data=price;  
model ULTA = BBWI AMZN ETSY LULU MGM NKE RL TPR WYNN  
YUM/VIF;  
run;
```

```
proc reg data=price;
```



```

model ULTA = BBWI ETSY LULU MAR MGM NKE RL TPR WYNN
YUM/VIF;
run;
proc reg data=price;
model ULTA = BBWI ETSY LULU MGM NKE RL TPR WYNN
YUM/VIF;
run;
proc reg data=price;
model ULTA = BBWI ETSY LULU MGM NKE RL WYNN YUM/VIF;
run;
/* R-Sq optimal*/
proc reg data=price;
model ULTA = BBWI ETSY LULU MGM NKE RL WYNN
YUM/selection=stepwise;
run;
proc reg data=price;
model ULTA = YUM LULU WYNN MGM NKE RL BBWI;
run;
/* PRESS optimal*/
proc GLMselect data=price;
model ULTA = BBWI ETSY LULU MGM NKE RL WYNN
YUM/selection=stepwise(choose=PRESS);
run;
proc reg data=price;
model ULTA = BBWI ETSY LULU MGM NKE RL WYNN YUM;
run;
/* CP optimal*/

proc reg data=price;
model ULTA = BBWI ETSY LULU MGM NKE RL WYNN
YUM/selection=CP;
run;

```

```

proc reg data=price;
  model ULTA = BBWI LULU MGM NKE RL WYNN YUM;
run;
/* R-sq adj optimal*/
proc reg data=price;
  model ULTA = BBWI ETSY LULU MGM NKE RL WYNN
YUM/selection=adjrsq;
run;

proc reg data=price;
  model ULTA = BBWI ETSY LULU MGM NKE RL WYNN YUM;
run;

/* Final first order model*/
proc reg data=price;
  model ULTA = BBWI LULU MGM NKE RL WYNN YUM/r;
run;

data price_2;
set price;

  BBWI_sq = BBWI**2;
  LULU_sq = LULU**2;
  mgm_sq = mgm**2;
  nke_sq = nke**2;
  rl_sq = rl**2;
  wynn_sq = wynn**2;
  yum_sq = yum**2;

  LULU_NKE = LULU*NKE;
  MGM_WYNN = MGM*WYNN;

```

```
run;
```

```
proc reg data=price_2;  
model ULTA = BBWI LULU MGM NKE RL WYNN YUM BBWI_sq  
LUIU_sq mgm_sq nke_sq rl_sq wynn_sq yum_sq;  
test BBWI_sq, LULU_sq ,mgm_sq ,nke_sq, wynn_sq, yum_sq;  
run;
```

```
proc reg data=price_2;  
model ULTA = BBWI LULU MGM NKE RL WYNN YUM BBWI_sq  
mgm_sq nke_sq wynn_sq lulu_nke/r;  
run;
```

```
proc GLM data=price_2;  
    model ULTA= LULU | NKE/ solution;  
    store GLMMODEL;  
run;
```

```
proc PLM restore=GLMMODEL noinfo;  
    effectplot slicefit(x=LULU sliceby=NKE);  
run;
```

```
proc GLM data=price_2;  
    model ULTA= MGM | WYNN/ solution;  
    store GLMMODEL;  
run;
```

```
proc PLM restore=GLMMODEL noinfo;  
    effectplot slicefit(x=MGM sliceby=WYNN);  
run;
```

```
proc reg data=price_2;  
model ULTA = BBWI LULU MGM NKE RL WYNN YUM BBWI_sq  
mgm_sq nke_sq wynn_sq lulu_nke MGM_WYNN;  
run;
```