BUAN 6337

Predictive Analytics

Predicting & Influencing Health & Wellness

Abhimithra Talla

Overview

We are given a dataset of information pertaining to quantitative research study in 2015 to better understand health and wellness. That dataset gives us observations of various participant's lifestyles, goals, and motivations. The questionnaire consists of 22 different questions surveyed to 1003 different respondents. Respondents provided personal health and wellness information as well as behaviors associated with activity, shopping behaviors, and eating habits.

The first 6 questions on the Survey are designed to gather quite a bit of information towards building a score, which we are calling Health score (wellness score, indicated as "POINTS" within the model results).

We have tried to think through the survey questions and the data provided from different aspects to see how we could find some interesting and useful insights. We came up with the following questions and have tried to answer them through our data investigation.

1. How does alcohol consumption affect health and wellness score?
    a. What about the age range of alcohol consumers?

The thought process behind this first question is to gain a better understanding of how living alone and age could be different in living a healthy life, and how that could potentially play a role in alcohol consumption. How could different grocery stores or smaller stores come up with a plan or a strategy to attract different customers based on their different living conditions?

2. Do dietary restrictions (column D1ra through D1ro) or specific lifestyle habits have anything to do with wellness score?

The reason we are asking this question separately on its own is to learn if there is any true to this or not. We also want to see how we could use some or all of these diet restrictions to think of ways to strategically position different product types throughout the stores to maximize the profits of the stores. Along with diet we want to see if there are specific habits in ones lifestyle that may contribute to a higher health score.

3. Does shopping at different grocery stores impact health scores?
    a. Example: promote certain protein shake at whole foods

The thought behind asking the question on grocery was to understand shopping from which stores are correlated with improving the wellness score. These grocery stores have different business models. For example, Whole Foods is big on Organic and healthy food tags, but not necessarily price. So does this mean that poorer people won't be able to improve their wellness score? We want to investigate this question along with business models of the grocery stores which have a significant positive or negative impact on wellness score.

4. Does healthy score have anything to do with:
    a. Education/race/age/sex/employment

With over 1000 participants in the survey, we want to digest the scale of participants using different backgrounds. This information will better serve SRG in knowing the average consumer base for more targeted and efficient marketing methods.

5. Health score & State correlation - health dept.
    a. Which state could be hit hardest by Corona if no social distancing?
        i. Of course this could be dependent on many other factors, and the score/points we are using as an indicator is not objective enough to be able to conclude this.
        ii. But it could be used to see which states need to proceed differently to encourage healthier lifestyles.

Using the location of our higher ranked health score consumers, we may be able to push for healthier lifestyles in certain areas, or even improve overall health across all states. America is known to be one of the unhealthiest countries in the world in terms of diet and obesity, so perhaps we can use this data to evenly distribute health practices all across the country.

6. Salary range with high or low health scores
    a. How could poor people live a healthy life? Which stores do they shop from?

It is clear that those in a lower wealth bracket will have less opportunities for healthier lifestyles. We want to corroborate this assumption and come up with ways to help out those with lower incomes.

7.

Alcohol consumption and health score:

The Regression output:

OLS Regression Results

Dep. Variable:	POINTS	R-squared (uncentered):	0.183
Model:	OLS	Adj. R-squared (uncentered):	0.182
Method:	Least Squares	F-statistic:	134.3
Date:	Mon, 27 Apr 2020	Prob (F-statistic):	3.58e-28
Time:	06:52:30	Log-Likelihood:	-2521.5
No. Observations:	601	AIC:	5045.
Df Residuals:	600	BIC:	5049.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
x1	0.0551	0.005	11.591	0.000	0.046	0.064

Omnibus:	3.089	Durbin-Watson:	1.654
Prob(Omnibus):	0.213	Jarque-Bera (JB):	2.694
Skew:	0.076	Prob(JB):	0.260
Kurtosis:	2.709	Cond. No.	1.00

- The F statistic for the model is 134.3 which tells that alcohol consumption is related to the health score.

- A 1 unit increase in the Alcohol consumption leads to a .055 increase in health score at 1% significance level.

It is not surprising that an increase in alcohol consumption leads to less healthy life, however there may be some types of alcohol that do promote good health, one such example is red wine. In fact, there are several medical studies and researches that show drinking a glass of red wine a day can lead to better cardiovascular health, blood pressure, and longer life in general. Let's analyze the relationship between specific alcohols below.

Type of alcohol consumption for health score:

The Regression output:

OLS Regression Results

Dep. Variable:	POINTS	R-squared (uncentered):	0.211
Model:	OLS	Adj. R-squared (uncentered):	0.195
Method:	Least Squares	F-statistic:	13.14
Date:	Mon, 27 Apr 2020	Prob (F-statistic):	3.94e-24
Time:	06:20:09	Log-Likelihood:	-2510.9
No. Observations:	601	AIC:	5046.
Df Residuals:	589	BIC:	5099.
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Domestic beer	0.0117	0.009	1.235	0.217	-0.007	0.030
Imported Beer	0.0577	0.015	3.740	0.000	0.027	0.088
Craft Beer	0.0056	0.017	0.326	0.745	-0.028	0.039
Sparkling Wine	-0.0659	0.023	-2.914	0.004	-0.110	-0.021
Red Wine	0.0464	0.013	3.486	0.001	0.020	0.073
White Wine	0.0935	0.019	4.906	0.000	0.056	0.131
Scotch	-0.0289	0.025	-1.174	0.241	-0.077	0.019
Gin	-0.0604	0.021	-2.894	0.004	-0.101	-0.019
Bourbon	0.0022	0.016	0.138	0.891	-0.029	0.033
Vodka	0.0583	0.020	2.851	0.005	0.018	0.098
Rum	-0.0009	0.018	-0.050	0.960	-0.036	0.034
Tequila	-0.0446	0.021	-2.175	0.030	-0.085	-0.004

Omnibus:	0.492	Durbin-Watson:	1.633
Prob(Omnibus):	0.782	Jarque-Bera (JB):	0.598
Skew:	0.038	Prob(JB):	0.742
Kurtosis:	2.865	Cond. No.	9.03

- The F statistic for the model is 13.14 which tells that types of alcohol consumption is related to the health score.
- At a 5% significance level, Imported Beer (Positive Cof), Sparkling Wine (Negative Cof), Red Wine (Positive Cof), White Wine (Positive Cof), Gin (Negative Cof), Vodka (Positive Cof), and Tequila (Negative Cof).
- Other ones do not have a significant impact on the wellness score, at least from the data we are studying.
- Compared to other alcohol types we have, White wine has the strongest effect on our wellness score.

- Adjusted R^2 is around .195, which explains the variation of y variable with respect to the predictors. Alcohol types can help explain roughly 20% of the variation on average in the wellness score

So what can we learn from all this summary? How could we use it? Why are we seeing these positives and negative coefficients? Can we explain the results?
Not all alcohols are associated with low wellness scores. People who drink Red & White Wine are more likely to have a higher wellness score. This one is not as surprising as seeing Vodka and Imported Beer with a positive significant impact on wellness score. Things to keep in mind is that many people who drink red/white wine and imported beers are not aiming to get close to the DUI eligible alcohol level. Most of the time, it's drinking for the taste. So how could this information be of use to us? One way is to put the wine or beer that goes well with specific types of healthy food choices close to each other. Another way is to create combo deals. I might go to the grocery store only to pick up steak, but if I see a red wine right next to the steaks that has a recommendation tag on it that goes well with steak, then why not get them both?
As for Vodka, if you google things to mix with Vodka, the first couple of options are all some sort of fruit juice. Hard liquors, which are not often drunk for the taste, are often mixed with something else. Vodka mixers are often healthier than other hard liquor options. Keep in mind that we are not concluding that drinking vodka will cause the wellness score to go up, we are simply stating a higher wellness score and Vodka consumption are positively correlated.

Alcohol consumption with Age: to update once I get interaction effect.

OLS Regression Results

Dep. Variable:	POINTS	R-squared (uncentered):	0.404
Model:	OLS	Adj. R-squared (uncentered):	0.391
Method:	Least Squares	F-statistic:	30.70
Date:	Mon, 27 Apr 2020	Prob (F-statistic):	7.07e-58
Time:	07:10:04	Log-Likelihood:	-2426.5
No. Observations:	601	AIC:	4879.
Df Residuals:	588	BIC:	4936.
Df Model:	13		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Age	0.2070	0.015	13.806	0.000	0.178	0.236
Domestic beer	-0.0178	0.009	-2.095	0.037	-0.035	-0.001
Imported Beer	0.0431	0.013	3.202	0.001	0.017	0.070
Craft Beer	0.0021	0.015	0.140	0.889	-0.027	0.031
Sparkling Wine	-0.0410	0.020	-2.075	0.038	-0.080	-0.002
Red Wine	0.0270	0.012	2.310	0.021	0.004	0.050
White Wine	0.0603	0.017	3.598	0.000	0.027	0.093
Scotch	-0.0041	0.021	-0.190	0.850	-0.046	0.038
Gin	-0.0256	0.018	-1.398	0.163	-0.062	0.010
Bourbon	-0.0056	0.014	-0.406	0.685	-0.033	0.022
Vodka	0.0350	0.018	1.962	0.050	-2.73e-05	0.070
Rum	-0.0035	0.016	-0.228	0.820	-0.034	0.027
Tequila	-0.0267	0.018	-1.496	0.135	-0.062	0.008

Omnibus:	6.765	Durbin-Watson:	1.958
Prob(Omnibus):	0.034	Jarque-Bera (JB):	5.556
Skew:	0.148	Prob(JB):	0.0622
Kurtosis:	2.634	Cond. No.	9.08

- The F statistic for the model is 30.70 which tells that age and types of alcohol consumption is related to the health score.
- Increase in 1year age leads to 20% increase in the health score at 1% alpha level. This makes sense as people tend to be more health conscious later in there years than when they were at youth.
- Still some of the types for alcohol are insignificant, lets get more deeper insights as we move further.
- R2 is around 40%, which explains the variation of y variable w.r.t predictors and we can see the value is increasing.

Does shopping at different grocery stores impact health scores?

OLS Regression Results

Dep. Variable:	POINTS	R-squared (uncentered):	0.323
Model:	OLS	Adj. R-squared (uncentered):	0.317
Method:	Least Squares	F-statistic:	47.41
Date:	Tue, 28 Apr 2020	Prob (F-statistic):	1.56e-47
Time:	02:59:05	Log-Likelihood:	-2464.8
No. Observations:	601	AIC:	4942.
Df Residuals:	595	BIC:	4968.
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Kroger/Safeway	0.1498	0.016	9.435	0.000	0.119	0.181
Target	0.0019	0.033	0.057	0.955	-0.063	0.067
Walmart	0.0458	0.020	2.304	0.022	0.007	0.085
wholefoods/trader joes	0.0862	0.036	2.418	0.016	0.016	0.156
sams/costco	-0.0022	0.036	-0.059	0.953	-0.073	0.069
dollar store	-0.0255	0.025	-1.016	0.310	-0.075	0.024

Omnibus:	2.551	Durbin-Watson:	1.814
Prob(Omnibus):	0.279	Jarque-Bera (JB):	2.343
Skew:	0.079	Prob(JB):	0.310
Kurtosis:	2.738	Cond. No.	5.09

- The F statistic for the model is 47.41 which tells that store to shop is related to the health score and other predictors are irreverent.
- At a 5% significance level, Kroger/Safeway, Walmart, and WholeFoods/TraderJoes all have positive and significant impact on the wellness score.
- Adjusted R^2 is roughly 32% which explains the variation of y variable w.r.t predictors.
    - That tells us by just using the alcohol types as independent variables, we are able to explain roughly 32% of the wellness score average behavior.

If we rank the Grocery stores by the ones that have the most significant positive impact on the wellness score, Kroger/Safeway comes first, then Whole Foods/Trader Joes, then Walmart. Traditional grocery stores like Kroger and Safeway have been expanding to accommodate those customers who are wanting to eat healthier. Although smaller than Whole Foods and Walmart, they are bringing a variety of products including natural, high quality, non-GMO, and healthy food. This aligns with a question we investigate further down in this paper, lifestyle and attitude of shoppers.

Few of the attitudes with positive significant impacts are looking for natural food, **preferring quality over price, variety in everyday life, price shoppers, and non-GMO products**. Whole foods offer all of those mentioned minus the "price shopper". When someone says Kroger, normally no one thinks of "oh god, it's expensive," but that does happen when thinking of Whole foods. Whole Foods still has a positive and significant impact with the Natural higher quality products.

An important fact about these three being the top contenders for having a positive and significant impact is that you do not have to be rich to live a healthy life. Traditional grocery stores and Walmart both offer a variety of natural healthy food options, which are not too pricey either. Something that the first two grocery store rankers have in common is selling Non-GMO products.

From the grocery stores listed on the provided data, we noticed that Sams/Costco and dollar stores do not have any significant impact on the score, and if any they might have some insignificant impact on the wellness score. Interestingly these two, Sams/Costco and DollarStores are on the exact opposite of the spectrum from the size of the shopping perspective. Sam's/Costco business model is to sell in bulk and larger than regular sizes, while Dollar store's goal is to sell individual items at a dollar value.

Sams/Costco surely have high quality items. They are membership based. Compared to traditional grocery stores, Whole Foods, and Walmart, they lack variety in their product. Costco has Kirkland for many of its products, and Sam's has Member's mark for many of its products. This lack of variety, and that they are both membership based, and that they are not aiming to attract customers who are price shoppers could be the very reason that they are not correlated with consumers with higher wellness scores. Another point to keep in mind is they sell in bulk, and that in itself might push healthier customers away. Main reason behind that is if it's natural fresh food, then it goes bad before the consumer has the chance to eat it all, or it will end up being unnatural GMO options, which again are not the ones that healthy customers would want. As for Dollar Store, surely they are aimed to attract the price shoppers, but they are missing all the other main things that healthier shoppers are looking for. Natural, Non-GMO, and quality over price.

Does healthy score have anything to do with Education/race/age/sex/employment:

- Race or sex should not affect the health score because health score is related to physical and psychological attitude. But lets just confirm them through our Analysis.

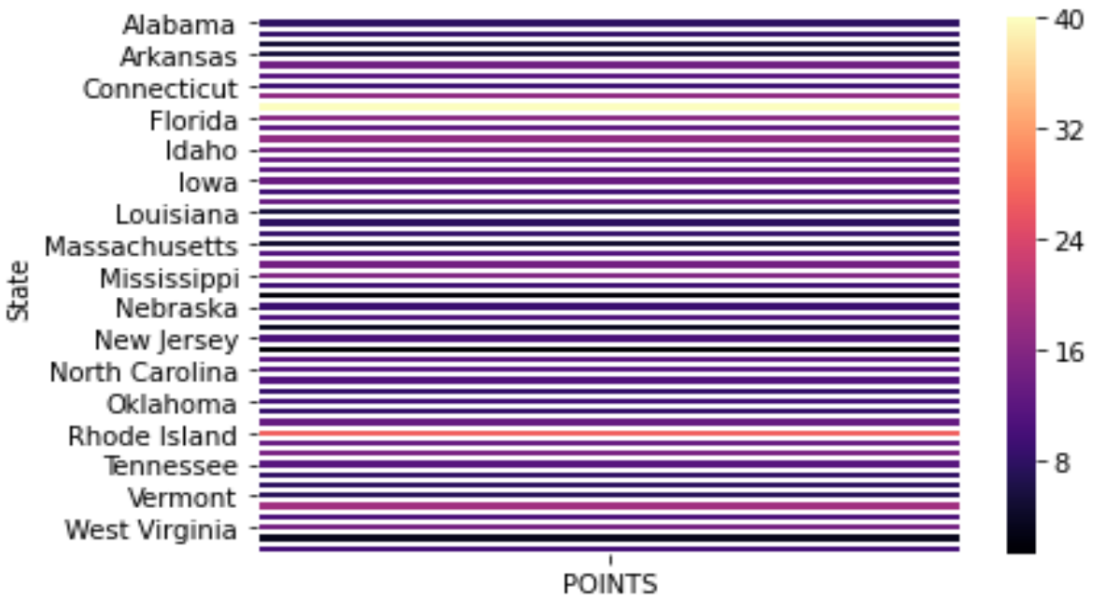
OLS Regression Results

Dep. Variable:	POINTS	**R-squared:**	0.104
Model:	OLS	**Adj. R-squared:**	0.062
Method:	Least Squares	**F-statistic:**	2.461
Date:	Tue, 28 Apr 2020	**Prob (F-statistic):**	6.27e-06
Time:	03:39:30	**Log-Likelihood:**	-3227.5
No. Observations:	802	**AIC:**	6529.
Df Residuals:	765	**BIC:**	6703.
Df Model:	36		
Covariance Type:	nonrobust		

- The F statistic shows that race and sex appear to be insignificant, which confirms our theory

Health score & state correlation - health dept:

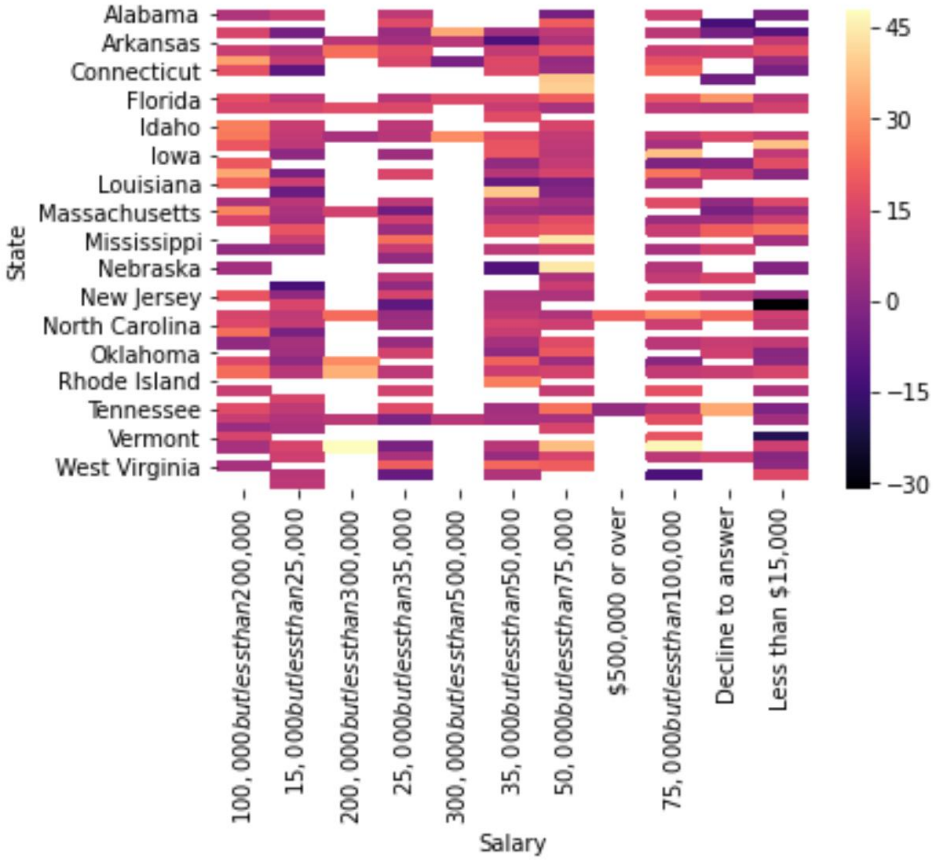
a. **Which state could be hit hardest by Corona if no social distancing?**



- Florida has the highest health score; perhaps they are more health conscious than other states and it would be profitable for the companies to sell more health conscious foods(e.g. low carb foods, protein shakes etc.). This could derive from a multitude of reasons including weather conducive for physical activity or coastal access to a wider variety of fresher foods.
- Wisconsin has the lowest health score, so they might prefer frozen foods or high calorific diets and it would be profitable for the companies to sell less health conscious foods(e.g. high carb foods etc.).

Salary range with high or low health score:

- We can clearly see that people with less income tend to have a lower health score and they are less health conscious and people with income range of 200k-3500k are having highest health scores, its because they might have more options than less income people to buy expensive health conscious foods.



- Least income people of New Jersey has the least health score and highest income people of Vermont has the highest health score.
- If food companies can target their customers according to type of diet and their state , then they can maximize their profits.

How the diet is affecting the health score:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 POINTS   R-squared (uncentered):                   0.266
Model:                            OLS   Adj. R-squared (uncentered):              0.252
Method:                 Least Squares   F-statistic:                              19.11
Date:                Mon, 27 Apr 2020   Prob (F-statistic):                    3.72e-41
Time:                        04:14:00   Log-Likelihood:                         -3137.5
No. Observations:                 752   AIC:                                      6303.
Df Residuals:                     738   BIC:                                      6368.
Df Model:                          14
Covariance Type:            nonrobust
==============================================================================
                                    coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Glten free?_Gluten free          1.6177      2.031      0.796      0.426      -2.370       5.606
Dairy free_Dairy free            4.2112      2.061      2.044      0.041       0.166       8.257
Nut free_Nut free                2.8409      2.577      1.103      0.271      -2.218       7.900
Diabetic_Diabetic                1.1134      1.528      0.729      0.466      -1.887       4.113
Plant based diet_Plant based diet   -4.3122      3.492     -1.235      0.217     -11.167       2.542
Vegan_Vegan                      2.7638      4.034      0.685      0.494      -5.157      10.684
Vegetarian_Vegetarian            2.2488      2.454      0.916      0.360      -2.569       7.066
Low sodium_Low sodium            7.3804      1.226      6.019      0.000       4.973       9.788
Low carb_Low carb                1.5979      1.583      1.009      0.313      -1.510       4.706
Low calorie_Low calorie          6.8514      1.522      4.502      0.000       3.863       9.839
Non-GMO_Non-GMO                  5.6225      1.839      3.057      0.002       2.012       9.233
Antibiotic free_Antibiotic free     -0.3120      2.753     -0.113      0.910      -5.717       5.093
Hormone free_Hormone free        8.7190      2.728      3.196      0.001       3.363      14.075
Allergen free_Allergen free      3.8894      2.775      1.402      0.161      -1.558       9.337
==============================================================================
Omnibus:                        0.777   Durbin-Watson:                   1.792
Prob(Omnibus):                  0.678   Jarque-Bera (JB):                0.625
Skew:                          -0.012   Prob(JB):                        0.731
Kurtosis:                       3.139   Cond. No.                         5.10
==============================================================================
```

- It's common sense that Diet should affect the health score in a positive way.
- As we ran regression through only these diet predictors, the regression output is somewhat controversial.
- The overall F statistic is 19.11, which is significant, but most of the predictors have insignificant p values except for Dairy Free, low sodium, Low calorie, Non-GMO, and Hormone free.
- These five have significant p values at 5% alpha level. They all have a positive impact on the wellness score. This means that the consumers who have these diet restrictions also tend to have a higher wellness score.

This could be a guide for the different Grocery stores on where to position different types of products. For example, putting some of the low calorie items and non-GMO items at one end of the store, and the Dairy free, low sodium and some other low calorie items at the other end of the store. Having other items on display in between the two for customers to pick up.

Another use of understanding the diet restrictions and its impacts on wellness score is for grocery stores to know how to advertise and what to advertise for. For example, Korger makes it known they offer a variety of Non-GMO and Organic options. Sams/Costco could start advertising more on the fact that they offer Low Sodium and Low Calorie products.

Lifestyle/attitude parameters affecting the health score:

OLS Regression Results

Dep. Variable:	POINTS	R-squared (uncentered):	0.600
Model:	OLS	Adj. R-squared (uncentered):	0.578
Method:	Least Squares	F-statistic:	27.37
Date:	Mon, 27 Apr 2020	Prob (F-statistic):	1.59e-115
Time:	04:25:17	Log-Likelihood:	-2909.7
No. Observations:	752	AIC:	5897.
Df Residuals:	713	BIC:	6078.
Df Model:	39		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
cooking score	0.9058	0.718	1.261	0.208	-0.504	2.316
No sale no shop	-1.5817	0.557	-2.838	0.005	-2.676	-0.488
save time over food choice in kitchen	-1.7191	0.605	-2.842	0.005	-2.907	-0.532
induldge in food	-0.5144	0.639	-0.805	0.421	-1.768	0.740
natural food	1.3307	0.734	1.814	0.070	-0.110	2.771
nat food better for me & envr	0.2195	0.705	0.311	0.756	-1.165	1.604
reward w/ sth tasty	0.1837	0.730	0.252	0.801	-1.250	1.617
seek out info on health/nutrition	1.1751	0.675	1.742	0.082	-0.149	2.499
satisfied w/ healthfulness of diet	6.7020	0.648	10.343	0.000	5.430	7.974
balance healthy & unhealthy food	1.0041	0.689	1.456	0.146	-0.349	2.358
give up good taste for healthy food	-0.6569	0.684	-0.960	0.337	-2.000	0.686
give up convinience for health	-0.2763	0.650	-0.425	0.671	-1.552	0.999
buy store brands	-1.3374	0.562	-2.382	0.017	-2.440	-0.235
eat organic	-0.0662	0.700	-0.095	0.925	-1.440	1.307
no junk food in home	0.4896	0.603	0.811	0.417	-0.695	1.674
diff types of food	-0.8558	0.678	-1.261	0.208	-2.188	0.476
fresh food over frozen/canned for cooking	-0.0787	0.699	-0.113	0.910	-1.452	1.294
new recipes	-0.5801	0.685	-0.847	0.397	-1.924	0.764
fam says am good cook	0.0253	0.776	0.033	0.974	-1.499	1.550
change stores to eat healthier	-1.2255	0.709	-1.729	0.084	-2.617	0.166
change stores to save money	-1.3328	0.673	-1.981	0.048	-2.654	-0.012
spend time to find new things in groceries	0.1431	0.617	0.232	0.817	-1.068	1.355
advance meal plans	0.3616	0.610	0.593	0.554	-0.836	1.560
brands to reflext my style	0.2655	0.674	0.394	0.694	-1.058	1.589
brands to support social causes	-1.2296	0.710	-1.733	0.084	-2.623	0.164
follow trends/fashion	0.5011	0.664	0.754	0.451	-0.803	1.806
brands I grew up with	-0.1395	0.582	-0.240	0.811	-1.283	1.004
price shopper	1.1733	0.655	1.792	0.074	-0.112	2.458
quality based shopping, not $	1.6637	0.637	2.610	0.009	0.412	2.915
recommend food I like to others	-1.2371	0.695	-1.781	0.075	-2.601	0.127
shopping to relax	0.0166	0.594	0.028	0.978	-1.149	1.182
strive to achieve high social status	-0.3614	0.674	-0.536	0.592	-1.685	0.962
wanting to impress others	0.7030	0.712	0.987	0.324	-0.696	2.101
prioritize time w/ fam	1.0545	0.602	1.752	0.080	-0.127	2.236
variety in everyday life	1.4255	0.780	1.829	0.068	-0.105	2.956
same life week to week	-0.3671	0.599	-0.612	0.540	-1.544	0.810
looking for a thrill	-1.4063	0.637	-2.209	0.027	-2.656	-0.156
consider myself sophisticated	0.8807	0.611	1.441	0.150	-0.319	2.081
taste over health	-1.0040	0.663	-1.515	0.130	-2.305	0.297

Omnibus:	0.485	Durbin-Watson:	1.985
Prob(Omnibus):	0.785	Jarque-Bera (JB):	0.534
Skew:	-0.060	Prob(JB):	0.766
Kurtosis:	2.947	Cond. No.	39.4

It is fascinating that from all different types of investigations we have done so far, lifestyle has had the most to do with explaining the wellness score, with adjusted R^2 of 58%.

Running regression using all the predictors:

- Let's run regression using all the predictors, those we feel are affecting the Health score and by using step-wise forward selection and lasso penalty for the predictors to get a subset of our best predictors for estimating the model leads to the following results.
- The purpose of running this model is to utilize all the variables and see which ones will remain as significant influencers on the wellness score. One thing to keep in mind is that we might be running the risk of omitted variable bias by running the models individually, so here we are running a more accurate model using the stepwise selection to keep only the ones that would have a significant impact.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	19	10085	530.80585	12.49	<.0001
Error	35	1487.41617	42.49760		
Corrected Total	54	11573			

Root MSE	6.51902
Dependent Mean	11.63636
R-Square	0.8715
Adj R-Sq	0.8017
AIC	278.36044
AICC	306.36044
SBC	261.50711
ASE (Train)	27.04393
ASE (Test)	482.59593
CV PRESS	2543.12331

- The F-statistic for our model is 12.49, which is significant and shows that our model is consistent.
- Adj R2 is 80.17, which is a good indication that our predictors might affect the health score.
- ASC(Train) is minimum, another good indication that the model may be consistent.

Regression output:

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-10.858465	9.508036	-1.14	0.2612
Age	1	-0.223141	0.083476	-2.67	0.0113
snacking_in_last_24_hrs__No	1	-5.093051	2.593895	-1.96	0.0576
wholefoods_trader_jo	1	-0.070306	0.044359	-1.58	0.1220
give_up_convinience_	1	-4.313020	1.389448	-3.10	0.0038
change_stores_to_sav	1	6.432415	1.388542	4.63	<.0001
advance_meal_plans	1	-2.401671	1.135296	-2.12	0.0416
brands_to_support_so	1	-1.062778	1.507811	-0.70	0.4856
quality_based_shoppi	1	10.876978	1.342865	8.10	<.0001
shopping_to_relax	1	5.159273	1.225867	4.21	0.0002
variety_in_everyday_	1	-2.124588	1.536826	-1.38	0.1756
Vegetarian_NO TO: Vegetarian	1	22.536357	4.395897	5.13	<.0001
Low_sodium_NO TO: Low sodium	1	-6.905127	2.479800	-2.78	0.0086
Allergen_free_Allergen free	1	15.995607	4.684479	3.41	0.0016
Locally_sourced_Locally sourced	1	14.983777	2.991918	5.01	<.0001
_of_household_member	1	-7.487569	2.381454	-3.14	0.0034
Marriage_status_Yes, married or living with partner	1	8.267937	2.367185	3.49	0.0013
__of_children	1	3.649507	3.107325	1.17	0.2481
Education_Some college or Associate degree	1	12.342386	2.389873	5.16	<.0001
weight__lb_	1	-0.096091	0.019969	-4.81	<.0001

The following are the variables that have a significant impact on the wellness score at 5% significance level with their relevant coefficient:

- age: -0.223
- snacking in 24 hrs: -5.1
- give up convenience: -4.31
- change stores to save: 6.43
- **advance meal plans: -2.4**
- **quality based shopping: 10.87**
- shopping to relax: 5.16
- no to vegetarian: 22.5
- no to low sodium: -6.9
- allergen free: 16
- **locally sourced: 15**
- #of household members: -7.48
- married or living w/ partner: 8.26
- **education - some college or associate: 12.34**
- **weight: -0.096**

Married or living w/ partner has a positive coefficient but then number of household members has a negative coefficient. A conclusion that could be drawn from this is that single parents with children and others who live with roommates run the chance of not being able to improve their wellness score. The more people, the more they could potentially encourage each other to give up on living a healthy life.

On the other hand, the ones that are living with their partner have a higher chance of living a healthier life to improve their wellness score.

Some familiar variables that have also gone through the filter in this model and have proven to have a significant impact on the wellness score are dietary restrictions. Those consumers who are not following a low sodium diet are more likely to be disproving their wellness score, low sodium and wellness score have proven to be negatively correlated. A mind blowing observation from this sample is the vegetarian diet. Believe it or not, those who specified No to vegetarian diet, are improving their wellness score by 22 points on average. And the last dietary restrictions concluded to have a positive significant impact on the wellness score is Allergen free. Low sodium and Allergen free could be the main points that a grocery store would want to highlight in advertising, to attract health conscious customers.

From the variables listed above, Age and Snacking within the last 24 hrs having a negative coefficient are not surprising, and do not need much explanation. Other interesting observations are negative coefficient for giving up convenience for health benefits and also positive coefficient for shopping to relax. These two together are weird combinations. This is showing that the customers who are striving to improve their wellness score are not necessarily running short on time. The ones that are shopping to relax and are investing their time to be convenient and also accomplish their grocery shopping are having a higher chance of improving their wellness score. Maybe giving up the convenience piece in itself is the cause of the wellness score going down. This could be of use for Grocery stores. For Grocery stores to focus on making the shopping process a rather enjoyable, calming, and convenient process. That way no customers will need to “give up convenience” and they will have to shop to relax at the place. That is definitely a common practice that traditional grocery stores such as Kroger/Safeway and Whole Foods have picked up.

Changing stores to save money has a positive coefficient. This could be used for the Grocery stores again. If those customers are willing to change to stores to save money are improving their health score, then why not have regular sales on some items in the stores to keep those customers keep shopping in the first place.

Final suggestion recommendation:

Eating healthy and improving the wellness score is not terribly dependent on how poor or rich someone is. On the contrary it is more dependent on their lifestyle, dietary restrictions, and education. Through this sample study, we are able to make a few recommendations in terms of what type of products stores should advertise more often and possibly how they should structure their products in the store (what products go where and what other products should be near each other). Following are just a few of those examples which would be good for stores to implement to higher profit:

- Snacks at the cashier. Perhaps the customers are aware of the fact that snacking might not be helping their wellness score improvement, and they often try to avoid the snack aisle when at the store to maintain a healthy diet. What can the stores do to go around this? Keep the snacks somewhere that customers can't skip seeing them, display it somehow for customers to not be able to resist the temptation.
- Low Sodium, Allergen Free, Low Calorie, and Non-GMO products are the products which are correlated with consumers with higher wellness scores. For these products, Featuring the products come into play.
    - One game plan would be putting some of these products together, and the stores would need to Feature only one or two of them to pull the customers into that aisle.
    - Another idea is to make combo deals for most and least popular Low Sodium and Low Calorie products.
    - From the first models, we learned that Red Wine is correlated with improving wellness score. Maybe Display Red Wine close by Low Sodium, Low Calorie, high protein products.
    - Other ways that we could use this finding is regarding the location of these products within a store. We see that both Low Sodium and No Allergen products have a positive impact on the wellness score. Locating these two products at two opposite ends of the store will make the customers walk from one end to another which would raise the chance of them picking up other items that were not initially on their list.
- Shopping to relax is correlated with higher wellness scores. This was a surprise to us, as one of our group members thought only he goes grocery shopping as a way to relax. Seems like that many people have the same thought, Grocery Shopping could be used as a way to relax. No wonder traditional grocery stores have been trying to maintain a homey calming style within their stores. Design of the store does matter. That is definitely a noticeable factor when one enters Whole Foods or majority of the traditional grocery stores including Kroger or Safeway. Kroger has made it an effort to change the looks and the design of its stores to make it more enjoyable to spend time in it.
- Locally sourced and Quality based shopping. Learning the importance of these two and their correlation with higher wellness score is indeed crucial when thinking of different ways of advertisement. Stores first need to pull the customers in, for them to enjoy the calming and relaxing design. How are the customers pulled in? By doing the right commercials and advertising, by letting the customers know that the products sold at the stores are locally sourced, and that they are the higher quality items.
- Not being a Vegetarian is positively correlated with wellness score. Who would have thought? This one was a surprise to us as being a Vegetarian has become one of the recent fads. This has grown to be a culture, although not a huge culture comparatively. Knowing about the products that have negative impacts on the wellness score is as important as knowing about the ones with positive impacts. If a store wants to be an advocate of wellness score and how their business model aligns with improving wellness of its customers, they need to know what products NOT to advertise for.