Mini project

Yar Moradpour

CS 131 - Sec 01

San José State University

28 November 2023

Objectives:

The dataset that I have decided to use for this mini project tends to me ‘cereal.csv’ from kaggle. This mini project tends to aim to conduct a comprehensive nutritional analysis of cereal products based on the provided dataset which includes macronutrients, and their distribution among different cereal types and manufacturers. The main objectives are based on assessing dietary patterns, focusing on calorie distribution and its relationship with other nutritional factors. it also seeks to explore consumer preferences concerning nutritional components and ascertain compliance with nutritional labeling. The goal is to derive insights into nutritional profiles, consumer behavior, and label accuracy to facilitate informed decision-making in the domain of cereal product consumption.

Link of the dataset: https://www.kaggle.com/datasets/crawford/80-cereals

Metadata:

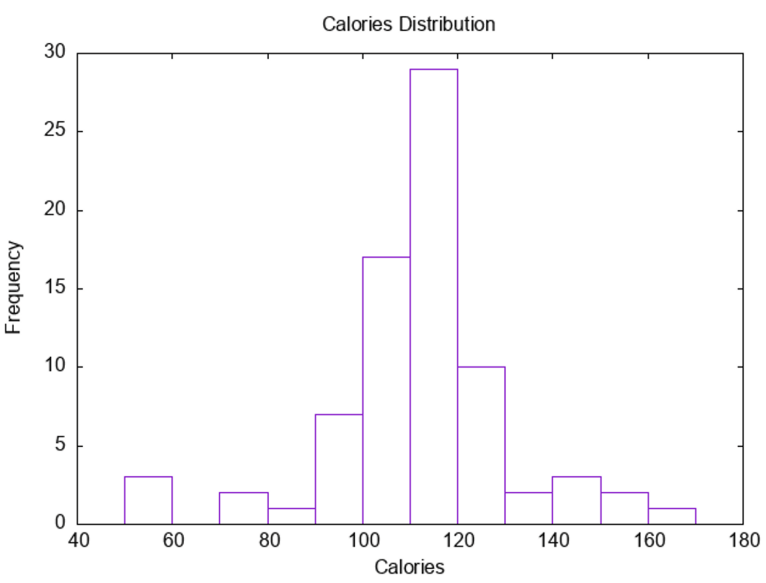
This dataset contains nutrition data of 80 known cereal products with 16 different columns including the columns as the following:

Columns present in the dataset:

1. name: Name of the cereal
2. mfr: The manufacturer of the cereal (encoded as a letter)
3. type: Type of the cereal
4. Calories: Number of calories per serving
5. Protein: Amount of protein in grams per serving
6. Fat: Amount of fat in grams per serving
7. Sodium: Amount of sodium in milligram per serving

8. Fiber: Amount of dietary fiber in grams per serving
9. Carbs: Amount of carbohydrates in grams per serving
10. Sugar: Amount of sugar in grams per serving
11. potass: Amount of potassium in milligrams per serving
12. Vitamins: Percentage of recommended vitamins and minerals
13. Shelf: Shelf which the cereal tends to be place at the store
14. weight: Weight of one serving per ounces
15. cups: Number of cups in on serving
16. rating: Rating of the cereal based on reviews of the customers

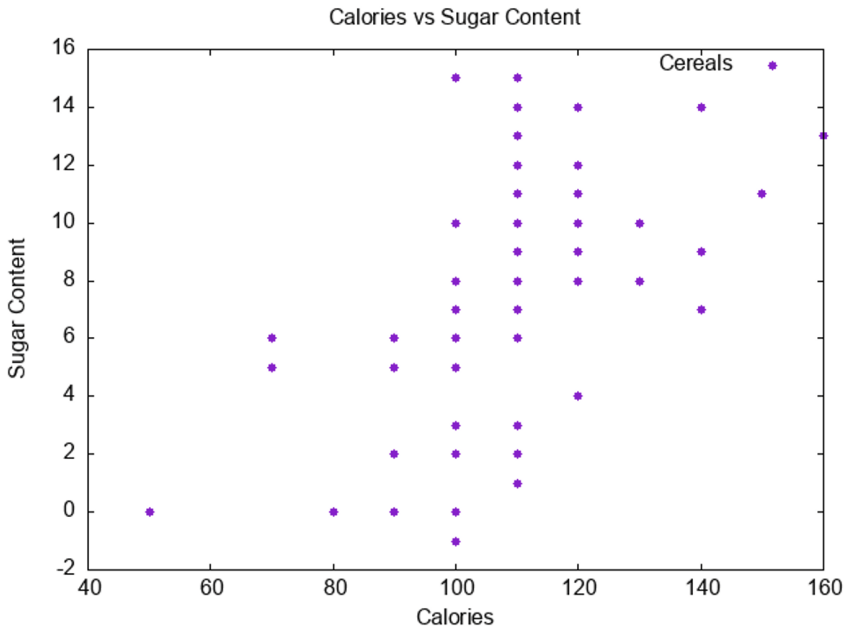
Calories Distribution Analysis



When looking at the histogram, it can be seen that most cereals in the bunch have calories between 110 and 120. The bars on the graph are the tallest around these numbers. This tells us that a bunch of cereals are pretty similar when it comes to calories. To figure out why some have

more or less, we might need to check things like the types of cereals or brands. Knowing this helps people pick cereals that suit them better for their diet.

Calories vs Sugar Content Analysis



Looking at the scatter plot comparing calories to sugar content, it's clear that most cereals fall within the range of 92 to 120 calories and have a sugar content between 4 to 12. This means that many cereals tend to have similar amounts of sugar for a particular calorie count. However, there are a few outliers that might be interesting to explore further, as they have either higher sugar or lower calorie content compared to the majority.

Tools and commands used for this mini project:

This mini project utilized awk to handle the dataset and gnu plot for plotting the final analysis. awk command in this case helped to handle extract specific columns for analysis, storing data subsets in separate files And gnu plots mainly for data visualization for the analysis.