**Data Cleaning of the Epicurious Dataset**

**Objective:**  
The goal was to clean the dataset to ensure accuracy and reliability for analysis by removing outliers, handling missing data, and making appropriate assumptions.

**Data Cleaning Steps:**

1. **Outlier Removal:**  
   Outliers can lead to distorted analysis, so the following thresholds were used to remove extreme values:
   * **Rating values** were assumed to range from 0 to 5. Any rating outside this range was removed as it was considered invalid.
   * **Calories:** Dishes with calorie values above 10,000 were identified as outliers and removed, as these values are implausibly high.
   * **Proteins, Fat, Sodium:** Any dish with protein, fat, or sodium levels above 5,000 was considered an outlier, as such values are unrealistic for standard recipes.
2. **Assumptions Made:**
   * **Rating Scale:** Ratings were assumed to be on a 0–5 scale.
   * **Nutritional Limits:** Upper bounds for calories, proteins, fats, and sodium were set to remove extreme outliers and maintain a realistic range for nutritional values.
3. **Handling Null Values:**  
   The dataset had missing values for some attributes. All records containing null values were removed to ensure data completeness and avoid skewing the analysis. This led to a clean and structured dataset, free of incomplete records.

**Data Analysis and Insights from the Epicurious Dataset**

**Objective:**  
The cleaned dataset was analyzed to extract meaningful insights regarding recipe ratings and nutritional content, and to explore trends based on dish categories (breakfast, lunch, dinner) and dietary preferences (vegetarian vs. non-vegetarian).

**Data Analysis Findings:**

1. **Relationship Between Rating and Nutrients:**  
   There was little to no correlation between a recipe’s rating and its nutritional components (calories, protein, fat, and sodium). This suggests that user ratings are more influenced by factors like taste, ease of preparation, or visual appeal rather than the healthiness or nutritional content of a recipe.
2. **Meal Categorization (Breakfast, Lunch, Dinner):**  
   Recipes were categorized into three meal types to study their nutritional properties. A few observations:
   * **Dinner** dishes tend to have slightly higher calorie content compared to breakfast and lunch dishes, possibly due to larger portion sizes and more complex ingredients.
3. **Vegetarian vs. Non-Vegetarian Nutritional Comparison:**
   * **Calories:** The average calorie content of vegetarian and non-vegetarian dishes is quite similar. This indicates that calorie-conscious individuals can opt for either type of dish without significant differences in energy intake.
   * **Protein and Sodium:** Non-vegetarian dishes tend to have higher levels of protein and sodium, which could be important for users monitoring their sodium intake or seeking high-protein meals.

**Additional Insights:**

* **Sodium Content in Non-Vegetarian Dishes:**  
  There is a tendency for non-vegetarian dishes to have higher sodium content, which might be a concern for individuals on low-sodium diets.
* **Caloric Trends by Meal Type:**  
  Dinner recipes generally have a higher caloric value than breakfast and lunch, suggesting that users may prefer more substantial meals later in the day.

**Recommendations:**

1. **User Feedback Incorporation:**  
   In future analyses, incorporating more qualitative data like user comments or feedback about recipe difficulty could provide better insights into the factors influencing higher ratings.
2. **Health-Focused Recipe Suggestions:**  
   Since nutritional values don’t significantly affect ratings, the platform could implement a recommendation system for healthier versions of highly-rated dishes, focusing on reducing sodium or fat content while maintaining taste.
3. **Detailed Nutritional Comparisons:**  
   Deeper analysis of other nutritional metrics (fiber, sugar, vitamins) between vegetarian and non-vegetarian dishes can offer more insights for health-conscious users.
4. **Customized Recommendations:**  
   A recommendation engine that suggests recipes based on dietary preferences (e.g., low-calorie, low-sodium, high-protein) could be beneficial for users with specific nutritional goals.