**Level 1 (Regular Data Science Questions):**

1. Calculate the average Cost per Carat across all gems.

446.72

1. Find the median Refractive Index for gems.

1.54

1. What is the range (difference between maximum and minimum) of the Average on the Vickers Scale across all gems?

11,010

**Level 2 (Multiple Step Hard Data Science Questions):**

1. For each Color, calculate the average Carat Weight. Which Color corresponds to the highest average Carat Weight?

Brown/Black with average Carat Weight of 20.0

1. Group gems by 'Year of Discovery' and calculate the median Cost per Carat for each group. Which group has the highest median cost?

400 BC with Median Cost per Carat of 13,000

1. Identify the gem with the highest Average on the Mohs Scale. What is the name and average Mohs Scale of this gem?

Diamond, 10.0

1. Calculate the average Cost per Carat for gems with a Refractive Index greater than 2.0. How does it compare to the average cost of gems with a Refractive Index less than or equal to 2.0?

Average Cost (Refractive Index > 2.0): 6502.5, Average Cost (Refractive Index ≤ 2.0): 364.89

1. For each distinct Carat Weight, calculate the total number of gems. Which Carat Weight has the highest count?

Carat Weight: 20.0 with 84 total gems

1. Divide gems into two groups based on the median 'Average on the Mohs Scale' (above or below). Compare the average 'Cost per Carat' between these two groups. Which group has a higher average cost?

Average Cost (Above Median Mohs Scale): 911.93, Average Cost (Below Median Mohs Scale): 161.59

**Level 3 (Multistep Data Analysis Aspects of data science/ML):**

1. Using a clustering algorithm (e.g., K-Means), cluster the gems based on Cost per Carat and Carat Weight. How many clusters provide the most meaningful grouping?

4 clusters

1. Predict the 'Color' of a gem using a decision tree classifier based on Cost per Carat, Refractive Index, and Average on the Vickers Scale. What is the accuracy of this model?

20%

1. Identify outliers in the dataset with respect to Cost per Carat and Average on the Mohs Scale using an Isolation Forest algorithm. How many outliers are detected?

24 Outliers

1. Implement Principal Component Analysis (PCA) on the numerical columns (Cost per Carat, Average on the Mohs Scale, Average on the Vickers Scale, Carat Weight, Refractive Index) and determine how many components explain 90% of the variance.

4 physical components

1. Using the features Cost per Carat, Carat Weight, Refractive Index, train a Support Vector Machine (SVM) classifier to predict whether a gem's Average on the Mohs Scale is above or below the median. What is the precision of this model?

0.0

1. Using the features 'Cost per Carat', 'Average on the Mohs Scale', 'Average on the Vickers Scale', 'Carat Weight', and 'Refractive Index', train a k-Nearest Neighbors (k-NN) classifier to predict the color of the gemstones. Divide the dataset into a training set (80%) and a test set (20%). What is the accuracy of the model on the test set?

26.67%