**Michael Pacheco**

**DSC 530**

**Statistical/Hypothetical Question**

The statistical/hypothetical question addressed in this analysis was: "Does the carat weight of a diamond have a significant impact on its price?" The hypothesis suggested that the carat weight would influence the price of diamonds.

**The outcome of EDA:**

During the EDA, we analyzed the "Price Prediction of Diamond" dataset and performed various statistical analyses. We examined the distribution of variables, identified outliers, calculated descriptive statistics, and explored relationships between variables. We also conducted a linear regression analysis to assess the impact of the carat weight on diamond prices.

The analysis revealed a positive correlation between the carat weight and the price of diamonds, indicating that as the carat weight increased, so did the price. The regression analysis confirmed the significance of carat weight as a predictor of diamond prices.

**Missed Aspects and Additional Variables:**

While the analysis focused on the carat weight, other variables in the dataset, such as cut, clarity, and color, could have provided valuable insights. Including these variables in the analysis would have allowed us to examine their individual and combined effects on diamond prices. By incorporating multiple explanatory variables, we could have obtained a more comprehensive understanding of the factors influencing diamond prices.

**Incorrect Assumptions:**

One potential incorrect assumption made during the analysis could be the linearity assumption in the linear regression model. Linear regression assumes a linear relationship between the independent and dependent variables. However, it is possible that the relationship between the carat weight and price may not be strictly linear. Using more advanced regression techniques, such as polynomial regression or nonlinear regression, could have accounted for potential nonlinear relationships and provided a better fit for the data.

**Challenges and Areas of Incomplete Understanding:**

During the analysis, challenges were encountered in interpreting certain statistical concepts and techniques. Understanding the nuances of regression analysis, including assumptions and model evaluation metrics, required careful consideration and further research. Additionally, exploring and selecting the most appropriate variables for analysis required domain knowledge and expertise in the diamond industry.

**Conclusion**

In conclusion, the EDA of the "Price Prediction of Diamond" dataset focused on examining the impact of carat weight on diamond prices. While the analysis provided valuable insights, there were missed opportunities to explore additional variables and consider potential nonlinear relationships. Further analysis incorporating variables like cut, clarity, and color, and utilizing advanced regression techniques could enhance the understanding of the factors influencing diamond prices. The challenges faced during the analysis emphasized the need for continuous learning and domain expertise to perform thorough and meaningful data analysis.