A1: SQL Analysis Assessment

Rhea Kapoor

Masters in Business Analytics

Hult International Business School

Business Analysis with Structured Data - DAT-7470 - FMBAN1

Prof. Chase Kusterer

December 09, 2022

Table of Contents

Introduction to the problem and definition of "healthy" and "cost"

This paper will explore the solution to the business question posed by the Whole Foods' stakeholders – "Do healthy foods cost less?" through performing a detailed SQL and statistical analysis on the database.

Healthy food

As generally defined by the U.S. Food & Drug Administration (FDA), a "healthy" product must meet two requirements: It must contain a meaningful amount of food, and it must not contain more than certain upper limits for saturated fat, sodium, and added sugars (Nestle, 2022).

Though the definition of healthy may differ from person to person, in context of the current database, falling under the dietary preferences of low fat, sugar conscious, whole foods diet or low sodium can be considered to be healthy. As far as desserts are concerned, if they meet the general requirement of not being high in saturated fat, sodium or added sugars, they can be considered to be healthy. While alcohol isn't a healthy choice in general, some alcohol like wine, whiskey, tequila, and hard kombucha are healthier options than beer and sugary drinks (Burch, 2021).

According to the FDA, since a calorie count of 400 or more per serving is considered to be high (Frey, 2020), we will assume calories per serving **less than 300 calories** to be healthy.

Cost

Cost or price refers to the amount of money for which something is sold or offered for sale (Cambridge Dictionary, 2022). According to research from the World Bank, the global

average daily cost of a healthy diet in 2020 was $ 3.35 or 335 cents for high-income countries like USA (World Bank, 2022). Accounting for the increase in inflation rate of 15.15% between 2020 and 2022, we will consider products costing more than **$ 3.85** to be costly for the purpose of our analysis ($1 In 2020 Is Worth $1.15 Today, n.d.).

Answering the business question

Based on the analysis performed in SQL using the criteria outlined above, it can be concluded that healthy foods cost less. When calculated at an overall level, the average cost per serving of the healthy products is $ 3.31 or 331 cents, which is below our established threshold of $ 3.85. Comparing the average cost of healthy foods ($ 3.31) vs unhealthy foods ($8.54) in our database further strengthens our conclusion.

However, when analyzed at the category level, while the average cost in most categories tends to be relatively inexpensive, the cost is high in categories such as Prepared Foods, Beer, and Wine (Figure 1).

Statistical Significance of the results

In order to ensure that the results are statistically significant, two tests were performed. The first approach was to identify a correlation (using SQL and Excel) between the price per serving and calories per serving to understand the relationship between the two variables. The resultant correlation coefficient was 0.18 which indicated a weak positive correlation thereby providing insufficient statistical significance (Figure 2 (a) and 2 (b)).

The second approach involved performing a Hypothesis testing (t-Test: Two-Sample Assuming Unequal Variances) in Excel which resulted in a p-value (0.01) less than alpha (α = 0.05) which in turn led to rejecting the null hypothesis (as outlined below):

H_0: Healthy foods cost more

H_1: Healthy foods do not cost more

Thus, the results obtained through our analysis are statistically significant based on the findings from Hypothesis Testing (Figure 3).

Actionable Insights

Insight- 1: Increase production and sale of Frozen Fruits and Vegetables

According to recent research, it has been revealed that frozen fruits and vegetables often retain the same nutritional content if not more as compared to their fresh counterparts (Bouzari et al., 2015). In fact, studies have shown that the vitamin content tends to be higher in case of frozen produce. This is majorly because before they are frozen, the fruits and vegetables are picked at their peak ripeness which is when they have maximum nutritional value, and the freezing process slows down the nutrient loss (Drayer, 2019). Fresh produce may lose nutritional value during transportation or lean sale periods and ultimately lead to food wastage and loss for Whole Foods.

Apart from being highly nutritious, frozen fruits and vegetables on an average cost much lesser than the fresh produce which can also be concluded from our database. The average price per serving of fresh fruits and vegetables was $ 3.15 (315 cents) while the frozen ones were priced at $ 1.18 (118 cents) [Figure 4].

If Whole Foods can run an awareness program for consumers highlighting the benefits of Frozen produce, it can achieve economies of scale and create a win-win situation for itself as well as the consumers.

Insight- 2: Unaligned proportion of Vegan and Meat products to dietary preference of Americans

According to the Statista Global Consumer Survey on diets and nutrition, around 5% of the population, i.e., 15.5 million people in the USA follow a vegetarian diet. Out of this, only 0.5% (2 million people) are pure vegans (Meyer, 2022). Additionally, the UMN Survey indicated that 80% of the American consumers prefer animal-based protein over plant-based (National Hog Farmer, 2022).

Analysis of the Whole Foods database revealed that the proportion of vegan products to total products was 34.04 % vs a mere 9.2% proportion of meat products to total products (Figure 5). Considering the major dietary preference of Americans as a whole, there exists a disparity in the proportion of products catering to the vegan and meat customer segments.

Thus, Whole Foods can consider diversifying further into the meat category by bringing in more products and intensifying its profit margins by targeting the meat-oriented customer segment.

Insight- 3: Adding more bottled water and coffee products to the Beverages category

The beverage industry is one of the largest industries in the USA. According to the data analysis team at Zippia, bottled water is the fastest growing segment in the beverage industry and was America's favorite beverage accounting for 24% of total beverage consumption in 2021 (Kolmar, 2022). It sold at a record volume of 15.7 billion gallons in 2021 (Redman, 2022).

Coffee made up the next most preferred and consumed beverage according to the statistical findings made by Zippia.

Drilling down into the beverages subcategories in the Whole Foods database, the ‘Water Seltzer & Sparkling Water’ subcategory made up only 9.09% of the total beverages category and the ‘Coffee’ subcategory made up 18.18% as against the major chunk of 27.27% for ‘Tea’ (Figure 6).

Hence, considering the scope of beverage industry and survey statistics, Whole Foods could deliberate over commensuration of goods across subcategories and add more products into the Bottled Water subcategory and Coffee subcategory.

Appendix

Excel file detailing statistical analysis of the database:

References

1. *$1 in 2020 is worth $1.15 today*. (n.d.). $1 In 2020 → 2022 | Inflation Calculator. Retrieved December 6, 2022, from https://www.in2013dollars.com/us/inflation/2020?amount=1&future_pct=0.02
2. Bouzari, A., Holstege, D., & Barrett, D. M. (2015). Vitamin retention in eight fruits and vegetables: a comparison of refrigerated and frozen storage. Journal of agricultural and food chemistry, 63(3), 957–962. https://doi.org/10.1021/jf5058793
3. Burch. (2021, March 24). *The healthiest, lowest calorie alcohol options for a less-bloated New Year's Eve*. Insider. Retrieved December 6, 2022, from https://www.insider.com/guides/health/diet-nutrition/healthiest-alcohol
4. Cambridge Dictionary. (2022). *price meaning: 1. the amount of money for which something is sold: 2. the unpleasant results that you must accept. . .. Learn more.* https://dictionary.cambridge.org/us/dictionary/english/price
5. Drayer, L. (2019, May 30). *Why frozen fruit and veggies may be better for you than fresh | CNN*. CNN. Retrieved December 9, 2022, from https://www.cnn.com/2019/05/30/health/frozen-fruit-vegetables-drayer-food/index.html
6. Frey. (2020, September 25). *What 2,000-Calorie Diet Means on a Nutrition Label*. Verywell Fit. Retrieved December 6, 2022, from https://www.verywellfit.com/what-does-based-on-a-2000-calorie-diet-mean-4099137
7. Kolmar. (2022, September 15). *15 U.S. Beverage Industry Statistics [2022]: Refreshing Trends, Facts, And Stats – Zippia*. 15 U.S. Beverage Industry Statistics [2022]: Refreshing Trends, Facts, and Stats – Zippia. Retrieved December 9, 2022, from https://www.zippia.com/advice/us-beverage-industry-

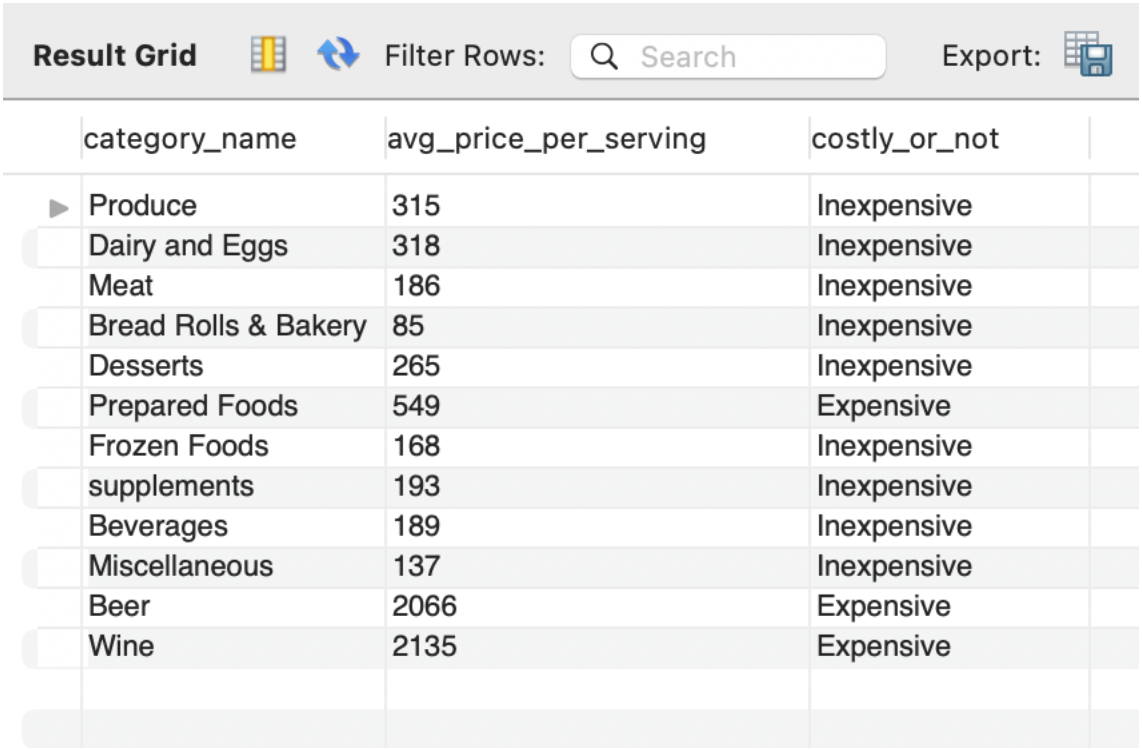
statistics/#:~:text=America's%20favorite%20beverage%20is%20bottled,a%20fifth%20of%20beverage%20consumption.

8. Meyer. (2022, September 30). Here's How Many VEGANS Are In The World (SEP 2022). The VOU. Retrieved December 9, 2022, from https://thevou.com/lifestyle/how-many-vegans-are-in-the-world/
9. National Hog Farmer. (2022, June 15). UMN survey: 80% of U.S. consumers prefer animal-based protein. National Hog Farmer. Retrieved December 9, 2022, from https://www.nationalhogfarmer.com/news/umn-survey-80-us-consumers-prefer-animal-based-protein
10. Nestle. (2022, October 7). *FDA's plan to define "healthy" for food packaging: Better than the existing labeling anarchy, but do we really need it?* STAT. Retrieved December 6, 2022, from https://www.statnews.com/2022/10/07/fda-plan-define-healthy-label-food-packaging/
11. Redman. (2022, June 3). *Bottled water sets all-time volume record*. Supermarket News. Retrieved December 9, 2022, from https://www.supermarketnews.com/consumer-trends/bottled-water-sets-all-time-volume-record
12. World Bank. (2022, July 14). *Food Prices for Nutrition DataHub: global statistics on the cost and affordability of healthy diets*. World Bank. Retrieved December 6, 2022, from https://www.worldbank.org/en/programs/icp/brief/foodpricesfornutrition

Figures

Figure 1

Average price per serving for healthy foods (category-wise)

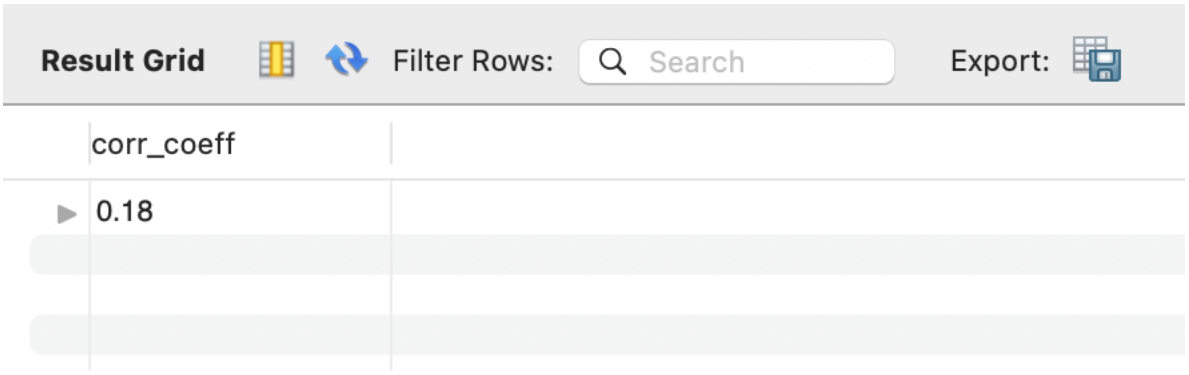


Result Grid Filter Rows: Search Export:

category_name	avg_price_per_serving	costly_or_not
Produce	315	Inexpensive
Dairy and Eggs	318	Inexpensive
Meat	186	Inexpensive
Bread Rolls & Bakery	85	Inexpensive
Desserts	265	Inexpensive
Prepared Foods	549	Expensive
Frozen Foods	168	Inexpensive
supplements	193	Inexpensive
Beverages	189	Inexpensive
Miscellaneous	137	Inexpensive
Beer	2066	Expensive
Wine	2135	Expensive

Figure 2(a)

Correlation calculation – SQL



Result Grid Filter Rows: Search Export:

corr_coeff
0.18

Figure 2(b)

Correlation calculation – Excel

Correlation coeff	**0.18**					
There is a positive correlation between calories and price per serving- i.e. both variables move in the same direction. But it is not very significant						
Correlation coeff calculation using Data Analysis Toolpak						
	caloriesperserving	*price_per_serving*				
caloriesperserving	1					
price_per_serving	**0.182027942**	1				

Figure 3

Hypothesis Testing – Excel

		H0: Healthier foods cost more					
Alpha: 0.05		H1: Healthier foods do not cost more					
t-Test: Two-Sample Assuming Unequal Variances							
	Healthy : price_per_serving	*Unhealthy : price_per_serving*					
Mean	331.32	854.03					
Variance	325,297.56	3,984,016.74					
Observations	178.00	97.00					
Hypothesized Mean Difference	-						
df	105.00						
t Stat	(2.52)						
P(T<=t) one-tail	0.01						
t Critical one-tail	1.66						
P(T<=t) two-tail	0.01		**p-value < alpha of 0.05 => Reject null hypothesis: Healthy foods do not cost more**				
t Critical two-tail	1.98						

Figure 4

Insight 1: Price of Frozen vs Fresh fruits and vegetables

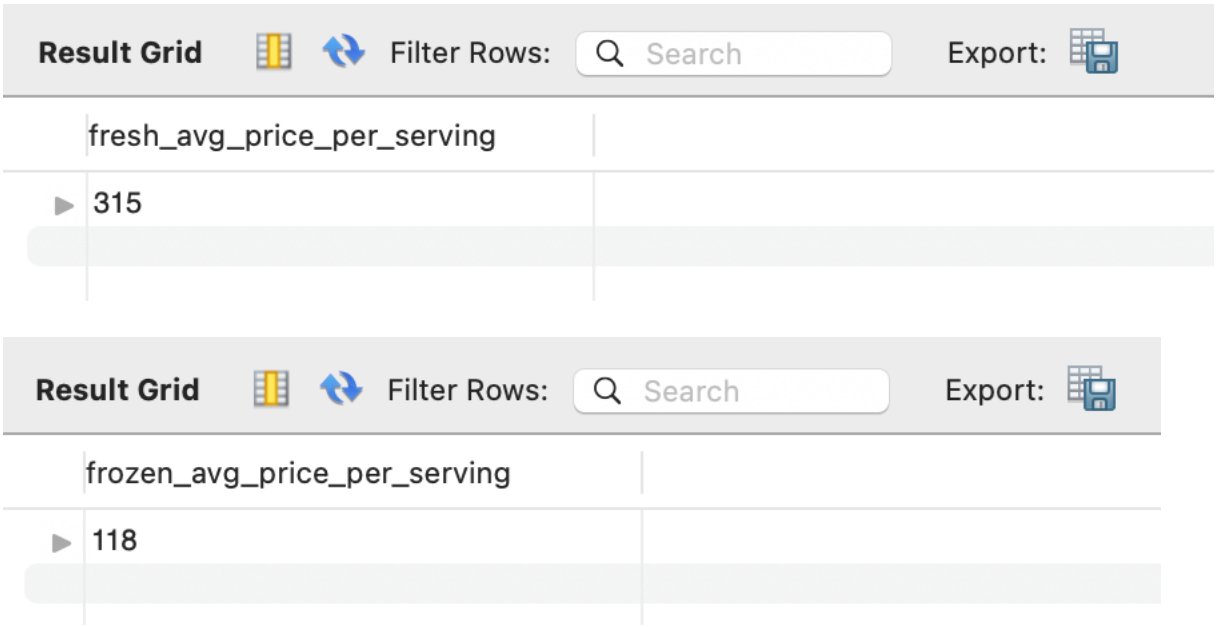


Figure 5

Insight 2: Vegan vs Meat proportion

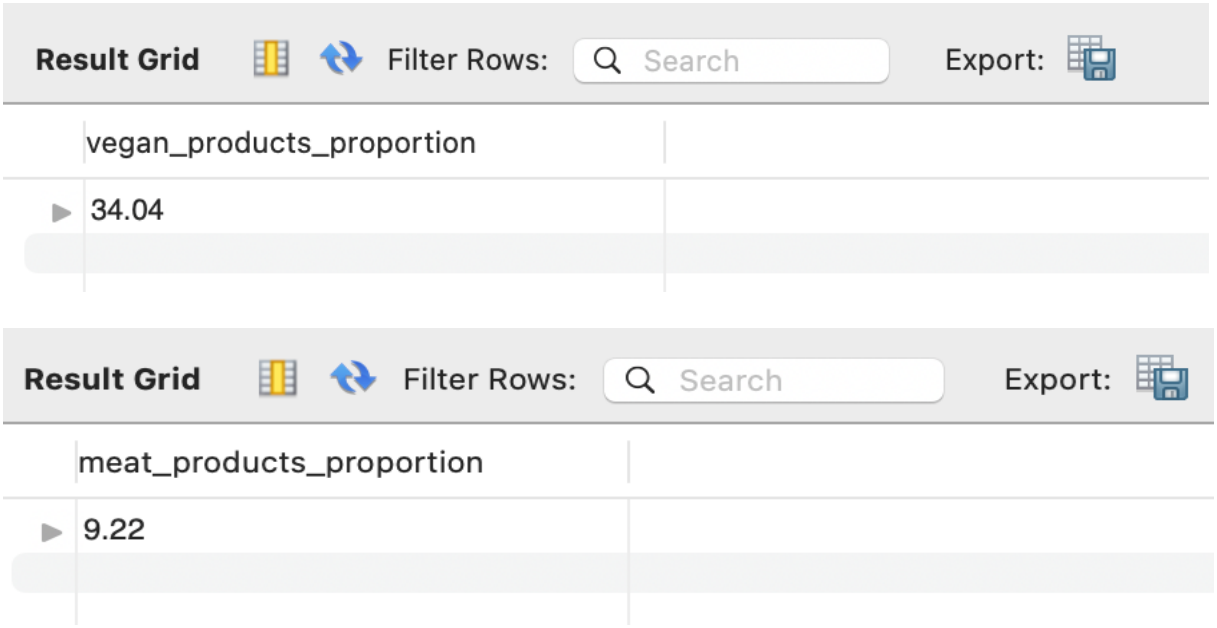
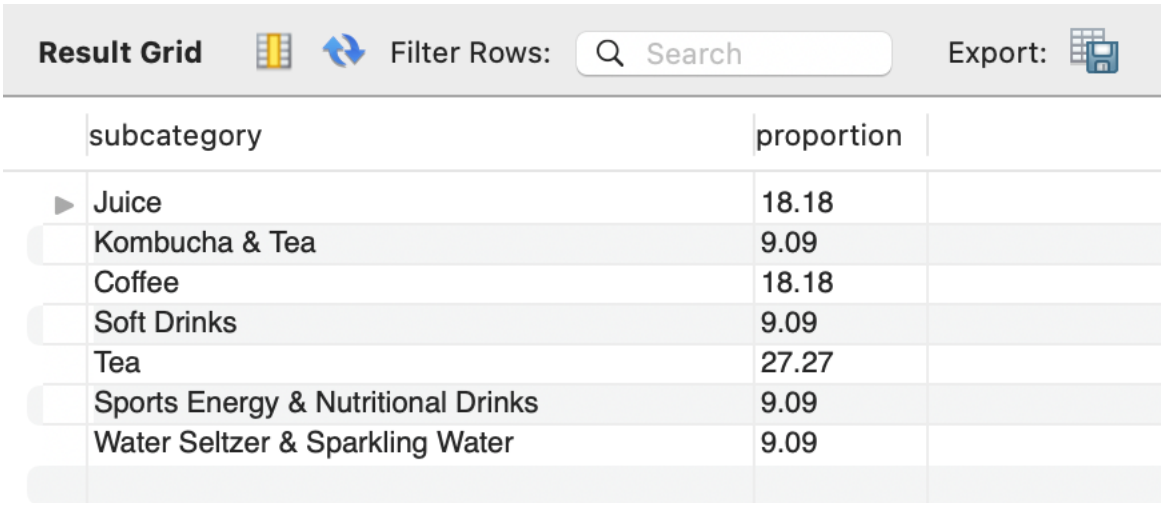


Figure 6

Insight 3: Beverages subcategory distribution



subcategory	proportion
Juice	18.18
Kombucha & Tea	9.09
Coffee	18.18
Soft Drinks	9.09
Tea	27.27
Sports Energy & Nutritional Drinks	9.09
Water Seltzer & Sparkling Water	9.09