**Documentation for Epicurious Recipe Analysis**

**Objective:**  
The purpose of this notebook is to analyze a dataset of recipes and provide insights based on their nutritional content. This analysis helps users identify recipes that are low in calories, fat, and sodium, making them suitable for a balanced diet.

**1. Data Import and Setup**

The notebook starts by loading a dataset into a pandas and NumPy Data Frame and this data set has contains total 20052 records and 680 columns. From the Epicurious dataset I have taken the following columns for my Analysis.

Columns and Columns details:

1. title: The name of the recipe (e.g., "Lentil, Apple, and Turkey Wrap").

2. rating: The rating assigned to the recipe, represented as an integer.

3. calories: The number of calories in the dish, stored as a float.

4. protein: The protein content (in grams), stored as a float.

5. fat: The fat content (in grams), stored as a float.

6. sodium: The sodium content (in milligrams), stored as a float.

7. 22\_minute\_meals: A binary integer indicating if the recipe can be prepared in 22 minutes or less (0 = no, 1 = yes).

8. 3\_ingredient\_recipes: A binary integer indicating if the recipe uses only 3 ingredients (0 = no, 1 = yes).

9. dessert: A binary integer indicating if the recipe is a dessert (0 = no, 1 = yes).

**Data Cleaning:**

***1.Handling missing values:***

In my dataset some 4 columns has missing values, The missing or null values can significantly affect my analysis. 1.Title has 0 null values, 2.Rating has 0 null values, 3.Calories has 4117 null values, 4.Protein has 4162 null values, 5.Fat contains 4183 null values, 6.Sodium contains 4119 null values, 7. 22\_minute\_meals contains 0 null values, 8. 3\_ingredient\_recipes contains 0 null values, 9.Dessert contains 0 null values.

I did fill this all null values with mean of the column values.

mean of the calories : 6322.958016943834

mean of the proteins : 100.16079295154185

mean of the fat : 346.8774970067427

mean of the sodium : 6225.974894872277

***2.Handling duplicate values:***

Duplicated entries in the dataset can distort analysis, so duplicates have been removed to ensure each recipe is unique.

data\_c = data\_1.drop\_duplicates()

***3.Removing Outliers:***

Outliers are extreme values that can skew analysis results. The dataset have been cleaned by identifying and removing outliers based on certain nutritional metrics like calories, fat, or sodium.

data\_c = data\_c[(data\_c['calories'] < data\_c.calories.quantile(0.95)) & (data\_c['calories'] >0)]

**Exploratory Data Analysis (EDA) for Epicurious Recipe Analysis:**  
The EDA in this notebook is aimed at gaining insights into the recipe dataset, understanding its distribution, identifying relationships between nutritional factors (such as calories, fat, sodium), and preparing the data for further analysis.

I**nitial Overview:**

* **Shape of the Dataset**:
  + Provides the number of rows and columns.

data\_c.shape

**Summary Statistics**:

* Descriptive statistics like mean, median, standard deviation for each numerical feature.

data\_c.calories.describe()

data\_c.dessert.describe()

**Data distribution:**

Plotted some plots for every column to know the distribution of the Data(ex. Pie, Histogram, Bar chart, Violin plot, Scatter plot, etc.)

**Data Analysis:**

Based on the exploratory analysis, provide at least three key insights. Use these insights to answer potential business questions, such as:

▪ What are the most common ingredients in highly rated recipes?

▪ Are there correlations between preparation time and recipe ratings?

▪ How can the data help improve the user experience for a recipe platform?

* **What are the most common ingredients in highly rated recipes?**

The most common ingredients with highly rated recipes data is:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | title | rating | calories | protein | fat | sodium | 22\_minute  \_meals | 3\_ingredient  \_recipes | dessert |
| 6724 | 3-Ingredient Peanut Butter Cookies | 5 | 100.0 | 3.0 | 6.0 | 5.0 | 0 | 1 | 1 |
| 7068 | 3-Ingredient Steak With Crispy Parmesan Potatoes | 4 | 831.0 | 46.0 | 54.0 | 426.0 | 0 | 1 | 0 |
| 7779 | 3-Ingredient Sweet Potato Casserole With Maple... | 5 | 598.0 | 7.0 | 21.0 | 655.0 | 0 | 1 | 0 |
| 8547 | 3-Ingredient Thanksgiving Turkey With Orange a... | 4 | 974.0 | 122.0 | 48.0 | 3152.0 | 0 | 1 | 0 |
| 10014 | 3-Ingredient Seeded Crackers | 4 | 33.0 | 1.0 | 2.0 | 40.0 | 0 | 1 | 0 |
| 12955 | 3-Ingredient Tomato Soup | 4 | 209.0 | 4.0 | 16.0 | 684.0 | 0 | 1 | 0 |
| 18220 | 3-Ingredient Caramel-Pear Galette | 5 | 234.0 | 2.0 | 7.0 | 213.0 | 0 | 1 | 1 |

* **Are there correlations between preparation time and recipe ratings?**

Their 0.017% relationship between preparation time and recipe rating column. It means there is no relationship between preparation time and recipe rating.

relation = data\_c[['rating','22\_minute\_meals']]

plt.figure(figsize=(10,5))

sns.heatmap(relation.corr(),annot = True)

plt.show()

* **How can the data help improve the user experience for a recipe platform?**

Provide recommendations based on balanced nutritional content (e.g., moderate calories, fat, and sodium).

# Function to recommend balanced diet recipes.

def balanced\_recipes():

balanced\_recipes = data\_c[(data\_c['calories'] <= 500) &

(data\_c['fat'] <= 20) &

(data\_c['sodium'] <= 300)]

return balanced\_recipes[['title', 'rating', 'calories', 'protein', 'fat', 'sodium']].head(10)

balanced\_recipes()

The function provides a preview of recipes that fit into a balanced diet based on the above criteria. Each result shows the recipe's title, rating, and its nutritional breakdown (calories, protein, fat, and sodium content).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Title | rating | calories | protein | fat | Sodium |
| 2 | Potato and Fennel Soup Hodge | 4 | 165.0 | 6.0 | 7.0 | 165.0 |
| 10 | Yams Braised with Cream, Rosemary and Nutmeg | 4 | 256.0 | 4.0 | 5.0 | 30.0 |
| 13 | Beef Tenderloin with Garlic and Brandy | 4 | 174.0 | 11.0 | 12.0 | 176.0 |
| 16 | Sweet Buttermilk Spoon Breads | 2 | 146.0 | 4.0 | 5.0 | 160.0 |
| 25 | Garlic Baguette Crumbs | 0 | 95.0 | 1.0 | 7.0 | 103.0 |
| 26 | Cucumber-Basil Egg Salad | 4 | 215.0 | 6.0 | 20.0 | 250.0 |
| 27 | Dried Pear Crisps | 3 | 14.0 | 0.0 | 0.0 | 0.0 |
| 28 | Green Bean, Red Onion, and Roast Potato Salad … | 4 | 351.0 | 6.0 | 19.0 | 79.0 |
| 29 | Apricot-Cherry Shortcakes | 4 | 311.0 | 5.0 | 5.0 | 226.0 |
| 39 | Better-Than-Pita Grill Bread | 3 | 145.0 | 3.0 | 6.0 | 208.0 |

**4. Conclusion**

This recipes that are considered healthy based on a defined set of nutritional thresholds. It allows users to easily find suitable recipes for a balanced diet by filtering for calories, fat, and sodium content.

**Declaration:**

After conducting a comprehensive exploratory data analysis (EDA) on the recipe dataset, we affirm the following insights:

1. **Nutritional Insights**: The dataset reveals a wide range of caloric values among the recipes, with the highest caloric recipe containing **800 calories** and the lowest containing **300 calories**. This indicates significant variability in the nutritional profiles of the recipes.
2. **Protein and Fat Content**: A positive correlation (approximately **0.85**) between protein and fat content was observed, suggesting that recipes with higher fat content tend to have increased protein levels. This relationship should be further explored in future analyses to understand its implications on dietary choices.
3. **Sodium Levels**: The sodium content ranged from **100 mg** to **400 mg**, indicating that some recipes are substantially saltier than others. This could have implications for individuals monitoring their sodium intake for health reasons.
4. **Rating Distribution**: The average rating of the recipes is approximately **4.5**, reflecting a general preference for the recipes included in the dataset. However, a notable percentage of recipes received lower ratings (below **3.0**), warranting further investigation into their attributes.

**Thank You.**

**Date: 06/10/2024**

**Place: Bangalore, Karnataka, IN.**

(Meeniga Nagendra)

(+91 6302399979)

meeniga4161@gmail.com