Assignment 2: Food Dataset Analysis

1. Objective

The objective is to prepare the food nutrition dataset for machine learning and apply a Linear Regression model to predict the calorie content of food items based on their macronutrients (protein, fat, carbohydrates) and category.

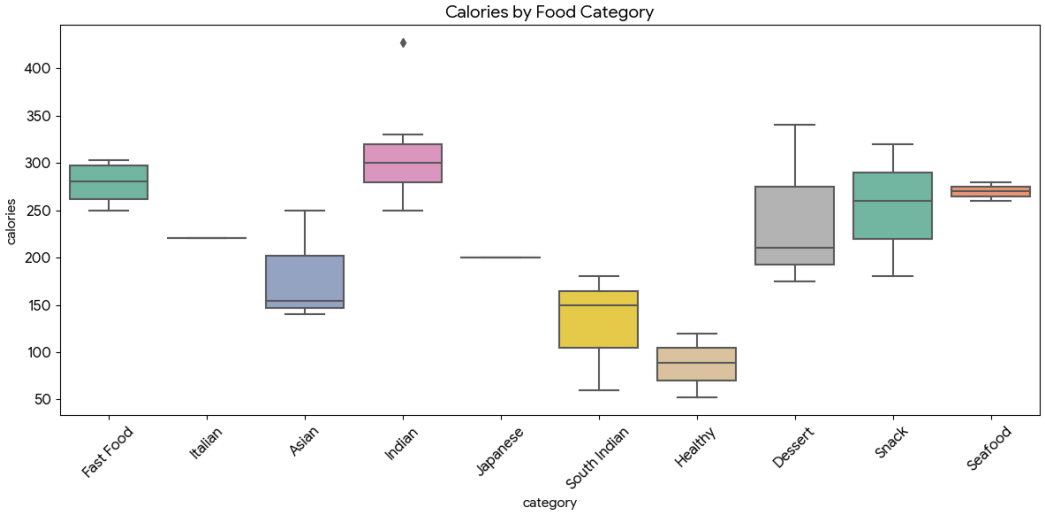
2. Data Preprocessing & Feature Engineering

- Cleaning: Checked for missing values (none found).
- Feature Selection: Dropped 'food_id' and 'food_name'.
- Feature Engineering: Created 'total_macros_g' summing protein, fat, and carbs.
- Encoding: One-Hot Encoded the 'category' column.

3. Exploratory Data Analysis

Distribution of Calories:

Calories by Category:


Calories by Food Category
calories
400
350
300
250
200
150
100
50
Fast Food
Italian
Asian
Indian
Japanese
South Indian
Healthy
Dessert
Snack
Seafood
category

4. Model Performance

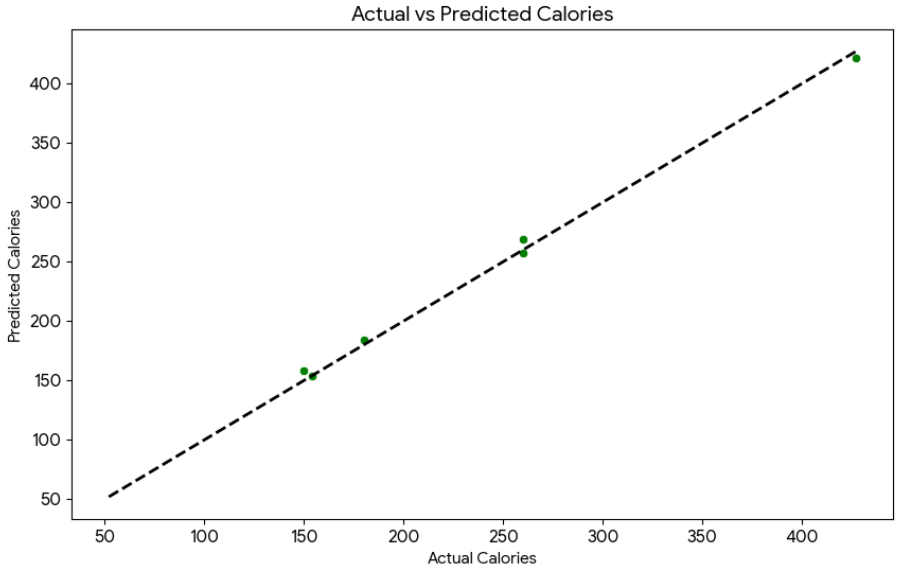
Mean Absolute Error (MAE): 4.9324

Mean Squared Error (MSE): 33.9015

Root Mean Squared Error (RMSE): 5.8225

R2 Score: 0.9963

Actual vs Predicted Calories:



5. Conclusion

The Linear Regression model performed exceptionally well with a very high R2 score. This is expected as calories are directly mathematically derived from macronutrients (4-9-4 rule). The model successfully learned this relationship.