Pichsinee Kalklod; ID #989451729

Minhua He; ID #989439989

Amina Tserendagva; ID #989446519

Diamond Report (Pears, 1.00 - 1.49ct)

I. Introduction

Fashion drives demand for luxury stones as society changes, and diamond became the most popular gemstone. Diamonds come in many shapes and sizes, it has Radiant, Emerald, Pear, Cushion, Round, and even customized shapes, out of all of those shapes Round and Pear are the main shapes that customers prefer. Since diamond is a growing industry of more than $90 Billion USD and is predicted to grow more in the long future, a lot of investors are interested in investing in the industry. There are already professional reports out there that educate and summarize the annual performance of the diamond industry, such as GIA, IGI, or Rapaport for pricing lists. For our report, we want to study the differences of the market prices for Natural diamonds versus Lab-grown diamonds and compare the prices between Natural diamonds versus the Rapaport report.

For a diamond's weight, bigger is not always better. The most important factors that decide the price of the diamond are the 4 Cs which are carat (weight), color, clarity and cut. On the wholesale level, diamond prices are first based on a diamond shape and size range (e.g. 1.00ct – 1.49ct). After that, it is broken down by color and clarity and then priced per carat. For our research, we chose to study the Pear-shaped diamonds with Ideal cut, the 4 Cs including carat weight of 1.00 to 1.49ct, color from D to K, clarity from SI2 to IF and finally ideal cut. Other qualities impact the price as well such as cut, fluorescence, types of imperfections, etc.

Finally, this project is collecting the data first and calculating the price divide by carat for natural diamonds and lab-grown diamonds.

II. Background

This project is about basic data analytics by using diamond price and its related 4Cs data which are the most significant factors that decide the diamond price including carat (weight) of the diamond, color, clarity, and cut. After this project, we comprehend step by step how to analyze specific data based on the real-world data.

This project is the combination of different tasks to find out the explore the diamonds' data to which we applied descriptive analytics, statistical inference, data visualization and linear regression to perform analysis and interpret diamond pricing data. Each task is approached differently and plays an important part in contributing to the result. First, we started with collecting raw data, or data acquisition. Raw data consists of Natural and Synthetic (lab made) diamonds, for which we chose to collect Pear-shaped diamond. Additionally, diamond color from D to K; diamond clarity from IF to S12; 1.00 to 1.49ct, Ideal cut array are considered as ranges for this project. We then filled those data in tables and created summary tables which combined all the collected information for making line and scatter charts as well as the heat maps for data visualization. Furthermore, we applied different methodologies for calculating either using Excel formula and normal computing in order to develop regression equation, covariance and interval estimate. And lastly, we designed the dashboard to navigate people through our project's findings.

III. Data Analysis:

Our group data criteria, we are interested in data collection of pear shaped diamonds with 1.00-1.49 carat and only in ideal cut. We collected data using several different suggested websites due to the limitation of data in each one, we were unable to collect all data from only one website. Since we have four members in the group, we divided data collection into four parts. One member collected data for color D, E, F, G for pear shaped diamonds with 1.00-1.49 carat and ideal cut for natural diamond, and another member collected data for color H, I, J, K and the

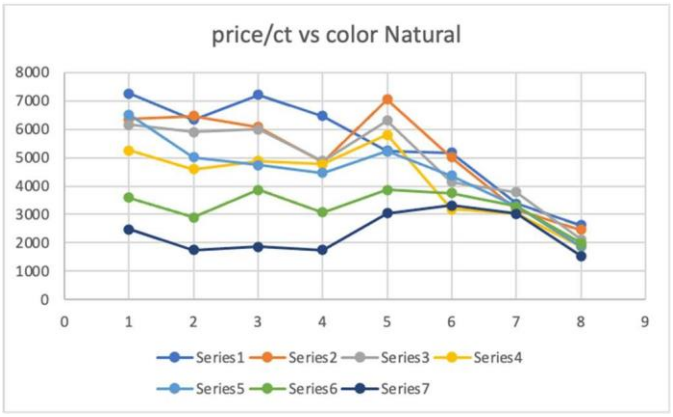
same for lab grown diamonds. By collecting three different prices of each clarity, we then calculated price/ct for each item. We collected data on April 15 to make sure that it is accurate in terms of times for more efficient and effective analysis. After completing the data collection, we created another table to find the average prices of each color and clarity, and in the result we have two tables containing average price/ct of all colors and clarity for both natural diamond and lab grown diamond (table 1, 2).

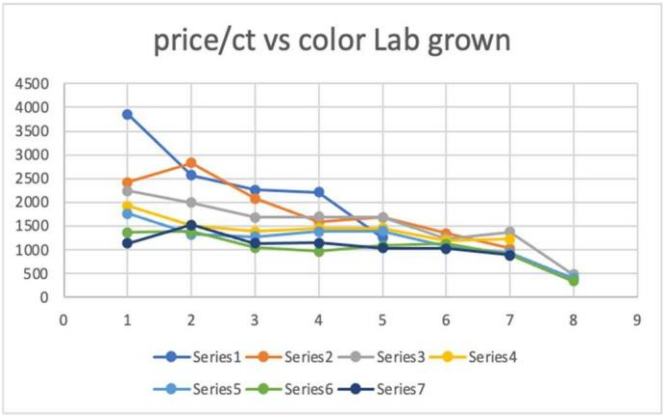
Table 1 Pear ideal cut Natural diamond price							
Pear	**IF**	**VVS1**	**VVS2**	**VS1**	**VS2**	**SI1**	**SI2**
D	7,254.44	6,356.68	6,162.42	5,252.24	6,514.80	3,596.20	2,481.95
E	6,343.52	6,461.83	5,893.07	4,596.82	5,013.09	2,892.13	1,746.61
F	7,200.06	6,074.92	6,003.04	4,881.52	4,740.78	3,869.17	1,862.87
G	6,468.55	4,832.38	4,882.98	4,765.14	4,459.41	3,067.47	1,754.45
H	5,243.39	7,043.01	6,305.00	5,802.99	5,226.37	3,856.67	3,040.00
I	5,168.84	5,003.79	4,141.38	3,183.77	4,358.59	3,750.67	3,306.67
J	3,373.12	3,164.03	3,789.34	3,040.59	3,275.73	3,276.00	3,025.89
K	2,626.78	2,456.52	2,112.92	1,866.81	1,881.69	1,985.87	1,539.63

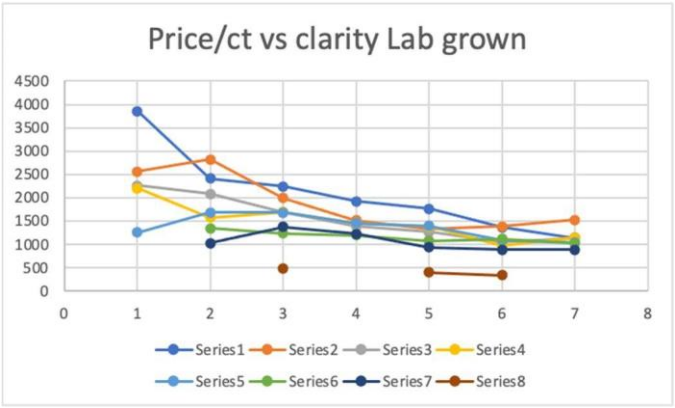
Table 2 Pear ideal cut Lab-grown diamond price							
Pear	**IF**	**VVS1**	**VVS2**	**VS1**	**VS2**	**SI1**	**SI2**
D	3,857.44	2,416.93	2,246.38	1,921.66	1,763.51	1,370.09	1,134.85
E	2,566.89	2,826.39	1,992.53	1,520.24	1,330.25	1,387.22	1,525.22
F	2,267.00	2,087.97	1,684.23	1,384.54	1,272.70	1,052.12	1,139.85
G	2,207.82	1,581.08	1,697.86	1,456.86	1,387.33	972.68	1,152.11
H	1,258.67	1,689.62	1,683.20	1,460.05	1,392.97	1,090.96	1,038.56
I		1,349.57	1,238.97	1,195.52	1,076.40	1,123.13	1,029.12
J		1,030.81	1,379.78	1,228.26	936.93	889.47	888.35
K			484.05		401.94	339.60	

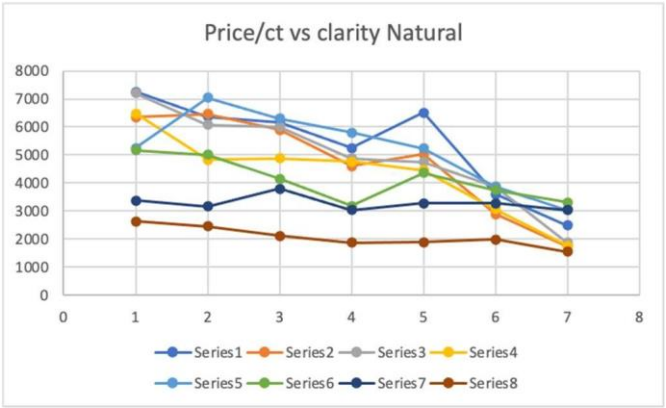
Please be aware that for Lab-grown diamond's prices in I, J, and K colors, there is not many data from the various websites that we have visited (Blue Nile, James Allen, Rare Carat, Brilliant Earth, etc.). This is because those colors have more yellow tone in the diamond, which leads to less demand and less profit. Laboratories can control the results of creating a diamond so they purposedly leave out some of the less-in-demand colors to focus more on the colors that consumers want.

Next, on regression model, we used Excel functions to create scatter charts to see the relationship between variables using colors and clarity as independent variables or X, and Price/ct as dependent variables or Y. Refer to the charts below.

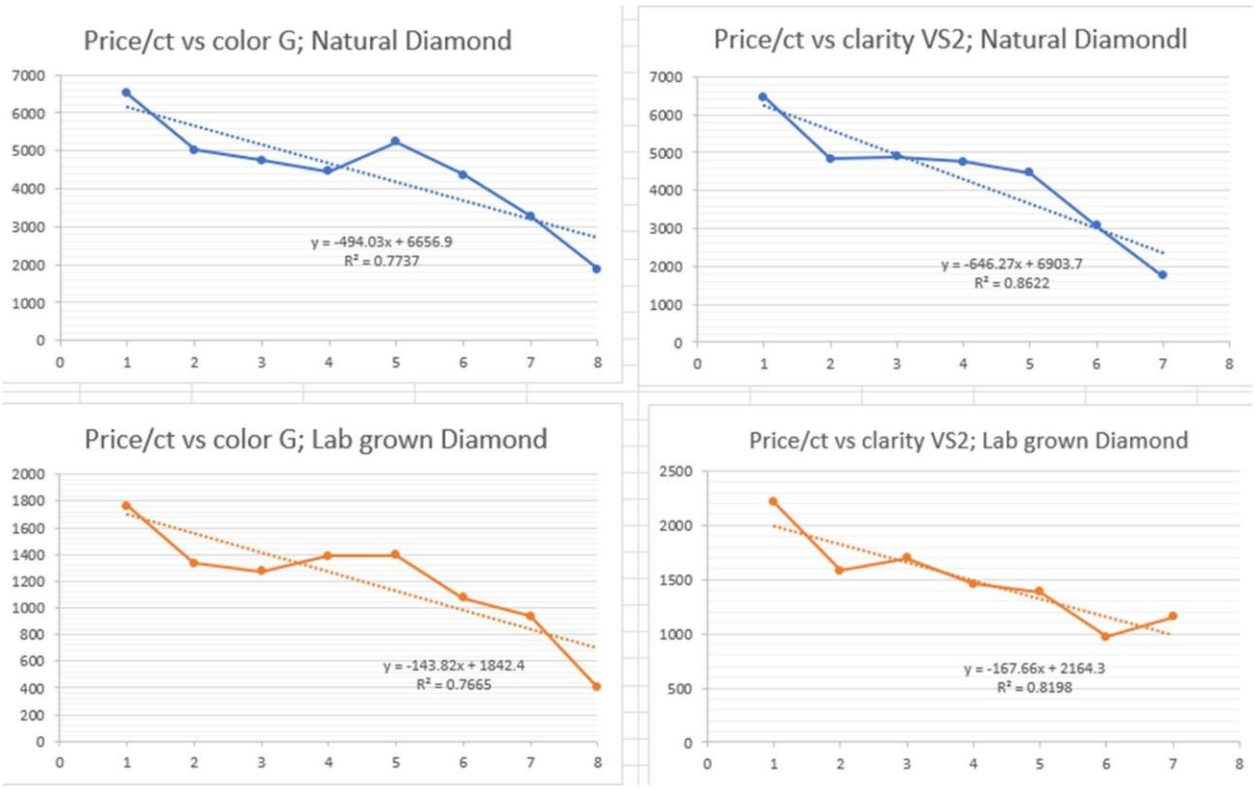








As seen from the charts, the color D and IF clarity has the higher prices, due to the rareness and demand in the market. While on the other hand, the color K and clarity SI2 has the lowest prices due to the unpopular demand. As seen from the charts, the prices of natural diamonds are in a higher range, almost double the price of lab grown diamonds. Also due to data limitations for lab grown diamonds, we were unable to collect all prices for diamonds in color K. Using these scatter charts to create regression models and find R squared as per details below.



After collecting data and creating the regression models and R^2, the equations for color G Natural diamond are Y = -71.297X + 4100.1 with R square = 0.0072 and Y = -50.564X + 4194.5 with R square = 0.005. The equations for color G Lab grown diamond are:

Y = 12.367X + 1251.8 with R square = 0.0022 and Y = -21.323X + 1169 with R square = 0.0114 as seen in the linear regression charts above. Now we will be able to analyze data mainly focusing on comparing the created tables with the Rapaport report since this is how the Rapaport report was created, along with analyzing the relationship between price/ct versus colors and clarity. There will be some inaccuracy regarding the time difference of when and where the data had been collected but we can summarize the usual trend between the two results. We compare our team's observed and collected data tables with Rapaport report by calculating the ratio in percentage and creating heat maps for both natural diamond and lab grown diamond to show the differences.

Heat map 1. Lab made diamond vs. Natural diamond							
Pear	**IF**	**VVS1**	**VVS2**	**VS1**	**VS2**	**SI1**	**SI2**
D	53.17%	38.02%	36.45%	36.59%	27.07%	38.10%	45.72%
E	40.46%	43.74%	33.81%	33.07%	26.54%	47.97%	87.32%
F	31.49%	34.37%	28.06%	28.36%	26.85%	27.19%	61.19%
G	34.13%	32.72%	34.77%	30.57%	31.11%	31.71%	65.67%
H	24.00%	23.99%	26.70%	25.16%	26.65%	28.29%	34.16%
I	0.00%	26.97%	29.92%	37.55%	24.70%	29.94%	31.12%
J	0.00%	32.58%	36.41%	40.40%	28.60%	27.15%	29.36%
K	0.00%	0.00%	22.91%	0.00%	21.36%	17.10%	0.00%

Heat map 2. Nature diamond vs. Rapaport diamond							
Pear	**IF**	**VVS1**	**VVS2**	**VS1**	**VS2**	**SI1**	**SI2**
D	58.04%	61.12%	65.56%	61.79%	86.86%	57.08%	47.73%
E	61.00%	68.74%	67.74%	57.46%	69.63%	48.20%	34.93%
F	76.60%	71.47%	75.04%	65.09%	68.71%	67.88%	38.81%
G	79.86%	62.76%	66.89%	68.07%	67.57%	56.81%	38.14%
H	75.99%	106.71%	100.08%	96.72%	91.69%	77.13%	70.70%
I	87.61%	89.35%	76.69%	61.23%	88.95%	85.24%	84.79%
J	70.27%	68.78%	86.12%	72.40%	81.89%	91.00%	94.56%
K	67.35%	66.39%	60.37%	56.57%	58.80%	68.48%	57.02%

Referring to ratio tables and heat maps above, we can describe that there are huge price differences mostly in natural diamonds. The highest ratio is diamond in color H and VVS clarity at 106.71%, followed by color H in VVS2 clarity at 100.08% and average ratio of this table is at 70.04% which is considered high. While for lab grown diamonds, the highest ratio is at 87.32% for diamond color E in SI2 clarity, followed by 65.67% for diamond color E in SI2 clarity and average ratio of this table is at 30.56%.

Another analytical aspect that we looked at was the sample covariance (Sxy). We collected data from Natural diamonds and on column X1 for Color, we calculated the average price/ct for every color (D to K) using table 1. The result we have shows the sample covariance is -2,274.70. This means that the two variables (Price/ct and Color) have an inverse relationship, the higher values of one associated with the lower values of the other. In this case, we can say that the price per

carat is less as the color becomes more yellow. Please refer to the image below for details on how we calculated the sample covariance:

Pear	X1	Y	*xi - xbar*	*yi - ybar*	*(xi - xbar)(yi - ybar)*	*(xi - xbar)^2*	*(yi - ybar)^2*
	Color	**Price/Ct**					
Natural Diamond	1	5,374.10	-3.5000	1,119.2011	-3,917.2039	12.2500	1,252,611.1614
	2	4,706.72	-2.5000	451.8199	-1,129.5498	6.2500	204,141.2545
	3	4,947.48	-1.5000	692.5749	-1,038.8624	2.2500	479,660.0419
	4	4,318.62	-0.5000	63.7216	-31.8608	0.2500	4,060.4426
	5	5,216.78	0.5000	961.8725	480.9362	0.2500	925,198.6870
	6	4,130.53	1.5000	-124.3732	-186.5598	2.2500	15,468.6998
	7	3,277.81	2.5000	-977.0896	-2,442.7240	6.2500	954,704.0745
	8	2,067.18	3.5000	-2,187.7273	-7,657.0454	12.2500	4,786,150.6045
Mean	4.5000	4,254.9027					
Std Dev	2.4495	1,109.8259					
sxy =	-2,274.6957	Sample Covariance					
sx =	2.4495	Sample Standard deviation of x					
sy =	1,109.8259	Sample Standard deviation of y					
rxy =	-0.8367	Sample correlation coefficient					
CORREL:	-0.8367						

Image 1. Sample covariance for Price/Ct vs Color (X1) of Natural Diamond

We also repeated the same steps for Natural diamonds but this time we will calculate Price/ct against Clarity (X2). The results we have is the same as sample covariance for Price/ct and Color (X1), an inverse relationship. We can say that the price per carat becomes smaller when the clarity has more imperfections. Image 2 below will show the details in calculations.

Pear	X2	Y	*xi - xbar*	*yi - ybar*	*(xi - xbar)(yi - ybar)*	*(xi - xbar)^2*	*(yi - ybar)^2*
	Clarity	**Price/Ct**					
Natural Diamond	1	5,459.84	-3.5000	1,204.9328	-4,217.2647	12.2500	1,451,862.9897
	2	5,174.14	-2.5000	919.2408	-2,298.1021	6.2500	845,003.7249
	3	4,911.27	-1.5000	656.3666	-984.5500	2.2500	430,817.1618
	4	4,173.74	-0.5000	-81.1672	40.5836	0.2500	6,588.1140
	5	4,433.81	0.5000	178.9034	89.4517	0.2500	32,006.4275
	6	3,286.77	1.5000	-968.1313	-1,452.1970	2.2500	937,278.3089
	7	2,344.76	2.5000	-1,910.1451	-4,775.3628	6.2500	3,648,654.3345
			-4.5000	-4,254.9027	19,147.0621	20.2500	18,104,196.8426
Mean	4.0000	4,254.9027					
Std Dev	2.1602	1,106.9636					
sxy =	-2,266.2402	Sample Covariance					
sx =	2.1602	Sample Standard deviation of x					
sy =	1,106.9636	Sample Standard deviation of y					
rxy =	-0.9477	Sample correlation coefficient					
CORREL:	-0.9477						

Image 2. Sample covariance for Price/Ct vs Clarity (X2) of Natural Diamond

IV. Conclusion

Our study of the diamond prices has helped to discover a lot of interesting information on diamond prices, including natural diamond and lab-grown diamond. From the data acquisition steps, we faced some challenges in finding diamond prices for lab-grown diamonds in these following categories: pear shaped, ideal cut, 1.00-1.49ct with color K. We believe this was due to the fact that even though synthetic diamond is indeed a growing industry worth billions of dollars, the lab diamonds are usually created to tailor to consumers' needs rather than mass production, so there are certain diamonds' weights, colors, or clarity that the lab will purposely avoid producing. Also, it is not profitable to create diamonds in H to K colors because they will have more yellow than colors from D to G, and so, prices would be cheaper. When clients pick out their diamonds, they are very attentive to the colors of the diamonds, most people would prefer a very colorless diamond with no to very little hint of yellow. So the lab-grown diamonds in colors H to K would not sell as well as the ones from D to G, hence, the production would be much less.

When we had finished gathering the data, our next step was to summarize the data into two tables, one for natural diamonds' prices and the other for synthetic diamonds' prices. After this was done, we went ahead and dive into our data analysis. Firstly, we utilized the heat map to do a simple comparison in prices for synthetic against natural diamonds. We found that the lab-grown diamonds' prices were nowhere near the earth-made diamonds' prices; most of the lab-grown diamonds' prices were around 20-35% of natural diamonds in the same categories. In fact, there were only four categories that show lab-grown diamonds were 50% the price of natural diamonds which are (1) color D with IF clarity, (2) color E with SI2 clarity, (3) color F with SI2 clarity, and (4) color G with SI2 clarity. And when we compare our natural diamonds with the Rapaport report, we found that most of our prices were less than Rapaport's prices. We could say that in comparison to the professional pricing report, we could find diamonds of the same categories for cheaper prices.

Moving onto the scatter and line charts, we can see that the regression line went downwards for both natural and lab-grown diamonds, this means that the prices tend to drop when we look for diamonds with more yellow, and less clarity. Next, we performed covariance calculations on (1) price/ct against the color (x1) and (2) price/ct against the clarity (x2). What we found from all 4 calculations shows negative sample covariance (Sxy): (1) -2,274.70 for natural diamond, -1,230.63 for lab-grown diamond, (2) -2,266.24 for natural diamond, -985.68 for lab-grown diamonds. This confirms the inverse relationship between the price/ct against the color and clarity of diamonds. Finally, to do the interval estimation, we gather data for 30 diamond prices in D color, VS1 clarity with ideal cut. Through standard statistical calculations, for natural diamond, we calculated the mean price of $8,966.21, the lower limit with 95% confidence level is $8,287.32 and the upper limit is $9,645.09. For synthetic diamonds, with the same calculations, the mean price is $2,737.73, lower limit is $2,513.25, and upper limit is $2,962.20. Through this, we can clearly see the big difference in pricing between natural and synthetic diamonds where lab made diamond's mean pricing is only 30.5% compared to earth grown diamonds.

Through this interesting study, we can conclude that diamond prices are always changing based on the market at the time. For natural diamonds, the price point, most of the time, is at least 50% higher than prices of synthetic diamonds in the same categories (cut, carat, clarity, color). If not for collection purpose or reselling purpose, we believe it is more advantage to buy synthetic diamonds for personal use as you can request the lab to customize the diamond to your personal preferences. The Rapaport report is a great source of pricing information for diamonds; however, it only serves as a reference list, not the exact reflection of prices as we have seen on the heat map 2. Overall, we want our report to serve as an informative study of the Pears diamond with ideal cut in the 1.00 - 1.49 ct range. Diamond industry is an ever-changing industry, but its pricing is still more stable than gold, however, we would like to ask

investors and readers to refer to professional reports and reliable websites in order to make final decisions on whether or not to invest in diamonds, or to buy between natural or lab diamonds.

V. Appendix – Please see attached Excel for all related data.