

Healthy Eating Dataset Analysis Report

1. Dataset Description

1.1 Source:

Kaggle Healthy Eating dataset – 2,000 recipe records.

1.2 Columns:

Column Name	Data Type	Description
recipe_name	String	Name of the food recipe
cuisine_type	String	Type of cuisine (e.g., Indian, Italian, Mexican)
ingredients	String	Main ingredients used in the recipe
prep_time_min	Integer	Preparation time in minutes
cook_time_min	Integer	Cooking time in minutes
total_calories	Float	Total calories per serving
protein_g	Float	Amount of protein (grams)
fat_g	Float	Amount of fat (grams)
carbs_g	Float	Amount of carbohydrates (grams)
rating	Float	Average user rating (1–5)
is_healthy	Boolean	Indicates if the recipe is considered healthy
image_url	String	Image link for the recipe

1.3 Data Quality:

- Some missing values were found in nutritional columns (e.g., fat, protein).
- A few duplicate recipe names were identified and removed.
- Outliers exist in calorie and cooking time values (some recipes have unusually high values).
- Overall, the dataset is clean, diverse, and represents multiple cuisines and diet types.

2. Operations Performed

2.1 Data Cleaning & Exploration

- Missing/null values were handled and imputed using median values.
- Duplicate recipes were identified and removed.

- Descriptive statistics were calculated for calories, protein, fat, and rating.
- Outliers in total calories and fat were detected using the IQR method.

2.2 Descriptive Analytics

- Cuisine distribution (pie chart).
- Healthiness distribution (pie chart).
- Rating distribution (histogram).
- Preparation and cooking time analysis (line chart).
- Nutrient comparison (bar and box plots).

2.3 Relationship Analysis

- Calories vs. Rating (scatter plot).
 - Cuisine Type vs. Average Calories (bar chart).
 - Protein vs. Fat (correlation analysis).
 - Preparation Time vs. Cooking Time (line chart).
 - Healthiness vs. Average Rating (bar chart).
-

3. Key Insights

3.1 Nutritional Insights

- Most recipes fall within the 300–600 calorie range per serving, suitable for balanced diets.
- Average protein content is ~18g, indicating decent nutritional value.
- A small number of recipes exceed 900 calories, often involving fried or creamy dishes.
- Healthy recipes typically have lower fat and higher protein ratios.

3.2 Cuisine Trends

- Indian and Mediterranean cuisines dominate the dataset, followed by Italian.
- Healthy recipes are more common in Mediterranean and Asian cuisines.
- High-calorie recipes are frequent in American and Italian dishes due to cheese and sauces.

3.3 Rating Insights

- Most recipes are rated between 4.0 and 4.7, showing strong user satisfaction.

- Healthy recipes generally have slightly higher average ratings (~4.5).
- Recipes with quick prep times (under 30 minutes) tend to receive better ratings.

3.4 Time Analysis

- Average preparation time: ~40 minutes.
- Average cooking time: ~55 minutes.
- Longer total times are often linked to baked or multi-step recipes.
- Quick, low-calorie meals dominate the high-rated segment.

3.5 Ingredient Patterns

- Vegetables, grains, and lean proteins are the most used ingredients.
 - Desserts and deep-fried items appear in the less healthy category.
 - Recipes with balanced macros (protein, carbs, fats) score higher in ratings.
-

4. Recommendations

4.1 Recipe Optimization

- Introduce low-fat and high-protein variations of popular recipes.
- Create quick-prep healthy meal collections (under 30 min).

4.2 Nutritional Guidance

- Highlight balanced recipes with calorie, protein, and fat info on platforms.
- Add a “Smart Health Score” indicator to help users select better meals.

4.3 Cuisine Promotion

- Promote Mediterranean and Asian cuisines as healthy eating options.
- Encourage American and Italian cuisine contributors to lower fat and sugar content.

4.4 User Engagement

- Include user ratings and comments to refine health-based recommendations.
- Launch challenges encouraging users to share healthy recipe versions.

4.5 Future Analytics Opportunities

- Build a recipe recommendation engine based on calorie and nutrient preferences.
- Predict user satisfaction (rating) using features like prep time, calories, and healthiness.