

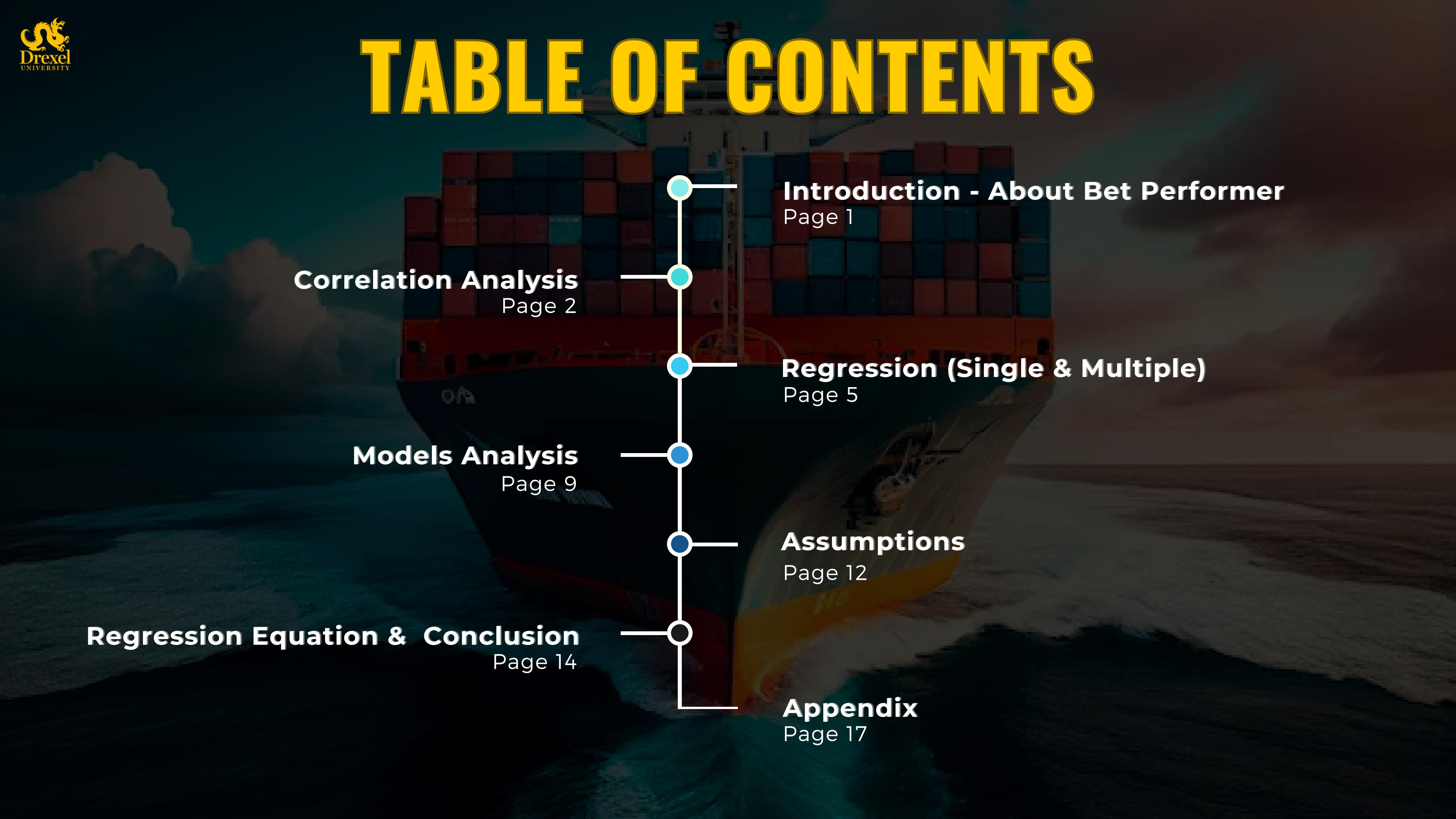




PRICE RECOMMENDATION FOR
BET PERFORMER

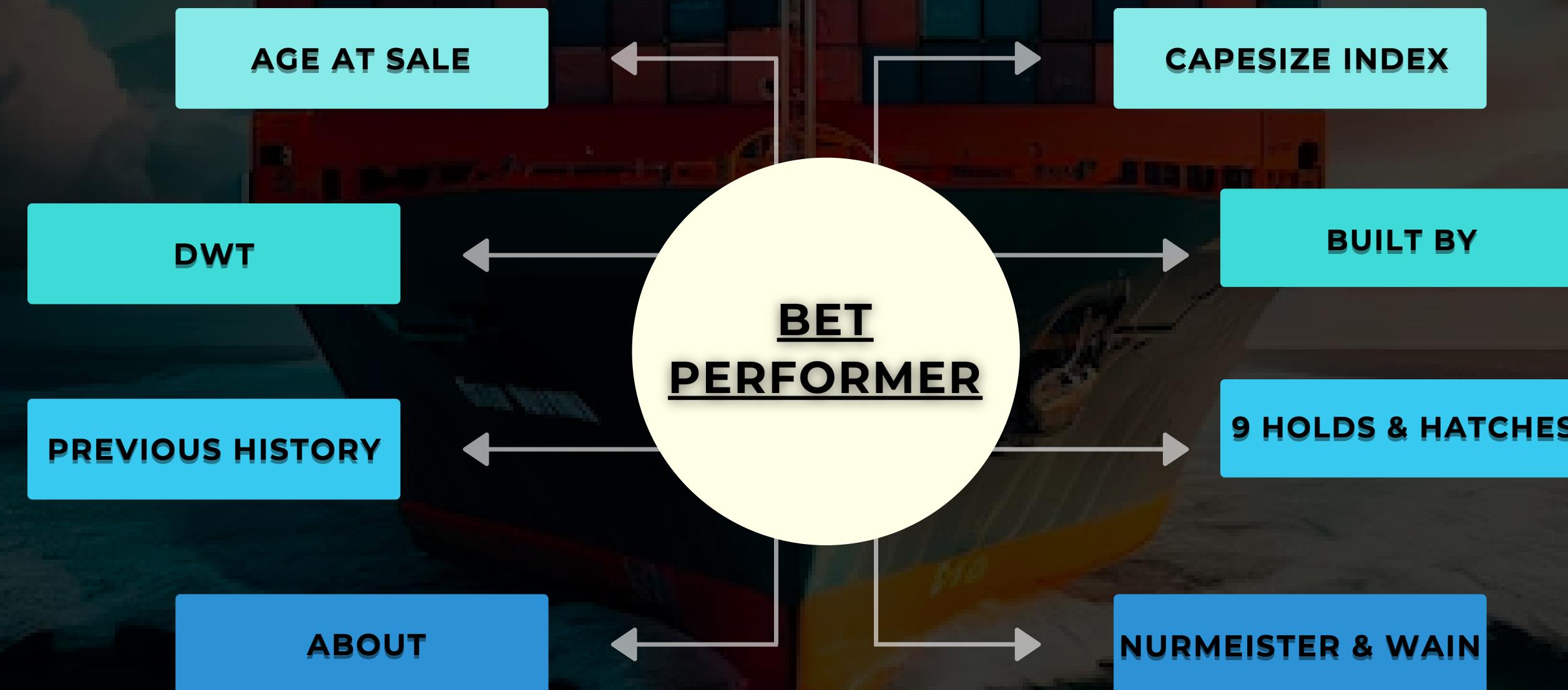
STATISTICS 610
GROUP 2

TABLE OF CONTENTS

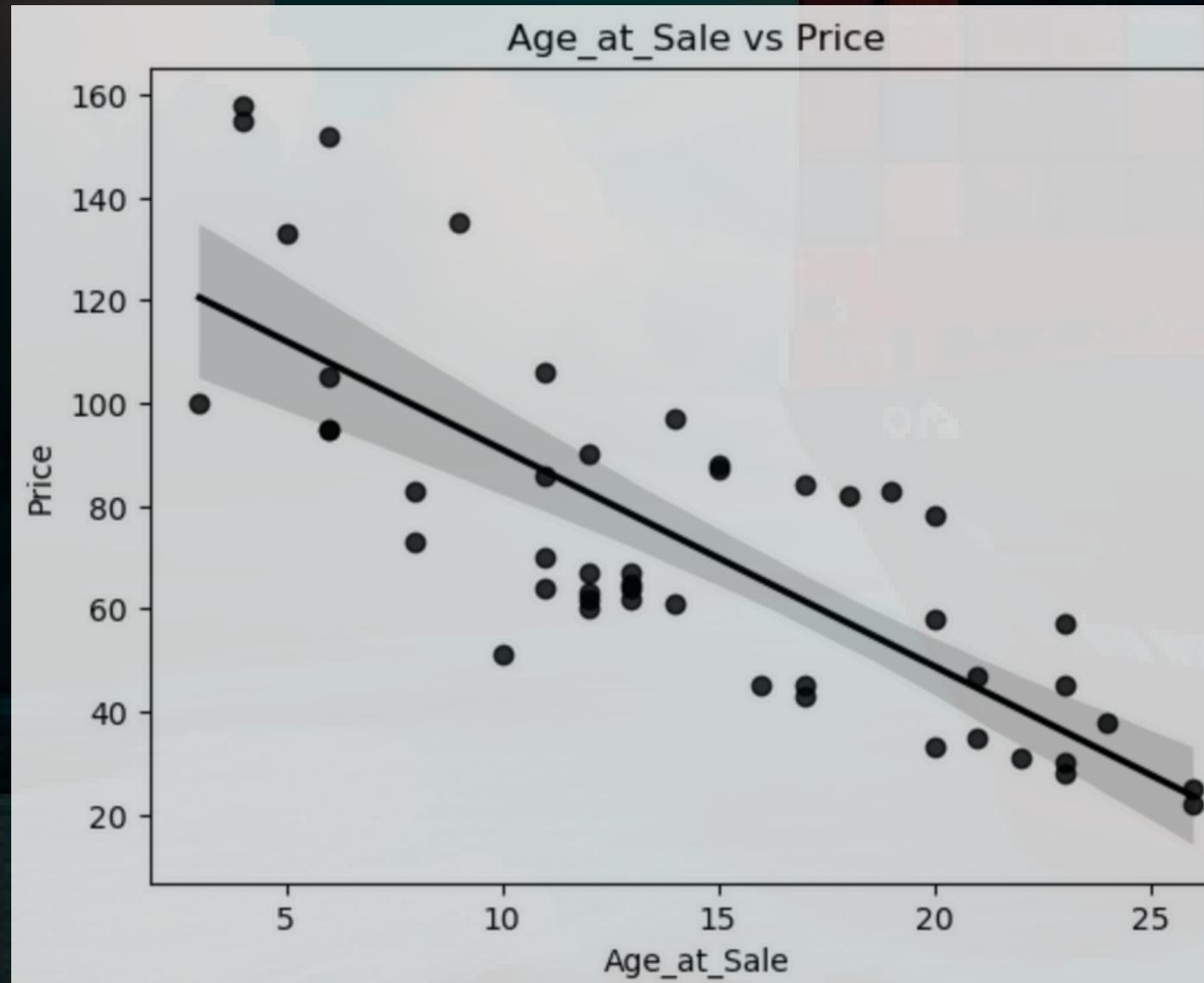


Introduction - About Bet Performer	Page 1
Correlation Analysis	Page 2
Regression (Single & Multiple)	Page 5
Models Analysis	Page 9
Assumptions	Page 12
Regression Equation & Conclusion	Page 14
Appendix	Page 17

WHAT DO WE KNOW ABOUT BET PERFORMER?

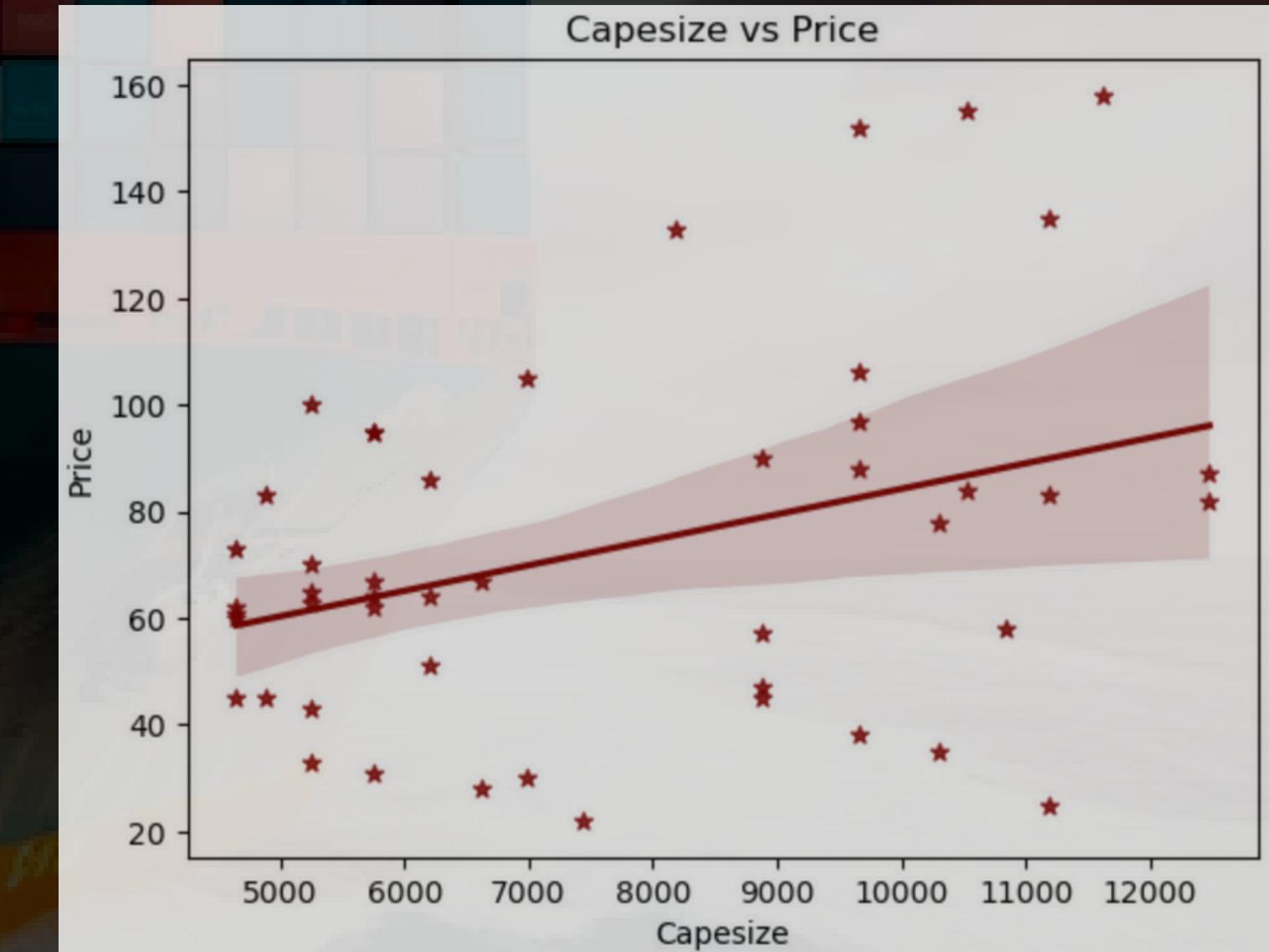


CORRELATION ANALYSIS



Correlation between Age at Sale & Price

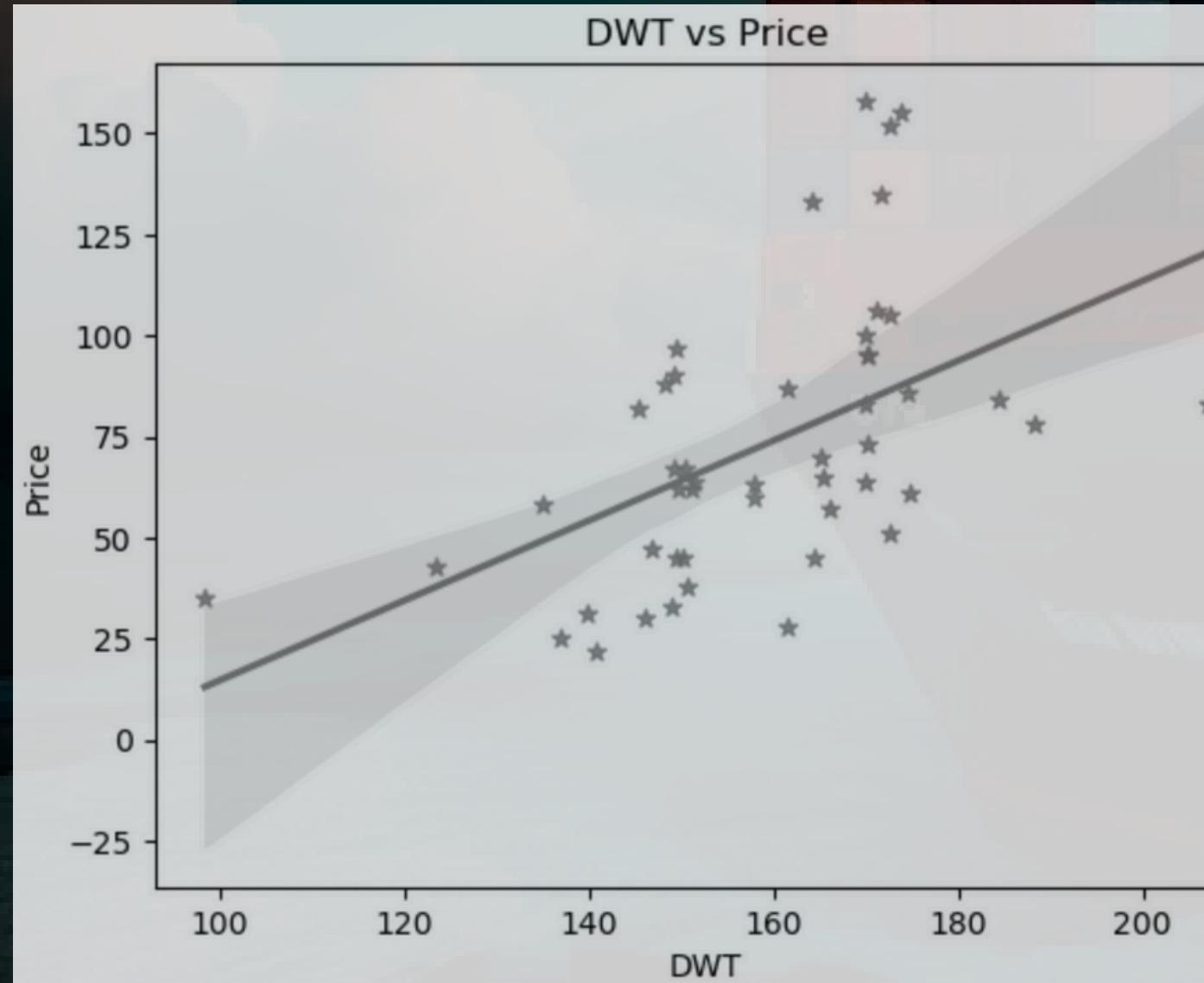
- Negative Correlation between the two variables
- Correlation = - 0.78



Correlation between Capesize & Price

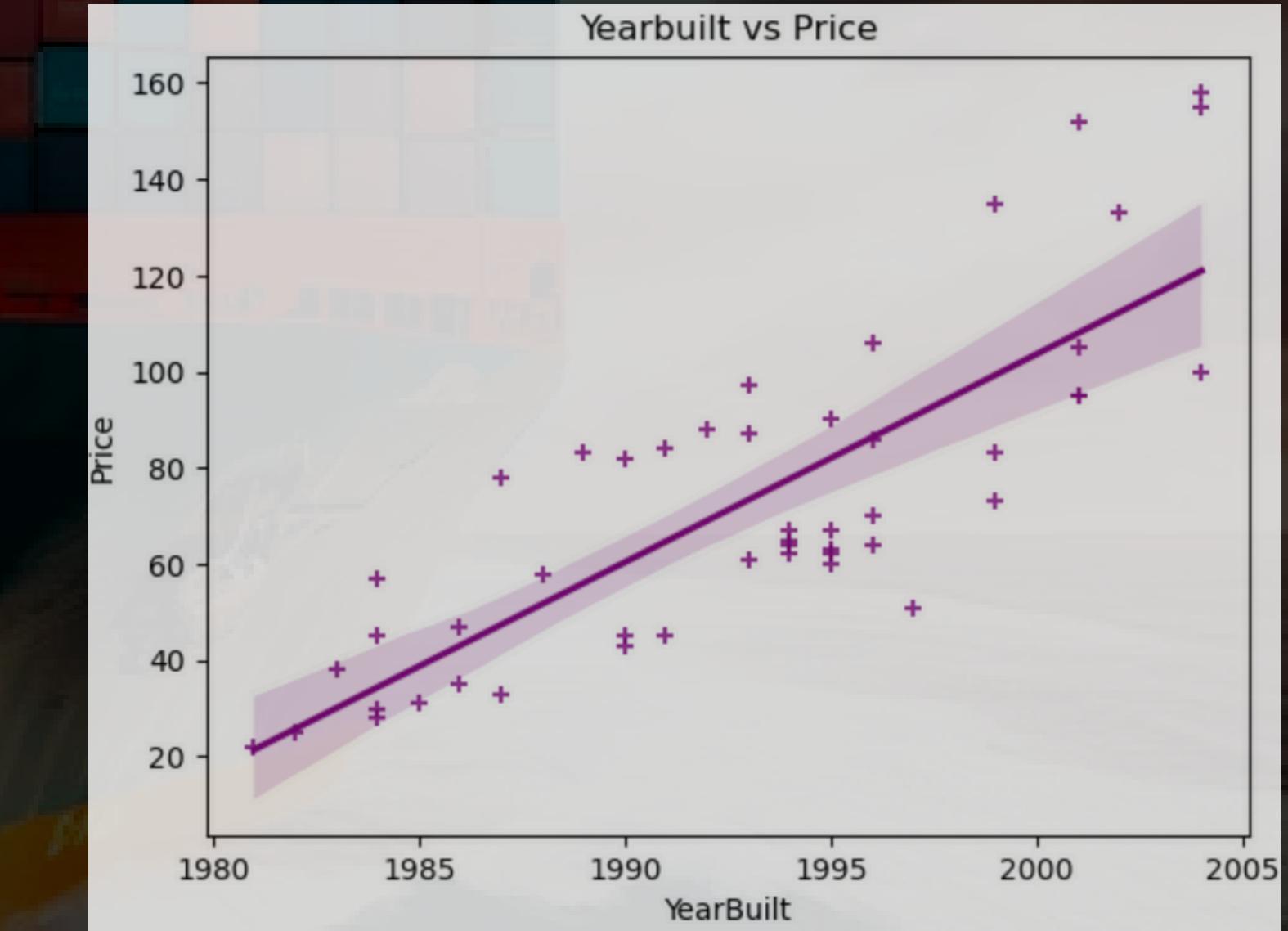
- Positive Correlation between the two variables
- Correlation = 0.35

CORRELATION ANALYSIS



Correlation between DWT & Price

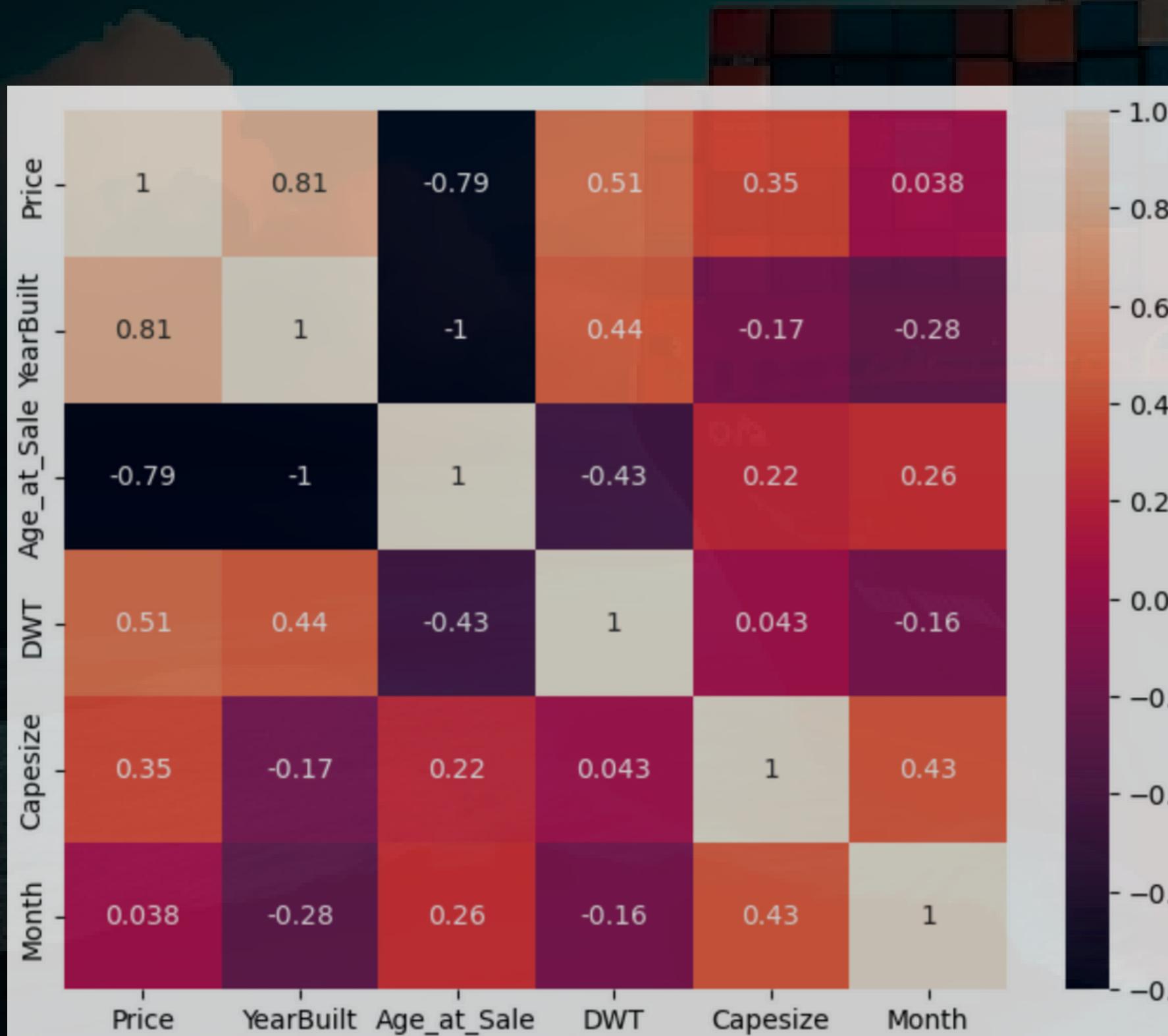
- Positive Correlation between the two variables
- Correlation = 0.514



Correlation between Year Built & Price

- Positive Correlation between the two variables
- Correlation = 0.808

CORRELATION ANALYSIS



	Price	YearBuilt	Age_at_Sale	DWT	Capesize	Month
Price	1.000000	0.808430	-0.787627	0.514754	0.352472	0.038315
YearBuilt	0.808430	1.000000	-0.998059	0.441826	-0.172633	-0.282364
Age_at_Sale	-0.787627	-0.998059	1.000000	-0.431264	0.217360	0.262640
DWT	0.514754	0.441826	-0.431264	1.000000	0.042766	-0.160653
Capesize	0.352472	-0.172633	0.217360	0.042766	1.000000	0.427984
Month	0.038315	-0.282364	0.262640	-0.160653	0.427984	1.000000

SINGLE REGRESSION ANALYSIS

OLS Regression Results

Dep. Variable:	y	R-squared:	0.620			
Model:	OLS	Adj. R-squared:	0.612			
Method:	Least Squares	F-statistic:	75.17			
Date:	Tue, 14 Nov 2023	Prob (F-statistic):	3.11e-11			
Time:	17:23:02	Log-Likelihood:	-213.47			
No. Observations:	48	AIC:	430.9			
Df Residuals:	46	BIC:	434.7			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	133.1600	7.581	17.566	0.000	117.901	148.419
x1	-4.2171	0.486	-8.670	0.000	-5.196	-3.238
Omnibus:	4.109	Durbin-Watson:	0.583			
Prob(Omnibus):	0.128	Jarque-Bera (JB):	3.293			
Skew:	0.522	Prob(JB):	0.193			
Kurtosis:	2.254	Cond. No.	38.9			

Single regression using Age at sales

- The R-squared value of 0.612 showcases a moderate fit of our single regression model.
- The p-value of 0 emphasizes a statistically significant relationship between age of sales and ship prices, instilling confidence in the predictive power of the variable.

SINGLE REGRESSION ANALYSIS

OLS Regression Results						
Dep. Variable:		y	R-squared:		0.265	
Model:		OLS	Adj. R-squared:		0.249	
Method:		Least Squares	F-statistic:		16.58	
Date:		Thu, 16 Nov 2023	Prob (F-statistic):		0.000182	
Time:		00:45:46	Log-Likelihood:		-229.33	
No. Observations:		48	AIC:		462.7	
Df Residuals:		46	BIC:		466.4	
Df Model:		1				
Covariance Type:		nonrobust				
	coef	std err	t	P> t	[0.025	0.975]
const	-84.1192	38.811	-2.167	0.035	-162.241	-5.997
x1	0.9884	0.243	4.072	0.000	0.500	1.477
Omnibus:		5.977	Durbin-Watson:		1.673	
Prob(Omnibus):		0.050	Jarque-Bera (JB):		5.369	
Skew:		0.816	Prob(JB):		0.0682	
Kurtosis:		3.135	Cond. No.		1.46e+03	

Single regression using DWT

- The R-squared value of 0.265 showcases a low fit of our single regression model.
- The p-value of 0 emphasizes a statistically significant relationship between DWT and ship prices, instilling confidence in the predictive power of the variable.

MULTIPLE REGRESSION ANALYSIS

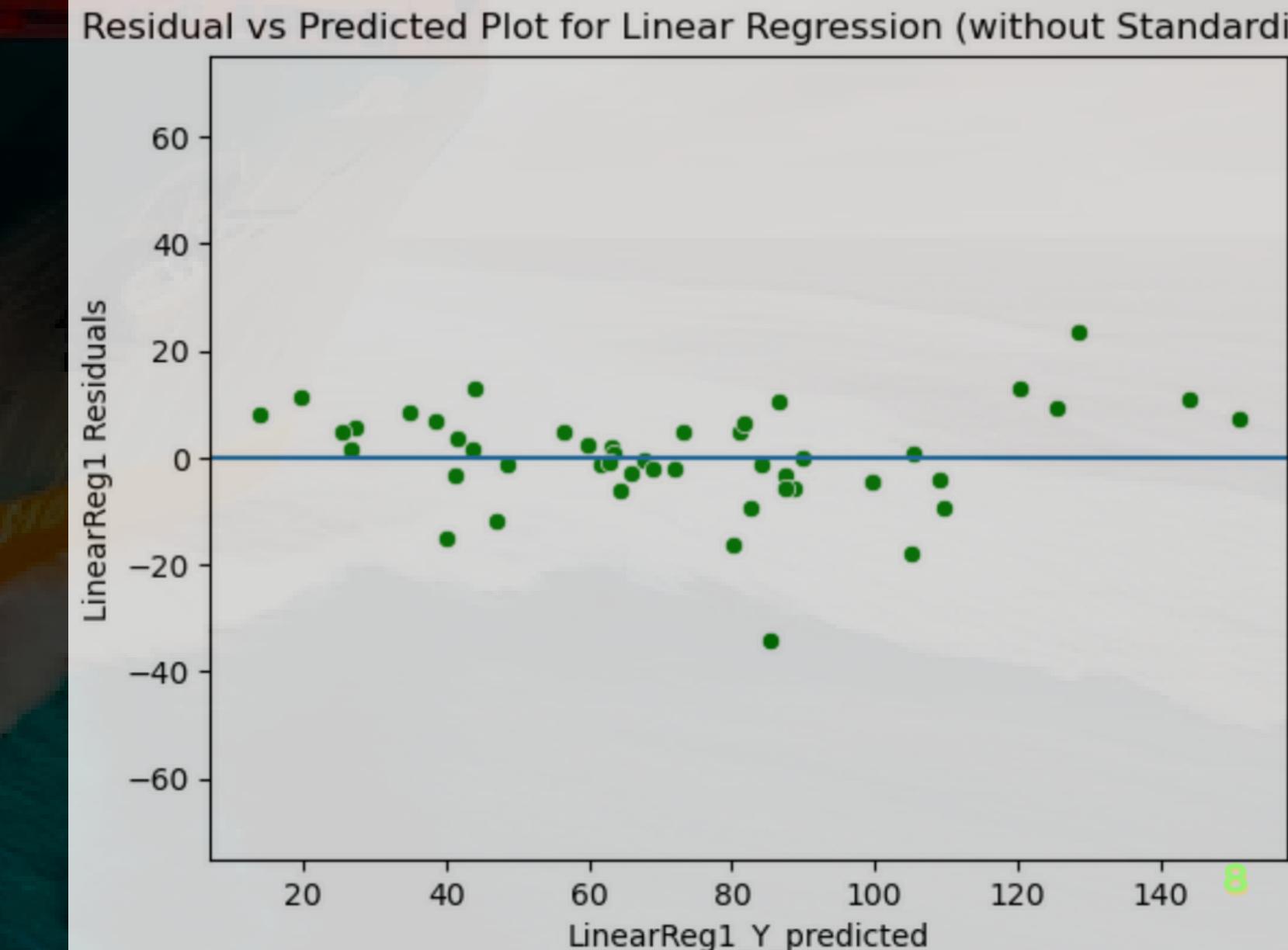
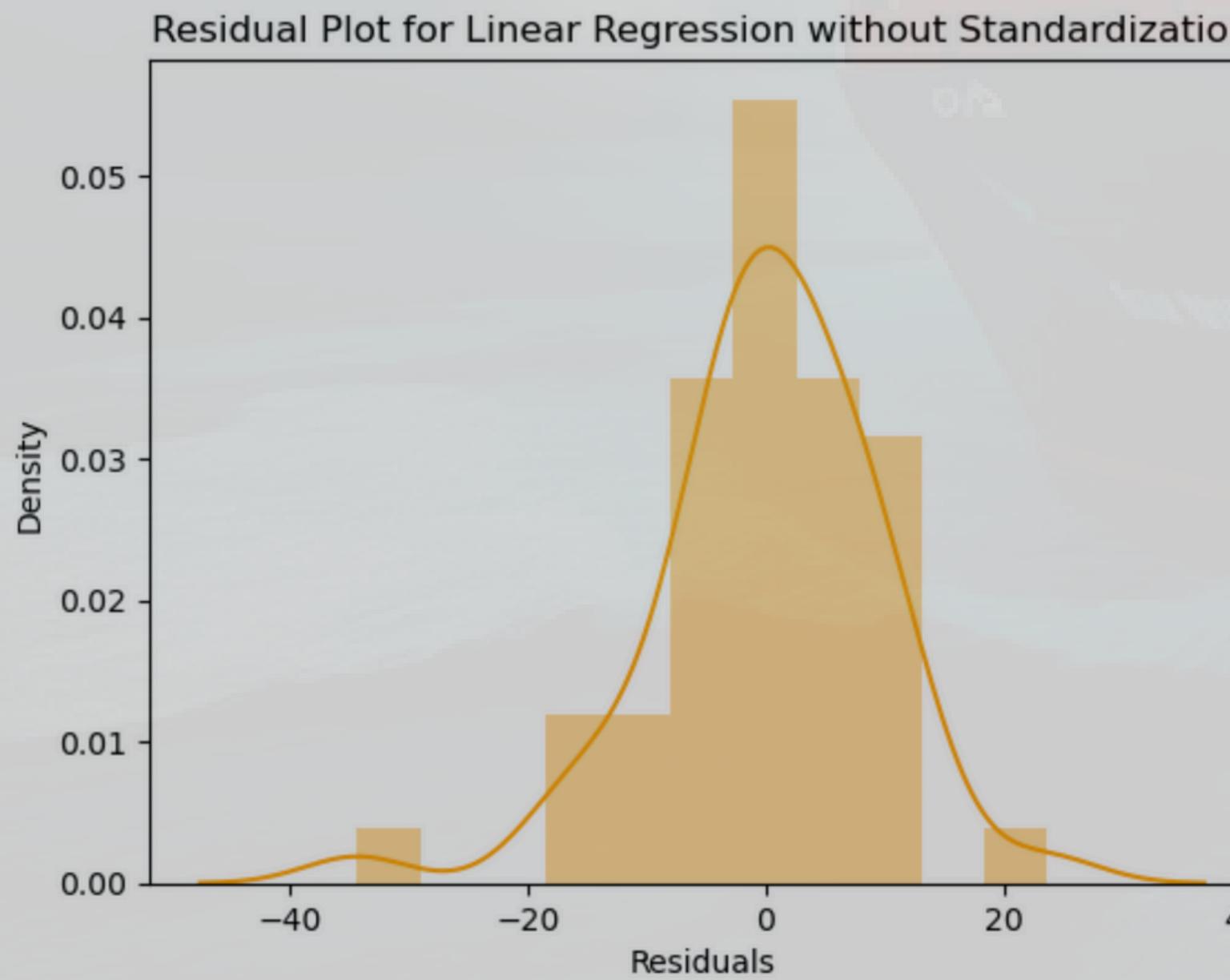
Dep. Variable:	y	R-squared:	0.921
Model:	OLS	Adj. R-squared:	0.915
Method:	Least Squares	F-statistic:	170.5
Date:	Sun, 12 Nov 2023	Prob (F-statistic):	3.07e-24
Time:	17:48:40	Log-Likelihood:	-175.86
No. Observations:	48	AIC:	359.7
Df Residuals:	44	BIC:	367.2
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P> t 	[0.02 5]	0.97 5]
const	44.3002	16.346	2.710	0.010	11.35 6	77.2 44
Age at Sale	-4.5449	0.261	-17.421	0.00 0	-5.07 1	-4.0 19
DWT	0.2418	0.091	2.646	0.011	0.05 8	0.42 6
Capesize	0.0072	0.001	12.081	0.00 0	0.00 6	0.00 8

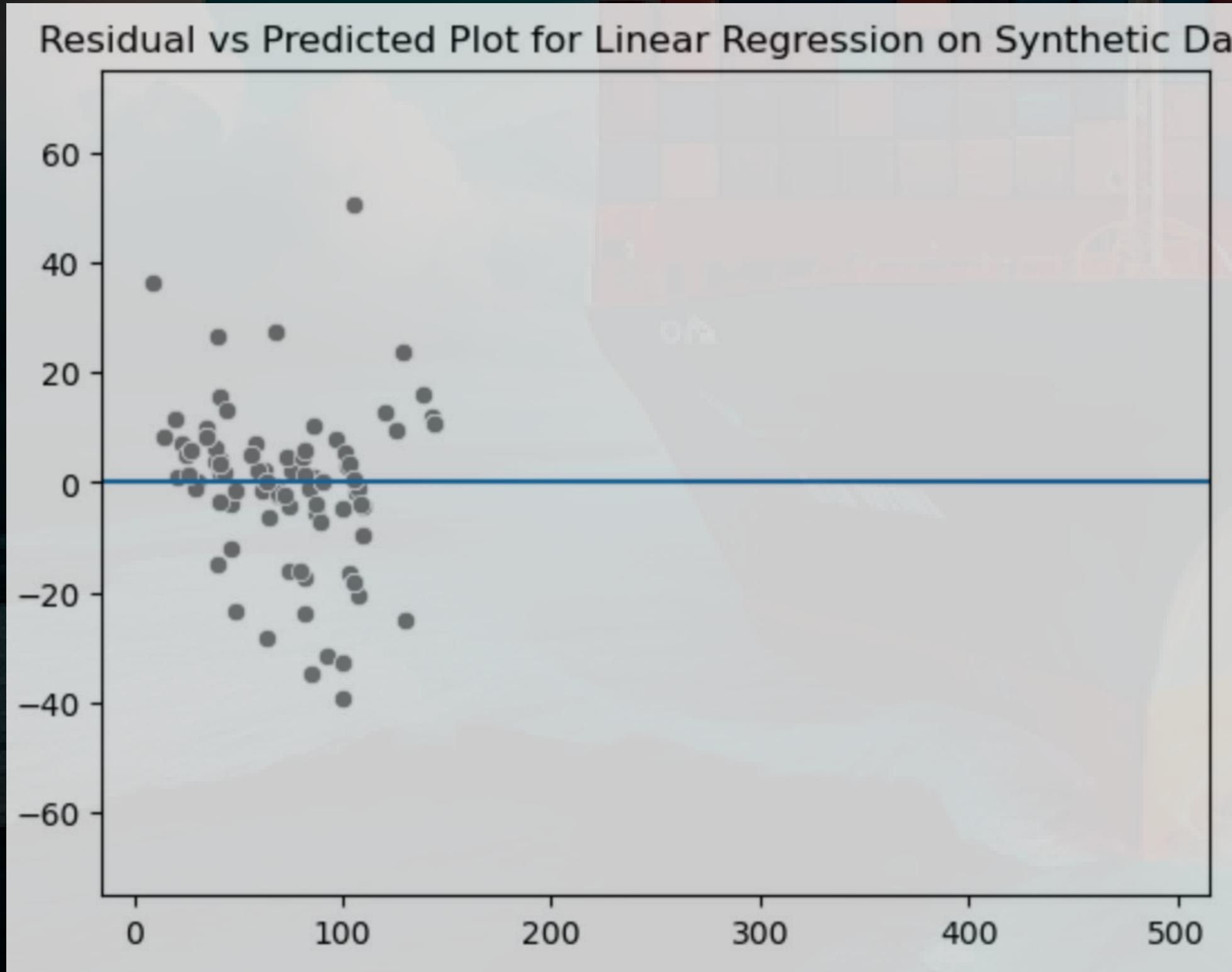
MULTIPLE REGRESSION ANALYSIS

WITH GIVEN DATA

As per our findings, the value of R squared, when using the existing variables, is **92.1**.



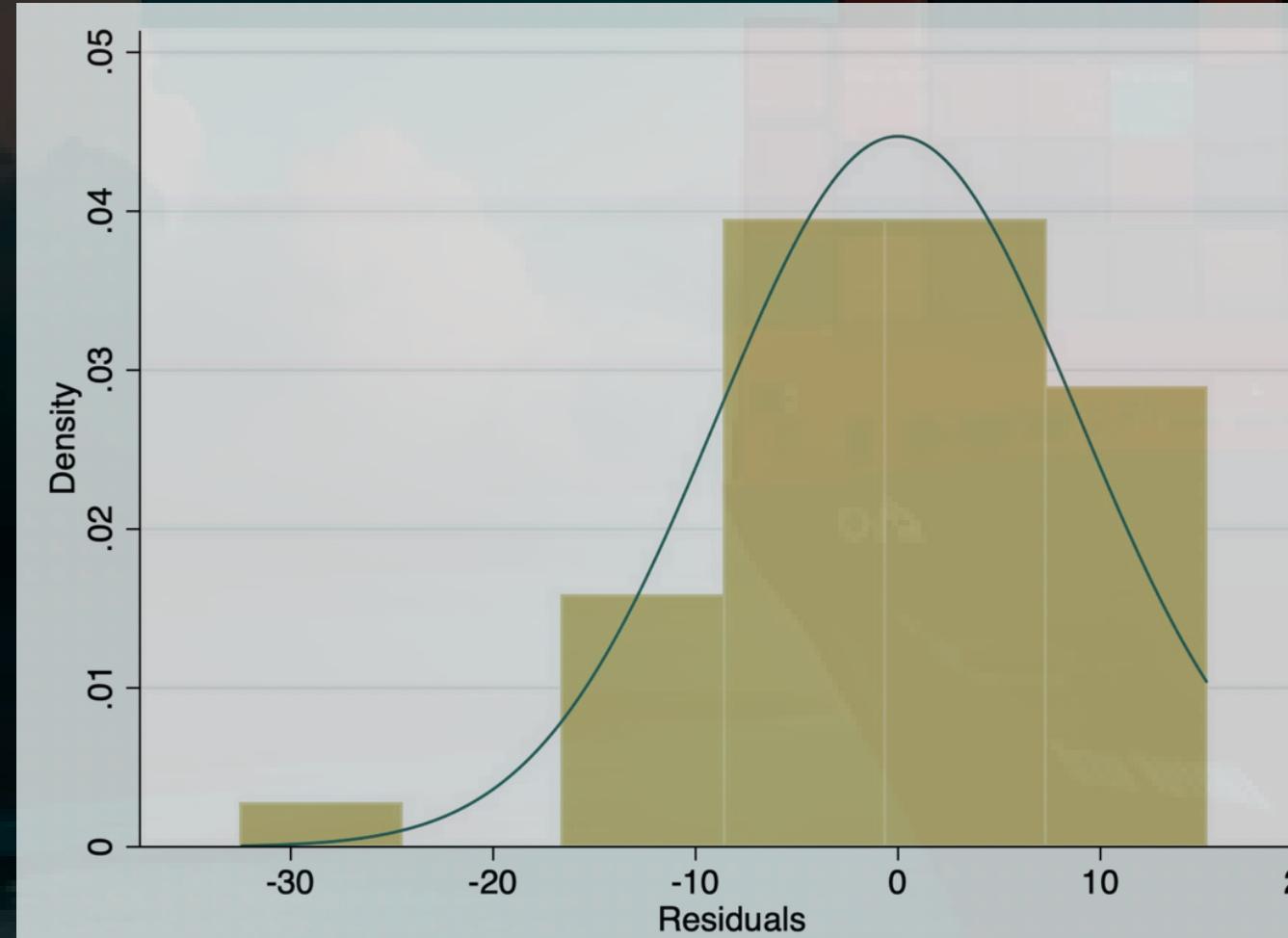
MODELS ANALYSIS - SYNTHETIC DATA



We generated a synthetic data which mimics our original data. We created 200 additional data points using Gretel.ai which uses ACTGAN (Anyway Conditional Tabular Generative Adversarial Network). We used our original dataset as train data and synthetic data as test data and ran multiple regression to see how our models performs on new data points.

The R-squared dropped to 67% from 92%. This resulted in overfitting of the model.

MODELS ANALYSIS - FUEL PRICES



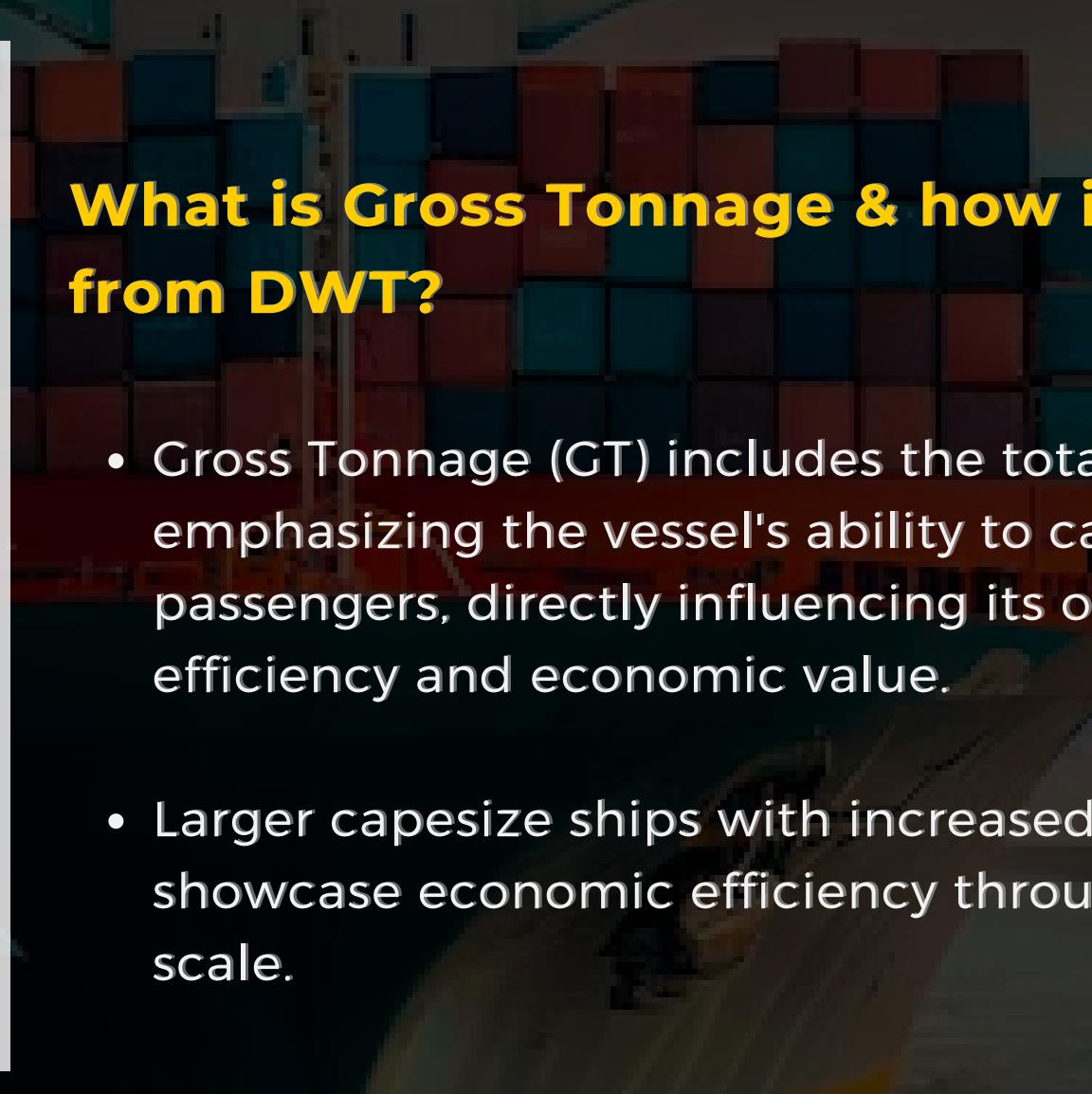
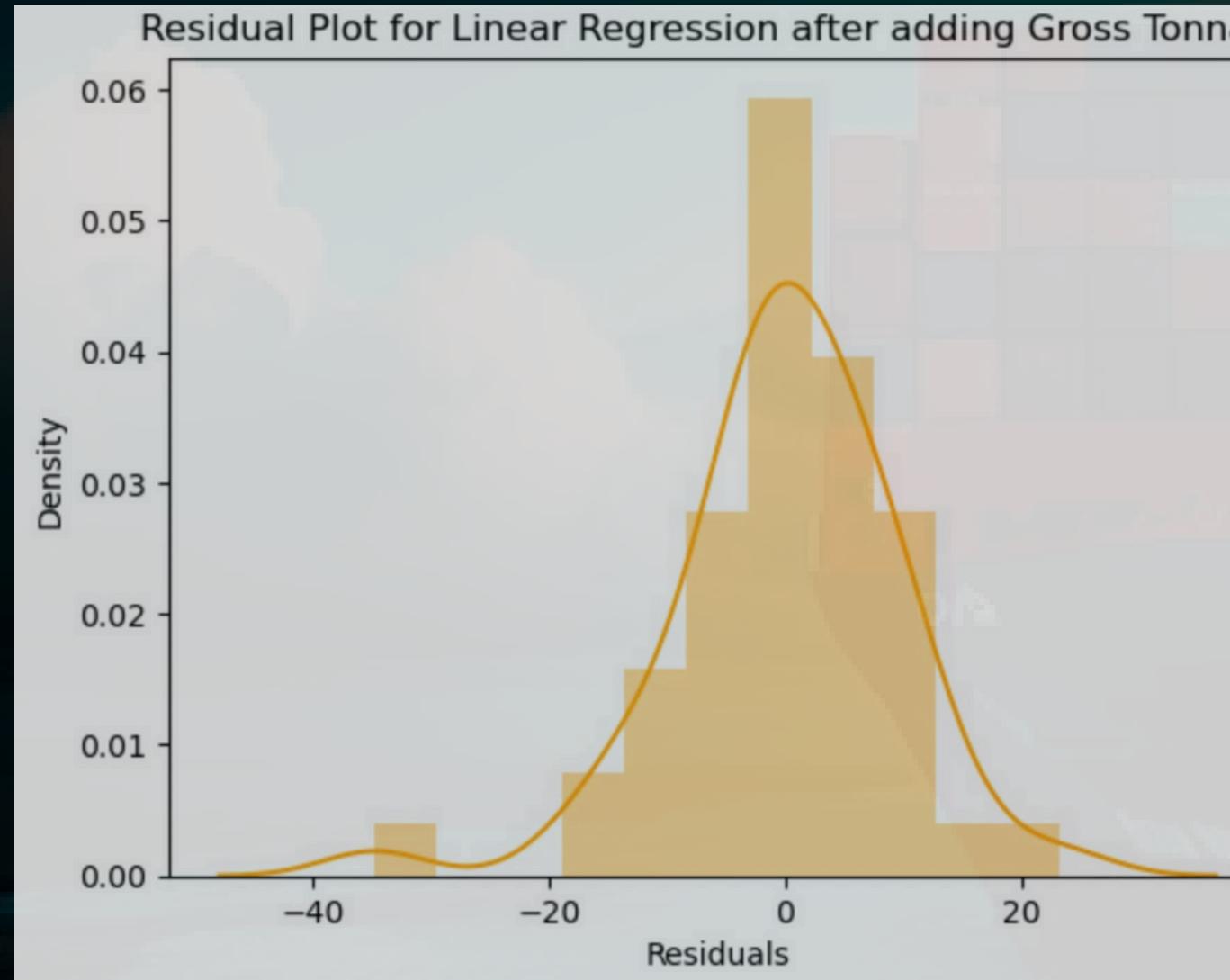
Source	SS	df	MS	Number of obs	=	48
Model	50256.092	4	12564.023	F(4, 43)	=	144.37
Residual	3742.02608	43	87.0238623	Prob > F	=	0.0000
Total	53998.118	47	1148.89613	R-squared	=	0.9307
				Adj R-squared	=	0.9243
				Root MSE	=	9.3287

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ageatsale	-4.582798	.2473236	-18.53	0.000	-5.081574 -4.084022
dwt	.2301844	.0866202	2.66	0.011	.0554982 .4048707
capesize	.0112493	.0016983	6.62	0.000	.0078244 .0146741
fuelprice	-22.17247	8.785222	-2.52	0.015	-39.88955 -4.455379
_cons	81.77965	21.46213	3.81	0.000	38.49713 125.0622

Did Fuel Price have any affect on the price of the ship?

- Analyzing the correlation between fuel prices and capesize ship costs to identify potential economic dependencies.
- Exploring how fluctuations in fuel prices might influence operational expenses and subsequently impact the pricing dynamics of capesize commercial ships.

MODELS ANALYSIS - GROSS TONNAGE



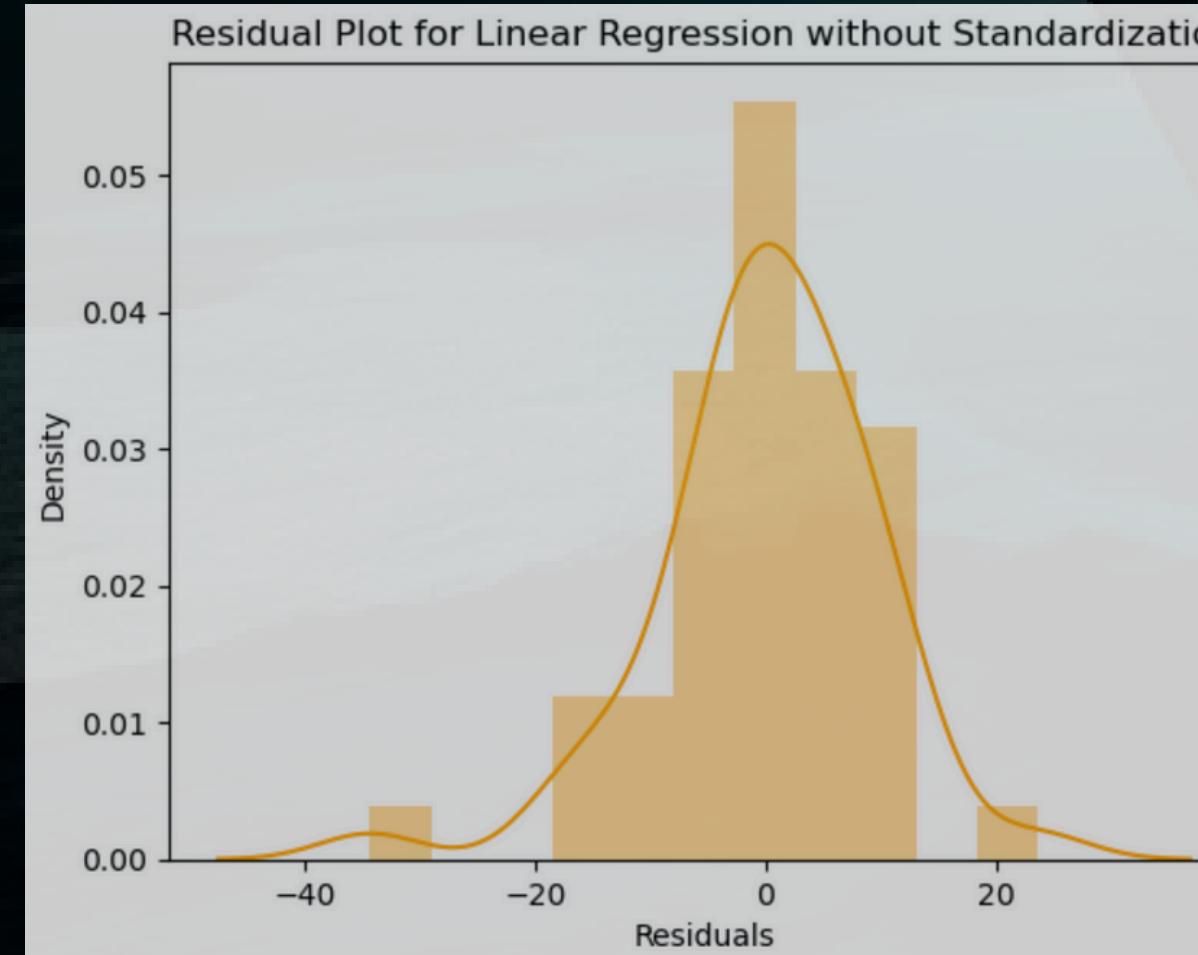
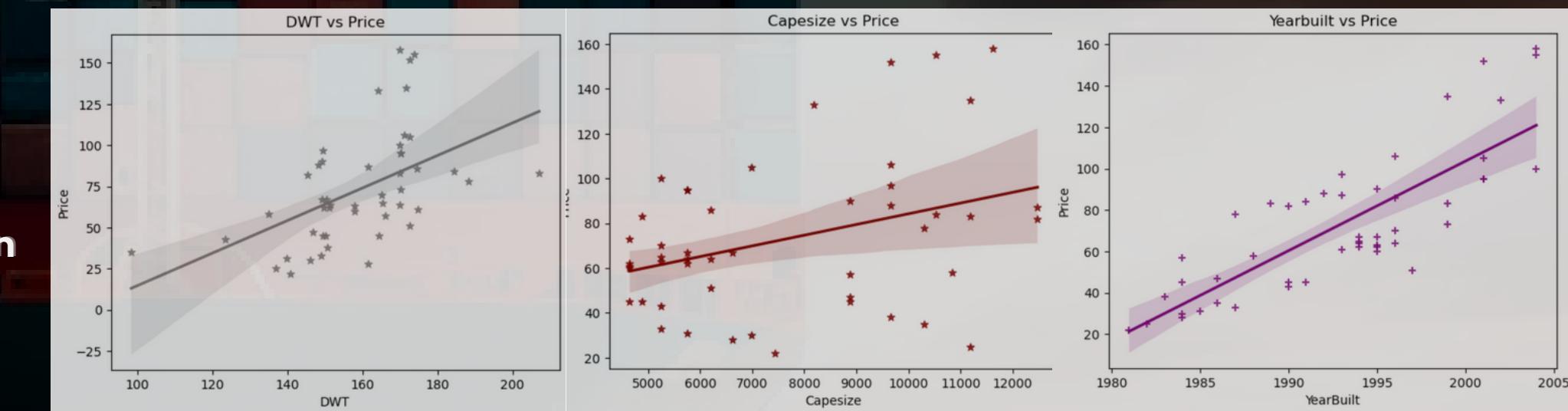
	coef	std err	t	P> t	[0.025	0.975]
const	43.3206	16.615	2.607	0.012	9.813	76.828
x1	-4.5678	0.267	-17.080	0.000	-5.107	-4.029
x2	0.2195	0.103	2.127	0.039	0.011	0.428
x3	0.0072	0.001	11.983	0.000	0.006	0.008
x4	0.0591	0.122	0.483	0.632	-0.188	0.306

Model:	OLS	Adj. R-squared:	0.914
Method:	Least Squares	F-statistic:	125.7
Date:	Wed, 15 Nov 2023	Prob (F-statistic):	3.91e-23
Time:	19:15:27	Log-Likelihood:	-175.73
No. Observations:	48	AIC:	361.5
Df Residuals:	43	BIC:	370.8
Df Model:	4		
Covariance Type:	nonrobust		

ASSUMPTIONS - MULTIPLE REGRESSION

Multiple Regression Assumptions:

1. Linearity - This states that the relationship between the dependent variable and the independent variables should be linear. The following plots show that the relationship between the variables is linear and hence it satisfies the criteria.



Residuals are normally distributed:

From the plot, we can see that residuals are nearly normally distributed.

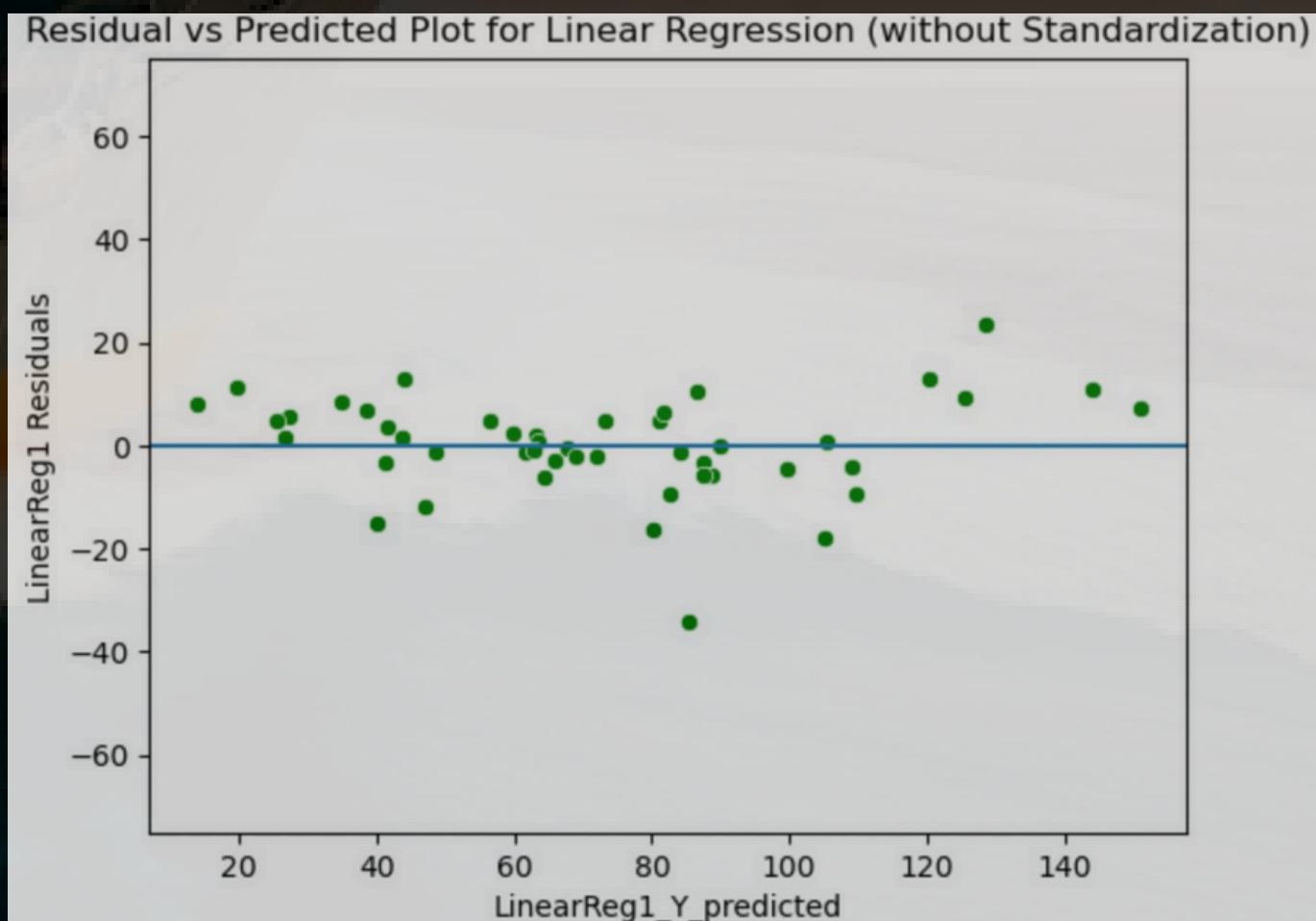
Multicollinearity:

We used Variance Inflation Factor (VIF) to conduct test for no multicollinearity. There appears to be no multicollinearity if the VIF for each variable is between 1 to 5. Below table shows the above-mentioned criteria is satisfied.

	VIF	variable
0	131.944976	Intercept
1	1.258727	DWT
2	1.318729	Age_at_Sale
3	1.075427	Capesize

Homoscedasticity:

It describes a situation in which the error term (that is, the “noise” or random disturbance in the relationship between the independent variables and the dependent variable) is the same across all values of the independent variables. The plot shows that the spread of error term is constant.



REGRESSION EQUATION

PRICE = 44.3 + AGE AT SALE*(-4.54) + DWT*(0.24) + CAPESIZE*(0.0072)

THE ULTIMATE CONCLUSION

$$\text{PRICE} = 44.3 + 11*(-4.54) + 172,000*(0.24) + 12,479*(0.0072)$$

\$125.7 MILLION

95% CONFIDENCE INTERVAL

\$116 MILLION

\$135.4 MILLION

SCENARIO ANALYSIS

	Original	Scenario 1 5 years younger	Scenario 2 20K DWT Light	Scenario 3 Capesize 30% lower
DWT	172	172	152	172
Age	11	6	11	11
Capesize	12,479	12,479	12,479	8,735
Price (in M)	\$125.49	\$148.19	\$120.69	\$98.53

Intercept	44.3
Coefficients	
DWT	0.24
Age	-4.54
Capesize	0.0072

APPENDIX

- Fuel Price Distribution -
https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EMD_EPD2DXL0_PTE_NUS_DPG&f=W
- info on Bet Performer-<https://www.shipspotting.com/photos/179754>
- Gross Tonnage distibution-<https://www.balticshipping.com/vessel/imo/9483243>
- Synthetic data generation- <https://gretel.ai>
- [https://www.breakwaveadvisors.com/insights/2022/3/4/1aozoe8qjjo8zf5mcznujk9a5sunpo#:~:text=As%20the%20price%20of%20oil,\(i.e.%20higher%20derived%20demand\).](https://www.breakwaveadvisors.com/insights/2022/3/4/1aozoe8qjjo8zf5mcznujk9a5sunpo#:~:text=As%20the%20price%20of%20oil,(i.e.%20higher%20derived%20demand).)
- <https://books.google.com/books?hl=en&lr=&id=HiVbBAAAQBAJ&oi=fnd&pg=PP1&ots=gTsQWnvksY&sig=UNoABwqgJR5uNxhqG9ys5t1b1HE#v=onepage&q&f=false>
- <https://link.springer.com/article/10.1007/s13437-022-00272-4#Sec2>
- For reference - <https://chat.openai.com/>

THANKYOU!

WE ARE OPEN TO QUESTIONS



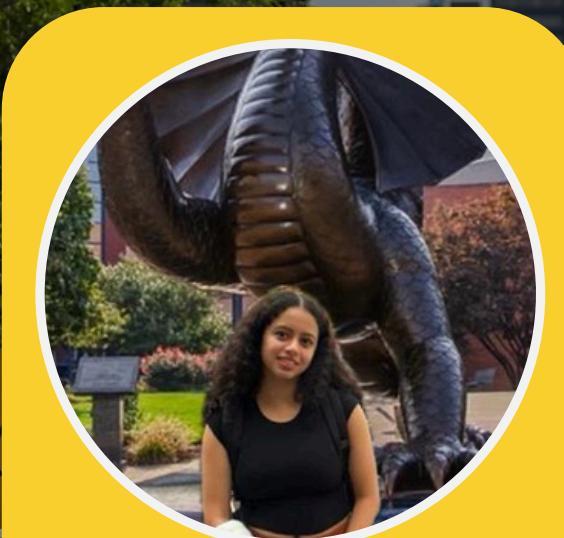
Naitik
Nayak



Gurpreet Singh
Chandok



Rishika Lall



Sana Parab



Ashikul Kabir



Rakshit Madan