FOOD NUTRITION ANALYSIS

ABSTRACT

In this project “FOOD NUTRITION ANALYSIS” we examine the nutritional composition of various food enabling a deeper understanding of their impact on human health.

INTRODUCTION

- Food nutrition analysis is a critical process that helps individuals, healthcare professionals and food manufactures understand the composition of food .
- Plays pivotal role in making informed dietary choices
- Helps in developing healthy eating plans

DATA PREPROCESSING

Brief description of the dataset

- The dataset contains information on various nutrients in different foods.
- It includes essential nutrients: calories, protein, fat, saturated fat, fiber, and carbohydrates.
- The dataset has categories of food items into groups like "Dairy products," "Vegetables," "Fruits," "Meat Poultry," etc.
- It has specific measurement units (e.g., "1 qt.," "1 cup," "3 oz.") for serving sizes. Data types include integers, floats, and object types.

Statistical summary

- Number of Food Entries: More than 300 food items.
- Nutrients Analyzed: Calories, Protein, Fat, Saturated Fat, Fiber, and Carbohydrates.
- Data Types: The dataset contains various data types, including integers, floats, and object types.

For each nutrient (Calories, Protein, Fat, Saturated Fat, Fiber, Carbohydrates), we provide a summary of basic statistics such as:

- Mean: The average value for each nutrient across all food items.
- Minimum: The smallest value observed for each nutrient in the dataset.
- Maximum: The largest value observed for each nutrient in the dataset.
- Standard Deviation: A measure of the spread or variability of each nutrient values.

THE PHASE OF DATA-PREPROCESSING

Replacing Trace Amounts

- Identified and replaced "t" values in the dataset, representing trace amounts, with zeros (0).
- Ensured consistency in data representation for minimal nutrient quantities.

Removing Unwanted Characters

- Removed unwanted characters like commas from numerical data.
- This step allows converting numerical values from string to integer or float types.

Converting Data Types

- Converted relevant columns to desired data types (int or float) for further analysis.
- Ensured that data types match the nature of the nutrient information for accurate computations.

Handling Null Values

- Performed data quality checks to identify and handle null or missing values.
- Rows with null values were typically removed to maintain data integrity.

Data Validation

- Validated the dataset for consistency and correctness after preprocessing.
- Ensured that nutrient values are meaningful and logical for each food item.

DATA VISUALIZATION

Here, we delve into the heart of our nutrition dataset, uncovering key insights through data visualization and analysis.

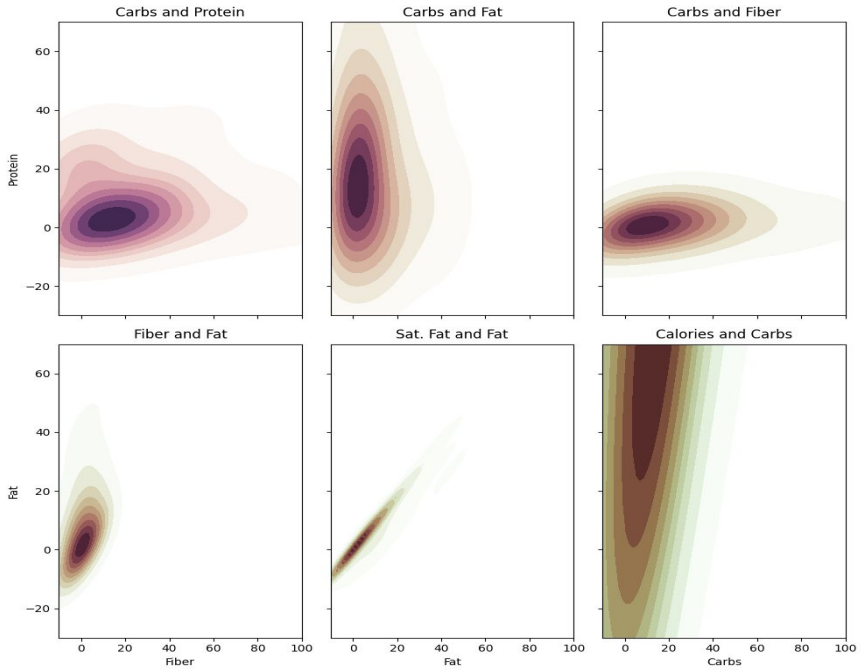
- Visual representations allow us to grasp complex patterns and relationships within the data, providing a comprehensive view of the nutritional landscape.
- We started by examining bar charts that highlight protein-rich and calorie-rich foods, shedding light on which foods pack the most nutritional punch.
- We explored the intricate interplay between fat content and calorie content in various food items, helping us make informed choices about dietary fats.
- Additionally, we presented category-wise distribution of nutrients using pie charts, offering a holistic perspective on nutritional patterns across different food groups.

We embark on this data-driven exploration, unraveling the nutritional secrets within our dataset through compelling visualizations and insightful analysis.

KERNEL DENSITY ESTIMATION (KDE) PLOTS

We began by creating KDE plots to explore the relationships between different nutrient variables.
These plots allowed us to visualize the distributions of nutrients and identify any patterns or correlations. We examined:

- Carbs vs. Protein
  Understanding the balance of carbohydrates and protein in various foods.
- Carbs vs. Fat
  Exploring the relationship between carbohydrates and fat content.
- Carbs vs. Fiber
  Investigating the connection between carbohydrates and dietary fiber.
- Fiber vs. Fat
  Analyzing the interplay between fiber and fat in foods.
- Fat vs. Saturated Fat
  Assessing the presence of saturated fat in relation to total fat content.
- Carbs vs. Calories
  Exploring the caloric content based on carbohydrate levels.



CATEGORY-WISE ANALYSIS

We delved deeper into specific food categories to answer important questions.

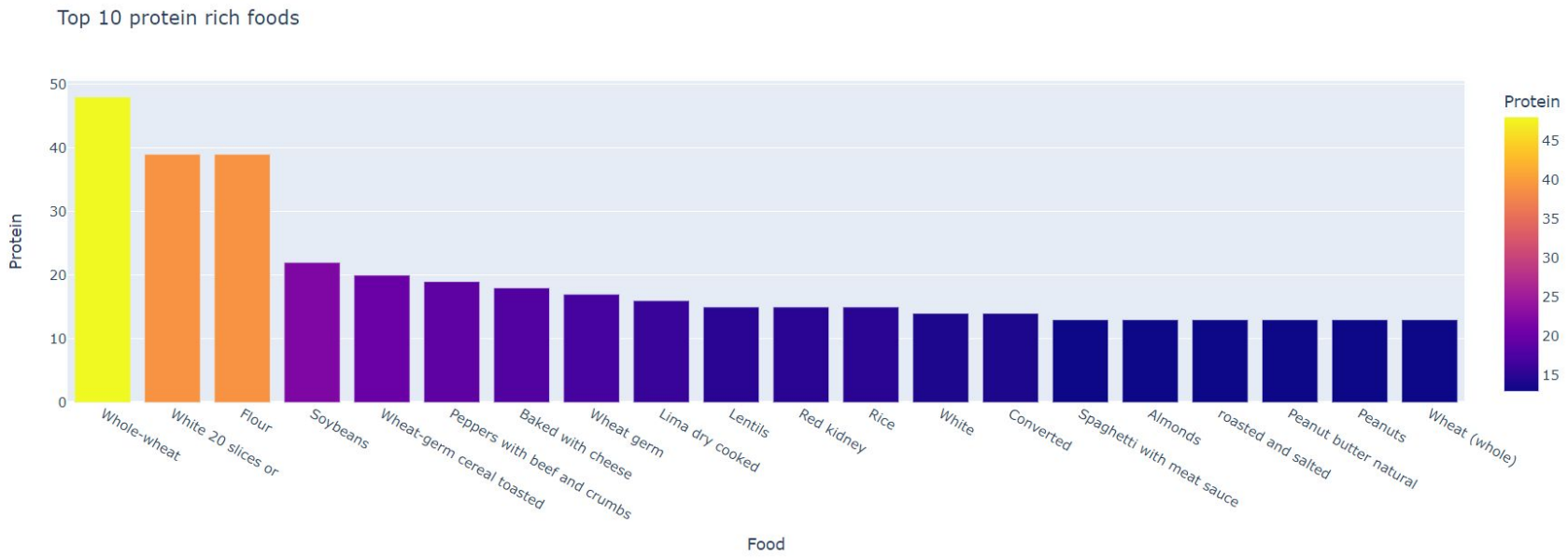
- Top 10 Protein-Rich Foods in Vegetables and Grains:
  We identified and visualized the most protein-rich foods within the vegetable and grain categories.
  This information can guide dietary choices for individuals seeking to increase their protein intake.

- Food with the Most Calories:
  We pinpointed the food item with the highest calorie content, providing valuable information for those monitoring calorie intake.

- Fat Content of Foods:
  We analyzed the fat content across various food items, shedding light on which foods are high or low in fat.

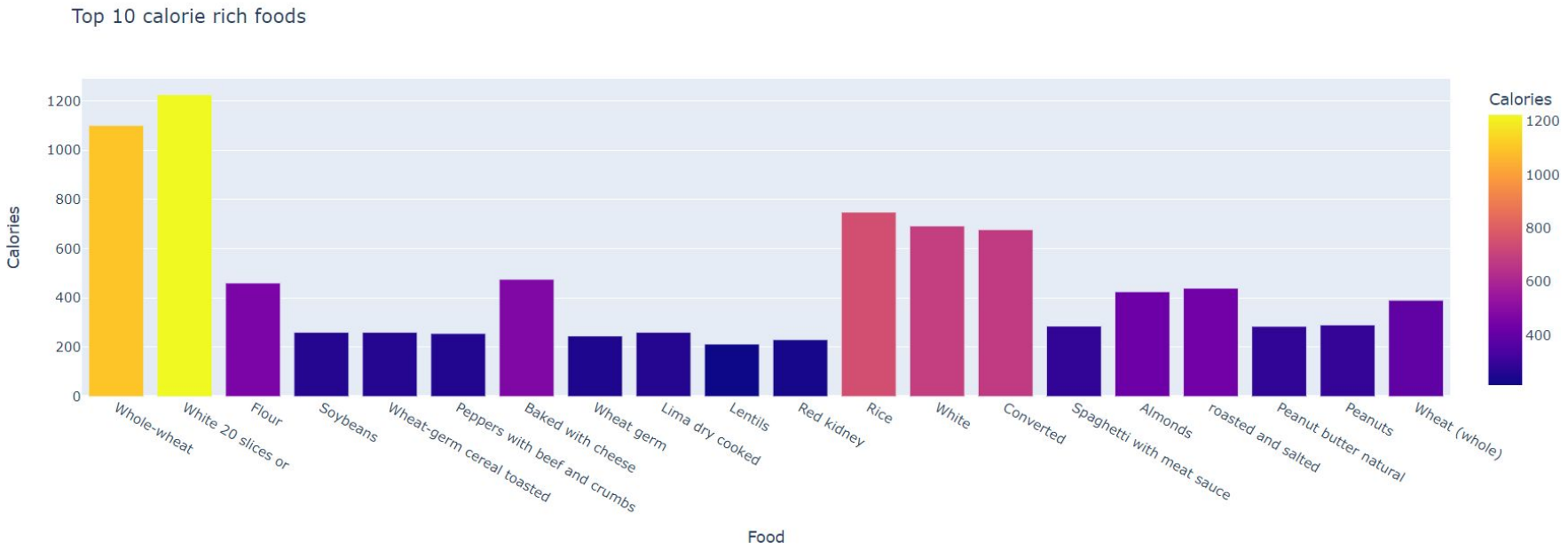
This insight can be crucial for individuals following specific dietary plans.

TOP 10 PROTEIN-RICH FOODS



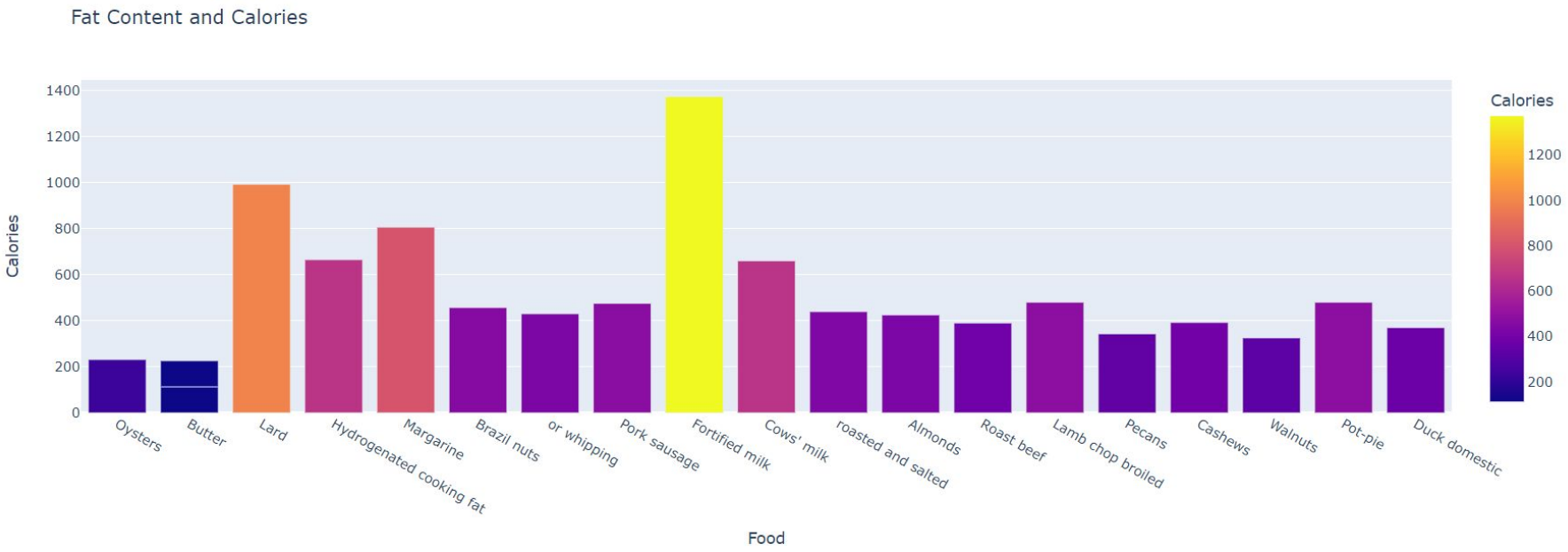
A bar chart highlighting the top protein-rich foods within the vegetable and grain categories, making it easy to identify protein sources.

CALORIE COMPARISON



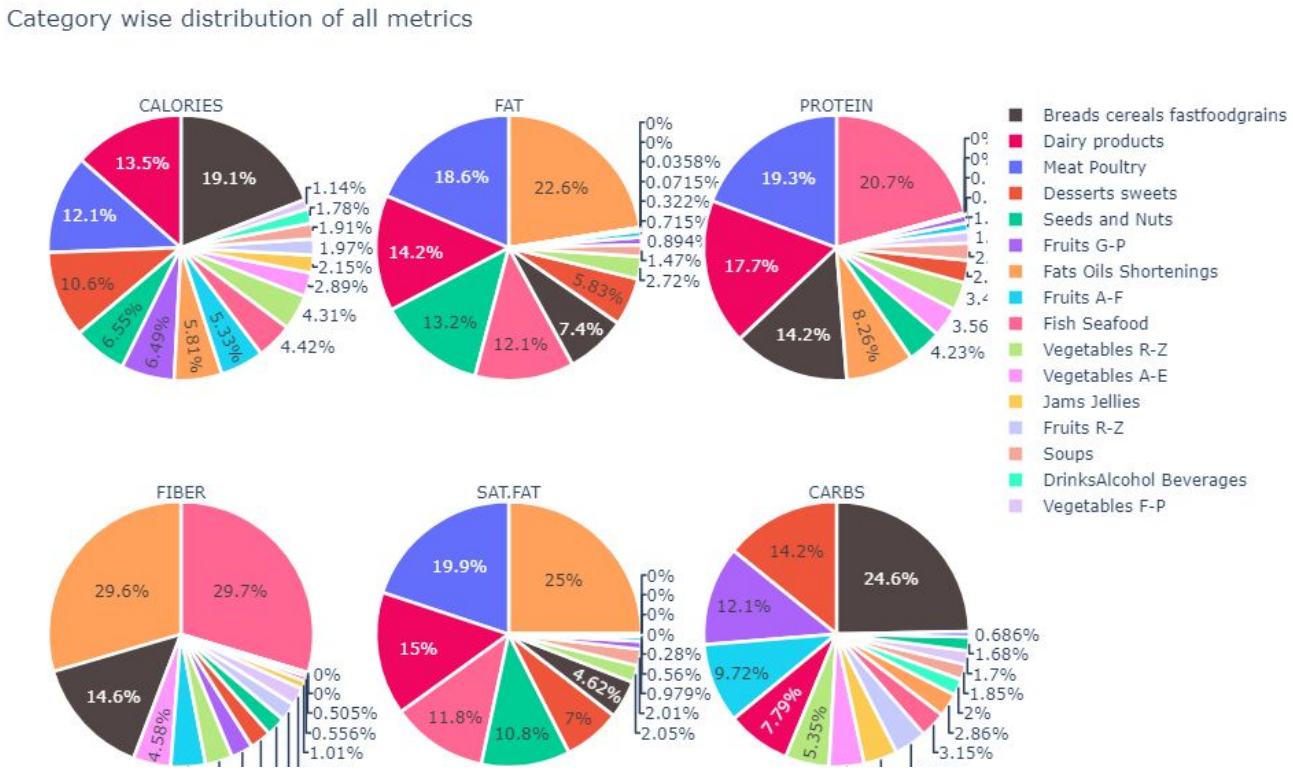
We used a bar chart to showcase the calorie content of different foods, revealing which food item is the highest in calories.

FAT CONTENT ANALYSIS



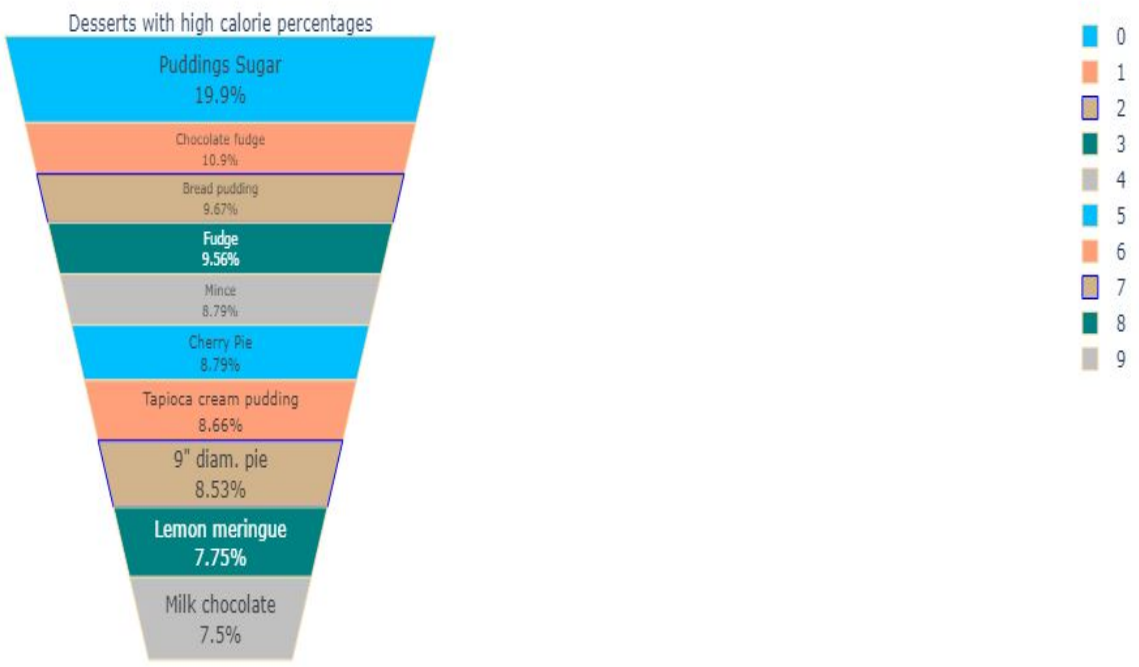
A bar chart helped us categorize foods based on their fat content, allowing for a quick assessment of fat-rich and low-fat options.

CATEGORY WISE DISTRIBUTION



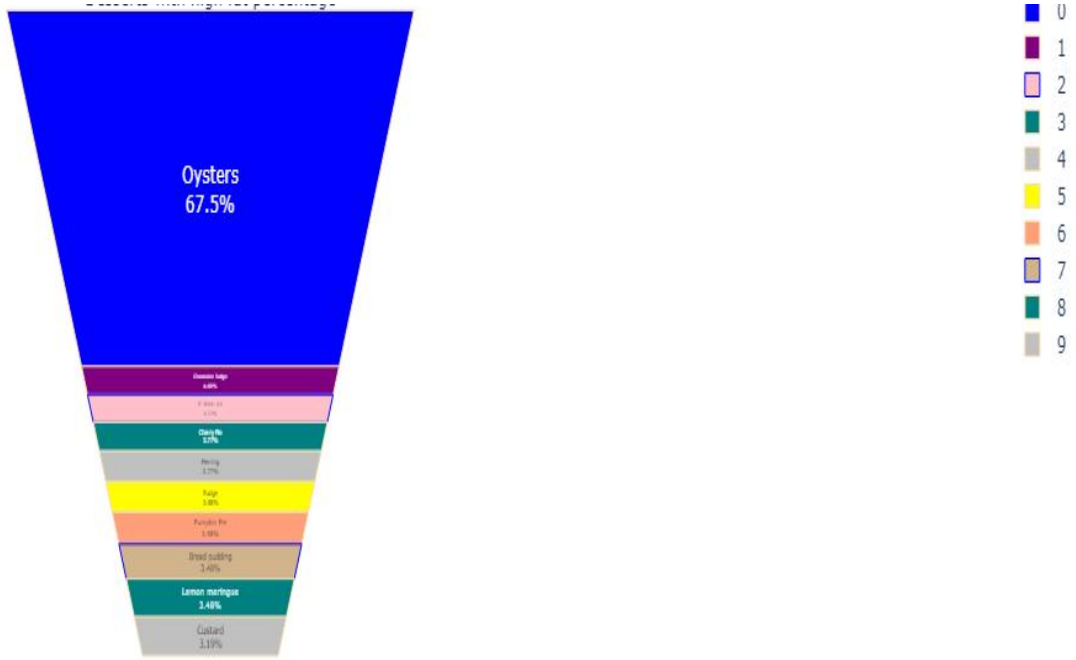
Distribution of food based on calories,fat,protein,fiber,sat-fat,carbohydrates using pie chart

Calorie distribution in Drinks, Alcohol, Beverages and Desserts



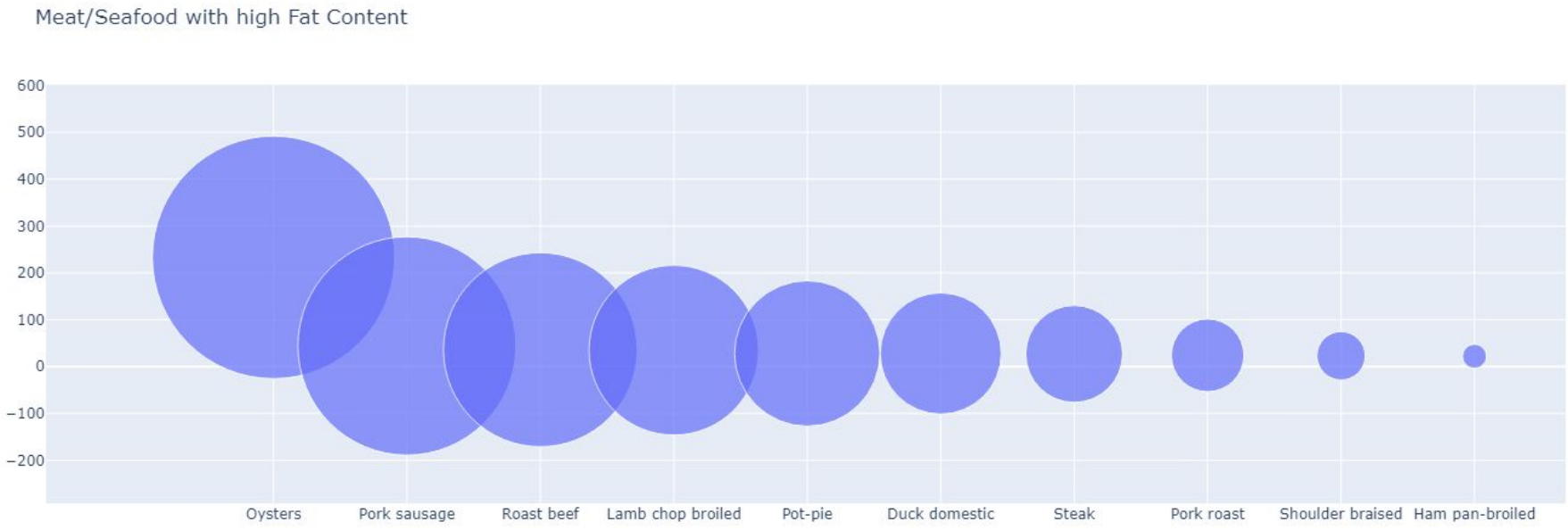
Funnel chart to understand the calorie distribution in drinks,Alcohol, Beverages and Desserts

Analysing fat distribution

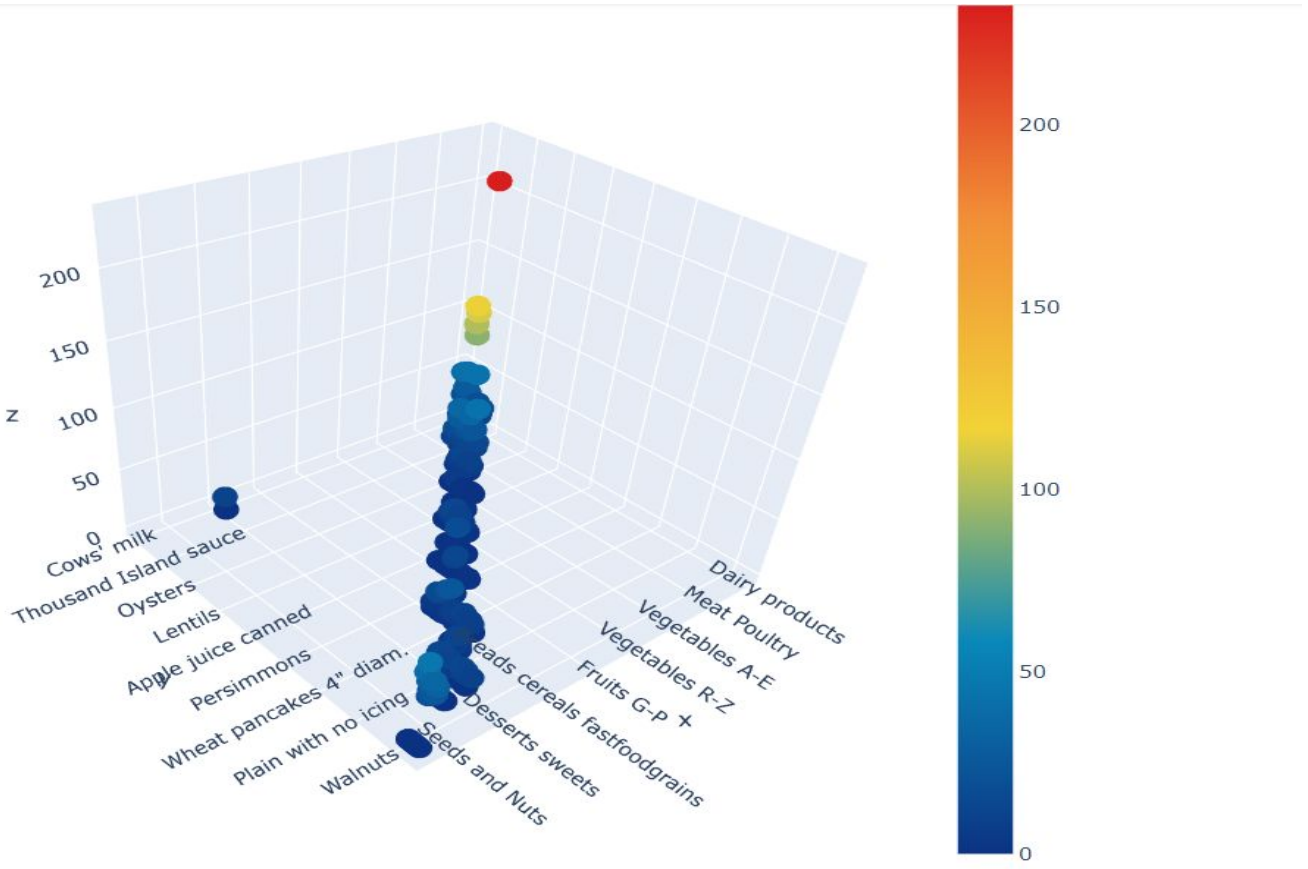


Protein Analysis in meat poultry and seafood.

Analysing Fat content in meat/seafood



3D Scatter Plot of Fatty foods (% Daily Value)



Model Building

Preprocessing Tasks:

- Unnecessary Columns
- Missing values
- Data Scaling
- Encoding Categorical Variables

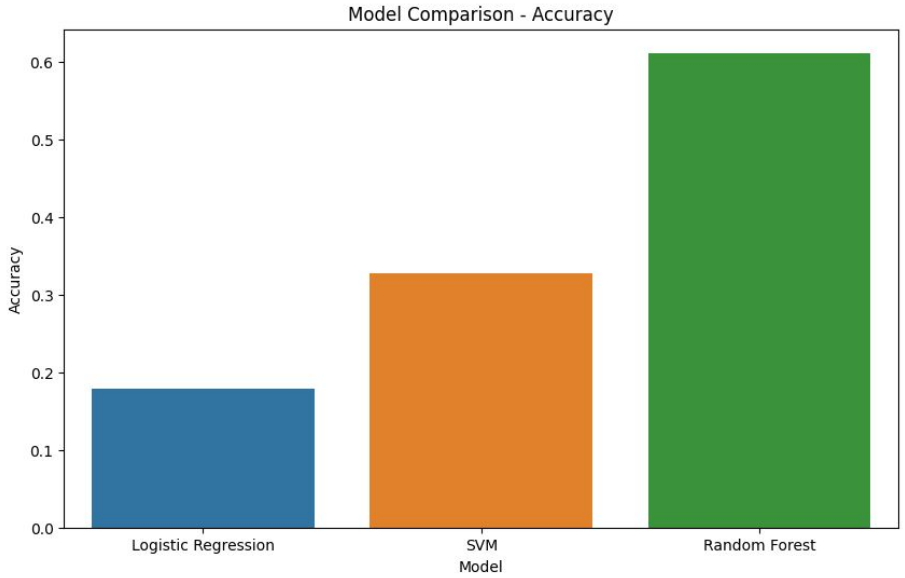
Model Training:

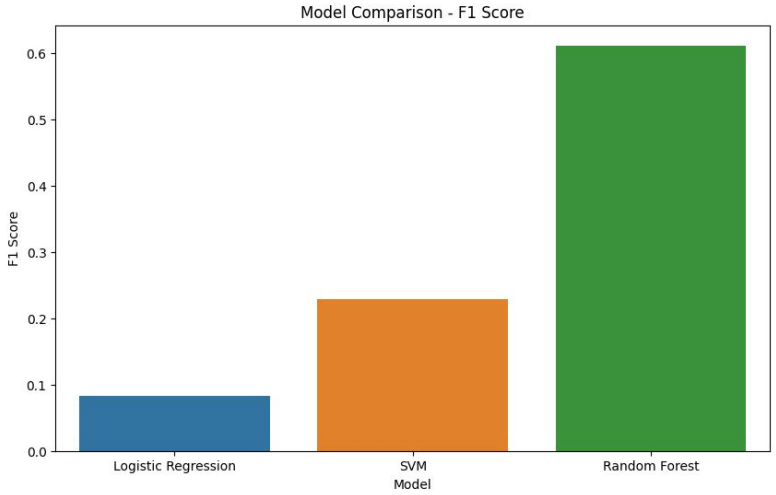
- Independent Variables: Grams, Calories, Protein, Fat, Saturated Fat, Fiber, Carbs
- Dependent Variable: Category

Test Size: 80-20

Algorithms

Logistic Regression
SVM
Random Forest Classification



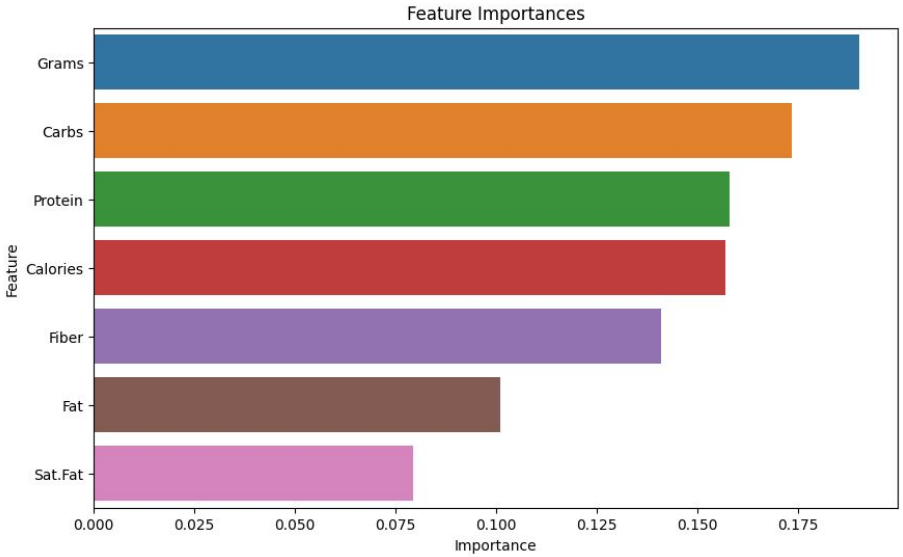


Improving the model

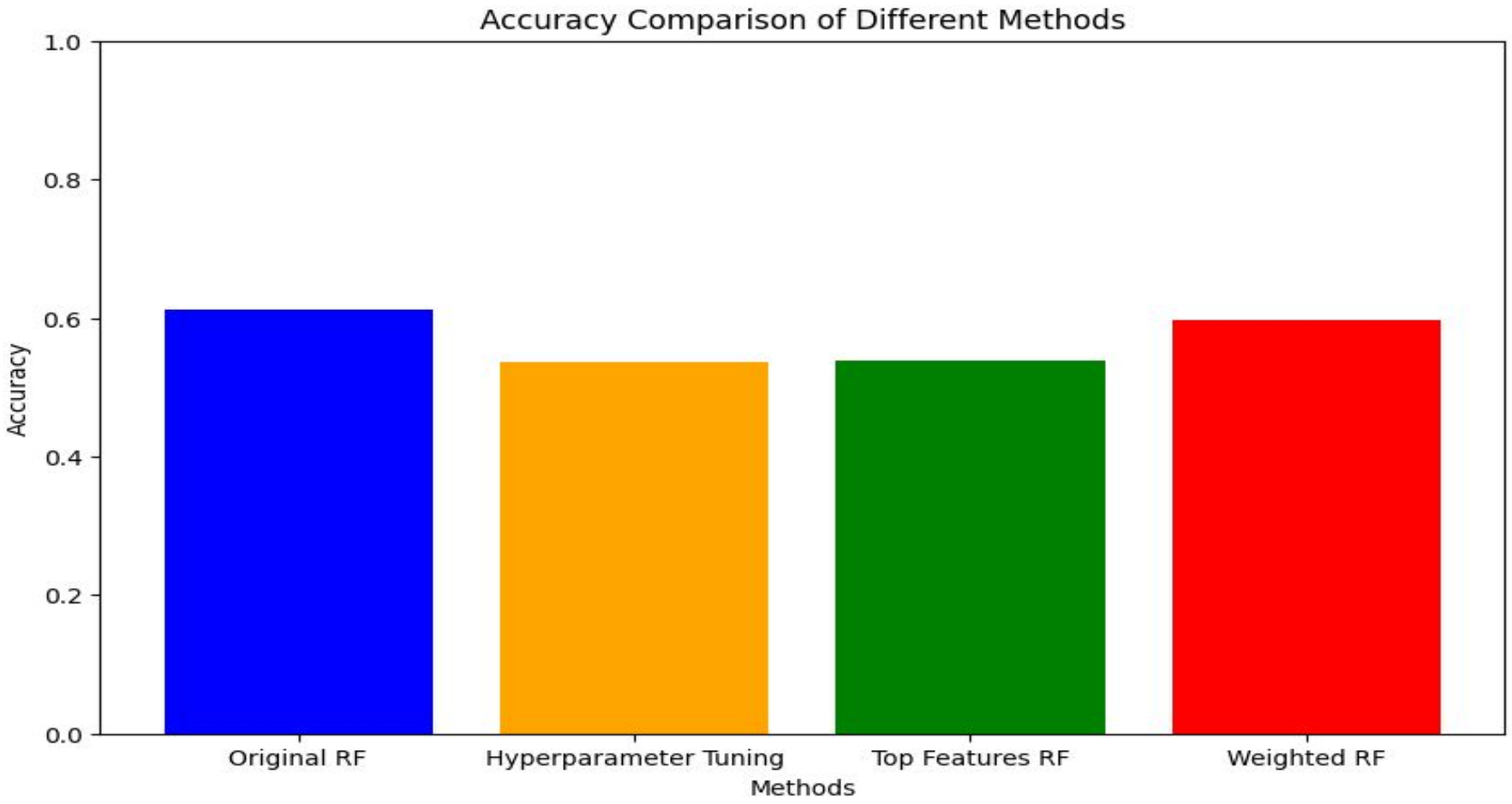
Hyperparameter Tuning:

- Best Hyperparameters: {'max_depth': None, 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 100}
- Accuracy: 0.6119402985074627

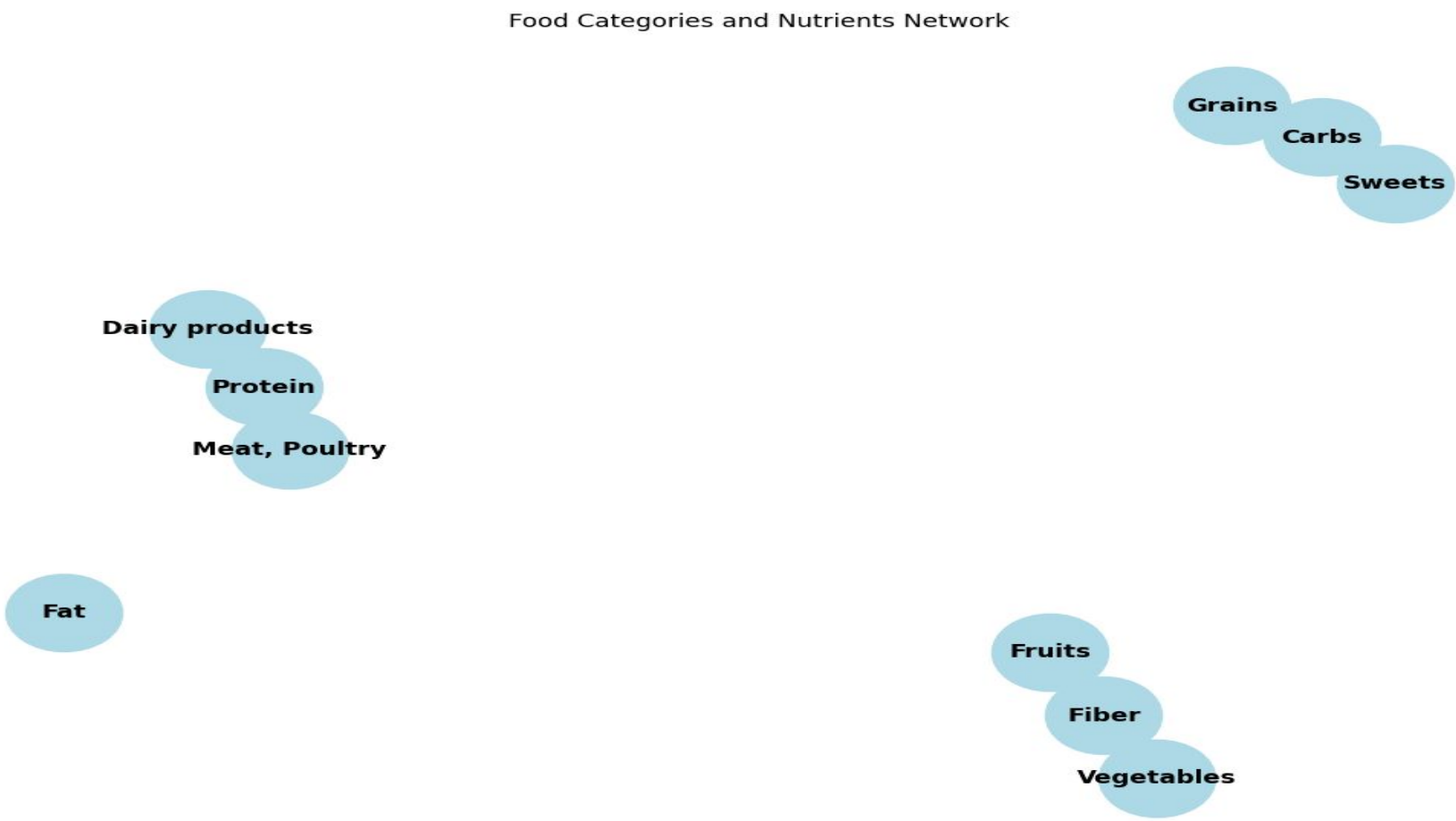
Feature Extraction:
Accuracy with Top Features: 0.53



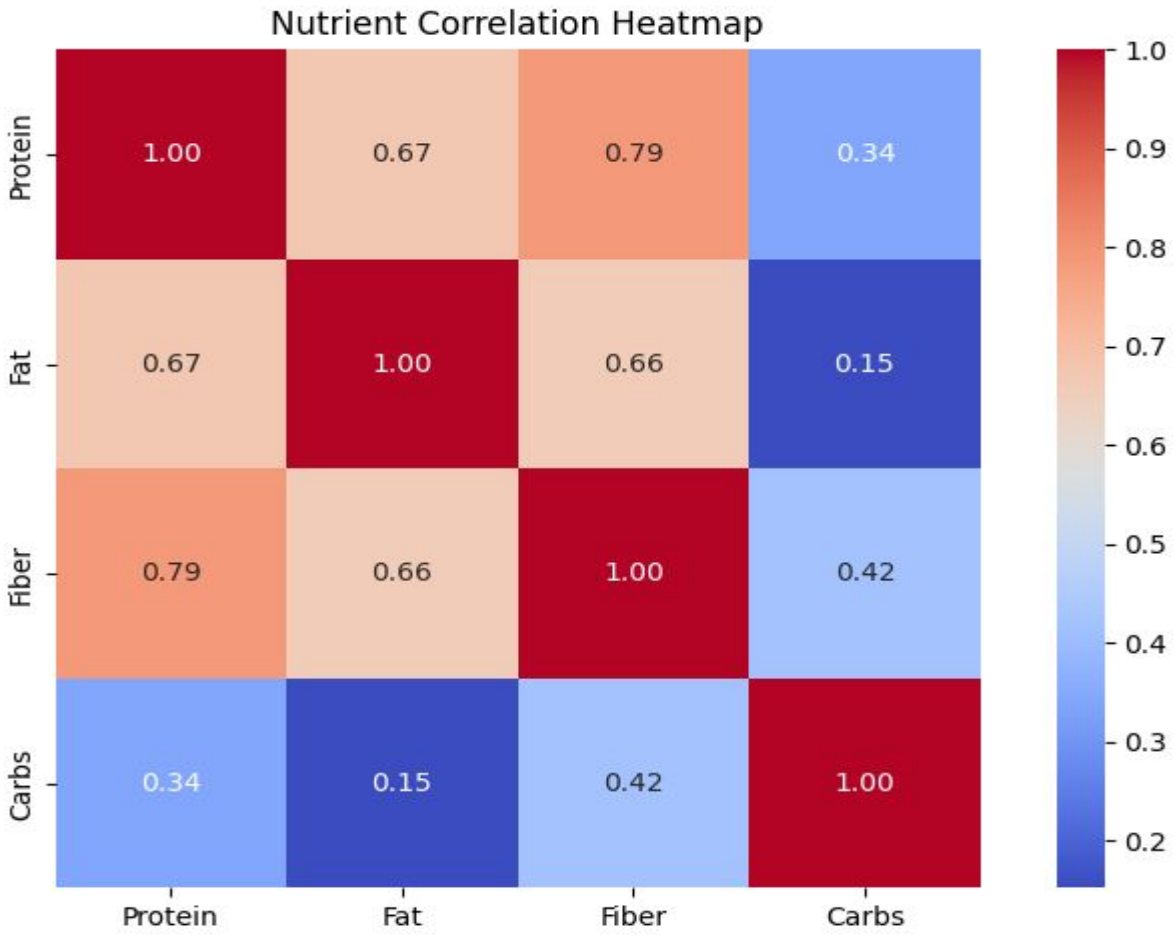
Comparative Analysis of Methods



Network Graph for Nutrients



Nutrient Correlation Heatmap



CONCLUSION

- Analysed data to gain meaningful insights.
- Built classification model using various ML Algorithms.

Future Scope:

To gather more quality data in order to build better performing model, which can be used to solve real world problems with respect to nutrients.