MGT 6203 Final Report – Team 9

Team Members:

1. Sloan Wheeldon; 903751655
2. Susannah Weaver; 903739278
3. Reagan Cook; 903158153
4. Caroline Schauder; 903569046
5. Ashianna Jetha; 903215020

Project Motivation:

Record breaking inflation in 2022 is affecting Americans around the country. We all have noticed our grocery bills increasing and were curious how this inflation is impacting multiple types of foods. We hypothesize that this inflation could be causing a change in consumer spending. We also hypothesize that inflation on processed foods may be lower than inflation on non-processed foods. Finally, we hypothesize that increased inflation on foods in general could be contributing to an increase in processed foods in American diets. We intend to test these hypotheses by answering the following questions:

- How has inflation impacted the consumption and affordability of processed vs. unprocessed foods?
- Is there a statistically significant difference between inflation on processed foods and fresh foods?
- At what level of inflation does consumer behavior change?
- How much does inflation explain the increase in processed food consumption over time?

To clearly set forth what we mean by "processed", we will establish this definition: Processed foods generally contain more ingredients, such as artificial flavoring and chemical additives. This matters because these kinds of additives are suggested to be a contributor to increased rates of obesity and chronic diseases such as heart disease and diabetes (Harvard School of Public Health). The NOVA food classification system is an internationally recognized method to classify foods into four categories: Unprocessed or minimally processed foods, processed culinary ingredients, processed foods, and ultra-processed foods. Of the categories, ultra-processed foods are linked the closest to "obesity and related chronic diseases" (Moubarac JC, 2017). Throughout this paper, anytime we refer to the term "processed" or "highly processed" we are specifically referring to the NOVA defined "ultra-processed" category. The other category of foods we will define as "less-processed," which includes all other NOVA defined food categories.

This has business ramifications for health insurance companies, which generally pay more claims for customers in poorer health. It is in the health insurance company's best interest to help its customers develop healthy habits in the interest of preventative care. Therefore, if a health insurance company can understand the link between inflation and consumption of processed foods, they can better understand how to help consumers afford fresh food. Health insurance companies often offer incentives, discounts, or access to certain vendors to encourage their members to have a healthier lifestyle or later avoid illness or disease. If our hypotheses are correct, it may be in the health insurance company's interest to subsidize or provide financial help for certain foods that are subject to high inflation. This will help to keep healthier options available to the country in the face of high inflation and wage stagnation.

Literature Review:

Research into the societal and business context of this problem was necessary to frame our analysis correctly, as we are interested in trends in the US specifically.

In America, people tend to over-consume reactive health care, and under-consume prevention (Gore, et al., 2017). American society prioritizes quick medical fixes (or sometimes not quick) over preventative care. In the end, medical interventions cost insurance companies and people more than preventative solutions, such as healthy eating, consistent exercise, and mental health prioritization. The health insurance industry has the unique opportunity to align financial incentives with positive public health outcomes by addressing this problem (Gore, et al., 2017).

In the US, nearly $173 billion per year is spent on health care costs for obesity (2022). 20% of young people from ages 2 to 19 and 42% of adults have obesity, which can put them at risk of heart disease, type 2 diabetes, and some cancers. An unhealthy diet can contribute to these health issues. Over 70% of the sodium that Americans eat comes from packaged, processed, store-bought, and restaurant foods (2022). Unhealthy food and beverages, such as sugar-sweetened beverages and highly processed food, can lead to weight gain, obesity and other chronic conditions that put people at higher risk of at least 13 types of cancer (2022). Therefore, it is in an insurance company's best interest to prioritize and incentivize preventative health. Healthy eating is an essential part of maintaining a healthy lifestyle.

Harvard Business Review cites the Vitality program, developed over 25 years ago by the South African insurance company Discovery (Gore, et al., 2017). The program provided Discovery clients with incentives to improve their health through gym memberships, discounts on healthy foods, and other awards based on achievement of personal health goals (Gore, et al., 2017). The program found that participating members reduced hospitalization costs by 30% and lived from 13 to 21 years longer than the rest of the population. Of course, the biggest successes of this program were the societal and health benefits to consumers, but it also financially benefited the insurance company by reducing their payouts.

Because ultra-processed foods are associated with poor diet quality and higher risk of several chronic diseases, it is important for insurance companies to understand who is consuming ultra-processed foods and why (2021). Our investigation attempts to address this question by analyzing the cost of healthy and unhealthy foods to determine whether consumers are being priced out of a good diet.

Overview of Project:

In short, our problem statement is that inflation in the US has risen significantly in the last two years and wages are not increasing proportionally. As a result, people are having to make budget cuts; this may include an increase in consumption of processed foods, which are typically more affordable and accessible. Therefore, we wanted to know how inflation has impacted the consumption and affordability of highly processed vs. less processed foods.

Our general approach was to first clean, join, and transform our chosen data sets to most ideal forms for creating analytical models. We then performed exploratory data analysis to get further insights about the data. Following this, we did CUSUM analysis to determine when a significant change was detected in consumer spending and prices of various foods. Additionally, we performed hypothesis testing to see if there was a statistically significant difference in inflation on highly processed foods compared to less processed foods. Finally, we created regression models to understand the relationship between prices of foods and overall grocery spending.

Overview of Data:

The key variables for this project were the binary variables "highly processed" and "less-processed", the specific food items: bacon, bananas, ground beef, white bread, butter, potato chips, chocolate chip cookies, whole milk, eggs and potatoes, and grocery store sales. During our exploratory data analysis, we saw some spikes in inflation and spending around the time of the Covid-19 pandemic, which we thought was interesting.

All of our food data came from FRED, our nutrient data came from the USDA website, and grocery spending data came from the Census Bureau website.

The food data contains pricing information in dollars and nutritional information for multiple types of food over a period of years. Most of the datasets chosen have a time frame from approximately 1980 through 2022, aggregated monthly. We chose ten specific foods that are common in many diets, making sure to choose some that are highly processed and some that are not. The ten food datasets were aggregated into one data frame for ease of use, and the data types were modified to ensure the dates are in a usable format. The final data frame contains the price in dollars of each food item in a given year.

The nutrient dataset is a table of macronutrients available in the food supply in the US (per capita, per day) by year from 1909-2010, aggregated yearly. We used it for exploratory data analysis and better understanding the nutrient makeup of the American diet. This data was for the most part already usable, other than renaming some of the factors.

The grocery sales dataset consists of total monthly sales in millions of dollars for various business types. The original spreadsheet contained several sheets worth of data, therefore requiring removal of irrelevant information like automotive business sales and the NAICS codes of businesses. Additionally, the data required pivoting and merging, as well as manipulation to obtain the proper date format for use in time series models. The final data-frame contains grocery sales by month with a column for the year as well, to facilitate comparison with the food datasets.

Overview of Modeling:

Our models consisted of CUSUM analysis, hypothesis testing, and regression models.

We first created change detection models to see when there was a significant change in prices of different foods and on grocery spending in general. We saw a positive change detected in the prices of different food items, which shows that inflation is affecting these prices. We also saw a positive change detected in overall grocery spending. This is interesting because we thought that people would have changed their spending habits to fit their budget. However, it does make sense because if prices dramatically increased, people are going to be forced to spend more money to afford food. It was also interesting because many of these changes that were detected seemed to be around the time or soon after the onset of the Covid-19 pandemic.

Date	Grocery Sales	
Jan. 2019	57392	0
Feb. 2019	51505	0
Mar. 2019	57486	0
Apr. 2019	56318	0
May 2019	59932	1850.761753
Jun. 2019	58025	1794.523507
Jul. 2019	59654	3367.28526
Aug. 2019	60065	5351.047014
Sep. 2019	56088	3357.808767
Oct. 2019	58132	3408.57052
Nov. 2019	59091	4418.332274
Dec. 2019	60682	7019.094027
Jan. 2020	58230	7167.855781
Feb. 2020	55239	4325.617534
Mar. 2020	73003	19247.37929
Apr. 2020	63454	24620.14104
May 2020	67331	33869.90279
Jun. 2020	62806	38594.66455
Jul. 2020	65575	46088.4263
Aug. 2020	63141	51148.18805
Sep. 2020	60565	53631.94981
Oct. 2020	62506	58056.71156

DATE	bacon	bananas	butter	chocolate_chip_	eggs	ground_beef	potato_chip	russet_potatoes	white_bread	whole_milk
2/1/19	0.06672250078	0.001579611842	0.6547426948	0.02352404665	0.3137402003	0	0	24.91645552	0	0
3/1/19	0.2394450016	0.01315922368	1.41648539	0	0.6144804005	0	0.03526944761	61.03291104	0	0
4/1/19	0.3561675023	0.01073883553	1.936228085	0	0.8342206008	0	0	97.44936655	0	0
5/1/19	0.7328900031	0.02031844737	2.547970779	0	0.9529608011	0	0	125.1658221	0	0
6/1/19	1.175612504	0.02489805921	2.969713474	0.01152404665	0.9127010013	0	0	160.6822776	0	0
7/1/19	1.441335005	0.02447767105	3.447456169	0	0.9124412016	0	0	200.2987331	0	0
8/1/19	1.582057505	0.0330572829	3.984198864	0.003524046646	0.8881814018	0	0	307.0151886	0	0
9/1/19	1.717780006	0.03263689474	4.438941559	0.09804809329	1.027921602	0.02713108737	0.009269447611	372.6316441	0	0.02923278547
10/1/19	1.935502507	0.03621650658	4.918684254	0.2075721399	1.066661802	0.04426217474	0	405.1480997	0	0.07546557094
11/1/19	2.005225008	0.03979611842	5.132426948	0.3180961866	1.228402003	0.03239326211	0.1402694476	477.1645552	0.02651286717	0.1916983564
12/1/19	2.043947509	0.04337573026	5.257169643	0.4096202332	1.520142203	0.07052434948	0.1205388952	552.3810107	0.05502573434	0.3069311419
1/1/20	2.113670009	0.04395534211	5.621912338	0.4421442799	1.737882403	0.1326554369	0.06280834283	633.1974662	0.07153860151	0.4871639274
2/1/20	2.17539251	0.04853495395	5.884655033	0.5006683265	1.943622603	0.1737865242	0.1050777904	714.3139217	0.1120514687	0.6103967128
3/1/20	1.997115011	0.05511456579	6.242397728	0.6411923732	2.225362804	0.2309176116	0.1833472381	795.4303773	0.1515643358	0.7856294983
4/1/20	1.907837512	0.06669417763	6.468140423	0.9747164198	3.001103004	0.459048699	0.5116166857	862.6468328	0.223077203	0.9798622838
5/1/20	1.820560012	0.08627378947	6.503883117	1.099240466	3.397843204	1.096179786	0.9448861333	891.9632883	0.3005900702	1.117095069
6/1/20	2.157282513	0.1048534013	6.495625812	1.323764513	3.708583404	2.009310874	1.480155581	925.2797438	0.4401029374	1.242327855
7/1/20	2.498005014	0.1154330132	6.568368507	1.52728856	3.866323605	2.449441961	2.032425028	1009.796199	0.5906158045	1.42456064
8/1/20	2.621727515	0.122012625	6.692111202	1.803812606	3.951063805	2.802573048	2.550694476	1114.012655	0.7511286717	1.757793426
9/1/20	2.805450016	0.1265922368	6.762853897	1.970336653	4.060804005	3.054704136	3.022963924	1143.62911	0.9086415389	2.133026211
10/1/20	3.092172516	0.1341718487	6.849596592	2.3258607	4.225544206	3.238835223	3.510233371	1165.645566	1.077154406	2.440258997
11/1/20	3.410895017	0.1427514605	6.675339286	2.600384746	4.432284406	3.441966311	4.043502819	1190.362021	1.257667273	2.792491782
12/1/20	3.806617518	0.1453310724	6.709081981	2.900908793	4.670024606	3.569097398	4.568772267	1206.478477	1.46118014	3.254724568
1/1/21	4.202340019	0.1729106842	6.853824676	3.07743284	4.892764806	3.710228485	4.959041714	1214.894932	1.672693008	3.649957353

Figure 1. CUSUM Change Detection Model

To determine whether there was a statistically significant difference in inflation on highly processed foods vs. less processed foods, we utilized hypothesis testing. We started by creating some exploratory plots of average price of different foods over time. As you can see, most food's prices are continually rising with sharp jumps in the last few years (Figure 2). Additionally, we explored the macronutrient dataset to see if this change coincided with any changes in the nutrients available to US citizens. Surprisingly, the supply appears relatively consistent (Figure 3). Finally, we categorized the foods into processed and less processed foods using the NOVA system of food categorization and plotted their group averages in Figure 4. As you can see, generally, processed foods are indeed more expensive.

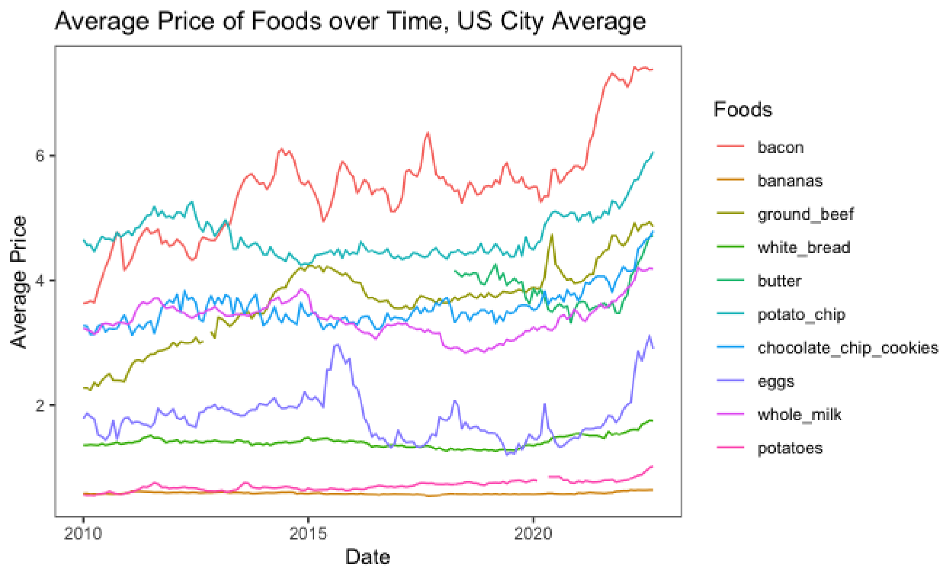

Figure 2. Exploratory Visualization of Food Price Over Time

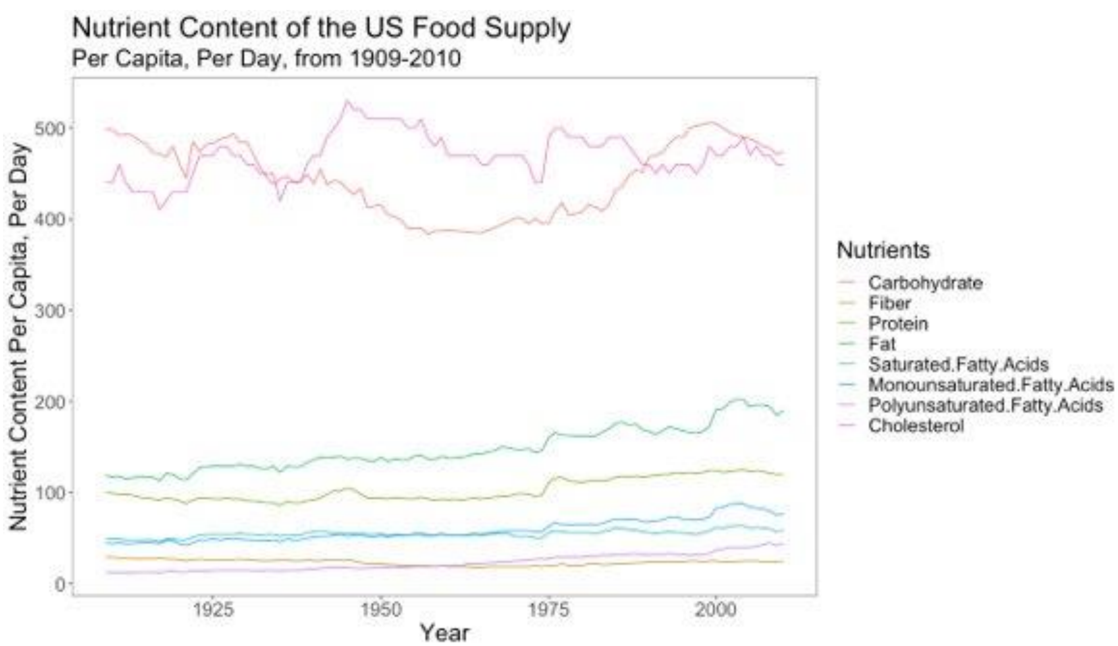

Figure 3. Exploratory Plot of Available Macronutrients

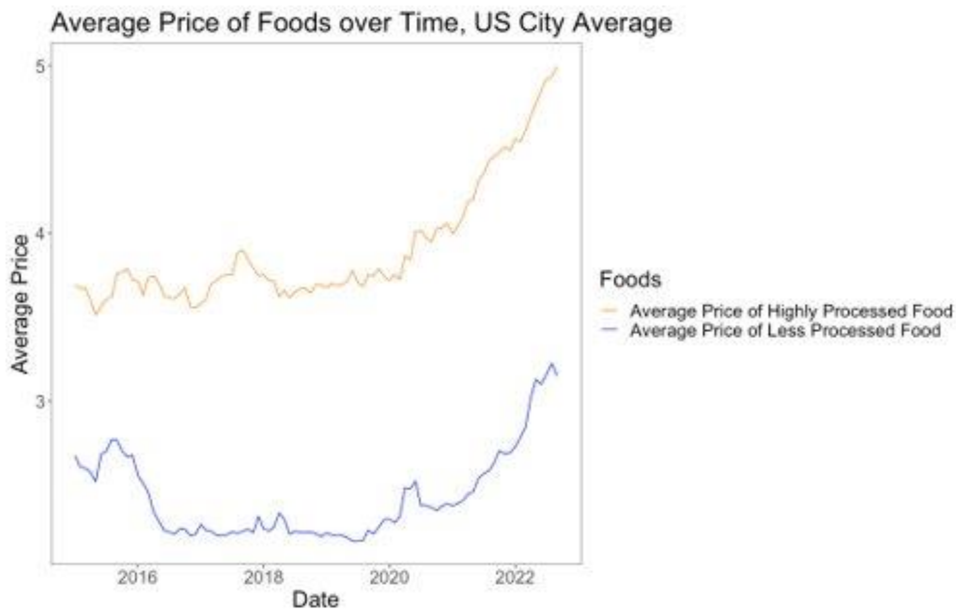

Figure 4. Processed Versus Unprocessed Average Price

We then prepared the data for hypothesis testing through a lag-transform. Lag transforming is often used with time series data due to the autocorrelation that happens from week to week. We started with a lag difference of 1, then 2, and found that a difference of 3 lags resulted in normal data. To confirm the data was approximately normal, we plotted its QQ plots (Figure 5 and 6) and applied the Shapiro-Wilkes test for normality, which both groups (Ultra-Processed and Less Processed) passed. We then did hypothesis testing using ANOVA to analyze this quantitatively. Despite our assumptions, hypothesis testing showed that there was no significant difference in the inflation rates of these two food categories. Notably, our small sample size of foods may have contributed to this; perhaps with more data the analysis would yield different results. Additionally, with further data from the end of 2022 and beyond, we may see significant changes in our conclusions.

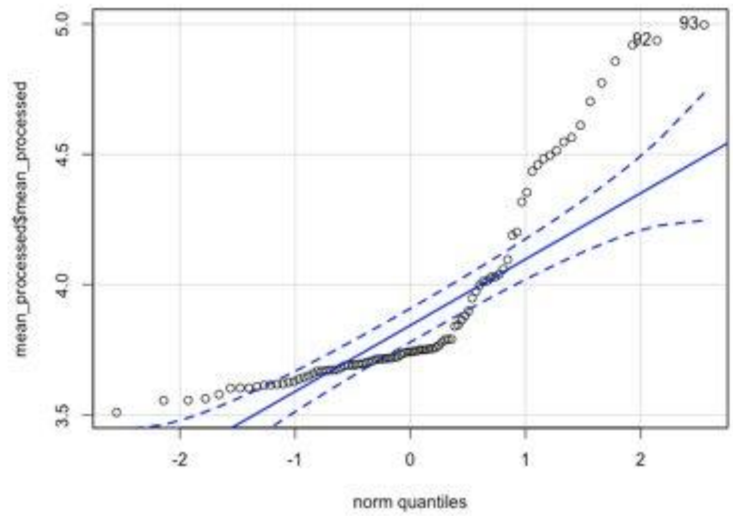

Figure 5. QQ Plot Before Lag Adjustment for Ultra-Processed Foods

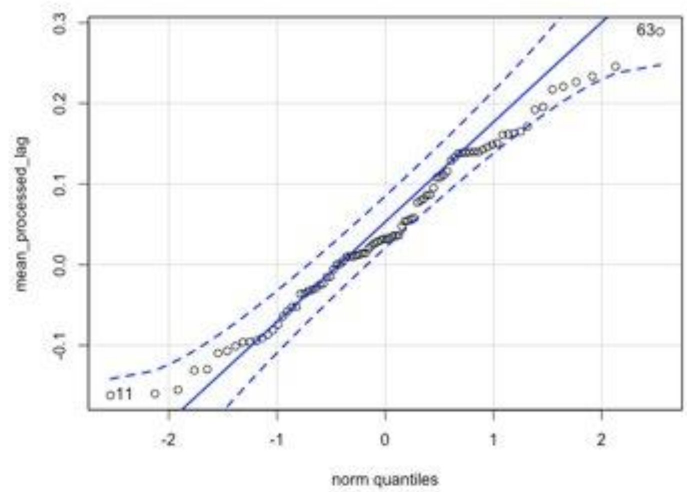

Figure 6. QQ Plot After Lag Adjustment for Ultra-Processed Foods

The last model we did was regression to look for correlation between the prices of food items and overall grocery spending. We fit a linear regression model using specific food prices to explain variability in overall grocery consumption. Based on the lack of statistical significance of several coefficients, variable selection was deemed appropriate. This was accomplished via bidirectional stepwise regression. We found that ground beef, white bread, chocolate chip cookies, eggs, whole milk, and potatoes were significantly significant and therefore, their prices had the biggest impact on grocery sales. The initial and final models are displayed below.

```
Call:
lm(formula = GrocerySales ~ bacon + bananas + ground_beef + white_bread +
    butter + potato_chip + chocolate_chip_cookies + eggs + whole_milk +
    potatoes, data = combined)

Residuals:
    Min      1Q  Median      3Q     Max
-3220.3  -634.7   -42.9   740.2  3734.7

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)               180.3     4043.0   0.045 0.964485
bacon                    -540.1      834.5  -0.647 0.518512
bananas                 -4344.6     6438.0  -0.675 0.500856
ground_beef              3427.8      860.6   3.983 0.000108 ***
white_bread             13574.2     3298.9   4.115  6.5e-05 ***
butter                    393.3      388.8   1.012 0.313450
potato_chip               945.3     1290.0   0.733 0.464888
chocolate_chip_cookies   3191.8     1166.5   2.736 0.006997 **
eggs                    -2322.7      806.3  -2.881 0.004575 **
whole_milk               2257.7      746.4   3.025 0.002946 **
potatoes                 6065.1     3637.6   1.667 0.097614 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1399 on 144 degrees of freedom
Multiple R-squared:  0.8719,	Adjusted R-squared:  0.863
F-statistic:    98 on 10 and 144 DF,  p-value: < 2.2e-16
```

Figure 7. Initial Model Results

```
Call:
lm(formula = GrocerySales ~ ground_beef + white_bread + chocolate_chip_cookies +
    eggs + whole_milk + potatoes, data = combined)

Residuals:
    Min      1Q  Median      3Q     Max
-3121.5  -619.1  -112.9   744.6  3672.9

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)              1097.5     2904.9   0.378 0.706106
ground_beef              3652.7      652.9   5.595 1.04e-07 ***
white_bread             12765.4     2208.3   5.781 4.26e-08 ***
chocolate_chip_cookies   2937.7     1078.7   2.723 0.007239 **
eggs                    -2521.2      668.4  -3.772 0.000234 ***
whole_milk               2526.3      696.0   3.630 0.000390 ***
potatoes                 6131.7     3474.2   1.765 0.079642 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1390 on 148 degrees of freedom
Multiple R-squared:  0.8701,	Adjusted R-squared:  0.8648
F-statistic: 165.2 on 6 and 148 DF,  p-value: < 2.2e-16
```

Figure 8. Final Model Results

Both models had high explained variability, which is useful in demonstrating that inflation is likely impacting grocery sales. Specifically, the price changes in white bread, ground beef, chocolate chip cookies, eggs, whole milk, and potatoes have had a statistically significant effect on overall grocery spending.

Conclusions

The results from hypothesis testing imply that inflation is affecting all foods. It is not necessarily affecting certain types, like processed foods, more than others. Therefore, this indicates that our initial hypothesis that inflation is affecting processed foods heavily was incorrect. Regression models indicate that price changes in white bread, ground beef, cookies, eggs, whole milk, and potatoes are significant in predicting overall grocery sales. Our change detection showed us that there was a positive change in grocery spending in May 2020. As of right now, it seems like consumers are not changing their spending habits due to inflation. They are still buying their normal items and in turn they are spending more money at the grocery store. This indicates that we also were wrong in our initial hypothesis of there being a change in consumer spending.

Potential Further Analysis and Business Impacts

Traditional economic theory would suggest that with the increase in inflation, we should be seeing effects in consumer spending. As we can see from our analysis, we have not seen those effects yet. With more time, we would love to dive into why we aren't seeing those effects and if inflation continues to increase, when we would observe changes in spending habits. We would also love to repeat the analysis with a larger sample size of highly processed foods and less processed foods. We would also like to perform forecasting on inflation of highly processed foods and less processed foods to better understand the potential future of inflation for these two categories.

We originally said that if our hypotheses were correct, health insurance companies could find ways to make less processed foods more affordable for their customers. This would help keep healthier options available in the face of high inflation. Also, because bad health leads to greater expected payouts for these companies, it would save them money by having less health insurance claims. Unfortunately, our hypotheses were incorrect. With that in mind, we think that health insurance companies should continue to monitor and analyze this situation to be preventative in unhealthy habits for the people they insure.

Knowing that certain types of food have price fluctuations that result in consumption changes, insurance companies could make various health campaigns during points that these foods have price fluctuations. Historically, it would be interesting to see if they saw any detrimental patient impacts during the time we detected noticeable changes in consumption of groceries, for example if those periods of time lead to increased obesity and therefore would want to put in preventative measures to hopefully save them possibly expensive health senarios.

Works Cited:

Gore, A., Harmer, P., Pfitzer, M., Jais, M. (2017). Can Insurance Companies Incentivize Their Customers to Be Healthier?. *Harvard Business Review*. https://hbr.org/2017/06/can-insurance-companies-incentivize-their-customers-to-be-healthier

(2022). Chronic Disease Fact Sheets. *National Center for Chronic Disease Prevention and Health Promotion*. https://www.cdc.gov/chronicdisease/resources/publications/factsheets/nutrition.htm

(2021). Americans Are Eating More Ultra-Processed Foods. *NYU*. https://www.nyu.edu/about/news-publications/news/2021/october/ultra-processed-foods.html#:~:text=Ultra%2Dprocessed%20food%20consumption%20grew,end%20(2017%2D2018)

(2022). Processed Foods and Health. *Harvard Health School of Public Health*. https://www.hsph.harvard.edu/nutritionsource/processed-foods/#:~:text=Often%20referred%20to%20as%20convenience,like%20heart%20disease%20and%20diabetes.

Moubarac JC. (2017). Ultra-processed foods in Canada: consumption, impact on diet quality and policy implications. *University of Montreal.* https://heartstrokeprod.azureedge.net/-/media/pdf-files/canada/media-centre/hs-report-upp-moubarac-dec-5-2017.ashx?la=en&rev=c844895571d0415cb2e0ab96e3512a46

(2018). The NOVA Food Classification System. *EduChange*. https://educhange.com/wp-content/uploads/2018/09/NOVA-Classification-Reference-Sheet.pdf

Appendix

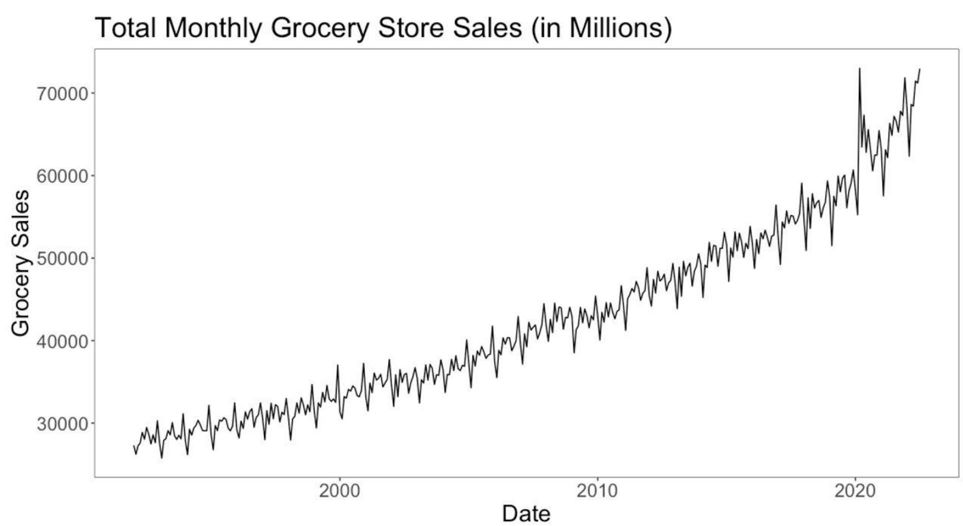

Figure 9. Grocery Sales Over Time