**Level 1 Questions (Basic Data Science Questions)**

1. Calculate the range of the 'Price' column.

550,000

1. Calculate the standard deviation of the 'Area' column.

465.24

1. Find the mean and median of the 'Rating' column.

Mean: 8.23, Median: 8.2

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

1. For each 'Property Type', calculate the skewness and kurtosis of the 'Price'. Which 'Property Type' has the most positively skewed Prices?

Single Family Home: Skewness = 1.15, Kurtosis = -0.30, Condo: Skewness = -1.10, Kurtosis = 0.14, Townhouse: Skewness = -0.72, Kurtosis = -1.24. The 'Property Type' with the most positively skewed Prices is the Single Family Home.

1. Calculate a “Desirability Score” for each property using a weighted combination of 'Rating' (40%), 'Bathrooms' (30%), and 'Bedrooms' (10%). For the top 5 properties by this score, sum up their Prices. What is this total?

$4,850,000

1. For each 'Location', calculate the correlation between 'Year Built' and 'Price'. Which 'Location' shows the weakest correlation?

City Center

1. Divide the dataset into three categories based on 'Area' (Small, Medium, Large). For each category, compute the average 'Distance to Nearest School'. Which category has the lowest average?

Small

1. Perform a rolling 3-property average on 'Garage Spaces' and identify the ID where this rolling average peaked. For that ID, what was the corresponding 'Property Type'?

Townhouse

1. Identify properties that have 'Price' within 10% of the median 'Price'. For these properties, what’s the average 'Area' and median 'Distance to Nearest School'?

Average ‘Area’: 1442.35; Median ‘Distance to Nearest School’: 0.3 miles

**Level 3 Questions (Multistep Data Analysis Aspects of Data Science/ML)**

1. By treating properties with a very high 'Price' as anomalies, can you use algorithms like Isolation Forest to detect these outlier properties? What is the most common 'Location' among the anomalies?

Suburban

1. The 'Year Built' column follows a pattern. Figure out that pattern and predict the next 5 values for the 'Year Built' column. Output the pattern found as well as the 5th generated value.

Pattern: Decrease 6 years per entry; 5th generated value: 1979

1. For each location type, what is the average 'Distance to Nearest School' and how does it relate to the average 'Price'?

It appears that properties in rural areas tend to have the longest distance to the nearest school and also the highest average prices. Properties in the city center have the shortest distance to schools and moderately priced properties, while suburban areas are in the middle in terms of both distance to schools and average property prices.

1. Can you use clustering algorithms like K-Means to categorize properties into distinct groups based on Price, Area, Bedrooms, Bathrooms, Garage Spaces, Year Built, and Distance to Nearest School? How many clusters are optimal, and what is the average 'Rating' for properties in the largest cluster?

Optimal Number of Clusters: 3; Largest Cluster: Cluster 0; Average ‘Rating’ for properties in that cluster: 7.94

1. Assume the data is sorted chronologically by ID, implement a time series forecasting model to predict the 'Price' of properties for the next 12 months based on their historical data. Assume that the dataset is sorted in chronological order. Use an ARIMA (AutoRegressive Integrated Moving Average) model for the forecast. What is the predicted price for the property listed in the 12th month from the last record?

$696490.02

1. Using the features 'Area', 'Bedrooms', 'Bathrooms', 'Garage Spaces', 'Year Built', and 'Distance to Nearest School', train a logistic regression model to classify properties as either 'City Center' or 'Suburban' based on their 'Location'. What is the accuracy and precision of the model?

98.25% and a precision of 100%