**NutriViz:-Nutrient Data Analysis and Visualization Documentation**

**1. Introduction**

This project involves analyzing and visualizing a dataset containing nutritional information about various food items. The dataset includes attributes such as food names, serving size, macronutrient content (calories, protein, fat, fiber, carbohydrates), and food categories. The visualizations are created using Python libraries **Matplotlib, Seaborn, Pandas, and NumPy**, and the final plots are stored in a **PDF file**.

**2. Dataset Overview**

**2.1 Columns in the Dataset**

* **Food** – Name of the food item
* **Measure** – Serving size
* **Grams** – Weight in grams
* **Calories** – Energy content
* **Protein** – Protein amount
* **Fat, Sat.Fat, Fiber, Carbs** – Macronutrient composition
* **Category** – Food category (e.g., Dairy, Meat, Vegetables, etc.)

**2.2 Data Preprocessing Steps**

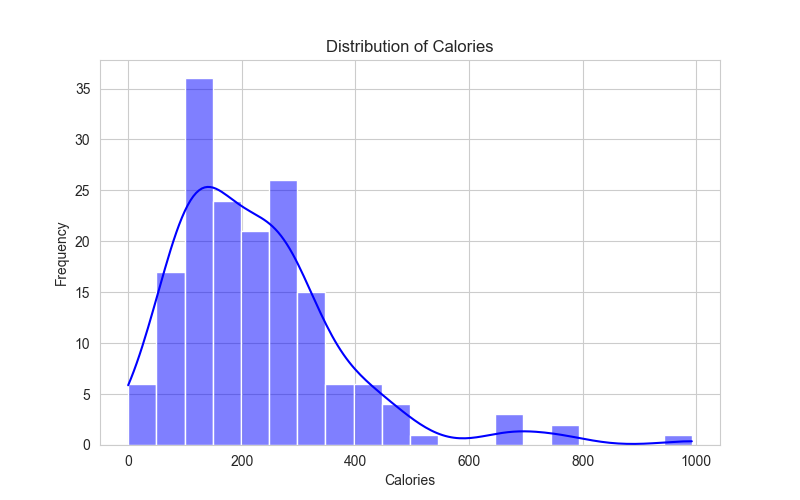
* Converted necessary columns to numeric values (handling missing or non-numeric values like 't').
* Dropped rows containing NaN values after conversion.
* Set an appropriate plotting style using **Seaborn**.

**3. Visualizations Implemented**

The following visualizations were generated and saved into a PDF file:

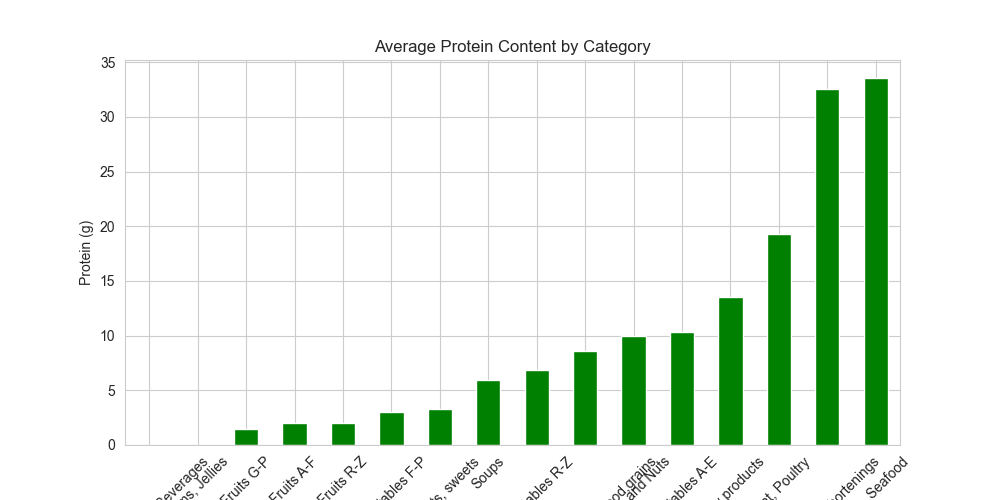
**3.1 Histogram: Distribution of Calories**

* Displays the frequency of different calorie ranges.
* Helps identify common calorie values in the dataset.



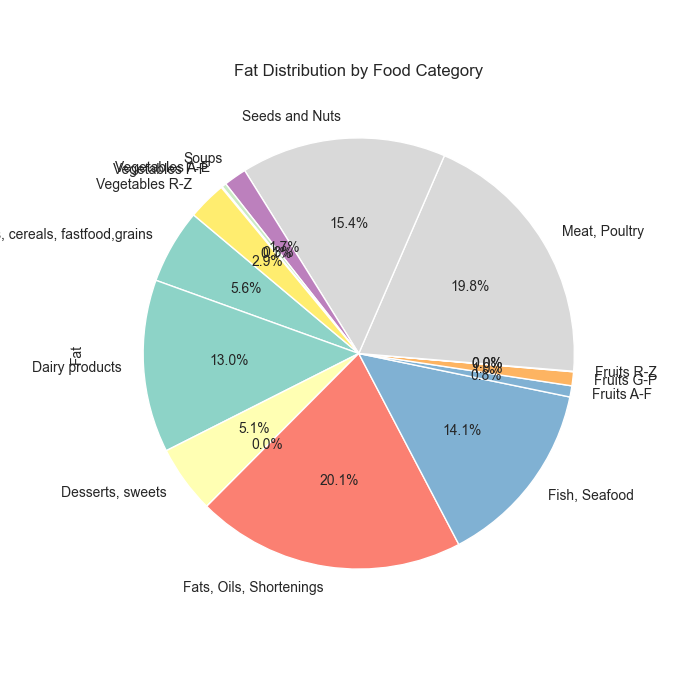
**3.2 Bar Plot: Average Protein Content by Category**

* Compares the average protein content across different food categories.
* Provides insights into which food groups are high in protein.



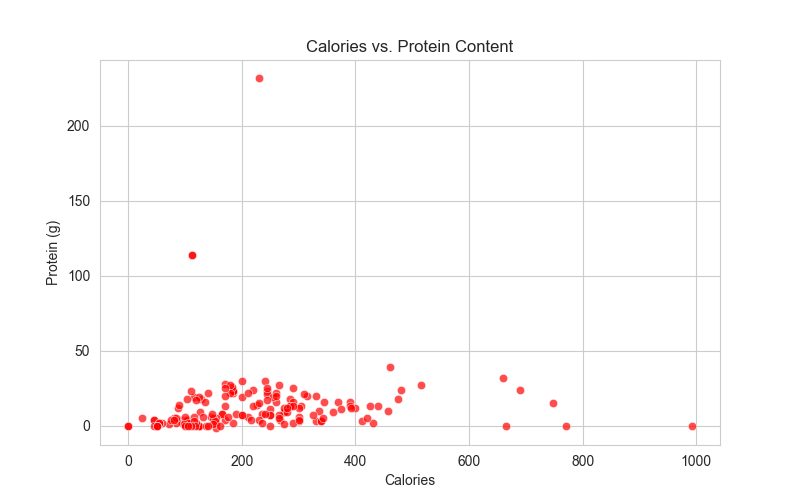
**3.3 Pie Chart: Fat Distribution Across Categories**

* Shows the proportion of fat contributed by each food category.
* Useful for understanding the primary sources of dietary fat.



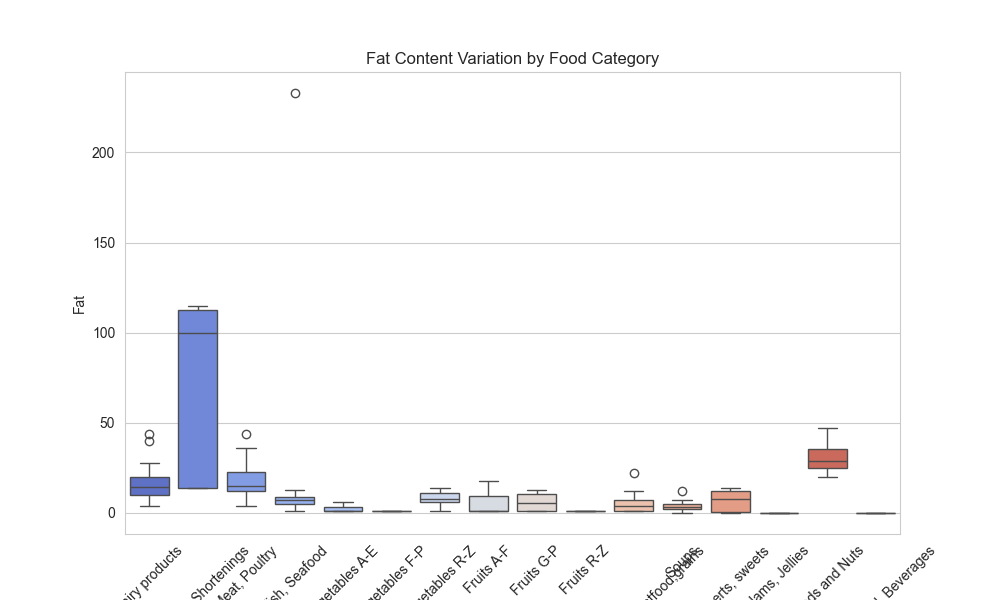
**3.4 Scatter Plot: Calories vs. Protein**

* Demonstrates the relationship between calorie content and protein.
* Helps identify high-protein, low-calorie foods.



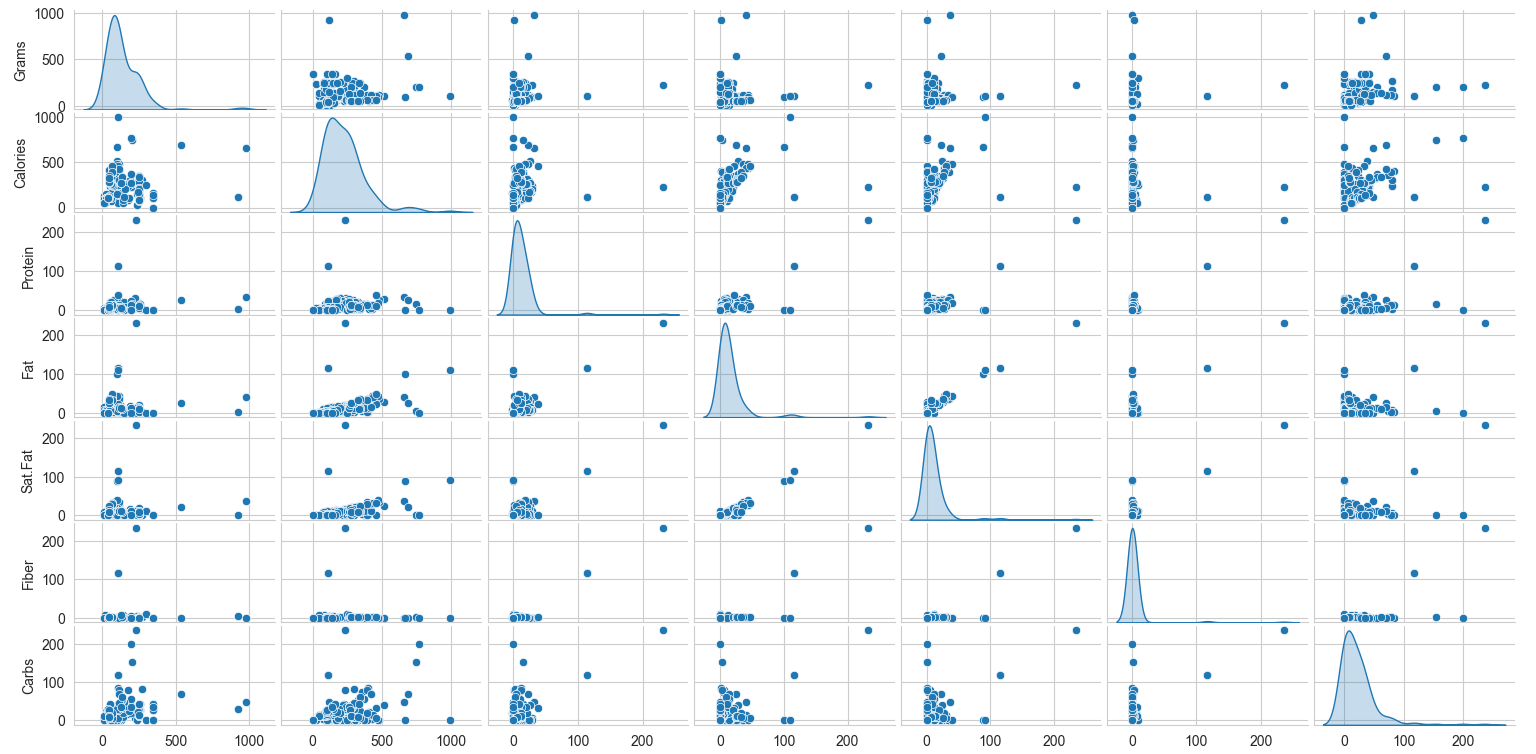
**3.5 Box Plot: Fat Content Variation by Category**

* Illustrates the spread of fat content within different categories.
* Useful for detecting outliers and variability.



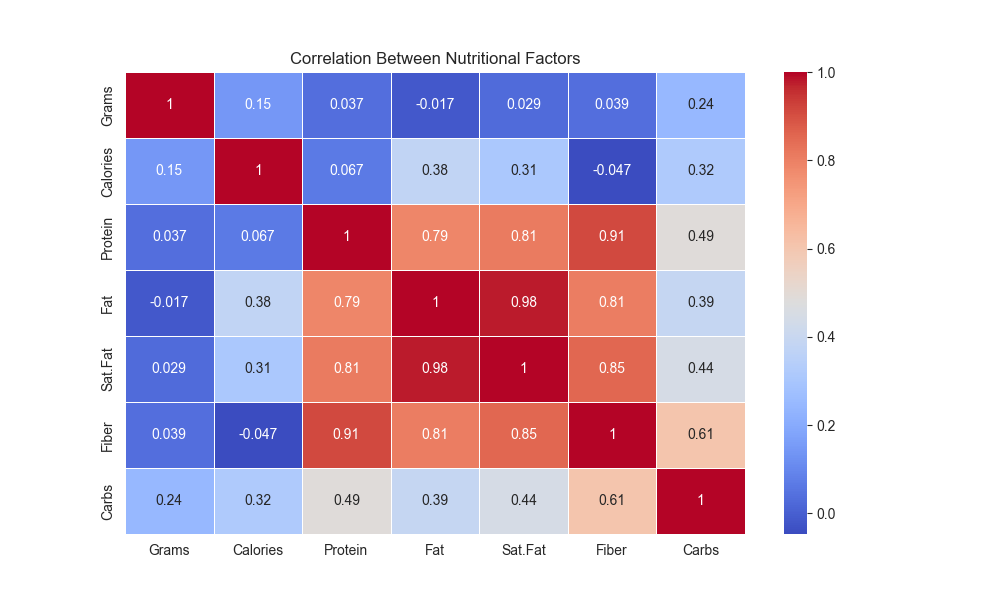
**3.6 Pair Plot: Relationships Between Nutritional Features**

* Helps visualize relationships between multiple numerical variables.
* Useful for detecting correlations and clusters.



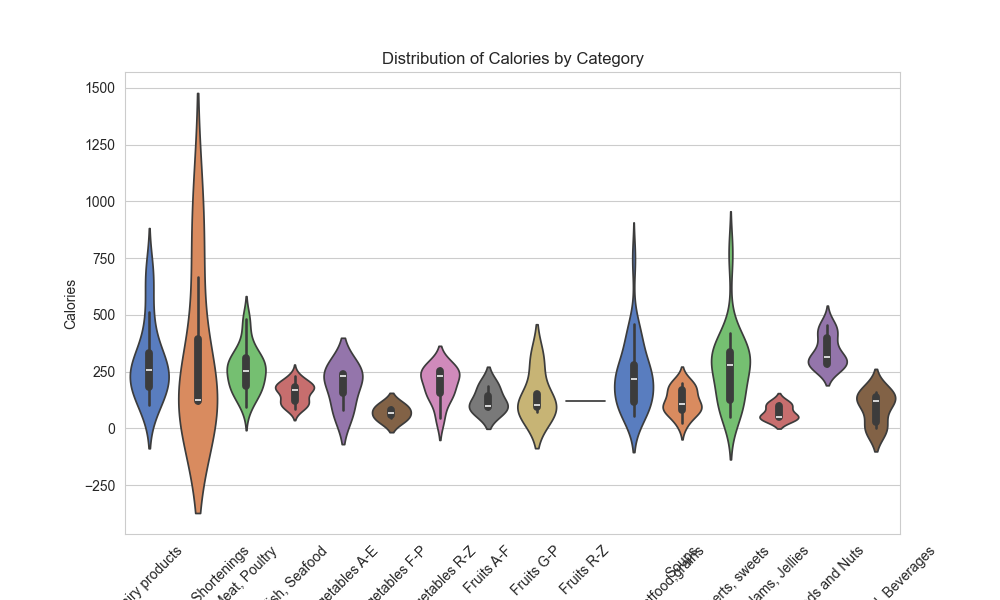
**3.7 Heatmap: Correlation Between Nutritional Factors**

* Displays correlation coefficients between numerical features.
* Helps understand how different nutrients relate to each other.



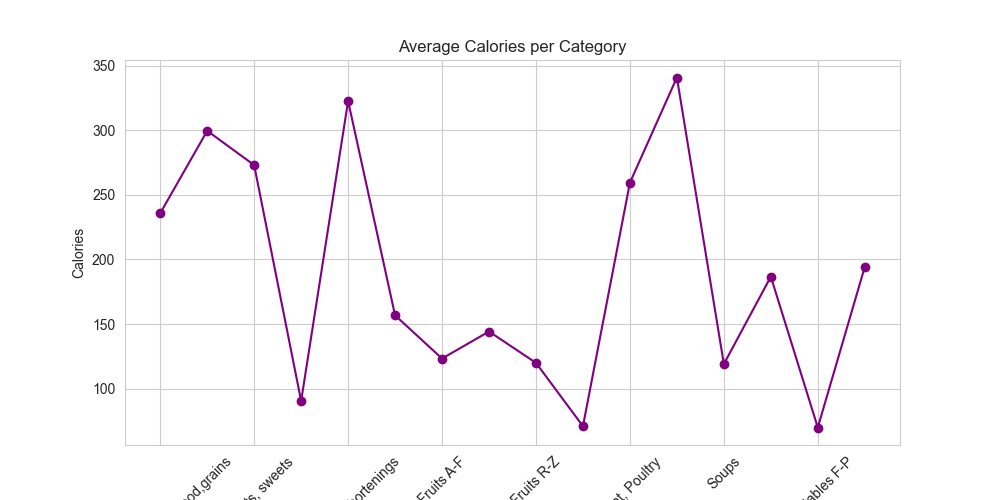
**3.8 Violin Plot: Distribution of Calories by Category**

* Shows the spread and density of calorie values for each category.
* Useful for comparing distributions across categories.



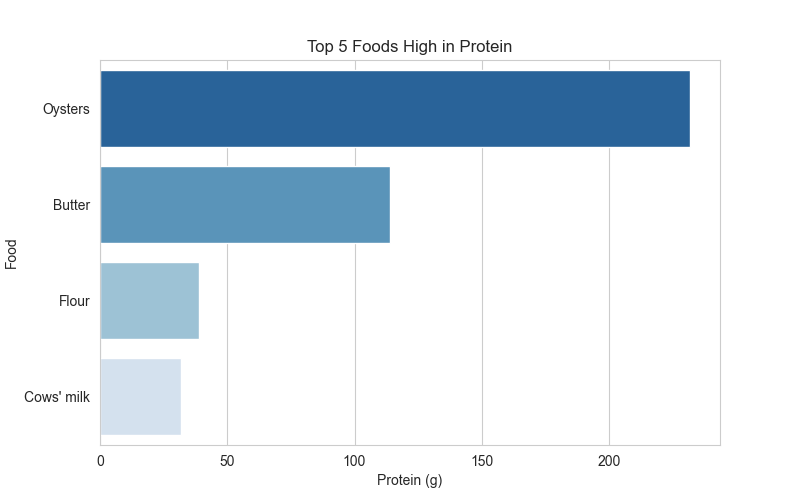
**3.9 Line Plot: Trends in Calories Over Food Categories**

* Displays the average calorie content for each category.
* Helps track variations across different food groups.



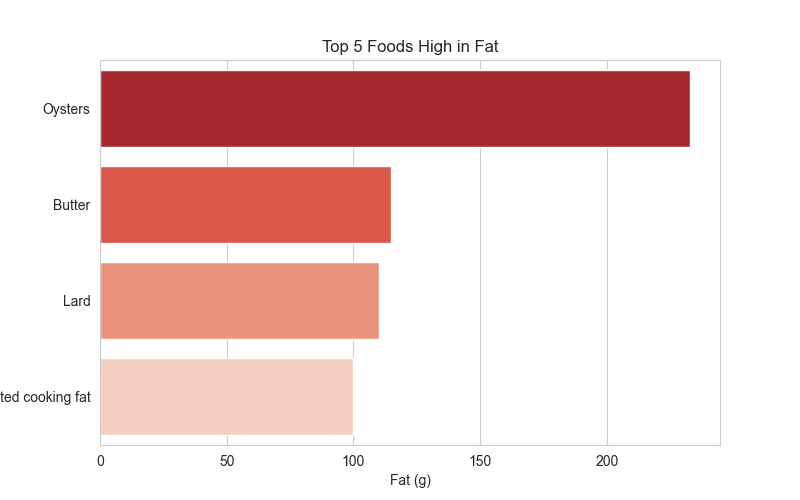
**3.10 Top 5 Foods High in Protein**

* Identifies the top 5 foods that contain the highest protein content.
* Uses a bar plot to visually compare these foods.



**3.11 Top 5 Foods High in Fat**

* Identifies the top 5 foods that contain the highest fat content.
* Uses a bar plot to visualize these high-fat foods.



**4. Saving the Visualizations**

* All visualizations were stored in a **single PDF file** (nutrient\_visuals.pdf).
* This ensures easy access and sharing of results without cluttering multiple image files.

**5. Technologies Used**

* **Python Libraries**: Matplotlib, Seaborn, Pandas, NumPy
* **File Formats**: CSV for input data, PDF for output visualizations
* **IDE**: Visual Studio Code (VS Code)

**6. How to Run the Script**

1. Ensure the dataset (nutrients\_csvfile.csv) is available in the working directory.
2. Install required dependencies using:

pip install pandas numpy matplotlib seaborn

1. Run the Python script to generate and save the visualizations:

python nutrient\_analysis.py

1. Open nutrient\_visuals.pdf to view the generated plots.

**7. Conclusion**

This project provides a structured way to analyze and visualize nutritional data efficiently. By leveraging **Matplotlib, Seaborn, and Pandas**, we created insightful visuals that help understand food composition in terms of calories, protein, and fat distribution across categories. The final output in PDF format ensures easy accessibility and sharing.

This project was built using **Visual Studio Code (VS Code)** as the primary development environment.

For future improvements, additional statistical analysis and more advanced machine learning techniques can be incorporated to derive deeper insights from the data.