## Scenario

Stable Diffusion w/dreamlike-photoreal-2  
“a business consultant looking at cereal boxes in a grocery store aisle, Cinematic, Canon50, Funny, detailed & intricate, HD, photorealistic”

A person in a grocery store

Description automatically generated with low confidence

As the lead analytics consultant at Rocket-AI, you have been tasked with performing exploratory data analysis on breakfast cereal nutrition facts for a new client, Breakfast Bliss Corp. The American food manufacturer is looking to release a new cereal called "Morning Munchies" in the US market and is seeking analytical insights to compete against major players such as General Mills and Kellogg's. To gain an edge in the market, the company's food science team wants to analyze existing cereal products and identify potential gaps or interesting insights in the market that the new product can fill. As the lead consultant, you will guide the exploratory data analysis process to uncover patterns and trends in the nutrition facts of various breakfast cereals. The goal is to identify key areas where "Morning Munchies" can stand out from the competition and appeal to a target market. *Although this is a case study, students will gain hands-on experience analyzing real-world data and applying analytical techniques to inform business decisions.*

The food science team at Breakfast Bliss Corp has tasked you with conducting a thorough exploratory analysis of the cereal database. The focus of the analysis is to identify popular, niche, and interesting segments within the market, as well as peculiarities and average nutritional values that can inform the development of the new product, Morning Munchies. It is important to note that the insights you uncover will be used to inform not only the food science team but also the marketing department, enabling better segmentation of the proposed cereal. As a highly paid consultant, your role is to provide a comprehensive recommendation that is grounded in data. While the food science team undoubtedly has a basic understanding of the market, you, as the analytical expert, are expected to find unique and valuable insights that go far beyond industry basics. By discerning key interactions within the data, your analysis will provide in-depth and actionable coverage of the breakfast cereal industry and a new product launch.

**You are asked to examine the data, clean it, use supplemental data and information to enrich the data/presentation then identify 4 or more interesting insights from the data. All relevant cleaning, enriching and EDA steps along with the 4+ insightful data nuances should be organized into a presentation. You will present jointly to the head of marketing who is looking for an “ah –ha” novel insight or previously unknown data relationship and to the food science team, who is also looking at specific data driven insights. Your presentation should include visualizations when appropriate. Your submission will include code, bulleted written document and PowerPoint slides with narration.**

## Data

Source: After collecting the cereal names from Wikipedia, the nutritional information was appended from the <https://www.edamam.com> API. Additionally, chatGPT was used to create a “rank” which [for the purposes of this case] equates to consumer preference. Additionally a description from chatGPT was also appended. Your submission does not need to use all of the variables, merely the variable that provide interesting and insightful outcomes for the audience.

## Data Table

|  |  |  |  |
| --- | --- | --- | --- |
| File Name | Rows | Columns | Description |
| cereals.csv | 185 | 66 | Consumer cereal information |

## Data Dictionary – all values are for 1 cup of cereal without milk

Example Data

|  |  |
| --- | --- |
| VarName | Example |
| cerealName | 100% Bran Cereal |
| parsedName | unparsed |
| brand | nabisco |
| dietLabels | HIGH\_FIBER, LOW\_SODIUM |
| healthLabels | LOW\_FAT\_ABS, VEGAN, VEGETARIAN, PESCATARIAN, MEDITERRANEAN,… |
| calories | 161 |
| Energy\_kcal | 161.2 |
| Total.lipid..fat.\_g | 3.04 |
| Fatty.acids..total.saturated\_g | 0.4 |
| Fatty.acids..total.trans\_g | 0 |
| Fatty.acids..total.monounsaturated\_g | 0.42 |
| Fatty.acids..total.polyunsaturated\_g | 1.31 |
| Carbohydrate..by.difference\_g | 46.03 |
| Carbohydrates..net.\_g | 27.86 |
| Fiber..total.dietary\_g | 18.17 |
| Sugars..total\_g | 9.73 |
| Protein\_g | 8.15 |
| Cholesterol\_mg | 0 |
| Sodium..Na\_mg | 159.96 |
| Calcium..Ca\_mg | 241.18 |
| Magnesium..Mg\_mg | 224.44 |
| Potassium..K\_mg | 632.4 |
| Iron..Fe\_mg | 10.91 |
| Zinc..Zn\_mg | 7.69 |
| Phosphorus..P\_mg | 713 |
| Vitamin.A..RAE\_µg | 323.64 |
| Vitamin.C..total.ascorbic.acid\_mg | 12.4 |
| Thiamin\_mg | 1.41 |
| Riboflavin\_mg | 1.68 |
| Niacin\_mg | 9.18 |
| Vitamin.B.6\_mg | 7.44 |
| Folate..DFE\_µg | 1362.76 |
| Folate..food\_µg | 25.42 |
| Folic.acid\_µg | 786.78 |
| Vitamin.B.12\_µg | 11.66 |
| Vitamin.D..D2...D3.\_µg | 2.67 |
| Vitamin.E..alpha.tocopherol.\_mg | 0.77 |
| Vitamin.K..phylloquinone.\_µg | 3.22 |
| Water\_g | 1.31 |
| Energy\_pct | 8.06 |
| Fat\_pct | 4.67 |
| Saturated\_pct | 1.98 |
| Carbs\_pct | 15.34 |
| Fiber\_pct | 72.66 |
| Protein\_pct | 16.29 |
| Cholesterol\_pct | 0 |
| Sodium\_pct | 6.66 |
| Calcium\_pct | 24.12 |
| Magnesium\_pct | 53.44 |
| Potassium\_pct | 13.46 |
| Iron\_pct | 60.62 |
| Zinc\_pct | 69.89 |
| Phosphorus\_pct | 101.86 |
| Vitamin.A\_pct | 35.96 |
| Vitamin.C\_pct | 13.78 |
| Thiamin..B1.\_pct | 117.28 |
| Riboflavin..B2.\_pct | 129.25 |
| Niacin..B3.\_pct | 57.35 |
| Vitamin.B6\_pct | 572.31 |
| Folate.equivalent..total.\_pct | 340.69 |
| Vitamin.B12\_pct | 485.67 |
| Vitamin.D\_pct | 17.77 |
| Vitamin.E\_pct | 5.13 |
| Vitamin.K\_pct | 2.69 |
| rawGPTRank | I couldn't find 100% Bran Cereal made by Nabisco, but if you meant "100% Bran Cereal" made by Post, then the rank is 6.5. |
| gptDescription | 100% Bran Cereal : is an all-natural bran cereal containing multiple grains, including wheat and oats, as well as natural sweeteners such as honey and molasses. This high-fiber cereal is an excellent source of essential vitamins and minerals including niacin, iron, and vitamin B6. |

Diet Labels

|  |  |  |  |
| --- | --- | --- | --- |
| **Type** | **Web Label** | **API Parameter** | **Definition** |
| Diet | Balanced | balanced | Protein/Fat/Carb values in 15/35/50 ratio |
| Diet | High-Fiber | high-fiber | More than 5g fiber per serving |
| Diet | High-Protein | high-protein | More than 50% of total calories from proteins |
| Diet | Low-Carb | low-carb | Less than 20% of total calories from carbs |
| Diet | Low-Fat | low-fat | Less than 15% of total calories from fat |
| Diet | Low-Sodium | low-sodium | Less than 140mg Na per serving |

Nutrient Guide

|  |  |  |
| --- | --- | --- |
| **NTR Code** | **Name** | **Unit** |
| SUGAR.added | Added sugar | g |
| CA | Calcium, Ca | mg |
| CHOCDF.net | Carbohydrate (net) | g |
| CHOCDF | Carbohydrate, by difference | g |
| CHOLE | Cholesterol | mg |
| ENERC\_KCAL | Energy | kcal |
| FAMS | Fatty acids, total monounsaturated | g |
| FAPU | Fatty acids, total polyunsaturated | g |
| FASAT | Fatty acids, total saturated | g |
| FATRN | Fatty acids, total trans | g |
| FIBTG | Fiber, total dietary | g |
| FOLDFE | Folate, DFE | µg |
| FOLFD | Folate, food | µg |
| FOLAC | Folic acid | µg |
| FE | Iron, Fe | mg |
| MG | Magnesium | mg |
| NIA | Niacin | mg |
| P | Phosphorus, P | mg |
| K | Potassium, K | mg |
| PROCNT | Protein | g |
| RIBF | Riboflavin | mg |
| NA | Sodium, Na | mg |
| Sugar.alcohol | Sugar alcohols | g |
| SUGAR | Sugars, total | g |
| THIA | Thiamin | mg |
| FAT | Total lipid (fat) | g |
| VITA\_RAE | Vitamin A, RAE | µg |
| VITB12 | Vitamin B-12 | µg |
| VITB6A | Vitamin B-6 | mg |
| VITC | Vitamin C, total ascorbic acid | mg |
| VITD | Vitamin D (D2 + D3) | µg |
| TOCPHA | Vitamin E (alpha-tocopherol) | mg |
| VITK1 | Vitamin K (phylloquinone) | µg |
| WATER | Water | g |
| ZN | Zinc, Zn | mg |

Health Labels

|  |  |  |
| --- | --- | --- |
| **Web Label** | **API Parameter** | **Definition** |
| Alcohol-Cocktail | alcohol-cocktail | Describes an alcoholic cocktail |
| Alcohol-Free | alcohol-free | No alcohol used or contained |
| Celery-Free | celery-free | Does not contain celery or derivatives |
| Crustcean-Free | crustacean-free | Does not contain crustaceans (shrimp, lobster etc.) or derivatives |
| Dairy-Free | dairy-free | No dairy; no *lactose* |
| DASH | DASH | Dietary Approaches to Stop Hypertension diet |
| Egg-Free | egg-free | No eggs or products containing eggs |
| Fish-Free | fish-free | No fish or fish derivatives |
| FODMAP-Free | fodmap-free | Does not contain FODMAP foods |
| Gluten-Free | gluten-free | No ingredients containing gluten |
| Immuno-Supportive | immuno-supportive | [Recipes which fit a science-based approach to eating to strengthen the immune system](https://www.edamam.com/covid-19/) |
| Keto-Friendly | keto-friendly | Maximum 7 grams of net carbs per serving |
| Kidney-Friendly | kidney-friendly | Per serving – phosphorus less than 250 mg AND potassium less than 500 mg AND sodium less than 500 mg |
| Kosher | kosher | Contains only ingredients allowed by the kosher diet. However it does not guarantee kosher preparation of the ingredients themselves |
| Low Potassium | low-potassium | Less than 150mg per serving |
| Low Sugar | low-sugar | No simple sugars – glucose, dextrose, galactose, fructose, sucrose, lactose, maltose |
| Lupine-Free | lupine-free | Does not contain lupine or derivatives |
| Mediterranean | Mediterranean | Mediterranean diet |
| Mollusk-Free | mollusk-free | No mollusks |
| Mustard-Free | mustard-free | Does not contain mustard or derivatives |
| No oil added | No-oil-added | No oil added except to what is contained in the basic ingredients |
| Paleo | paleo | Excludes what are perceived to be agricultural products; grains, legumes, dairy products, potatoes, refined salt, refined sugar, and processed oils |
| Peanut-Free | peanut-free | No peanuts or products containing peanuts |
| Pescatarian | pecatarian | Does not contain meat or meat based products, can contain dairy and fish |
| Pork-Free | pork-free | Does not contain pork or derivatives |
| Red-Meat-Free | red-meat-free | Does not contain beef, lamb, pork, duck, goose, game, horse, and other types of red meat or products containing red meat. |
| Sesame-Free | sesame-free | Does not contain sesame seed or derivatives |
| Shellfish-Free | shellfish-free | No shellfish or shellfish derivatives |
| Soy-Free | soy-free | No soy or products containing soy |
| Sugar-Conscious | sugar-conscious | Less than 4g of sugar per serving |
| Sulfite-Free | sulfite-free | No Sulfites |
| Tree-Nut-Free | tree-nut-free | No tree nuts or products containing tree nuts |
| Vegan | vegan | No meat, poultry, fish, dairy, eggs or honey |
| Vegetarian | vegetarian | No meat, poultry, or fish |
| Wheat-Free | wheat-free | No wheat, can have gluten though |

## The Submission

* The submission will include consultancy slides covering the problem, data and at least 4 insights. Without a presentation, the “organization” section of the rubric will be 0. Exceptional submissions are well ordered and provide a coherent narrative covering all insights while covering the business implications.
* The submission will include a written supplemental representing the insights identified and described in the business presentation. The written portion can be 3-5 sentences for each insight in a bulleted list format. Exceptional submissions include statistics from external credible sources that support the identified personas or insights. For example, “…the children’s cereal market is expected to grow XXX over the next YYY timeframe…”. Without a written supplemental that coincide with the narration and supported by code the “written supplemental” section will be 0.
* The submission will include either a recorded screen narration of the business presentation, a text file with a URL to a recording (like youtube video) or audio that is embedded into the slide deck. Tone, volume, cadence, use of filler words and pronunciation will be accounted for in this section. No points will be deducted based on English proficiency (ie ESL) but technical descriptions that are incorrect will be detrimental. Failure to submit a narration means the “delivery” section of the rubric will be 0.
* An R script covering all data munging, modeling (if applicable), evaluation (if applicable) and visualization construction used to create the presentation artifacts (you do not need to use R to construct the slides, but it is possible!). Your code must use the following R functions at least once throughout your code, group\_by, aggregate & subset. Make sure that your code contains ample comments. Logical ordering, consistency, commenting, and efficiency are evaluated in the documentation section. Failure to turn in an R script will result in a “Documentation” score of 0.

## Criteria for Success

The submission will be evaluated on an equal weighted scale with the following criteria. For example, 20 points per each category [depends on the individual course weighting found in Canvas]

* **Organization** – Was the presentation well organized?
* **Delivery** – Was the content delivered clearly and persuasively with the audience in mind?
* **Code Documentation** – Was the data mined to support the conclusion?
* **Written Supplemental** – Are the bullets clear and supported in narration and code?
* **Data Mining Process** – Overall, as a complete portfolio of work, is the topic interesting, organized, researched, supported and delivered effectively?