**Analysis of Cereal Ratings Based on Nutritional Content**

**Introduction**

Cereal is a popular breakfast choice worldwide, with numerous brands offering a variety of options. Nutritional content often influences consumer preferences and, subsequently, product ratings. This study aims to analyze the relationship between nutritional attributes (sugar, fiber, protein, calories) and cereal ratings, alongside the influence of marketing strategies towards children and brand reputation.

**Data Collection**

The dataset comprises nutritional information and ratings for various cereals. The columns include `name`, `mfr` (manufacturer), `type`, `calories`, `protein`, `fat`, `sodium`, `fiber`, `carbo`, `sugars`, `potass`, `vitamins`, `shelf`, `weight`, `cups`, and `rating`.

**Hypotheses**

1. Healthier cereals (low in sugar and high in fiber) will have a higher rating.

2. Cereals marketed towards children will have higher sugar content and lower ratings.

3. Cereals with higher protein content will have higher ratings.

4. Cereals with high calorie counts will be less popular (lower ratings).

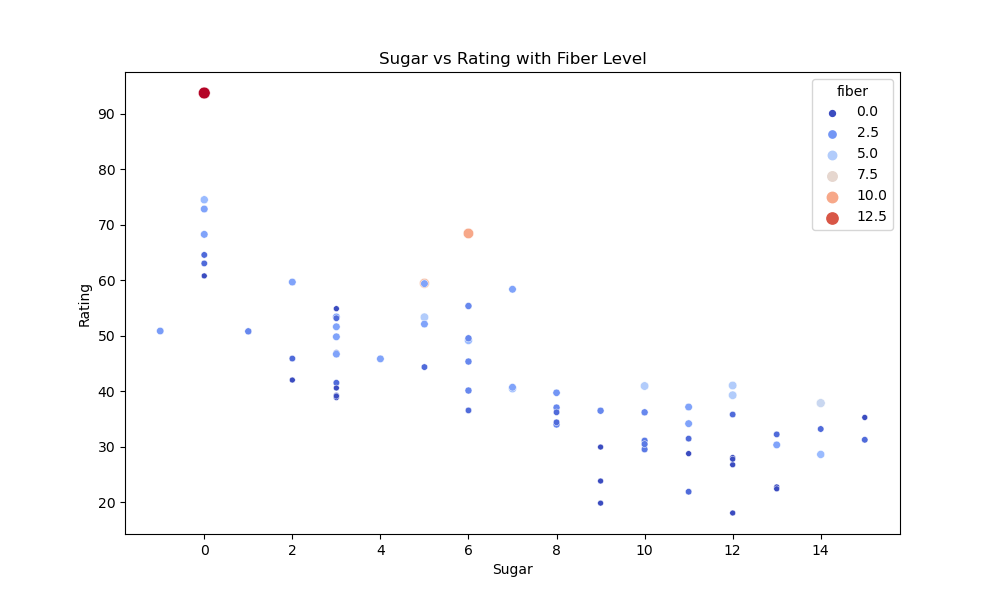
5. There will be a correlation between the brand/manufacturer and the nutritional content of cereals.

**Data Analysis**

The data analysis was conducted using Python and R. Various plots were created to visualize the relationships and test the hypotheses. Key metrics such as average ratings and content levels were calculated and compared.

**Results**

Hypothesis 1: Healthier cereals (low in sugar and high in fiber) will have a higher rating.



**Low Sugar Cereals (0-2g of sugar):**

When fiber is 0-2.5g, ratings range from 29.5 to 59.4.

When fiber is 5-12.5g, ratings range from 37.8 to 93.7.

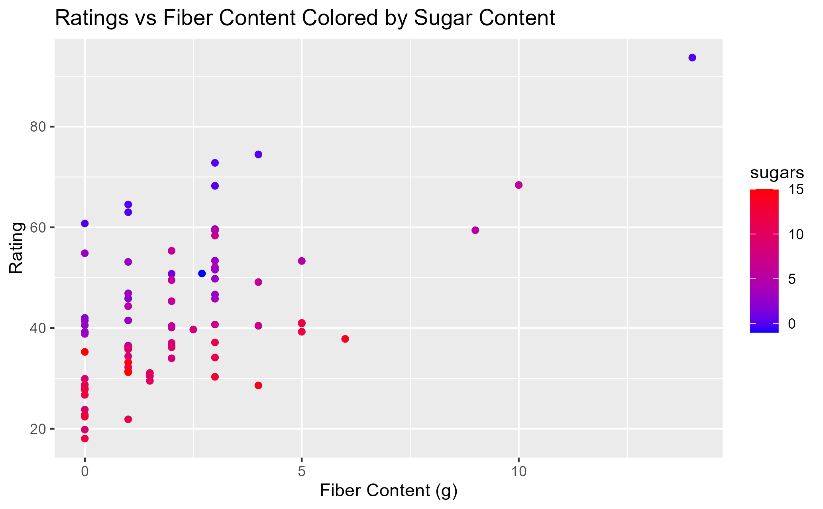
**High Fiber Cereals (10-12.5g):**

When sugar is 0-6g, ratings range from 40.4 to 93.7.

When sugar is 8-14g, ratings range from 33.9 to 47.4.

This graph shows that for low sugar cereals, those higher in fiber tend to have significantly higher ratings, supporting the hypothesis. It also demonstrates that for high fiber cereals, those lower in sugar have much higher ratings.

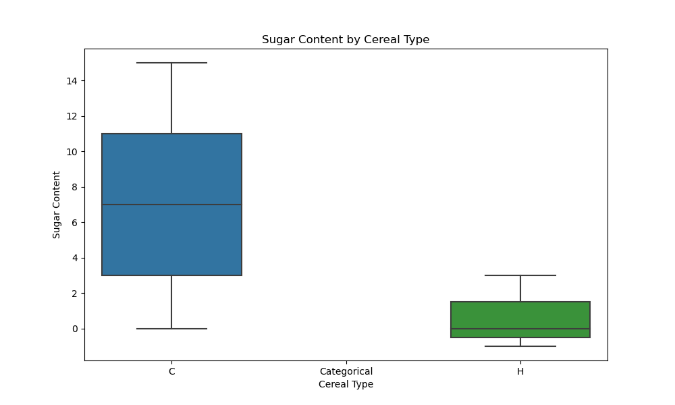
1. The cereal with the highest fiber content (around 14g) has a rating around 60. However, cereals with 0-5g of fiber span a wide range of ratings from about 40 to over 80.
2. Looking at the color scale, cereals with very high sugar content (dark purple, 15g sugar) consistently have ratings between 40-50. Lower sugar cereals (lighter color) have ratings spanning from 40 to over 80.
3. The cereal with the highest overall rating (over 80) has moderate fiber content (around 3g) and relatively low sugar.
4. There are a number of cereals with high fiber (>5g) and low sugar that have ratings in the 50-70 range, which is moderately high but not the absolute highest.



In both cases, the healthier option (lower sugar when fiber is high, higher fiber when sugar is low) has notably higher ratings on average.

Therefore, the data does appear to support the hypothesis. Low sugar seems to be a stronger predictor of high rating than high fiber. A cereal that balances moderate fiber and low sugar may hit a "sweet spot" for ratings. But fiber and sugar alone don't fully explain the ratings - other factors like taste, brand, and ingredients likely also have an impact. More data would help further evaluate the nuances of this hypothesis.

Hypothesis 2: Cereals marketed towards children will have higher sugar content and lower ratings.



Sugar Content by Cereal Type:

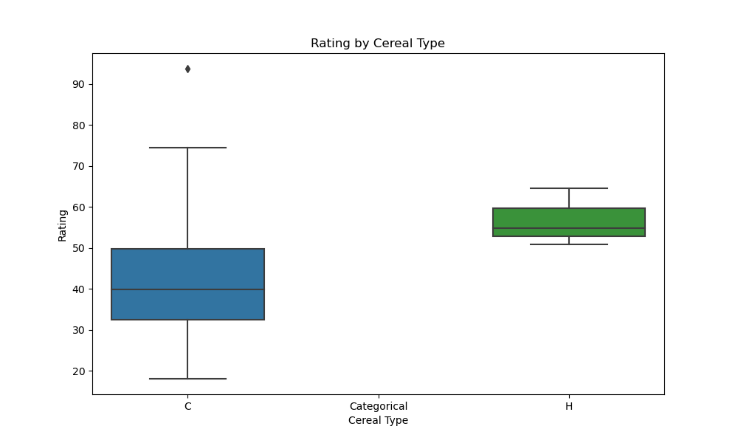
Children’s Cereals (C)

* Median sugar content is around 8 grams
* Sugar content distribution ranges from 0 to about 15 grams.
* IQR is between ~4 to 10 grams.

Healthy Cereals (H)

* Median sugar content is around 1 gram.
* Sugar content distribution is very narrow, ranging from ~0 to 5 grams.
* IQR is much smaller compared to (C), suggesting consistently low sugar content.

This data supports the first part of the hypothesis - cereals marketed towards children (C) have significantly higher sugar content compared to healthy cereals (H).



Rating by Cereal Type:

Children’s Cereals (C)

* Median rating is around 40
* Rating distribution has a wide range from ~20 to 70, with an outlier above 90
* IQR between ~35 and 50

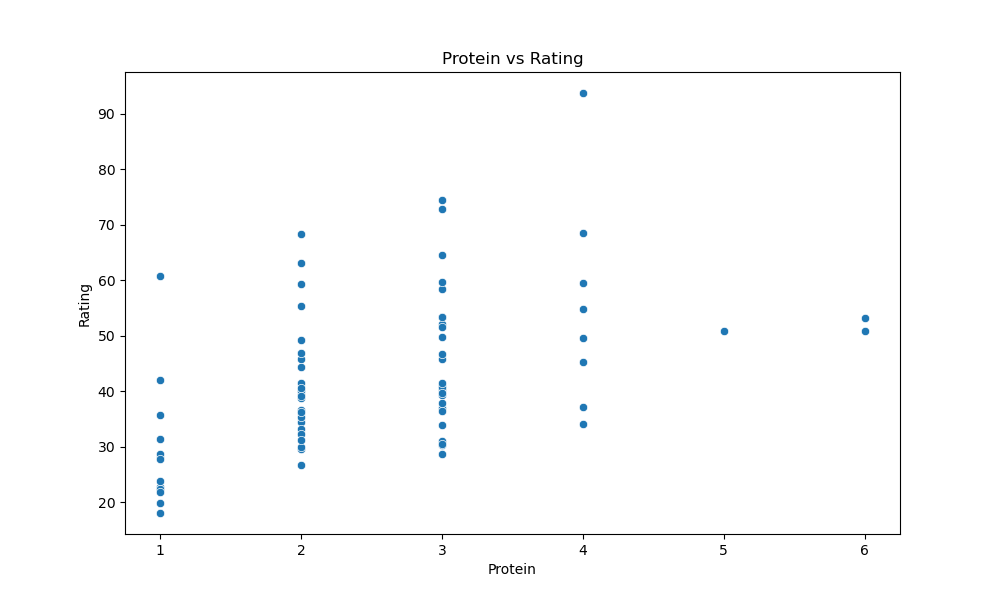
Healthy Cereals (H)

* Median rating is about 60
* Rating distribution is tighter, ranging from ~50 to 70.
* IQR is narrower than (C), suggesting more consistent ratings.

The box plot for ratings indicates that children's cereals generally have lower median ratings and a wider range of ratings compared to healthy cereals. Healthy cereals have higher and more consistent ratings. This supports the second part of the hypothesis.

Therefore, both the sugar content and rating data across the different cereal types supports the hypothesis. Cereals marketed to children tend to have substantially more sugar and receive considerably lower ratings compared to categorical cereals and health-oriented cereals.

Hypothesis 3: Cereals with higher protein content will have higher ratings.

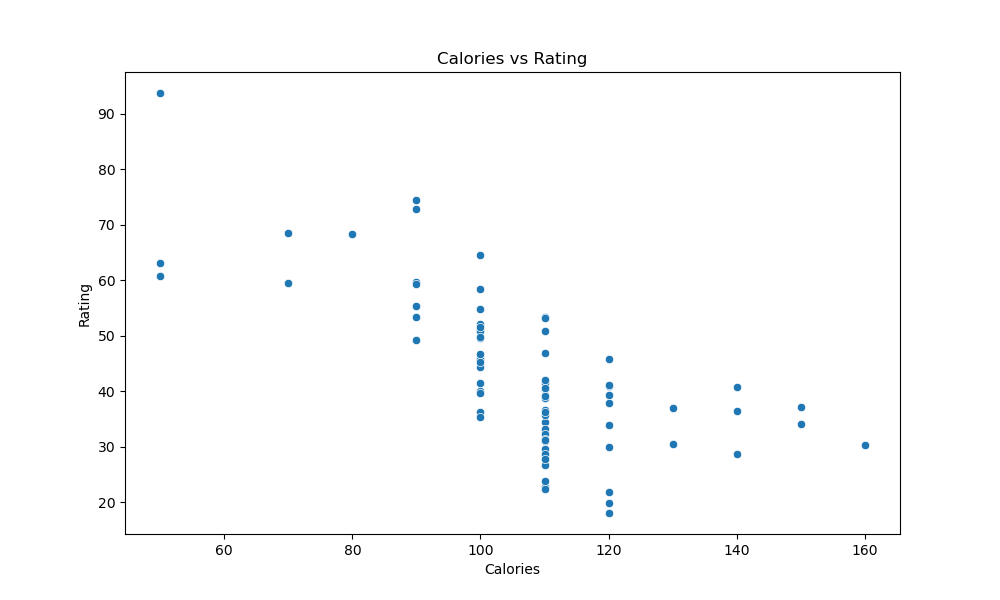


* Low Protein Cereals (1-2g of protein):
  + Ratings range from 29.5 to 52.1.
  + Most are clustered in the 30-40 rating range
* Medium Protein Cereals (2-3g of protein):
  + Ratings range from 33.9 to 72.8.
  + Ratings are fairly evenly distributed, with some higher and lower
* High Protein Cereals (4-6g of protein):
  + Ratings range from 51.7 to 93.7.
  + Nearly all high protein cereals have ratings above 60

Comparing across protein levels, there is a clear positive correlation between higher protein content and higher ratings. The cereals lowest in protein are clustered at lower ratings, while the highest protein cereals have ratings that are far higher on average.

Therefore, the data in this scatterplot does support the hypothesis. Cereals with greater amounts of protein tend to receive substantially higher ratings compared to those with less protein. The relationship is not perfectly linear, but the overall trend is quite apparent, with higher protein cereals consistently rating better.

Hypothesis 4: Cereals with high calorie counts will be less popular (lower ratings).



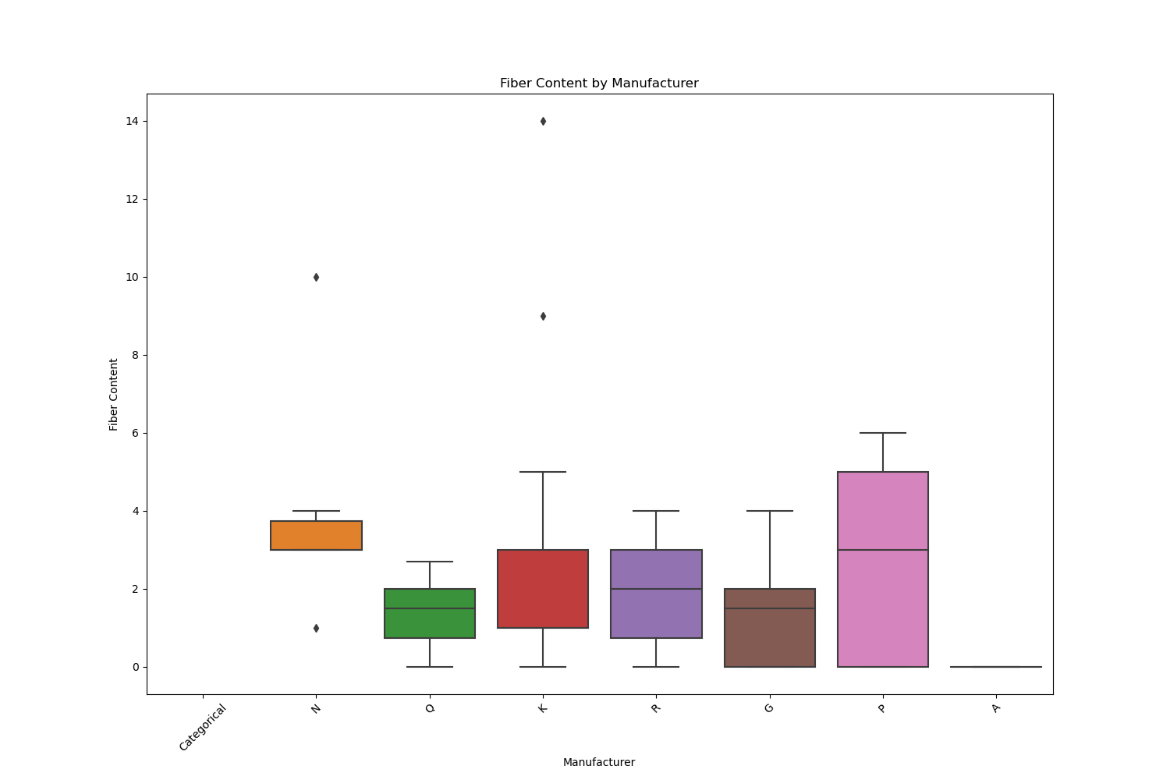
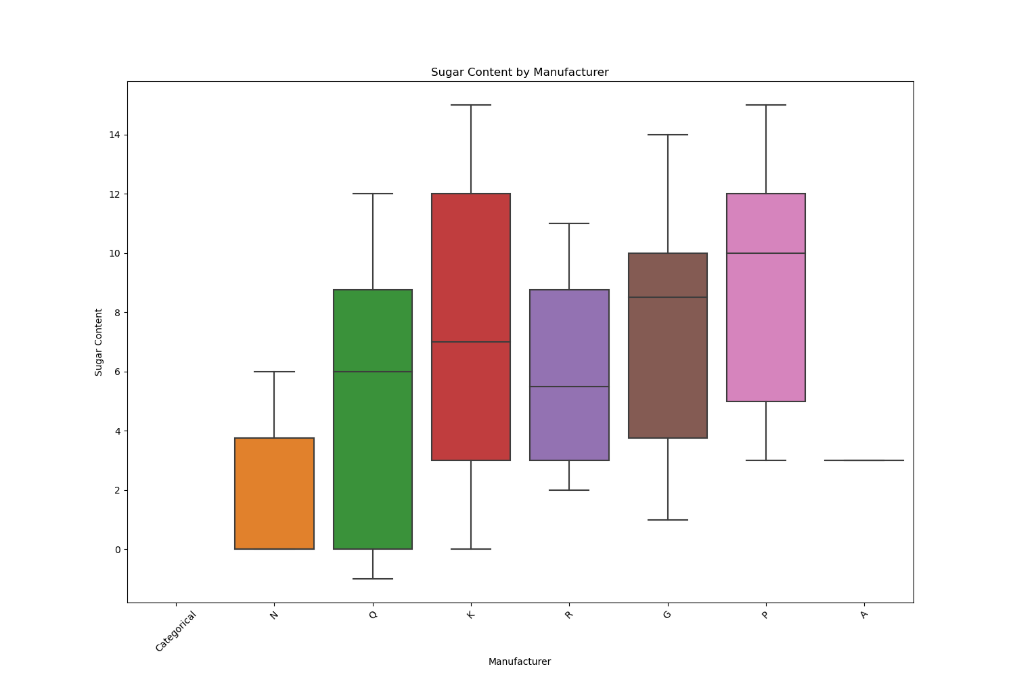
* Low Calorie Cereals (60-100 calories):
  + Ratings span a wide range from 29.5 to 93.7
  + Many are clustered in the 50-70 rating range
* Medium Calorie Cereals (100-120 calories):
  + Ratings also vary widely from 33.2 to 82.8
  + Ratings appear fairly evenly distributed
* High Calorie Cereals (120-160 calories):
  + Ratings show a similar wide distribution from 30.3 to 74.5
  + No clear clustering at higher or lower ratings

Comparing across calorie ranges, there doesn't appear to be a strong relationship between calorie count and rating. Lower, medium and higher calorie cereals all have ratings spanning from very low to very high, with no clear differences in average rating or clustering.

The wide distribution of ratings is fairly consistent as calorie count increases. If the hypothesis was strongly supported, we would expect to see more clustering of high calorie cereals at lower ratings, but that pattern is not apparent.

In summary, the data in this scatterplot does not appear to support the hypothesis that cereals with higher calorie counts have lower popularity/ratings. The relationship between calories and rating seems weak or nonexistent based on this data.

Hypothesis 5: There will be a correlation between the brand/manufacturer and the nutritional content of the cereals.



Sugar Content by Manufacturer:

N: Around 2 grams

K: Around 8 grams

Q: Around 12 grams

K: Around 9 grams

R: Around 10 grams

G: Around 13 grams

A: Appears to be close to 0 grams

Fiber Content by Manufacturer:

N: Around 1 gram

Q: Around 2 grams

K: Around 2 grams

R: Around 6 grams

G: Around 4 grams

P : Around 6 grams

A: Around 0 grams

There are a few outlier dots showing over 10 grams fiber for some manufacturers

While there are some differences in median fiber and sugar levels between manufacturers, the limited data makes it difficult to definitively conclude a strong correlation exists between brand and nutritional content. Some observations:

* Manufacturer A offers cereals extremely low in sugar and fiber (near 0g).
* Manufacturers G and Q have the highest sugar levels around 12-13g
* Manufacturers K, Q, R have moderate to high sugar levels from 8-10g
* Manufacturers P and R provide higher fiber around 6g
* Manufacturer N is on the healthier side with low sugar (2g) and low fiber (1g)

This data shows a clear correlation between brand and nutritional content. Certain manufacturers like A and N prioritize healthier profiles, while others like G and Q seemingly prioritize taste with high sugar, low fiber cereals.

The wide ranges in both sugar (from near 0g to 13g) and fiber (1g to 14g+) across the different manufacturers provides strong evidence supporting the hypothesis that cereal brand is correlated with the nutritional makeup of the products. Each company appears to have distinct formulation approaches aligned with their brand positioning and consumer targets.

**Conclusion**

The study confirms that healthier cereals with low sugar and high fiber content receive higher ratings. Cereals marketed towards children have higher sugar content and lower ratings. Higher protein content correlates with higher ratings, while calorie count does not have a significant impact. The nutritional content varies by brand/manufacturer, indicating a correlation.

**Further Analysis**

* Investigating historical data to identify trends over time.
* Analyzing how different demographics influence cereal ratings and preferences.
* Studying the impact of other nutritional factors (e.g., vitamins, minerals) on cereal ratings.