**Group 1: Regular Data Science Questions**

1. Calculate the average Sodium content for foods with a 'Good' Typical Nutritional Value.

44.6 mg

1. What is the correlation coefficient between Calories and Sugar content?

0.115

1. Find the median Protein content for foods that have a 'Negative Health Stigma'.

20.0g

**Group 2: Multiple Step Hard Data Science Questions**

1. For each 'Typical Nutritional Value' category, calculate the range of Sugar content. Which category has the highest range?

Good

1. Create a 'Caloric Density' score (Calories per gram of food). Rank the foods based on this score and identify the top 3 foods.

Coconut oil, Avocado oil, Canola oil

1. Compare the average Protein content between foods with and without a 'Negative Health Stigma'. Which category has a higher average?

Negative Health Stigma

1. Identify foods that are outliers in terms of Caloric content using the IQR method.

Olive oil, Avocado oil, Canola oil, Flaxseed oil, Sesame oil, Peanut oil, Almond oil

1. Perform a linear regression analysis between Sodium and Sugar content. What is the R-squared value of the model?

0.00445

1. Cluster the foods into 3 groups based on their nutritional content (Calories, Sodium, Sugar, Protein). What are the central values of each cluster?

Cluster 0: Calories 678.17, Sodium 64.23 mg, Sugar 7.36 g, Protein 11.99 g; Cluster 1: Calories 63.55, Sodium 41.83 mg, Sugar 3.01 g, Protein 5.67 g; Cluster 2: Calories 339.79, Sodium 38.45 mg, Sugar 20.06 g, Protein 12.28 g

**Group 3: Multistep Data Analysis and Machine Learning Questions**

1. Predict whether a food has a 'Negative Health Stigma' using a logistic regression model based on its nutritional content. What is the accuracy of the model?

95.56%

1. Build a decision tree classifier to predict the 'Typical Nutritional Value' based on other features. What is the precision of the model?

95.56%

1. Use principal component analysis (PCA) to reduce the dimensionality of the nutritional content data. How much variance is explained by the first two principal components?

First component: 38.48%, Second component: 27.89%

1. Develop a neural network to classify foods based on their 'Calories', 'Sodium', 'Sugar', and 'Protein' levels. Report the F1-score of the model.

93.38%

1. Implement a random forest regressor to predict the 'Sugar (g)' based on other nutritional values. What is the mean squared error of the model?

0.0337

1. Use a Gradient Boosting Machine (GBM) to predict the 'Protein (g)' based on other nutritional values (Calories, Sodium, Sugar). What is the model's Root Mean Squared Error (RMSE)?

0.25153293769775026