

Are all cereals created equal?

[team 18]

Ayon Biswas

Nitesh Gundavarapu

Zongcheng Wang

Yaocong Hu

Aniket Tiwari

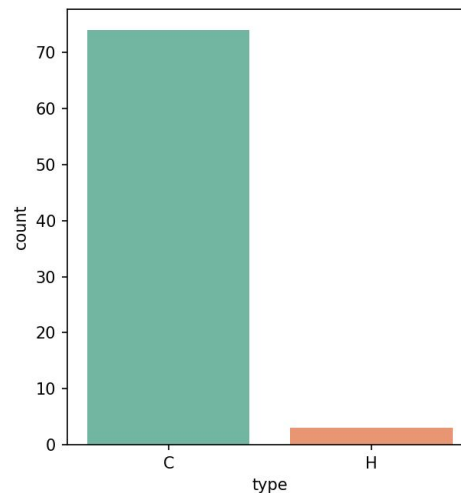
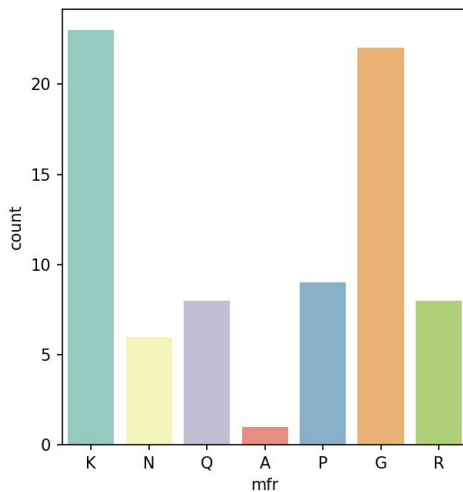
Overview

- Who are the major cereal manufacturers and what are their offerings? What are their ratings?
- What are their macro and micro nutrient stats?
- Which manufacturer covers a wide variety of products and which manufacturer pertains to a niche market?
- Are the consumer ratings based on health or taste?
- What are the factors that made each cereal different from the others
- Given the nutrient value for a cereal and name of the manufacturer, can we predict the rating of the cereal?

About Dataset

16 columns - calorie value, manufacturer, various nutrient contents and customer ratings.

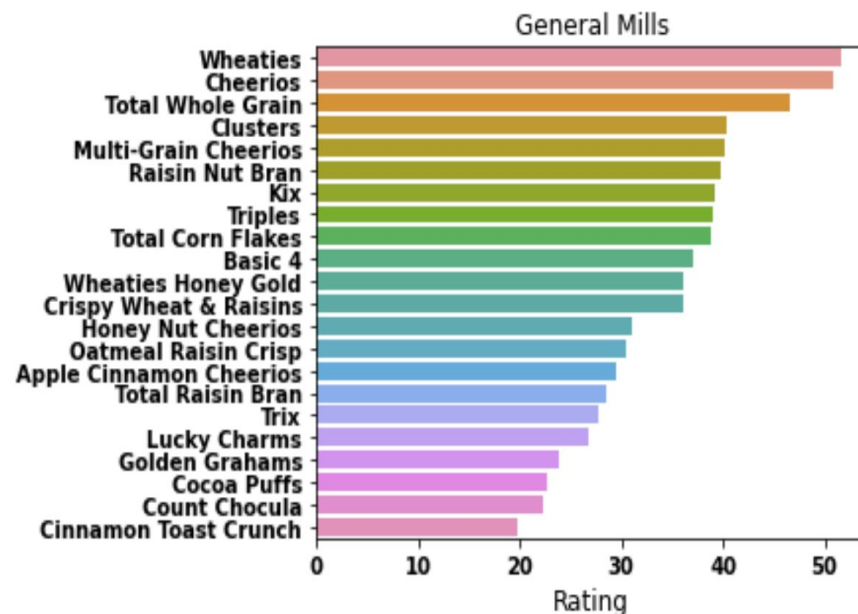
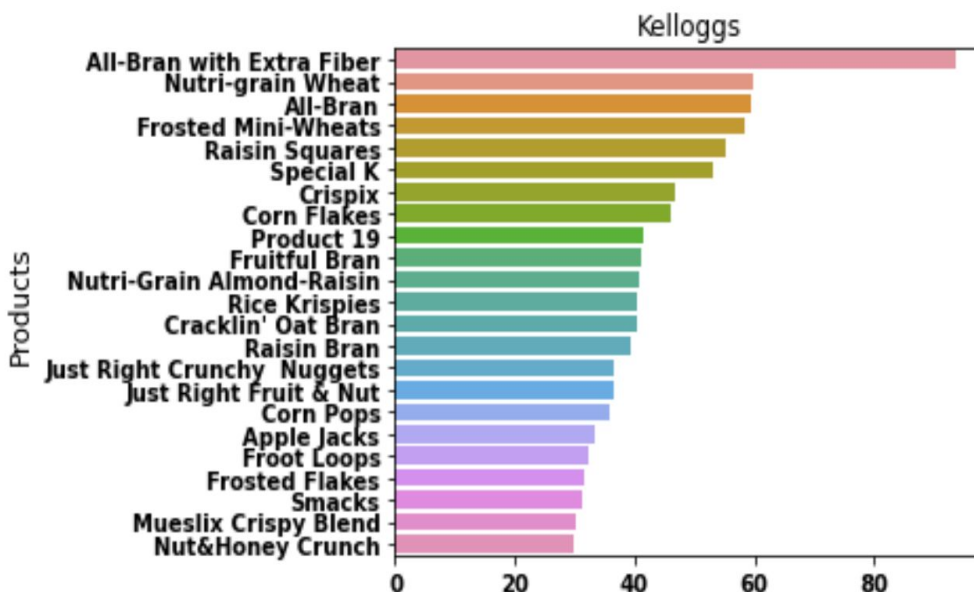
	name	mfr	type	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	weight	cups	rating
0	100% Bran	N	C	70	4	1	130	10.0	5.0	6	280	25	3	1.0	0.33	68.402973
1	100% Natural Bran	Q	C	120	3	5	15	2.0	8.0	8	135	0	3	1.0	1.00	33.983679
2	All-Bran	K	C	70	4	1	260	9.0	7.0	5	320	25	3	1.0	0.33	59.425505
3	All-Bran with Extra Fiber	K	C	50	4	0	140	14.0	8.0	0	330	25	3	1.0	0.50	93.704912
4	Almond Delight	R	C	110	2	2	200	1.0	14.0	8	-1	25	3	1.0	0.75	34.384843



- 96% cereals - hot and 4% - Cold
- 28% - General Mills and 29% - Kelloggs

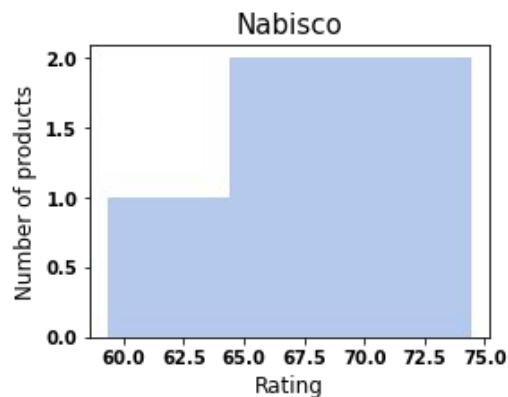
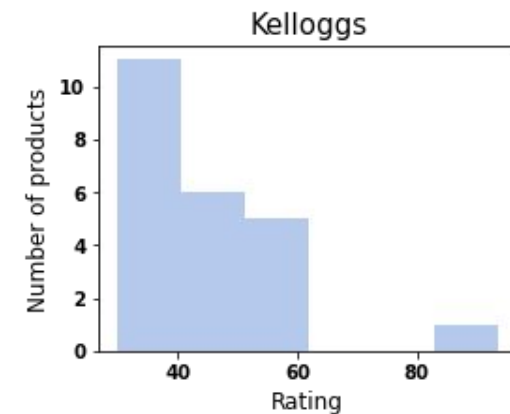
Cereals and their ratings

- 7 manufacturers with a total of 80 cereals
- The cereals are rated 1-100 by users
- The highest rated cereal is *All-Bran with Extra Fiber*



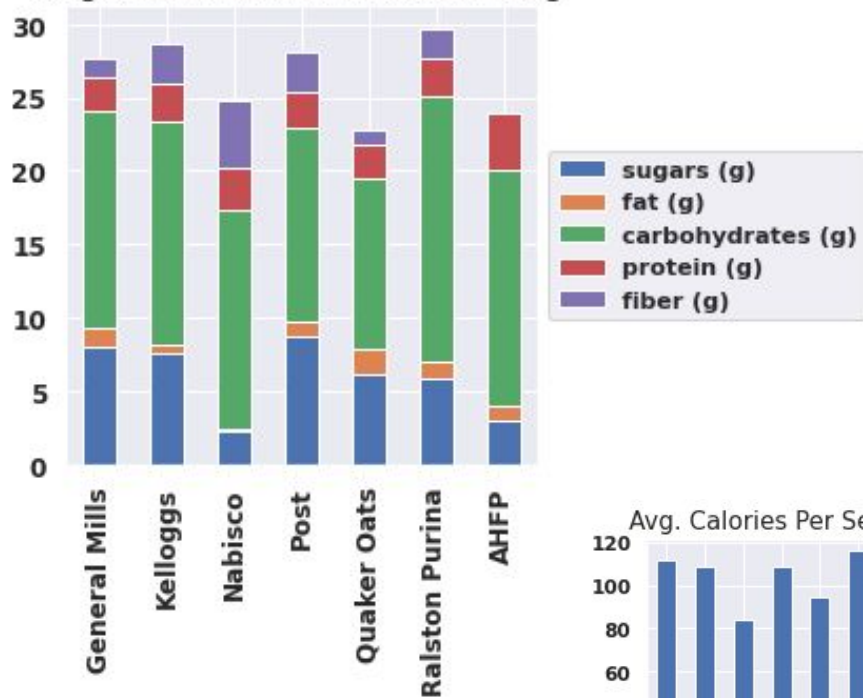
Manufacturer wise rating statistics

- Although Kelloggs makes the best product, its products have a high variance in the rating as compared to Nabisco or AHFP

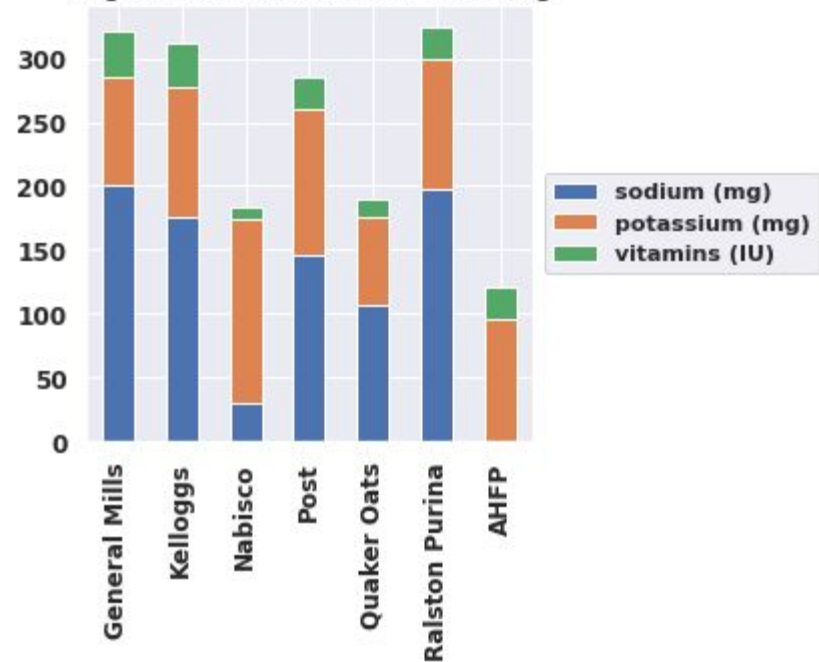


Manufacturer wise nutrient statistics

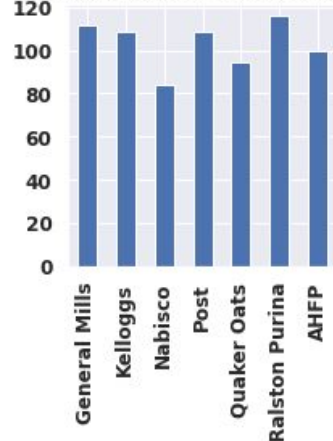
Avg. Macro Nutrients Per Serving



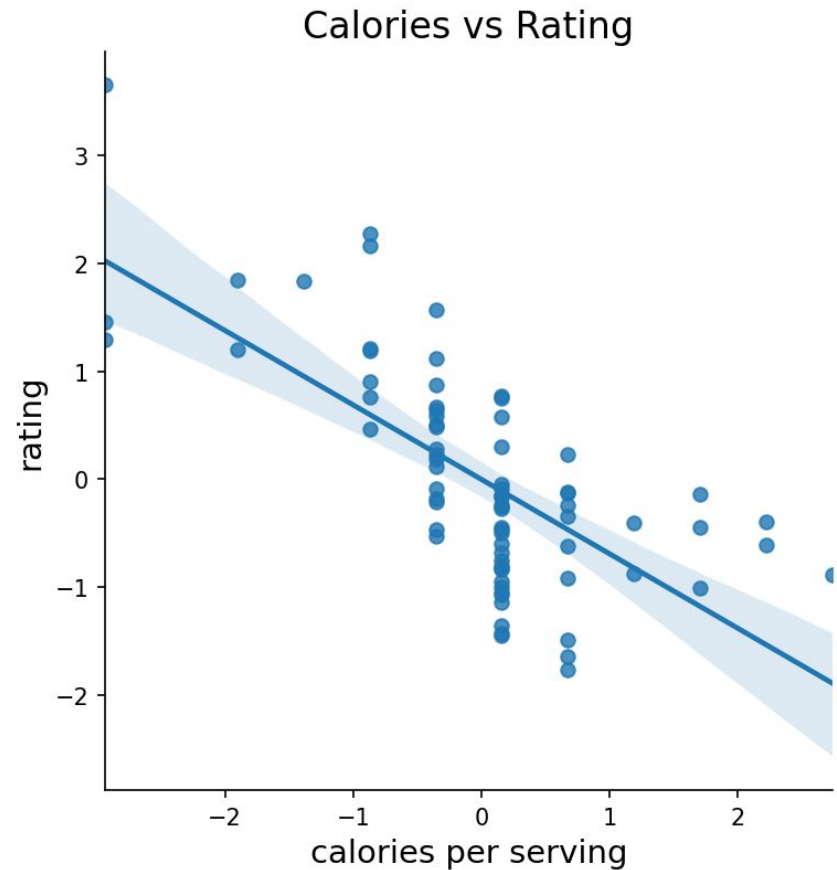
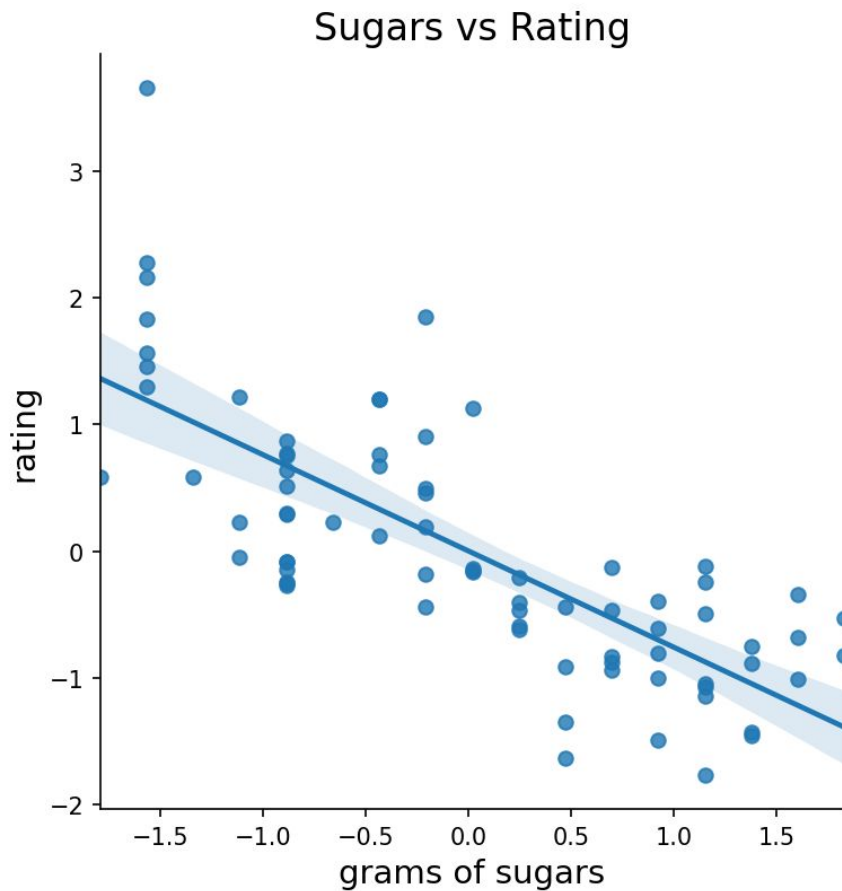
Avg. Micro Nutrients Per Serving



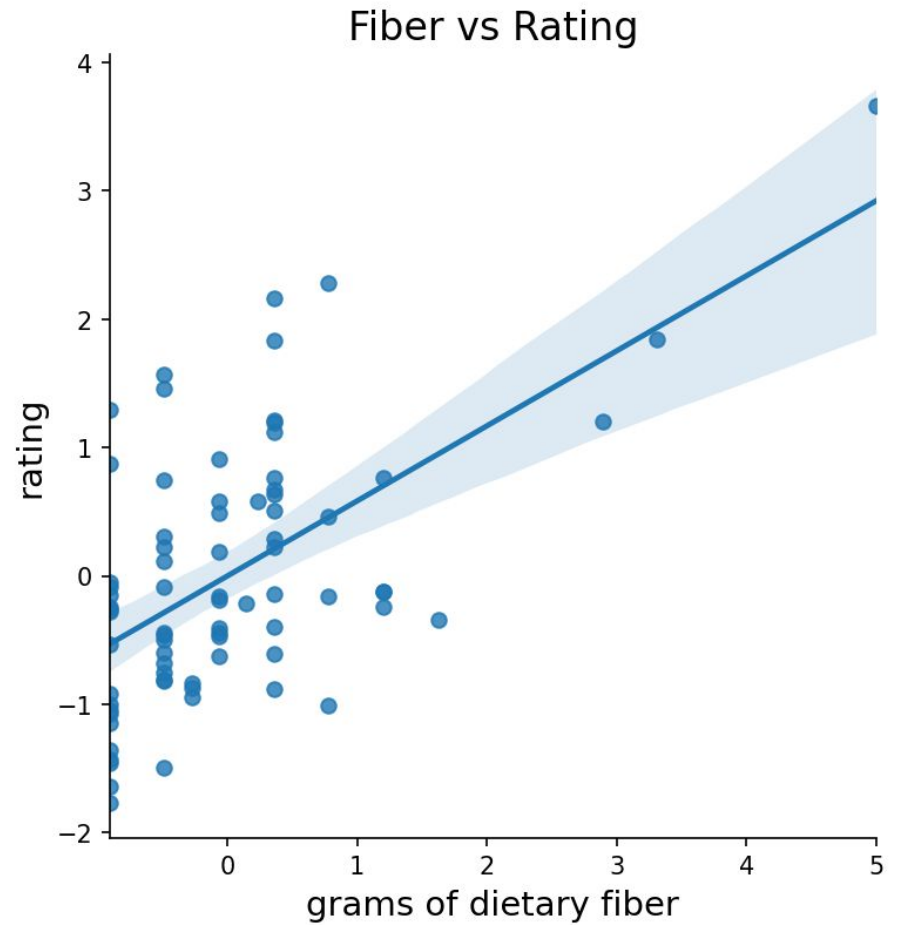
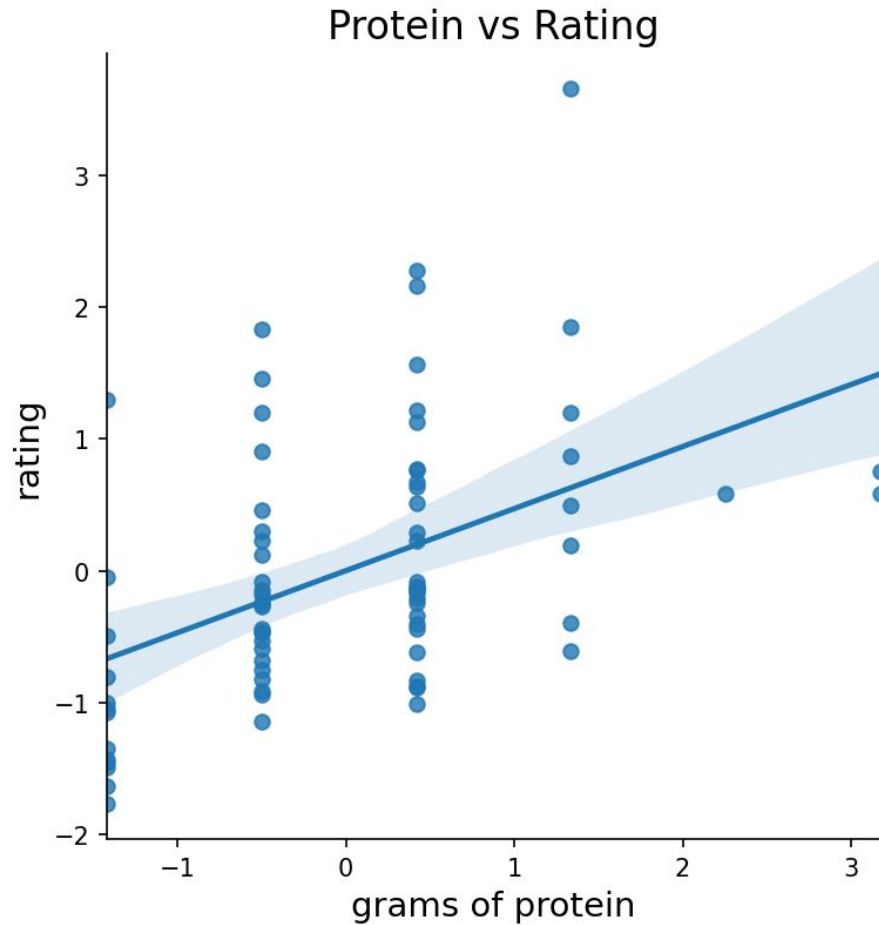
Avg. Calories Per Serving



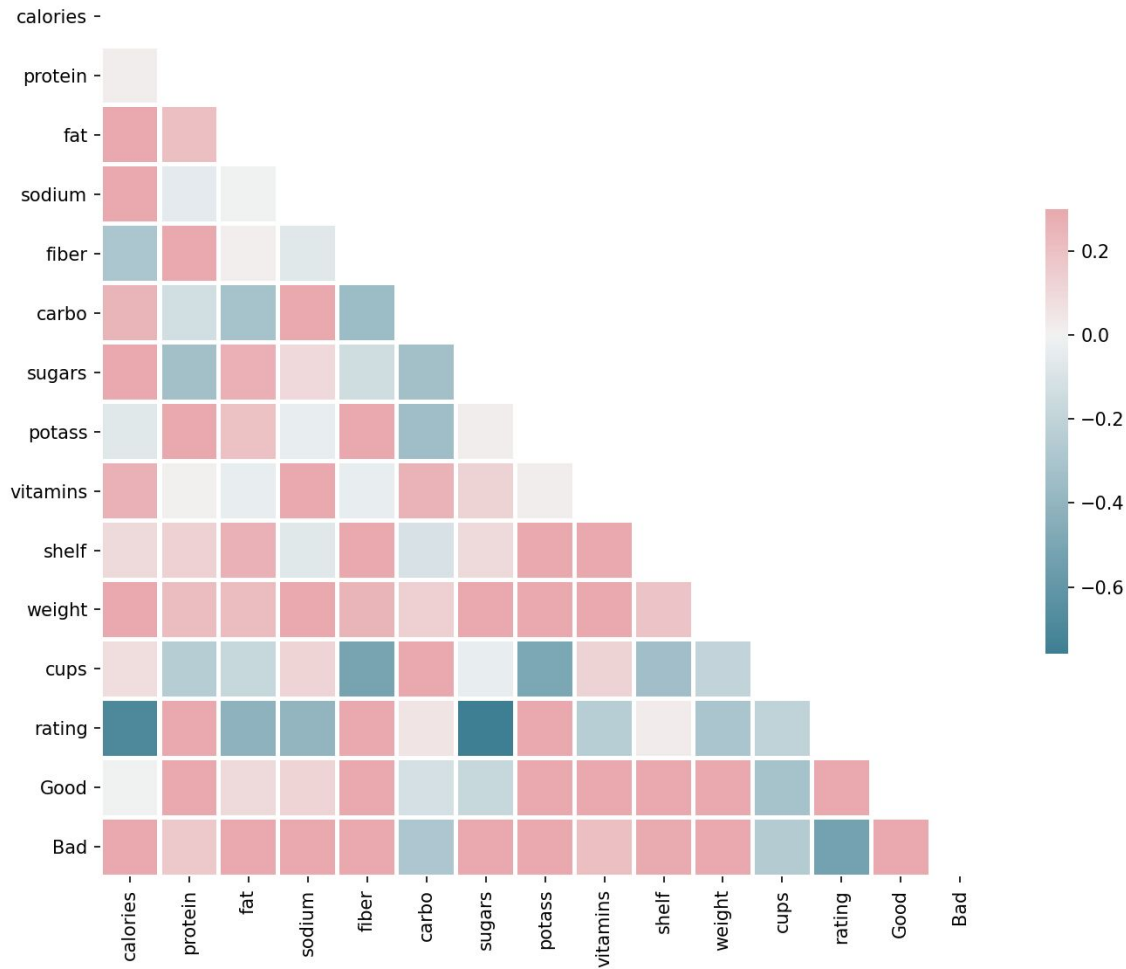
Do customers prefer sugar or calories ?



Do customers prefer proteins or fibers ?

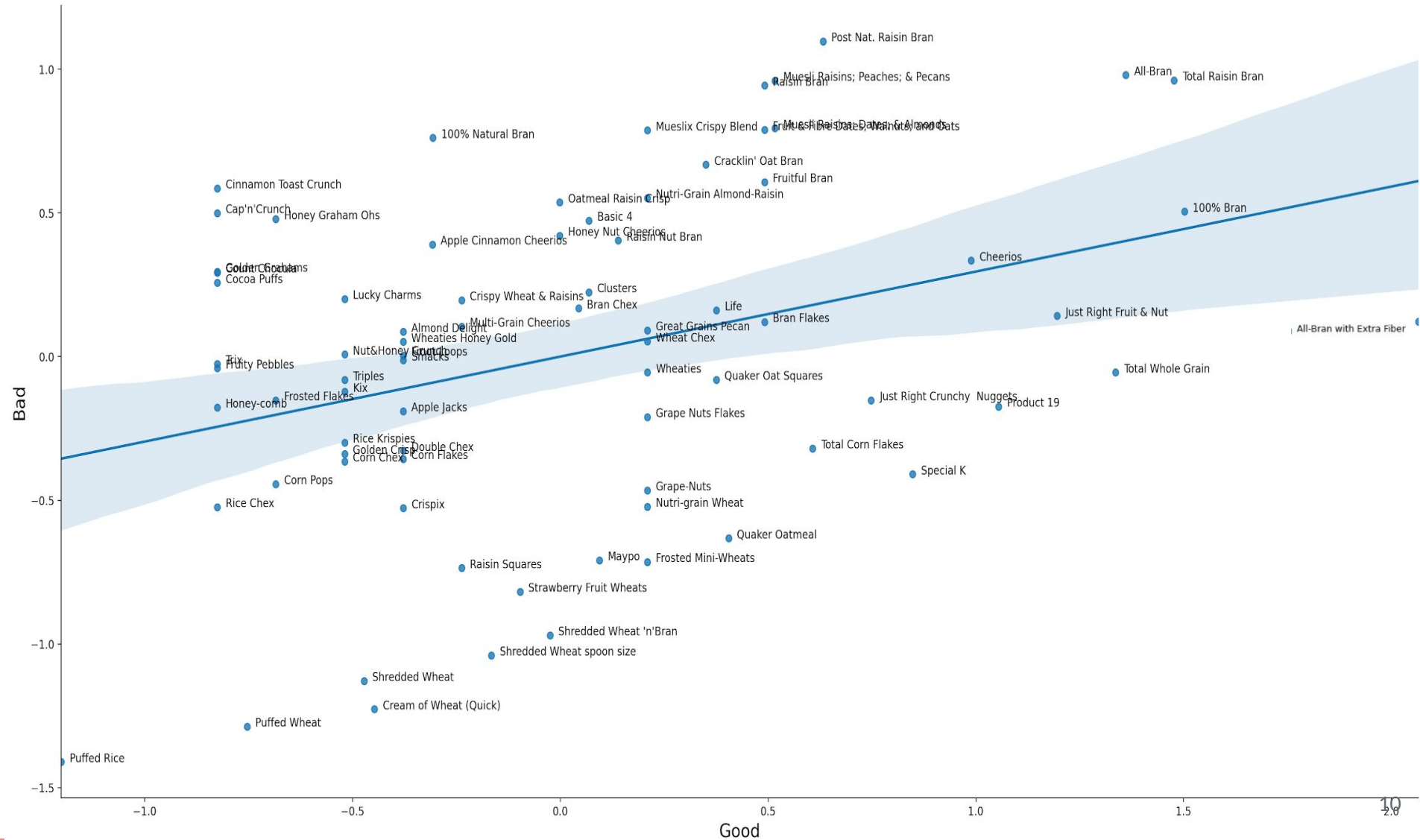


How features correlate ?



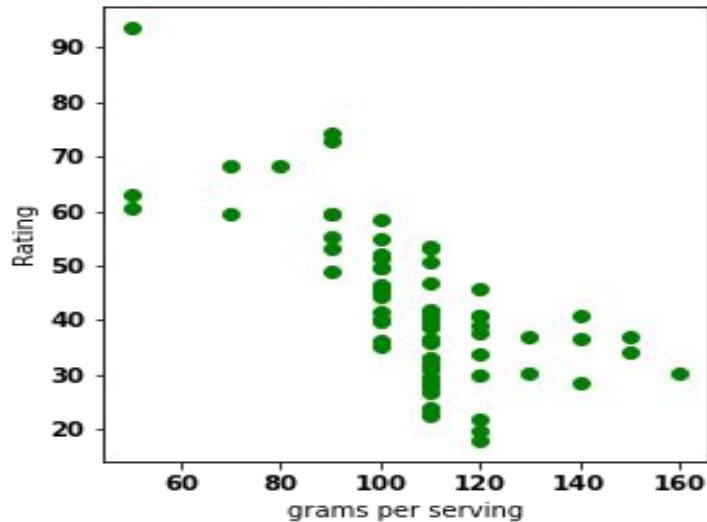
- strongest positive correlation is between fiber and potassium
- Calories have positive correlations with weight, sugars and fat.
- Strong negative correlations between sugars and rating, and calories and rating.

Healthy vs unhealthy

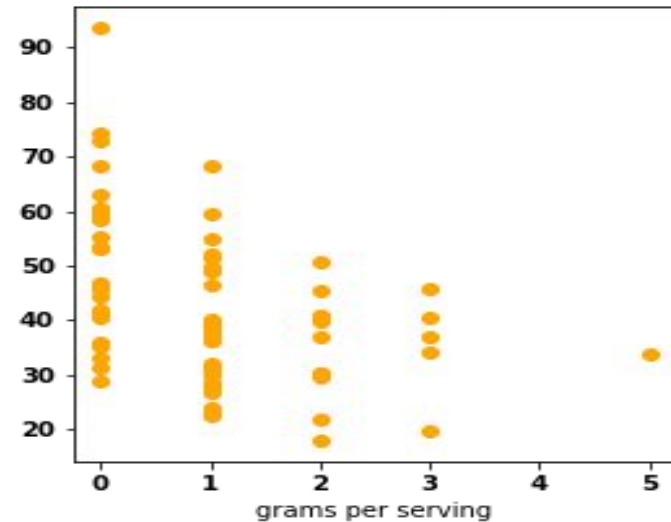


Can macronutrients and calories affect ratings?

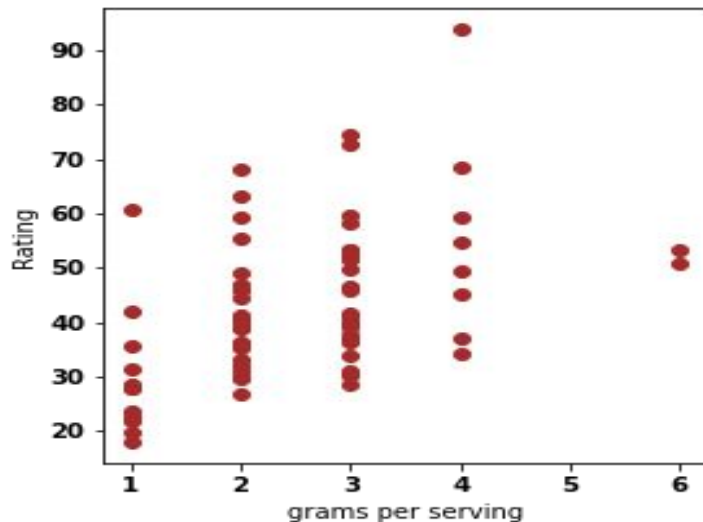
Calories



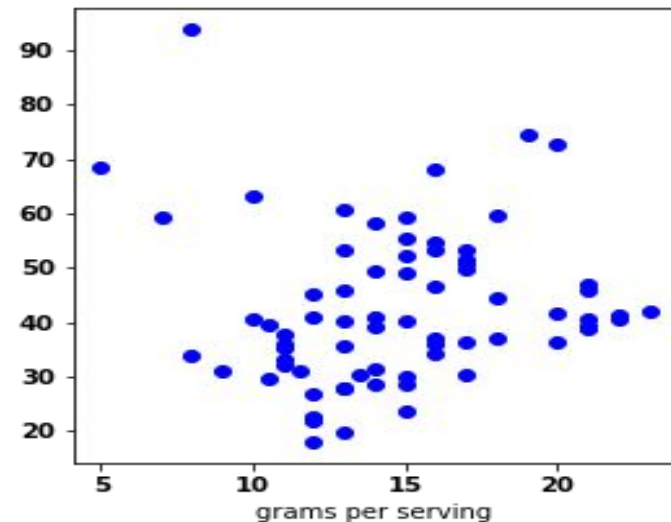
Fat



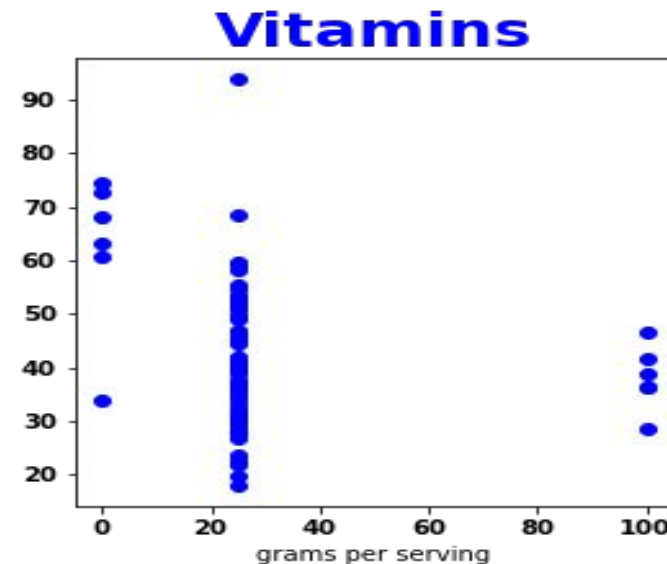
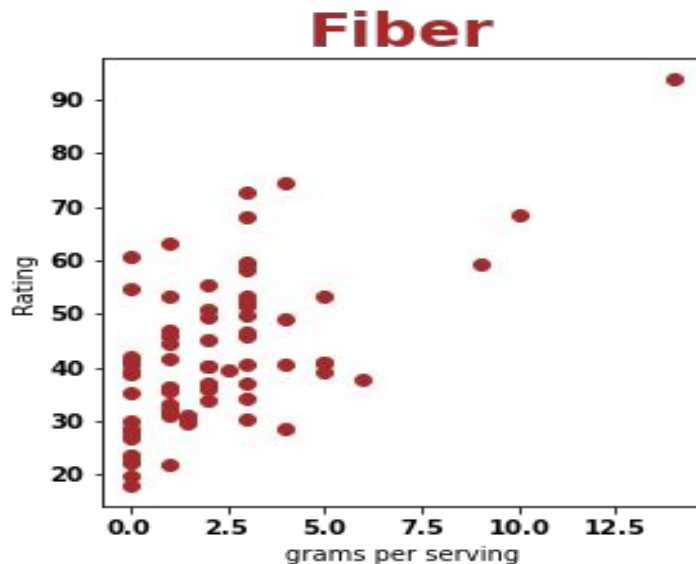
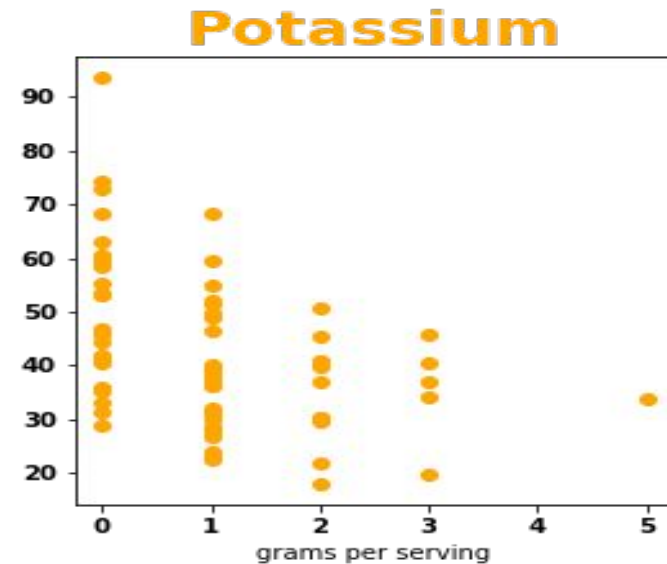
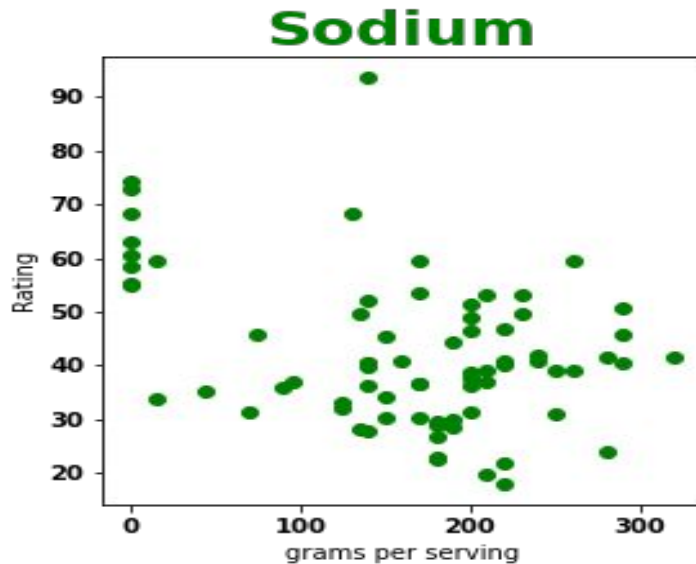
Protein



Carbs



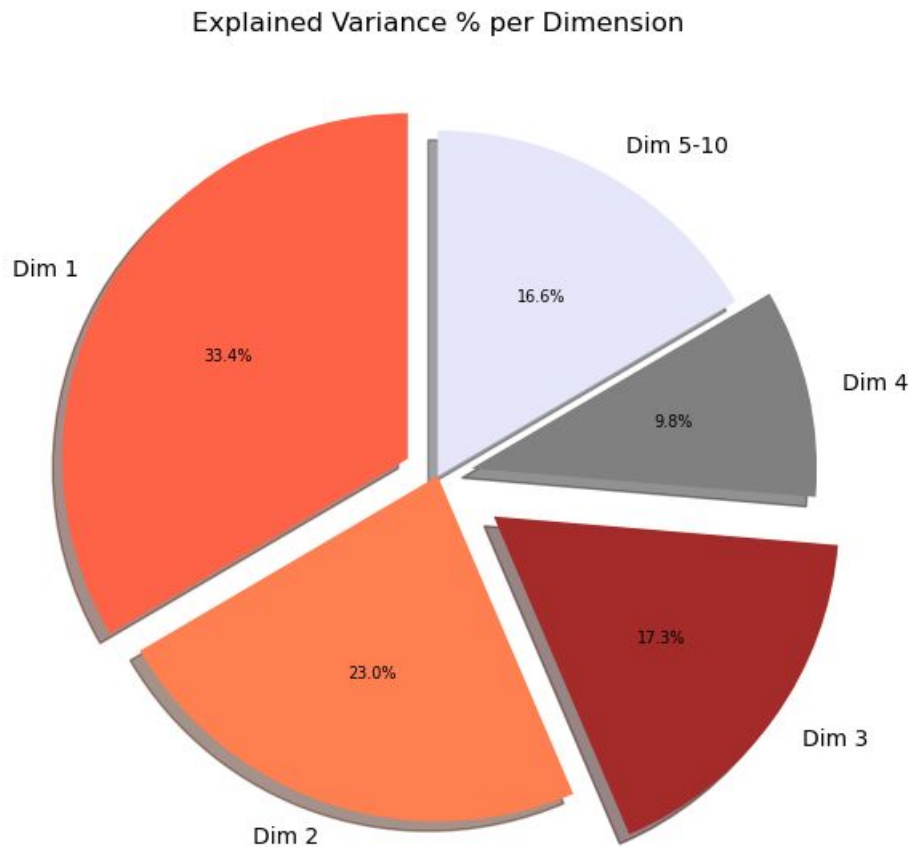
Can micronutrients affect ratings?



Factors making each cereal different

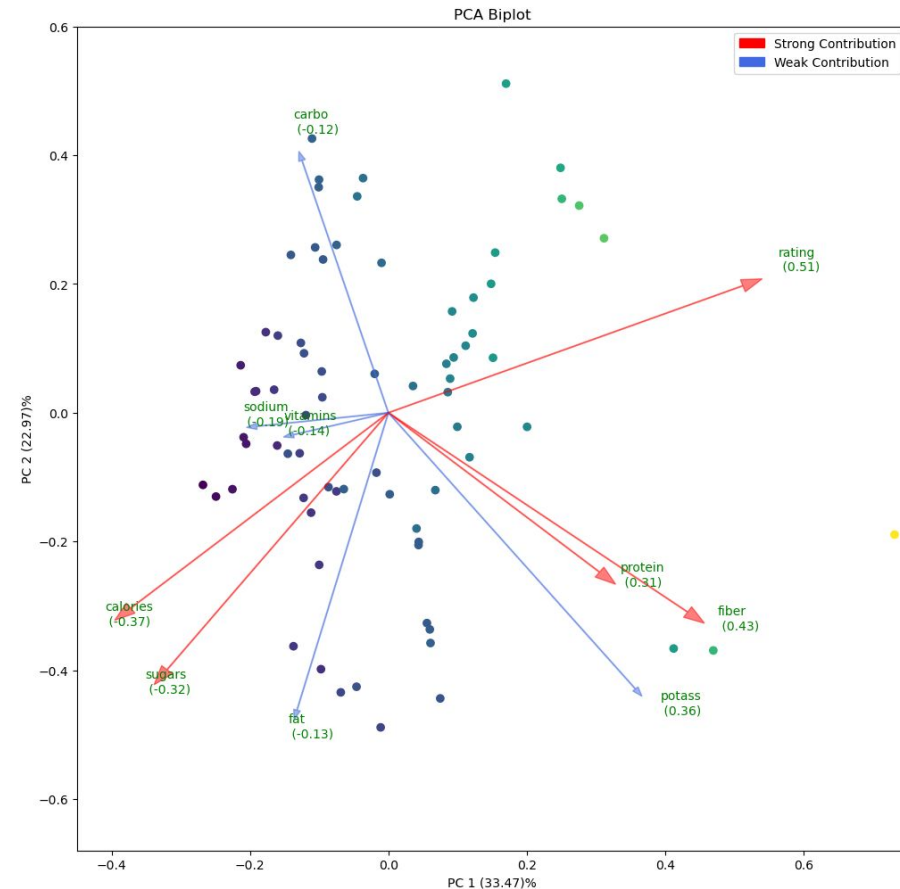
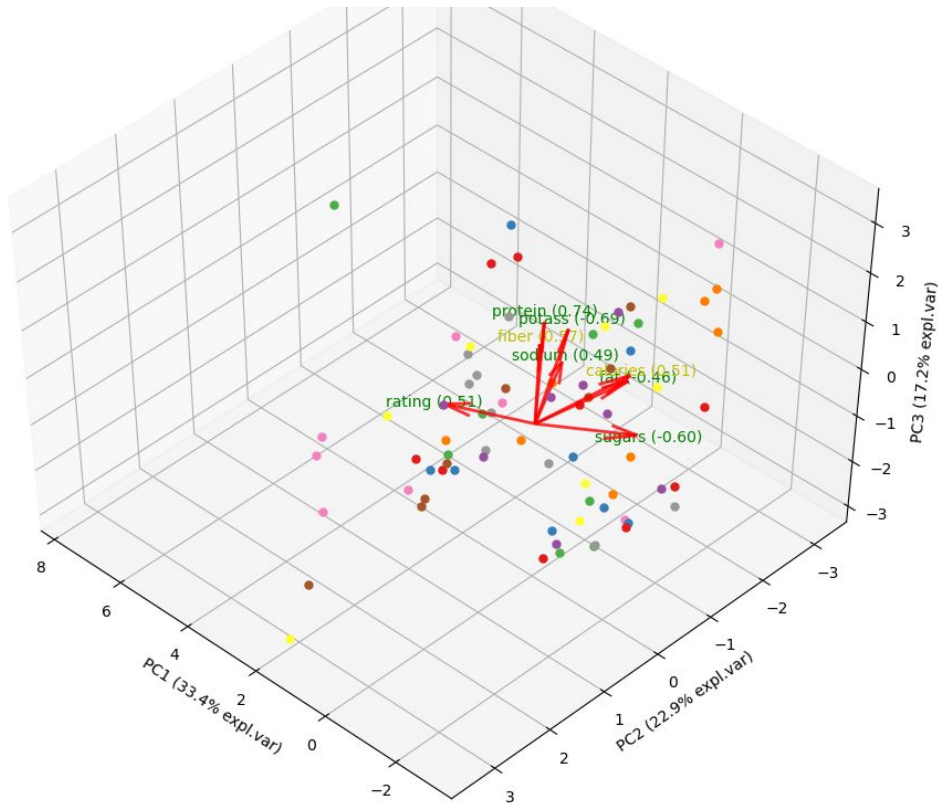
Objective: Find features that contributes to the variance the most

Solution: PCA



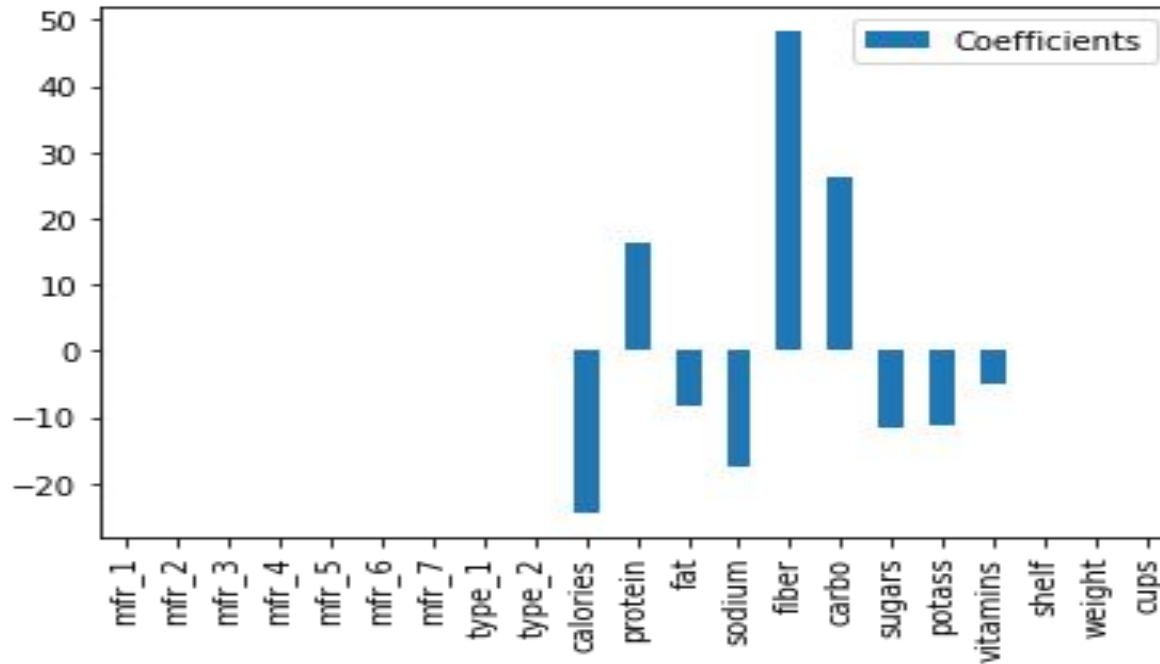
- 6 Principal Components Explains 95 % of the variance
- Majority of the variance can be explained using first three Principal Components

Identifying Top Features



- Importance of each feature is reflected by the magnitude of the corresponding values in the eigenvectors (higher magnitude - higher importance)
- Rating, Fiber, Protein, Calories and Sugars have the strongest contribution to overall variance in the dataset

Predicting the rating



- We tried the Linear regression algorithm with accuracy of 99%
- The following are the coefficients for the linear regression

Conclusion

- Nabisco is healthier than Kelloggs and has the highest average rating
- Customers prefer health over taste
- Macronutrients effect has more influences on rating than micronutrients
- PCA shows that Fiber, Protein, Calories and Sugars are the key nutrients to explain why each cereal is different from the others
- The coefficients of the linear regression algorithm sustain the previous results.

Questions?

Thank You