Presented By – Nidhi Sehgal, DAB July 2025

Breakfast cereal database from Kaggle.com https://www.kaggle.com/datasets/sahirmaharajj/breakfast-cereals

For the raw data from Kaggle.com

**Dataset Documentation**

|  |  |  |  |
| --- | --- | --- | --- |
| Column Name | Data Type | Description | Potential Business Value |
| name | String | Name of the cereal product | Used to identify and group products for brand analysis or comparisons |
| Mfr = Manufacturer | String | Manufacturer code  A = American Home Food Products; G = General Mills; K = Kelloggs; N = Nabisco; P= POST; Q= Ouaker; R=Ralston | Helps analyze performance by brand and market share |
| type | String | C = Cold cereal, H = Hot cereal | Useful for segmenting products and understanding customer preferences |
| calories | Integer | Number of calories per serving | Important for health-conscious marketing and product positioning |
| protein | Float/Integer | Grams of protein per serving | Key for targeting high-protein diet consumers |
| fat | Float/Integer | Grams of fat per serving | Identify high fat products |
| sodium | Float/Integer | Milligrams of sodium per serving | Important for health ratings, low-sodium product promotion and for heart healthy product |
| fiber | Float/Integer | Grams of dietary fiber per serving | Used to highlight “high-fiber” cereals |
| carbo | Float/Integer | Grams of carbohydrates per serving | Useful for diabetic-friendly or low-carb product development |
| sugars | Float/Integer | Grams of sugar per serving | Critical for compliance with sugar regulations and for heart healthy product |
| potass | Float/Integer | Milligrams of potassium per serving | Can be used to market heart-healthy products |
| vitamins | Integer | % of FDA recommended daily vitamins (0, 25, or 100) | Helps position fortified cereals in the market |
| shelf | Integer | Store shelf display level (1 = bottom, 3 = top) |  |
| weight | Float | Weight of cereal box in ounces |  |
| cups | Float | Serving size in cups | Supports portion control recommendations |
| rating | Float | Nutritional rating score of the cereal | Can be used to rank products and recommend the healthiest options |

\*American Home Food Products was a manufacturer of hot cereals, but the company no longer exists under that name. now it is Hometown Food Company in 2025

**Data Assessment and Cleaning**

* I used both SQL and Excel to clean the data.
* Most of the excel tables are made using SQL codes.
* Duplicates were checked using DISTINCT function in SQL and conditional formatting in Excel
* There were 4 Missing/ NULL values found, did not removed from the data, rather replaced with corrected values
* In carb table- quaker oatmeal, NULL changed to 27 gm of carbohydrate (checked from internet) (But actually did not use that table for analysis)
* The other 3 is changed to 0

DATA ANALYSIS:

* Finding which type of Breakfast cereal (Hot or Cold) dominate the market
* From the data it is clear that the “Cold” type of cereal dominates the market (96%) and people prefer a cold breakfast to begin their day.

Pivot Table analysis-

|  |  |
| --- | --- |
| **Row Labels** | **Sum of cereal\_count** |
| Cold | 74 |
| Hot | 3 |
| **Grand Total** | **77** |

|  |  |
| --- | --- |
| **Row Labels** | **Sum of cereal\_count** |
| Cold | 96.10% |
| Hot | 3.90% |
| **Grand Total** | **100.00%** |

* This Pivot table answers that Consumers prefer to have cold cereals than hot ones.
* Also, there is opportunity for manufacturers to launch more hot cereals in the market as there are very less competeitors.

2. Average Count and Rating of cereals by Manufacturer

|  |  |
| --- | --- |
| **Row Labels** | **Sum of cereal\_count** |
| General Mills | 28.57% |
| Hometown Food Company | 1.30% |
| Kellogg's | 29.87% |
| Nabisco | 7.79% |
| Post | 11.69% |
| Quaker Oats | 10.39% |
| Ralston | 10.39% |
| **Grand Total** | **100.00%** |

* The Data shows that Kellog’s and General Mills have maximum number of products/cereals in the market (about 60% market coverage combined), whereas Hometown Food company manufactures only one cereal (Maypo).

|  |  |
| --- | --- |
| **Row Labels** | **Sum of avg\_rating** |
| General Mills | 10.53% |
| Hometown Food Company | 16.75% |
| Kellogg's | 13.45% |
| Nabisco | 20.75% |
| Post | 12.73% |
| Quaker Oats | 13.10% |
| Ralston | 12.68% |
| **Grand Total** | **100.00%** |

* **Highest Rated:** " Nabisco and Hometown brands cereals have the highest average rating, followed by Kellogg’s, making up **about 50% of all top-rated cereals** in the dataset "
* HomeTown FC produces only 1 cereal and it is highly rated. (One good Product that generates revenue)
* **Lowest Rated:** "General Mills cereals have lower ratings compared to competitors, suggesting room for improvement."
* **Middle Range:** "Post, Quaker and Ralston brands fall in the middle, with a mix of high- and low-rated cereals."

3. Compare Nutrients between different brands(manufacturers)and which brand has the lowest sugar on average or the highest sodium

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Row Labels** | **Sum of rounded\_calories** | **Sum of rounded\_sugars** | **Sum of rounded\_fiber** | **Sum of rounded\_sodium** |
| General Mills | 111.4 | 8 | 1.3 | 200.5 |
| Hometown FC | 100 | 3 | 0 | 0 |
| Kellogg's | 108.7 | 7.6 | 2.8 | 174.8 |
| Nabisco | 86.7 | 1.9 | 4 | 37.5 |
| Post | 108.9 | 8.8 | 2.8 | 146.2 |
| Quaker Oats | 95 | 6.2 | 1.4 | 92.5 |
| Ralston | 115 | 6.2 | 1.9 | 198.2 |
| **Grand Total** | **725.7** | **41.7** | **14.2** | **849.7** |

Healthy criteria followed–

1. **Sugar <10g** , minimal or no sugar is ideal
2. **Sodium <200mg /100 g,** minimal or no sodium
3. **Fiber** - **3g or more per serving** is generally considered a “good source of fiber,”
4. **Protein >10 g** or higher

* From the data we can infer that for **Sugar** levels, all the brands fall under the healthy criteria (<10 grams), with Nabisco and Hometown Food Company being on the minimum side. Post and General Mills have sweeter products than the other brands.
* For **Sodium** content – General Mills products have the highest sodium content, making it unhealthy for those with higher blood Pressure values.
* For **Fiber** content – Nabisco had added the maximum fiber in their products (4gm), making their consumer feel much fuller than others. This can help in controlling the portion size of their meal.
* For **Protein** content, none of the brands can provide the required amount of Protein(>10g). So there is a need to add protein rich ingredients to these cereals like Greek yogurt, cottage cheese, protein powder, or a splash of milk, nuts and seeds like [chia seeds](https://www.google.com/search?sca_esv=9349b9a2334e1130&rlz=1C1JZAP_enUS933US933&sxsrf=AE3TifNc9148FqvZzIhqspvr8yf4Owr97w%3A1758708199790&q=chia+seeds&sa=X&ved=2ahUKEwj4hozMkvGPAxXiE1kFHQl2DzEQxccNegQINBAB&mstk=AUtExfBu16KYaWMZ2I8IG4a4qQK7N0Rgpg22nvLJZN9K1gdOvAdt2A7DRB7HCS_B9NMTtKtdnfIbNYkPOOgJvDEM7sJwzUuIrrZn1V0MmyfuGC1z17xOp94HNQ9M3F3iGPXcuEJueX-9mksG5L9vkR_nkLlpaDuDjrqF6iQNUTjpXif8JM3LCc-rKlun7gZwTYfA8zc5wIecB3fh4vLFx_Suxf_QjYRZJUb62M1M8RclIW5k9RYLxIfdexjAen8uQRGxRYlWJZQq0-TrzyjhD1nScZdo&csui=3) or [flaxseeds](https://www.google.com/search?sca_esv=9349b9a2334e1130&rlz=1C1JZAP_enUS933US933&sxsrf=AE3TifNc9148FqvZzIhqspvr8yf4Owr97w%3A1758708199790&q=flaxseeds&sa=X&ved=2ahUKEwj4hozMkvGPAxXiE1kFHQl2DzEQxccNegQINBAC&mstk=AUtExfBu16KYaWMZ2I8IG4a4qQK7N0Rgpg22nvLJZN9K1gdOvAdt2A7DRB7HCS_B9NMTtKtdnfIbNYkPOOgJvDEM7sJwzUuIrrZn1V0MmyfuGC1z17xOp94HNQ9M3F3iGPXcuEJueX-9mksG5L9vkR_nkLlpaDuDjrqF6iQNUTjpXif8JM3LCc-rKlun7gZwTYfA8zc5wIecB3fh4vLFx_Suxf_QjYRZJUb62M1M8RclIW5k9RYLxIfdexjAen8uQRGxRYlWJZQq0-TrzyjhD1nScZdo&csui=3), or a spoonful of nut butter. For a savory twist, try crumbled [paneer](https://www.google.com/search?sca_esv=9349b9a2334e1130&rlz=1C1JZAP_enUS933US933&sxsrf=AE3TifNc9148FqvZzIhqspvr8yf4Owr97w%3A1758708199790&q=paneer&sa=X&ved=2ahUKEwj4hozMkvGPAxXiE1kFHQl2DzEQxccNegQINxAB&mstk=AUtExfBu16KYaWMZ2I8IG4a4qQK7N0Rgpg22nvLJZN9K1gdOvAdt2A7DRB7HCS_B9NMTtKtdnfIbNYkPOOgJvDEM7sJwzUuIrrZn1V0MmyfuGC1z17xOp94HNQ9M3F3iGPXcuEJueX-9mksG5L9vkR_nkLlpaDuDjrqF6iQNUTjpXif8JM3LCc-rKlun7gZwTYfA8zc5wIecB3fh4vLFx_Suxf_QjYRZJUb62M1M8RclIW5k9RYLxIfdexjAen8uQRGxRYlWJZQq0-TrzyjhD1nScZdo&csui=3) (Indian cottage cheese) or sprouted [moong](https://www.google.com/search?sca_esv=9349b9a2334e1130&rlz=1C1JZAP_enUS933US933&sxsrf=AE3TifNc9148FqvZzIhqspvr8yf4Owr97w%3A1758708199790&q=moong&sa=X&ved=2ahUKEwj4hozMkvGPAxXiE1kFHQl2DzEQxccNegQINxAC&mstk=AUtExfBu16KYaWMZ2I8IG4a4qQK7N0Rgpg22nvLJZN9K1gdOvAdt2A7DRB7HCS_B9NMTtKtdnfIbNYkPOOgJvDEM7sJwzUuIrrZn1V0MmyfuGC1z17xOp94HNQ9M3F3iGPXcuEJueX-9mksG5L9vkR_nkLlpaDuDjrqF6iQNUTjpXif8JM3LCc-rKlun7gZwTYfA8zc5wIecB3fh4vLFx_Suxf_QjYRZJUb62M1M8RclIW5k9RYLxIfdexjAen8uQRGxRYlWJZQq0-TrzyjhD1nScZdo&csui=3).

4. Top 10 Heart healthy Cereals

* The data shows that 3 products of Quaker oats (Quaker Oatmeal Puffed rice and Puffed Wheat) and 3 products of Nabisco brand are considered excellent for heart health-conscious consumers.
* Hometown FC ‘s only product Maypo is another competitor.

--6. Measuring Healthiness of a cereal using a formula (fiber + protein) - (sugars/2.0 + fat) AS Healthy Score) ranking from the healthiest to least healthy option

* The data shows that according to the given formula, 3 varieties/ products of Nabisco are top Healthy options with zero sugars and zero sodium.
* Also, Nabisco has products that are very low on the healthy scores chart, catering to the needs of different consumers.
* Similarly, Quaker oats are another such brand.

7. Ranking by healthy score and identify which cereals are both healthy and popular

**My Hypothesis- healthy score cereals are more popular**

* Analyzing the dataset reveals distinct patterns: cereals with **high Healthy Scores and high ratings** represent the ideal combination of nutrition and taste, making them highly recommended for regular consumption.
* Cereals with **high Healthy Scores but lower ratings** are nutritious but may not appeal to all consumers due to taste preferences.
* Conversely, cereals with **low Healthy Scores but high ratings** are popular for taste but less healthy, suggesting occasional consumption.
* Finally, cereals with both **low Healthy Scores and low ratings** are neither nutritious nor favored by consumers and are generally the least desirable choice.
* Visualizing cereals on a scatter plot of Healthy Score versus Rating, reveals that there is no specific pattern and the cereals are all scattered on the plot. They are not on a line.
* Some cereals are clustered on top of the graph (healthy and popular) whereas some are down below.
* This means my hypothesis was proven wrong that healthy score cereals are more popular. There are consumers that like unhealthy options as well. So the there is no correlation between popularity and Healthy options.
* This analysis can guide consumers in choosing cereals that balance **healthiness and enjoyment**, and assist manufacturers in product development and marketing strategies.

**My data analysis also showed that** :-

* General mill brand should be avoided by consumers who want to limit Sodium intake as well as sugar conscious consumers. Their preferred brand should be Quaker and Nabisco products.
* In order to make this cereal the healthiest, consumers should add fiber externally in the form of Chia seeds, Ground flaxseed, Psyllium husk, Shredded coconut (unsweetened), Berries (raspberries, blackberries), Sliced banana, Almonds, walnuts
* From the Manufacturer’s perspective, there is a wide room for expansion in Hot cereal market. Overall, there are ways to produce newer products that are both healthy and tasty, by adding more protein and fibers and reducing sodium.
* These manufacturers have done a good job in maintaining a low levels of sugars on an average.