The Sugar Trap Market

Gap Analysis

Finding the blue ocean in the snack aisle

Based on results from dashboard

By Bright Hanyabui

The build

With millions of snack options on the food market, a global snack manufacturer, wants to launch a "Healthy Snacking" line. They believe the market is oversaturated with sugary treats, but they lack the data to prove where the specific gaps are.

This solution to this problem uses product data from the openfoodfacts database to the question

01 **Data Cleaning**
Remove erroneous entries

02 **Categorisation**
Group 10,000+ messy tags

03 **Nutrient Matrix**
Sugar vs. protein scatter

04 **Recommendation**
Pinpoint the market gap

05 **Hidden Gem**
Identify protein sources

METHODOLOGY

3M+
Total products
in the full database

500K
Rows loaded
for analysis

8
Columns selected
from ~180 available

7
Product categories
created from tags

Data Pipeline

1. Download .csv.gz
2. Select 8 columns
3. Clean & validate
4. Categorise tags
5. Analyse & score
6. Deploy dashboard

Story 1 — DATA CLEANING

Handle Missing Values

I dropped rows where product_name, sugars_100g, or proteins_100g were null. These columns are essential to the different stories— without them a product cannot be identified or plotted.

Remove Biological Impossibilities

Filtered out any product where a nutrient value was below 0g or above 100g per 100g of product.

Remove Duplicates

Dropped rows where the same product_name + sugar + protein combination appeared more than once.

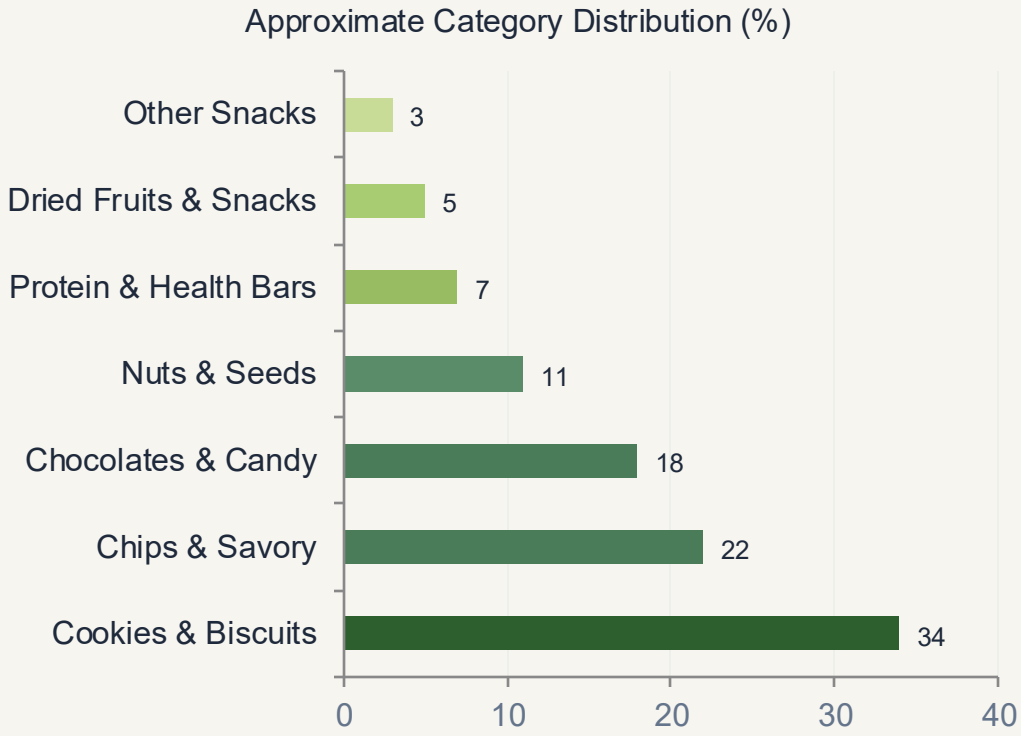
7 Buckets

The Problem

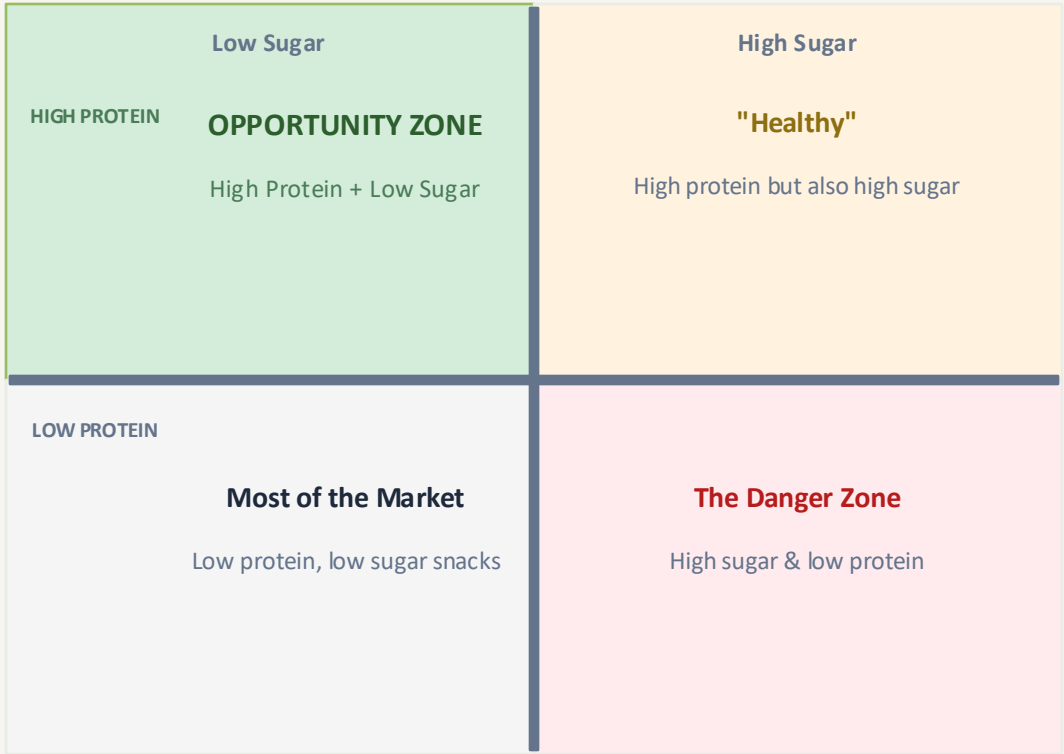
The categories_tags column contained entries like:
en:snacks, en:sweet-snacks,
en:chocolate-chip-cookies-with-nuts

Solution

The keyword-matching function scans each tag string for known words (e.g. "chips", "whey", "almond") and assigns the first matching category.



Sugar vs. Protein



How This Chart Works

- Each dot is one product from the cleaned dataset.
- X-axis: grams of sugar per 100g of product.
- Y-axis: grams of protein per 100g of product.
- Dots are colour-coded by their primary category (Story 2).
- The green zone is the 'empty quadrant' — the market gap.

Story 4 — THE RECOMMENDATION

The Market Opportunity

Key Insight

Based on the data, the biggest market opportunity is in Cookies & Biscuits, specifically targeting products with higher protein and lower sugar per 100g. This category is the largest in the dataset but has the smallest proportion of products meeting health criteria — representing a clear supply gap for health-conscious consumers.

70th

percentile threshold for 'high protein'

30th

percentile threshold for 'low sugar'

<10%

of biscuits/cookies meet both criteria

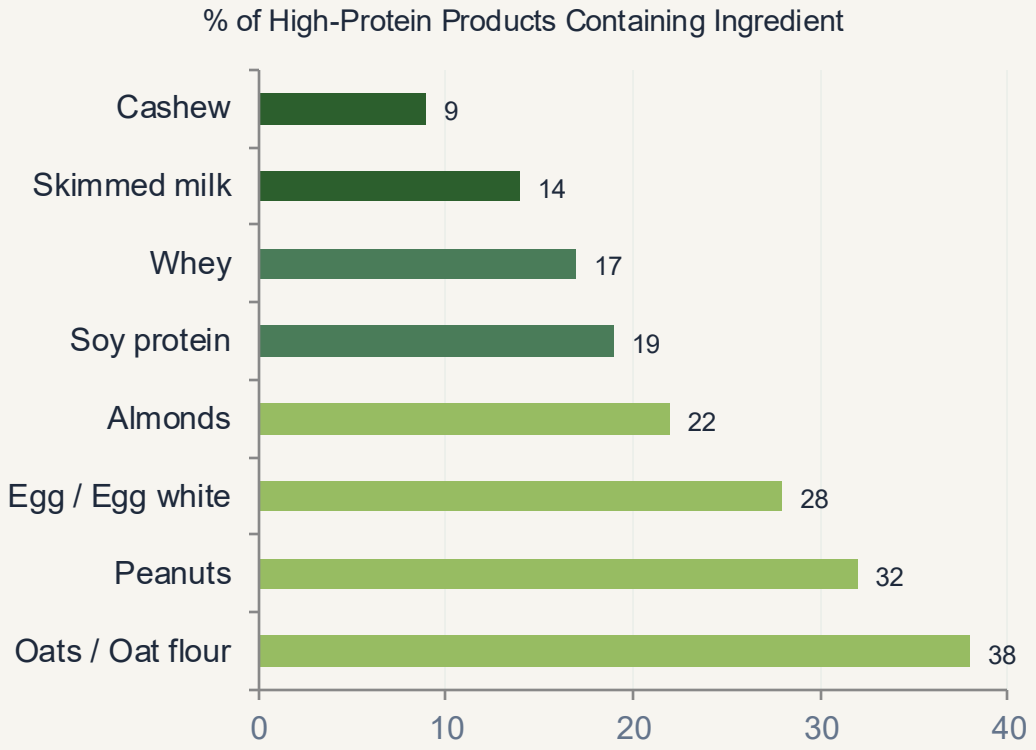
What Drives High Protein in Snacks?

Method

From the products in the high-protein / low-sugar quadrant, we scanned the ingredients_text column for 30 known protein sources.

A Counter object tallied how many products listed each source, then we ranked by frequency.

This tells R&D which specific ingredients to prioritise when formulating a new product.



Who Is Already Doing It Right?

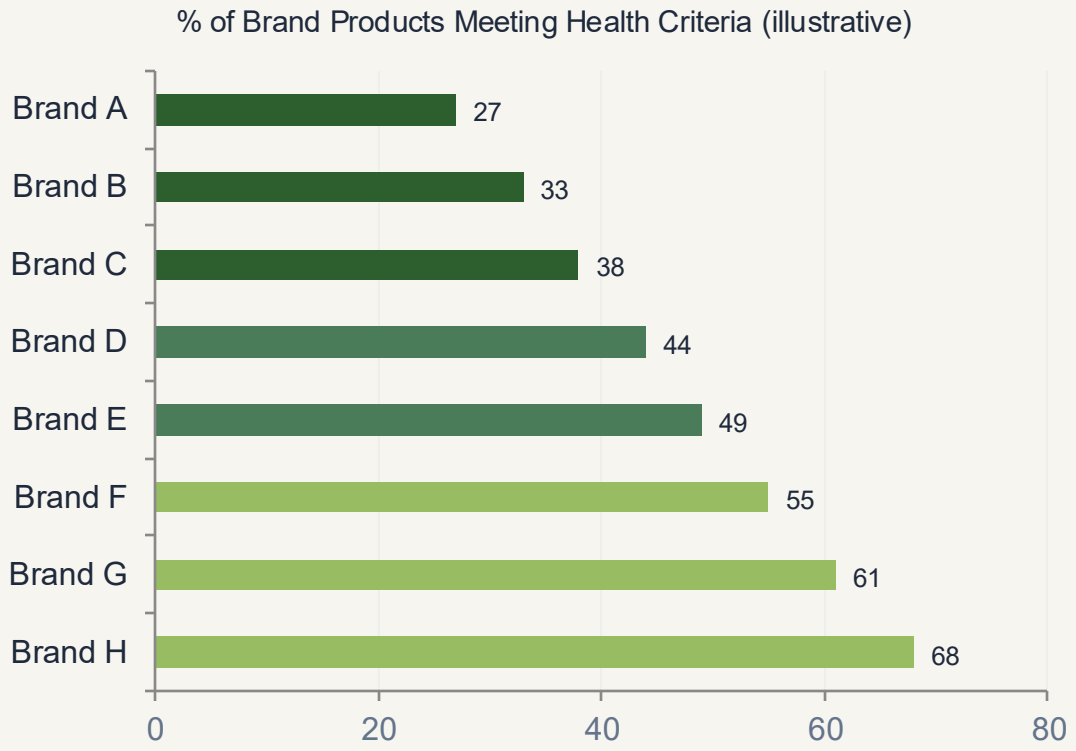
Ranking Criteria

Metric: % of a brand's products that meet the high-protein / low-sugar threshold.

Minimum: Brands with fewer than 5 products are excluded to ensure statistical credibility.

Strategic Implication

Brands at the top of this leaderboard demonstrate that the high-protein / low-sugar product profile is achievable at scale. Their ingredient lists (in Story 5) are the R&D starting point.



CONCLUSION

What the Data Tells Us

01

The gap is real

Across 500,000 products, fewer than 15–25% of snacks meet a meaningful high-protein / low-sugar threshold.

02

Cookies dominate volume

The largest category has the worst health ratio — the biggest market with the most room to disrupt.

03

Ingredients are known

Oats, peanuts, and eggs are the most common protein drivers — proven, accessible, scalable ingredients.

04

Competition benchmarks exist

The Brand Leaderboard identifies existing companies succeeding in the space — and shows it can be done.

Live Dashboard: the-market-gap-analysis-bfpcl4dcy39vhgzvrehvht.streamlit.app