**Summary Report: What is Good food Project**

**Background**

The project investigated the nutritional composition of foods to inform a debate on diet and nutrition. Data from the USDA's Food Data Central database was analyzed to understand the impact of various nutrients on calorie content.

**Data Preprocessing**

Nutritional values were cleaned, converted to numeric types, and missing values were imputed with the mean. The dataset included a wide range of food items with detailed nutritional information.

**Analysis Findings**

* Highest Vitamin C Content: Acerola (West Indian Cherry) was identified as the fruit with the highest Vitamin C content. Other significant sources include Spices and Herbs, Beverages, Baby Foods, and Vegetables.
* Calories vs. Water Content: An inverse relationship was observed between calorie content and water content in food items, suggesting that foods with higher water content are lower in calories.

**Modelling**

Three models were developed to predict the calorie content of food items based on their nutritional composition:

* Linear Regression Model: The first model used protein, carbohydrate, and total fat as features and achieved a certain level of accuracy, with predictions made for the calorie contribution of each macronutrient.
* Enhanced Linear Regression Model: The second model included additional features such as alcohol, fiber, and water content. This model provided a more comprehensive understanding of the factors contributing to the caloric content of foods.
* Polynomial Regression Model: The third model applied a polynomial transformation to capture non-linear relationships between macronutrients and calories. This model aimed to improve the fit and accuracy of the predictions.

Each model's performance was evaluated using the root mean squared error (RMSE), and predictions were made for the calorie contribution of 1 gram of protein, carbohydrate, and fat.

**Conclusion**

The analysis and modelling efforts provided data-driven insights into the nutritional composition of foods and their caloric impact. The findings support the importance of a balanced diet and highlight the role of water content in reducing calorie density. The models developed offer valuable tools for predicting calorie content from nutritional information.