UNIVERSITY OF THE PACIFIC

MSBA 210 (Spring 2023) Final Project Assignment

This report assignment is to let students accomplish basic data analytics tasks using diamond price and the related 4Cs data. Students will use descriptive analytics, data visualization, statistical inference, and linear regression to perform analysis and interpretate real-world data. These tasks include data acquisition, data analysis and interpreting analytical reports.

Assignment Instructions

The most important factors that decide the price of the diamond are the 4Cs including carat (weight) of the diamond, color, clarity, and cut. In this assignment, you are required to analyze diamond's price and 4Cs through collecting raw data, completing several tables and charts, and using Rapaport report as reference for comparison. At the end of the semester, you are required to submit a project summary report.

Each MSBA 210 student needs to collaborate with other classmates to work on this project as a group. You will form project groups with no more than 4 group members. Each group will submit a project summary report not longer than 10 pages together with the appendix of the raw data you collected in excel format. In this project, each team will need to complete the following tasks:

1. **Data Acquisition**.

Collect natural and synthetic (lab made) diamond price (in $) of assigned carat range and deliver a serial of tables (sample table listed below) to help analyst to:

(1) refer to specific numerical values (such as diamond price).
(2) to make precise comparisons between different price/ct values of different 4Cs.

Attach the data you collected in excel, name the file in following format:

Group No., Cut(round/pears)-Carat Range.

Example: Group A-2, Round-0.9-0.99Ct

For each 4Cs, you will collect at least **3 different diamonds' price data**. For example, you need to find out the prices of at least 3 different diamonds which have 4Cs of idea cut, 0.9-0.99 ct, E color, and IF clarity to calculate the corresponding price/Ct. (shown in red rectangle in table Below). Each team collects the 4Cs and price data of diamond of your assigned weight range. For example, if you were working on 0.9-0.99 carat range, you can only collect 4Cs data and price of this weight range.

If there is no price data for a specific 4Cs setting, leave it blank. You need to provide your explanation for why you cannot find the related data in your report. Refer to Sample Raw Data table as follow:

Group A-5 Round					
0.9-0.99 Carat. E color, Idea cut, **(round,natural diamond , Bluenile,2/13)**					
Price ($)	Carat	Cut	Color	Clarity	Price/Ct
3,175	0.9	ideal	E	SI2	3,528
3,626	0.9	ideal	E	SI2	4,029
4,557	0.98	ideal	E	SI2	4,650
3,692	0.9	ideal	E	SI1	4,102
4,032	0.9	ideal	E	SI1	4,480
4,085	0.9	ideal	E	VS2	4,539
4,450	0.9	ideal	E	VS2	4,944
6,095	0.96	ideal	E	VS2	6,349
4,524	0.9	ideal	E	VS1	5,027
5,742	0.9	ideal	E	VS1	6,380
5,733	0.9	ideal	E	VVS2	6,370
6,056	0.9	ideal	E	VVS2	6,729
6,668	0.92	ideal	E	VVS2	7,248
11,633	0.93	ideal	E	VVS1	12,509
11,629	0.92	ideal	E	VVS1	12,640
12,819	0.99	ideal	E	VVS1	12,948
13,218	0.92	ideal	E	IF	14,367
13,281	0.99	ideal	E	IF	13,415
13,613	0.97	ideal	E	IF	14,034
14,129	0.95	ideal	E	IF	14,873

Sample raw data table developed for idea cut, 0.9-0.99 ct, E color, round, natural diamond with price/ct calculated for each sample.

2. Descriptive Analysis and Data Visualization

Develop tables to describe the cross-sectional data of natural diamond and lab made diamond.

(1) Each group needs to develop **2 tables** to describe how does price/ct (dependent variable, y) change with the change of clarity and color (independent variables, Xi) for idea cut, **natural diamond** (table 1 as an example below) and **Lab made diamond** (table 2).

Table 1. Round ideal cut Natural diamond Price ($/Ct)							
Round							
	IF	VVS1	VVS2	VS1	VS2	SI1	SI2
D							
E	14,172	12,699	6,782	5,703	5,277	4,291	4,069
F							
G							
H							
I							
J							
K							

(2) Compute the ratio of the following cases in percentage form and develop the **heat map** to compare the price/Ct for:

(a) Lab made diamond vs. natural diamond.

(b) Natural diamond vs. Rapaport data.

(3) Develop the **line/scatter charts** of

(a) Price/ct (y) vs. color (X1).
(b) Price/ct (y) vs. clarity (X2).

Use the following table to convert letter variables to numeric variables.

Letter	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**
Numeric	1	2	3	4	5	6	7	8
Letter	**IF**	**VVS1**	**VVS2**	**VS1**	**VS2**	**SI1**	**SI2**	
Numeric	1	2	3	4	5	6	7	

(4) Develop the **regression equation** using the scatter charts in (3), and $\mathbf{R^2}$.

(5) Calculate the **covariance (Sxy)** of (measure of association between two variables)

(a) Price/ct (y) vs. color (X1).

(b) Price/ct (y) vs. clarity (X2).

(6) Collect 30 Data from your assigned diamond (both natural and lab made) weight range with the other 3Cs set as: D color, VVS1, idea cut.

(a) Calculate the sample mean and the sample standard deviation.
(b) Develop the **interval estimate** of the price/ct with 95% confidence level.

0.9-0.99 Carat. E color, Idea cut, (round,natural diamond , Bluenile,3/27)						
Price ($)	Weight (Ct)	Price/Ct				
6,449	0.9	7,166			*Column1*	
6,807	0.9	7,563				
6,950	0.9	7,722			Mean	10213.03433
7,058	0.9	7,842			Standard Error	377.2729417
7,370	0.91	8,099			Median	9763.186499
7,335	0.96	7,641			Mode	#N/A
7,878	0.91	8,657			Standard Deviation	2066.409005
8,234	0.91	9,048			Sample Variance	4270046.177
8,282	0.95	8,718			Kurtosis	-0.892614379
8,299	0.92	9,021			Skewness	0.463387759
8,409	0.95	8,852			Range	7046.532357
8,746	0.95	9,206			Minimum	7165.555556
8,987	0.92	9,768			Maximum	14212.08791
9,136	0.96	9,517			Sum	306391.0299
9,164	0.98	9,351			Count	30
9,270	0.95	9,758			Confidence Level(95.0%)	771.6098036
9,585	0.93	10,306				
9,533	0.97	9,828			Mean	10,213
9,777	0.95	10,292			lower limit	9,441
10,523	0.91	11,564			upper limit	10,985
10,641	0.92	11,566				
10,670	0.98	10,888				
10,930	0.96	11,385				
11,711	0.95	12,327				
11,908	0.97	12,276				
12,135	0.9	13,483				
12,367	0.95	13,018				
12,500	0.92	13,587				
12,933	0.91	14,212				
13,180	0.96	13,729				
	Sample Mean	10,213				

(7) Design a dashboard to summarize your findings.

3. Final Report

In Final report, you must:

(1) Describe your research problem, your dependent variable (price/ct) and independent variables.

(2) Your report must contain at least: your data tables, charts to describe main data, one dashboard to summarize the project, the trendlines which describe your regression model together with the R^2.

(3) Describe the process of your data collection, data processing and your regression model.

(4) Provide your data analysis results.

(5) Present your report in class. Your group presentation time is 10 minutes or less.

(6) Your Final Written report must contain the following sections:

I. **Introduction** – Start with a general description of your research problem area, then narrow down to a clear statement of the purpose of your assigned project and list your specific research question(s). Try to write for a general audience of business analysts or managers.
II. **Background** – Provide background information in this section. You should clearly describe what do you cover in this project. Try to help readers understand what you try to find out/achieve.
III. **Data Analysis**: Explain what data you have collected, how it was collected and calculated/processed. For the regression model you proposed, the **independent** and **dependent** variables should be clearly stated. A description of how the collected data was analyzed should be provided.
IV. **Conclusion** – Clearly summarize your report. Your conclusion should "stand on its own." By this I mean that it should not be necessary for readers to have to go back and read the entire report in order to understand your conclusion
V. **Appendix –** Attach all your raw data in excel in Appendix.

NOTE: Plagiarism may result in a grade of zero on this assignment and a failing grade in the course.

Website Example:

www.bluenile.com/
www.brilliantearth.com/
www.jamesallen.com/
www.withclarity.com/
.....

Grading:

Your Final Written Research Proposal will be graded using the rubric below.

Component	Requirements
Teamwork (15 Points)	The team worked well together to achieve project goals. Each member contributed in a valuable way to the project. All supporting materials and sources indicated a high level of mutual respect and collaboration.
Final Report I-IV (60 Points)	All requirements and objectives are identified and completed. The presentation and report offered valuable information on the topic and demonstrated knowledge and evidence of extensive research effort and a depth of thinking about the topic. The data analysis is accurate, and conclusion is persuasive.
Raw Data & supporting Material (15 Points)	All relevant data and information were obtained, and information sources were valid and accurate. Analysis and applications were well supported by the supporting materials. Identify data sources.
Composition (10 Points)	The presentation and report were well organized and clearly written. The underlying logic was clearly articulated and easy to follow. The research shows proper usage of vocabulary and grammar, and free from errors. Diagrams or analyses enhanced and clarified presentation of ideas.