Acme Jewelry Company
Estimating the Impact of Synthetic Diamond's Weight on Sale Price

University of California, Berkeley | School of Information

DATASCI 203-003 Statistics for Data Science - Fall 2022

Team 2: Ivy Chan, Jonathan Hodges, Dipika Kumar, Christian Lee

December 6, 2022

Berkeley
UNIVERSITY OF CALIFORNIA

AGENDA

1. Introduction
2. Data & Methodology
3. Results
4. Limitations
5. Conclusion

Berkeley
UNIVERSITY OF CALIFORNIA

INTRODUCTION

- Technological advances in synthetic diamond manufacturing has led to increased carats of produced synthetic diamonds.
- Should Acme upgrade their old manufacturing process? Is it economically viable to do so?
- Focus of Study: How large is effect of synthetic diamond carat on the diamond's sales price?

Data & Methodology

Dataset

- Diamonds dataset on Kaggle from 2021 by Abhijit Singh

Cause	Number of Samples Available for Analysis (after removal for cause)	Removed Number Samples for cause
Start	53,940	0
Remove samples with dimension variables with value of 0	53,920	20
7 of 10 variables transformed to natural log scale	53,920	0
Remove samples where natural log values of volume are infinite	53,920	17
Split into 30% EDA set (16,171) and 70% confirmation set (37,732)	53,903	0

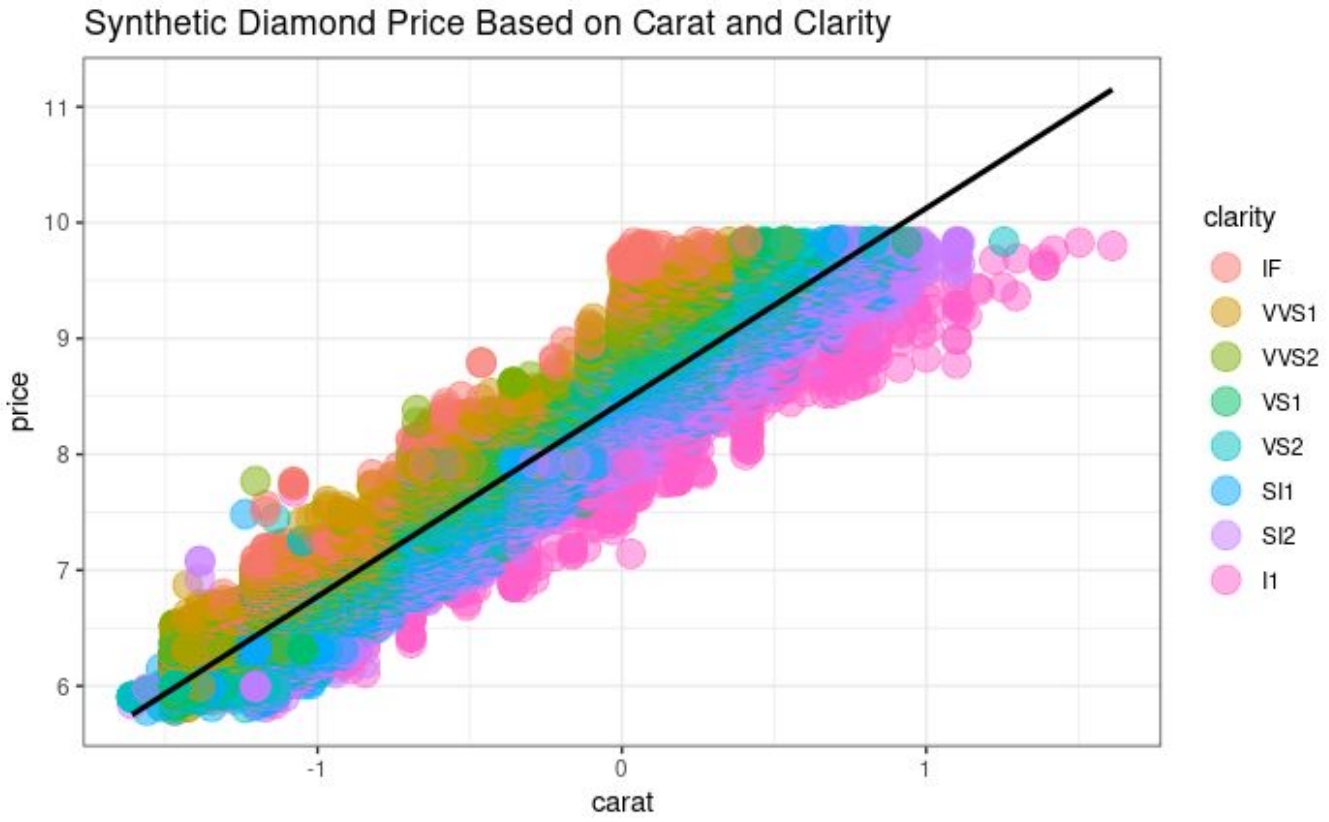
Model Variables

- To predict the price model we had to operationalize varying features.

Operationalized	Variable	Type	Manipulation
Response	Price	Metric	Log
Predictor	Carat		
Predictor	Volume		Log (Length_in_mm * Width_in_mm * Depth_in_mm)
Predictor	Clarity	Categorical	Hot encoded to ordinal
Predictor	Color		

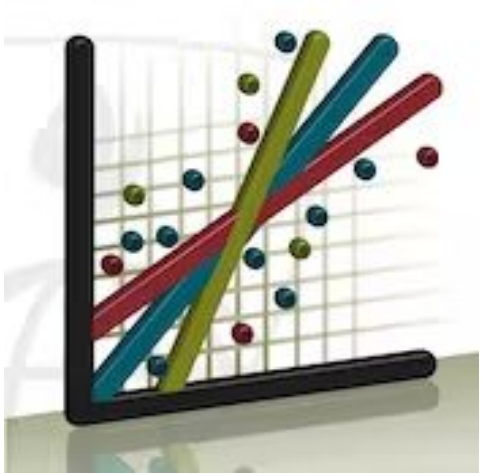
OLS Regression Model

$$price = \beta 0 + \beta 1C + Z\gamma$$



Result

Three Regression Models



- Model 1: price ~ carat
- Model 2: price ~ carat + color + clarity
- Model 3 : price ~ carat + color + clarity + volume

Stargazer: Model Comparison

Estimated Regressions

	Output Variable: price per carat		
	(1)	(2)	(3)
Constant	8.45***	9.43***	6.89***
	(0.002)	(0.004)	(0.13)
Carat	1.68***	1.87***	1.38***
	(0.002)	(0.001)	(0.03)
Color		-0.08***	-0.08***
		(0.0005)	(0.0005)
Clarity		-0.13***	-0.13***
		(0.001)	(0.001)
Volume (Length * Width * Depth)			0.50***
			(0.03)
Observations	37,895	37,895	37,895
R2	0.93	0.98	0.98
Adjusted R2	0.93	0.98	0.98
Residual Std. Error	0.26 (df = 37893)	0.15 (df = 37891)	0.15 (df = 37890)
Note:			*p<0.05; **p<0.01; ***p<0.001
			HCrobust standard errors in parentheses.

Berkeley
UNIVERSITY OF CALIFORNIA

Limitations

OLS Regression Limitations

✅ Independence and Identically Distributed (I.I.D.)
- ✅ Independence → Data randomly collected
- ✅ Identically Distributed* → Insignificant data removal

* Geographic sampling location unknown

🚫 Unique BLP Exists
- → Bimodal distributions
- → Heavy tails
- → No perfect collinearity

❓ Omitted Variable Bias
- → Unknown variables i.e economic conditions

Conclusion

Recommendation

- Acme Synthetic Diamond Company should upgrade to the new equipment

References

References

Content:

- https://www.kaggle.com/code/abhijit10singh/eda-applying-multiple-linear-regression/data
- https://medium.com/@kyawsawhtoon/log-transformation-purpose-and-interpretation-9444b4b049c9
- https://instoremag.com/its-real-acceptance-of-lab-grown-diamonds-is-expanding-exponentially/
- https://www.nationaljeweler.com/articles/11284-looking-ahead-4-factors-that-could-shape-lab-grown-diamonds-future
- https://www.cnn.com/2022/04/27/business/diamonds-manmade-demand/index.html
- https://indianewengland.com/why-lab-grown-diamonds-are-a-rage-amongst-millennials/
- https://www.professionaljeweller.com/industry-view-the-future-of-lab-grown-diamonds/
- https://www.professionaljeweller.com/feature-will-lab-grown-diamonds-ever-overtake-natural-sales/
- https://www.jewellermagazine.com/Article/10945/Lab-created-diamond-sales-reaching-new-heights-in-US
- https://www.prnewswire.com/news-releases/igi-certifies-worlds-largest-lab-grown-diamond-30-carats-301564929.html

References

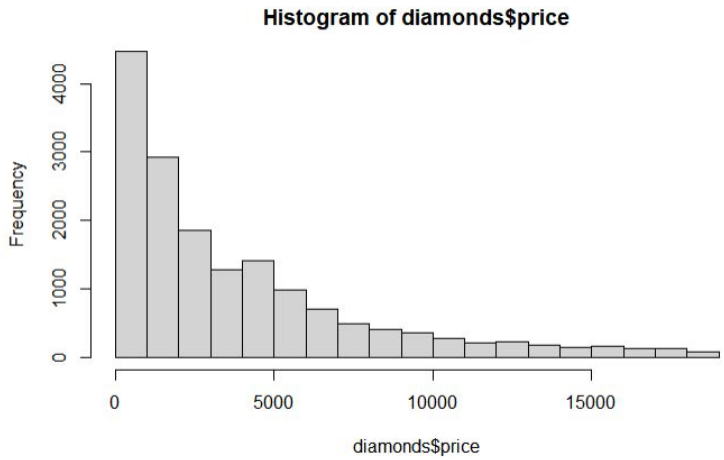
Images:

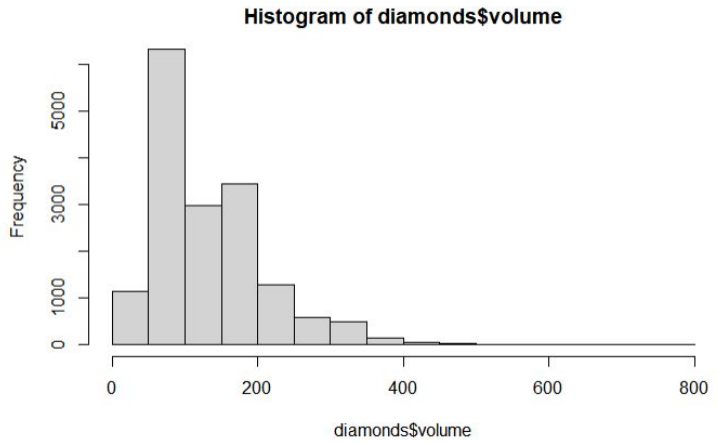
- https://acmejewelry.com.ph/
- https://www.google.com/url?sa=i&url=https%3A%2F%2Fnews.mit.edu%2F2020%2Fdiamond-metal-conductor-1005&psig=AOvVaw2rWiEPRKgRIp_ShOueKsqk&ust=1670034104273000&source=images&cd=vfe&ved=0CA8QjRxqFwoTCOj1mv7v2fsCFQAAAAAdAAAAABAa
- https://www.nationaljeweler.com/articles/10953-a-milestone-in-the-lab-grown-industry-the-3-largest-certified-diamonds-over-15-carats-coming-to-jck/gallery
- https://d3njjcbhbojbot.cloudfront.net/api/utilities/v1/imageproxy/https://coursera-course-photos.s3.amazonaws.com/4e/2b9450fd5011e88a28fd978cb69b7d/Public-Health-Biostatistic_Logo5_Multiple-Regression-Methods-04.png?auto=format%2Ccompress&dpr=1&w=175&h=175&fit=fill&bg=FFF
- https://medium.com/@shuklapratik22/linear-regression-from-scratch-a3d21eff4e7c

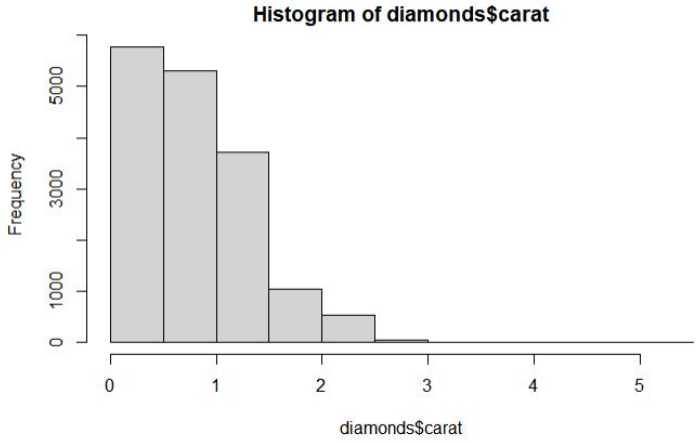
Appendix

Metric Variables

- Data set variables:

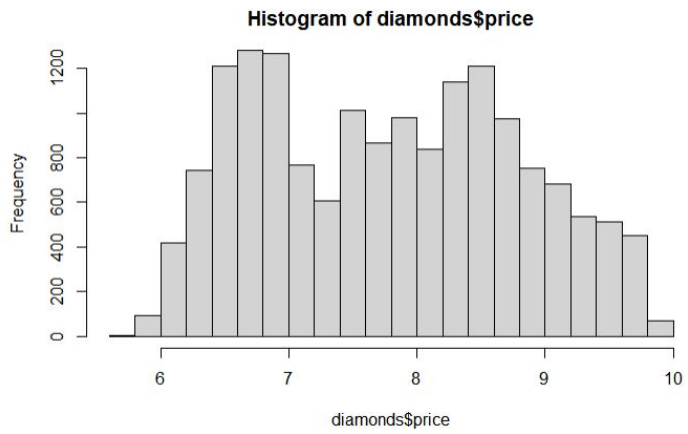


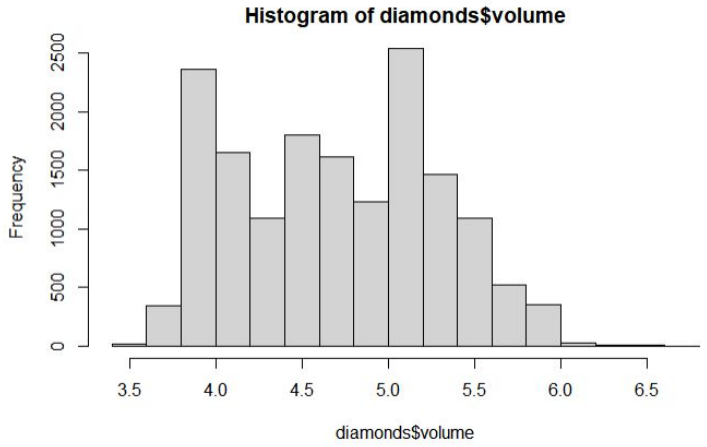


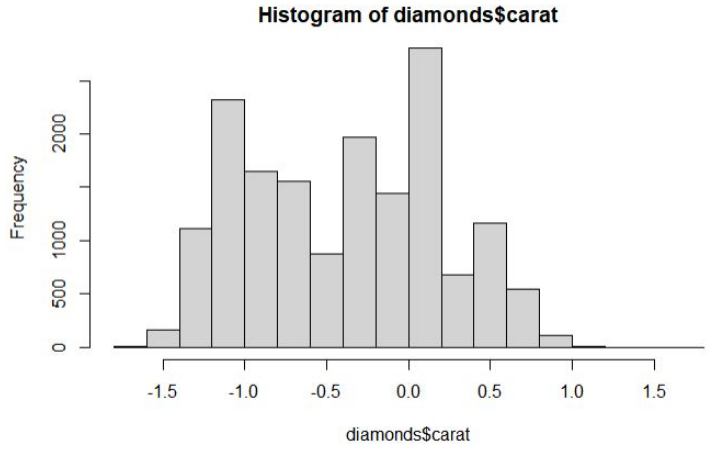


Metric Variables after Data Transformation

- Data set variables:







Berkeley
UNIVERSITY OF CALIFORNIA

Stargazer: Testing Dataset

Table 1: Estimated Regressions

	Output Variable: price per carat		
	(1)	(2)	(3)
Constant	8.45***	9.43***	7.91***
	(0.002)	(0.01)	(0.18)
Carat	1.67***	1.87***	1.58***
	(0.004)	(0.002)	(0.03)
Color		-0.08***	-0.08***
		(0.001)	(0.001)
Clarity		-0.13***	-0.13***
		(0.001)	(0.001)
Volume (Length * Width * Depth)			0.30***
			(0.03)
Observations	16,159	16,159	16,159
R2	0.93	0.98	0.98
Adjusted R2	0.93	0.98	0.98
Residual Std. Error	0.26 (df = 16157)	0.15 (df = 16155)	0.15 (df = 16154)

Note: *p<0.05; **p<0.01; ***p<0.001
HCrobust standard errors in parentheses.

Berkeley
UNIVERSITY OF CALIFORNIA

VIF

Model 2

Variables <chr>	Tolerance <dbl>	VIF <dbl>
carat	0.7825562	1.277864
color_in_num_order	0.9151113	1.092763
clarity_in_num_order	0.8378022	1.193599

3 rows

Model 3

Variables <chr>	Tolerance <dbl>	VIF <dbl>
carat	0.002039398	490.340678
color_in_num_order	0.914761112	1.093182
clarity_in_num_order	0.831394943	1.202798
volume	0.002048253	488.221014

4 rows