Web Analytics – Grocery Store Price Comparison

Team:

Tyler (Ty) DiNapoli-Chiappelli MSBA,

Timothy (Tim) Welles MSBA,

Yinzhe (Ricardo) Lu MSBA,

Nicholas (Nick) Mercogliano MSBA

Fordham Graduate School of Business

December 2020

Executive Summary

The grocery store industry is a great example of monopolistic competition so that there are still some differences across various companies in regard to prices, products, and quality. Any one grocery store has an acute interest in knowing, across all their products, how they compare to competitors.

This project aims to gather timely data corresponding to the prices of groceries from major competitors' websites and looking at how the prices vary across the spectrum. This tool allows our client to make informed changes in regard to pricing schemes, marketing efforts, and any other business ideas in order to get as close to their maximum profit margin.

In 4 weeks[1], we collected pricing information from 4 major grocery stores: Whole Foods, Walmart, Wegmans and Stop and Shop from 5 sections: Dairy, Fresh Fruit, Fresh Vegetables, Meat, and Seafood, by using web crawling tools selenium and beautiful soup. We chose these stores because they represent 4 different demographics when looking at the grocery store landscape[2].

Based on the data analysis, we learnt that Whole Foods constantly had the highest price in Dairy and the lowest sale. Walmart, on the other hand, had the lowest price per unit on these sections. Wegman has the highest average prices for seafood and meat. Last but not least, the number of products on Sale reached a high point on Thanksgiving Day and dramatically dropped during the recess.

In the future, if another research is going, we would recommend:

- Conduct the crawling process for at least two weeks to better understand the market research
- As most website cannot only use beautiful soup, try to use Selenium first
- Adapt the code to track real time data, or crawl the website at the same time each day
- Focus on one specific region to narrow on geographic scope

[1] From November 5, 2020 to December 5, 2020
[2] For details, see next section Business Analytics

Business Analysis

As mentioned above, we chose Whole Foods, Walmart, Wegmans, and Stop and Shop as our four stores to profile. We chose these four because they represented different sections of the grocery store environment. Walmart is your end-all-be-all big box store where you can get everything from flat-screen TVs to produce. Wegmans covers your regional grocery store, primarily located in upstate New York. Stop and Shop cover the regional "big box" space focusing on the northeast. Then there is Whole Foods falling into the specialty space as they target healthy living shoppers.

We decided to look at these four to get an idea of what type of model our client should try and emulate, because at this point there is nothing new that is happening in the space apart from reworking the actual looks of stores. From here we had to decide what products we had to focus on.

We decided to narrow our search to what we thought were the five staple products across the board: Dairy, Fresh Fruit, Fresh Vegetables, Meat, Seafood. As noted above Dairy, Fruit, and Vegetables remained consistent within each store, but we did see a difference in price across them. Again, as noted above, Whole Foods had the highest prices per unit across each section, while Walmart remained the cheapest option out of the four. We found this interesting because both Walmart (#1) and Whole Foods (#10) were both in the top 10 grocery stores based on revenue.[3]

With this in mind, we started to make assumptions about why Walmart's revenue eclipsed Whole Foods by more than $200 billion in 2019. Our main assumption was that since Walmart offered other items because of their size, they were able not worry about margins as much as Whole Foods would. The mark-up price on a flat-screen TV could cover the cost of multiple items. That being said, the more amazing thing is that Whole Foods made the top 10 by doing $16 billion across 500 stores. That is roughly $32 million per store while it took Walmart almost 4,300 stores to double that number (just shy of $68 million per store).

The graphs below will show prices across these four stores throughout the five sections that were mentioned above. They will look at price over time (individual and comparison), any trends that came up regarding pricing of certain products, Sales vs no sales, etc. From here we

[3] https://www.foodindustry.com/articles/top-10-grocers-in-the-united-states-2019/

will give our client the best road path to success in regard to following the blueprint of any of these stores.

Crawling Techniques

When we started this project, our initial thought was to use Beautiful Soup to help with the crawling, but we ran into trouble because the websites appeared to be too complicated for it. So, we had a quick crash course in Selenium and hit the ground running.

Working with Selenium was great because most of us encountered pop-ups once we ported into each of our respective sites. With a few simple lines of code, we were able to just "x" out of it and proceed to rip the page of data. From here it was just straightforward Python with multiple "for" loops to help with the cleaning every step of the way.

There were some other interesting/smart uses of Selenium throughout the project. The first example would be regarding the Whole Foods code. Since the page kept expanding when you reached the bottom, code was written to continue to scroll until the bottom of the page was met and then the data would be ripped. Allowing for the most data to be gathered from each section at Whole Foods.

Another was a workaround for sites that weren't set up like Whole Foods. Some group members were able to write code to have it navigate to the next section within the same browser, while others had to close and reopen a new browser on the correct section.

The real important thing for us was to be able to clean the data while we were ripping it. We figured that the end product would not only be better, but it would be easier for us to work with if it was as close to clean as possible. Within the first few days there were of course some hiccups, but we ironed everything out and ended up with a great dataset across these four stores.

One of the first things we had to do was figure out the metrics that we wanted to focus on. From here we could write code to strip the numerical data out of the strings that made up the product descriptions. These gave us the metrics of Price, Quantity, Units, and Price Per Unit. This also allowed us to further clean the data for one off products that could potentially mess up the dataset as a whole.

Standardization and Dataframe Congregation

Our team crawled pricing data every day from the 5 sections (mentioned above) for each of the grocery store websites over a period of 4 weeks. Each team member was responsible for one store's data collection. From there, each member crawled and saved all the data collected in folders that were broken up by their sections.

After data collection, we standardized the weights and liquid volume measurements into "each (EA)" and "pounds (LBS)" [4]. We have listed our unit conversions below. We used the below figures to create a new feature called "price per unit":

```
gallons_to_pounds = 8.35
liters_to_pounds = 2.205
ounces_to_pounds = .0625
pints_to_pounds = 1.043
quarts_to_pounds = 2.09
```

From here, our next step was to congregate all the data two-fold. First, we had to combine our five separate grocery section data (i.e., Dairy, Fruit, Veggie, Meat, Seafood) into one master store csv. Then we each took our individual master files and created the total master file containing all the data from the above five sections at each of our stores (i.e., Wegmans, Whole Foods, Walmart, Stop and Shop). The next step was taking the data into Tableau to provide the below visuals to present to our "client."

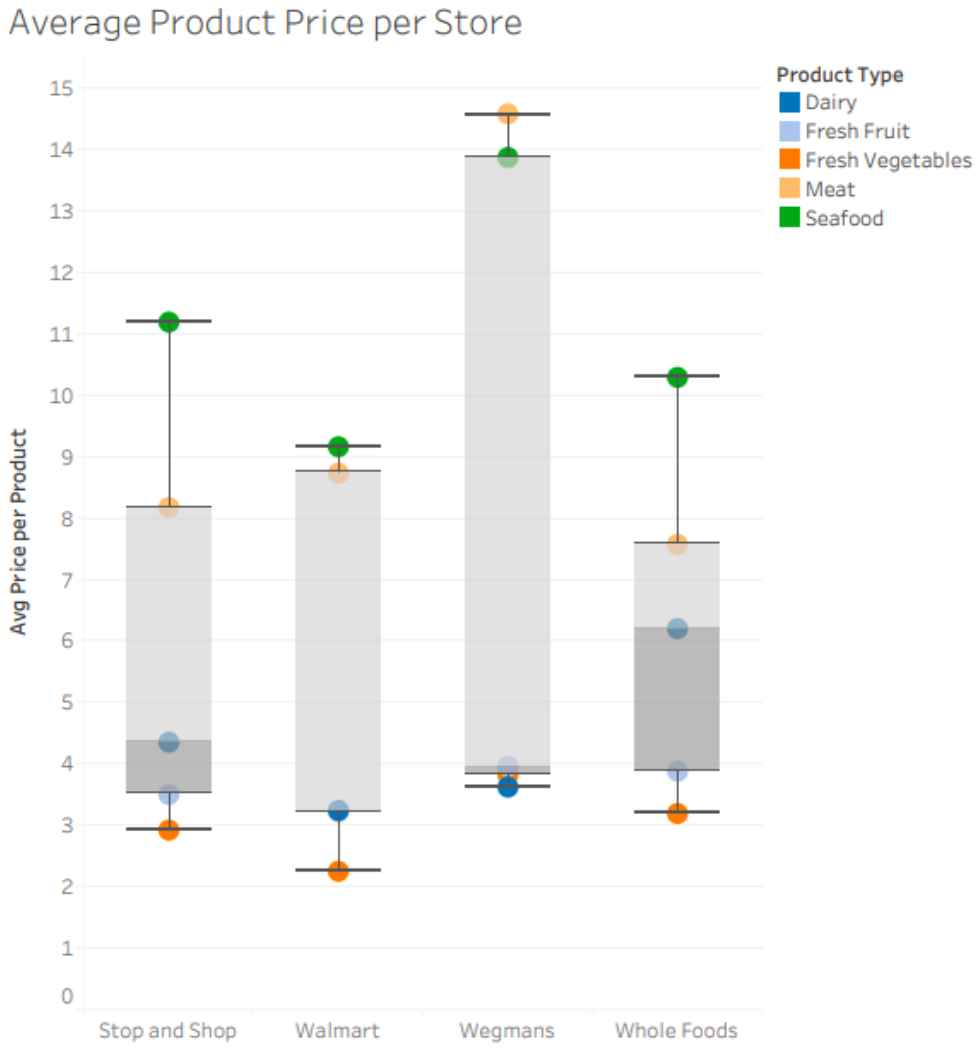
Data Interpretation

Among five separate grocery section data (i.e., Dairy, Fruit, Veggie, Meat, Seafood), the box plot below describes the median product section prices distribution over a month period. In other words, we take the median prices of certain product types into account of each store. Whole Foods has the highest median average prices[5] of about $6, which is double the lowest median average prices of about $3 in Walmart. Among other stores, Stop and Shop has slightly

[4] Our conversion of liquid volume to pounds is based on water's feature. We realized that the measurement could potentially cause confusion and affect the results and comparisons. We grouped the product by their types to prevent this issue.

[5] To clarify median average prices, we take the median prices of a product type in a month period and calculate each section's prices within a store.

higher median average prices than Wegmans. However, we find a skewed distribution of average prices within Wegmans and Walmart. Wegmans has the most extreme distribution of average prices. Average prices of lower 3 sections are in a range of $3.5 - $ 4, while another 2 of their sections are widely distributed within $4 - $14.5. Surprisingly, Wegmans also has their lowest price and highest average prices above all other stores. With the highest median average, Whole Foods has the narrowest price distribution of 5 sections. Even Stop and Shop has their highest price section higher than Whole Foods'.

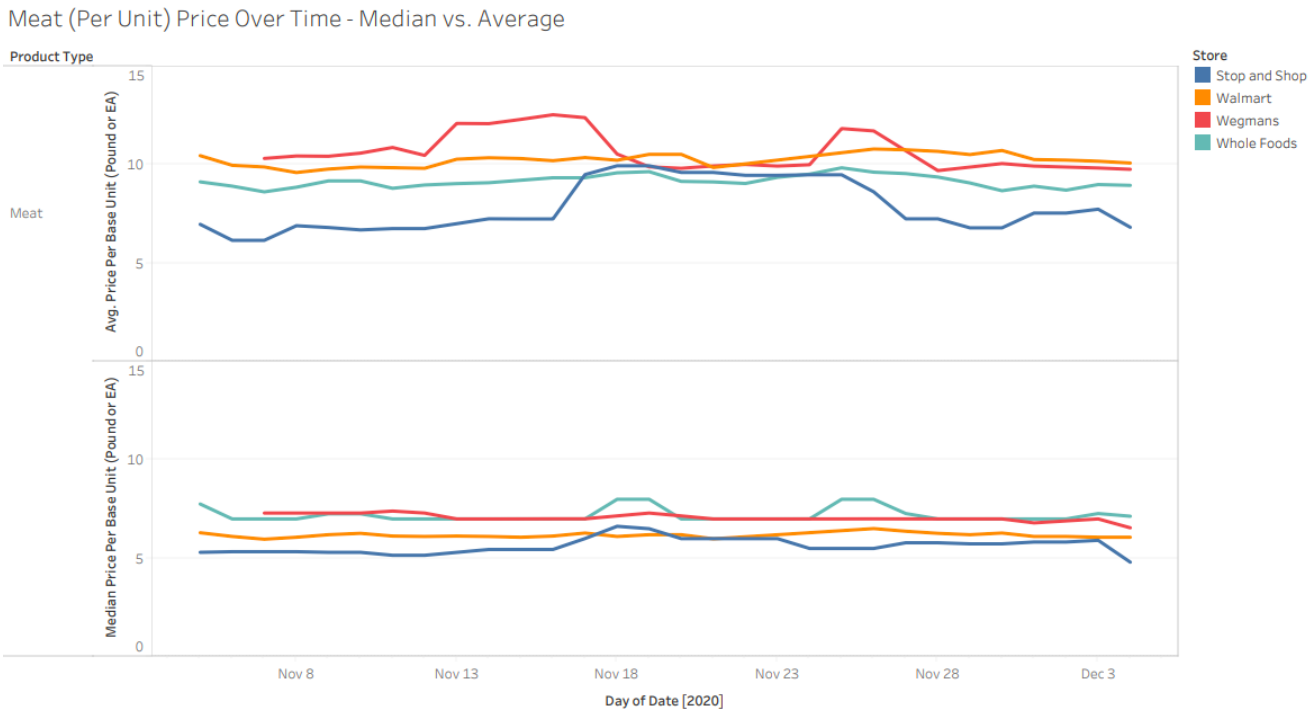


To have a closer look at the product types, the bar chart below shows the average product price of 5 sections in each store. Among all grocery stores, meat and seafood were the most expensive sections. Wegmans had the highest average price of meat and seafood sections throughout the month. Whole Foods has the highest average price of dairy, which is almost double the prices of the dairy section in Stop and Shop and Wegmans. Among all stores, Fresh

Fruit and Fresh Vegetables are two sections that have consistently low average prices that are below $4.

One of the main reasons we looked at both the meat and seafood sections of each store, is because we knew that this would be where some of the big-ticket items would be. With this comes the potential for big margins and big profits. For Wegmans in particular they had some meat products that cost upwards of $50, and seafood to match it.

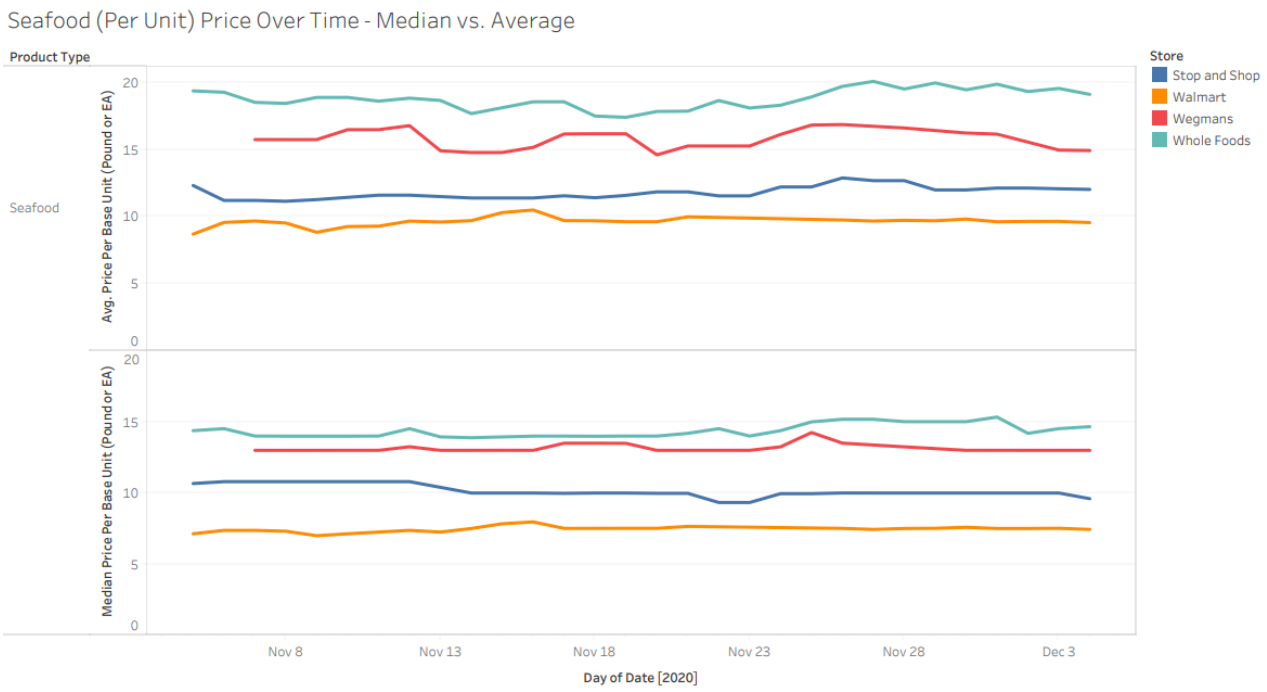
That being said, in looking at the graph below, you will see that the average price of meat is pretty consistent across the board (roughly $10/lb.). It is also interesting to note some of the fluctuation for both Stop and Shop, and Wegmans. One could make the assumption that these spikes and plateaus are a correction back to normalcy (regarding price). Or, it could have been a stocking issue. Much like in the beginning of quarantine when there was a run-on chicken, and chicken prices skyrocketed while quantity decreased, the same could be happening here. But with all that in mind, it is still interesting that across all four stores, the price per pound remains pretty consistent.



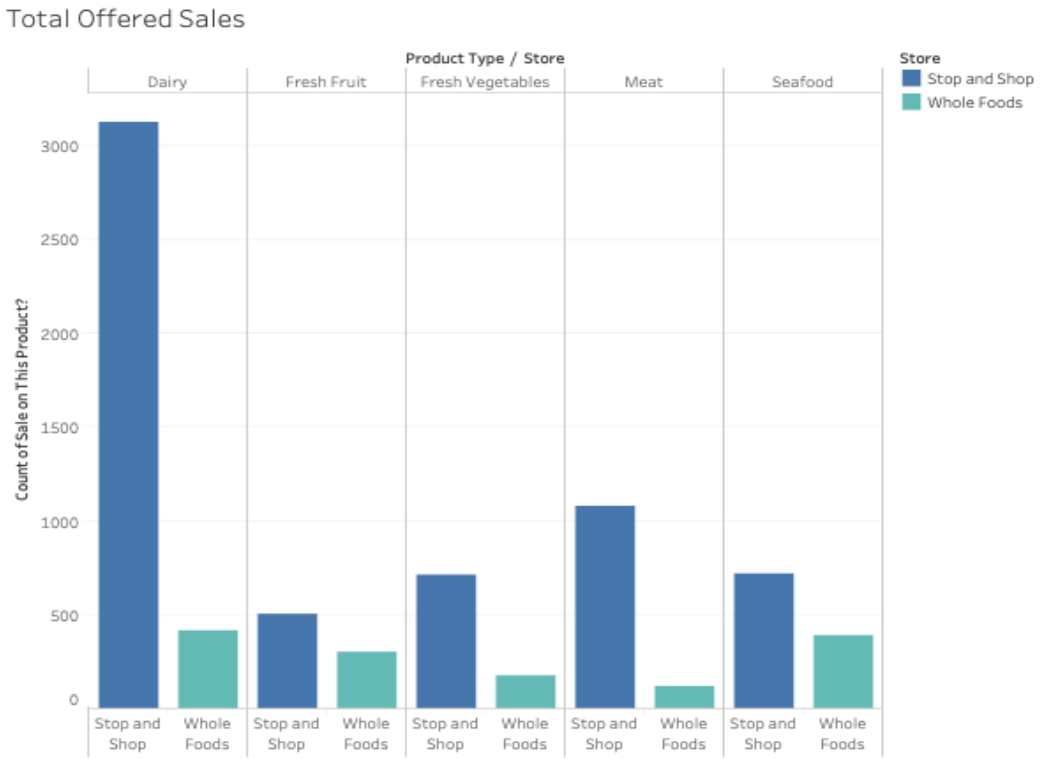
Now we move to seafood! No surprise, Whole Foods (the healthy/fresh choice) has the highest priced fish, while Walmart has the cheapest. Here is where the idea of margins comes into play. We would assume that since Walmart can offset the cost of fish with sales of massive TVs and other high budget items, their price can very easily be the lowest.

Whole Foods, on the other hand, is trying to be the "purpose-driven company that aims to set the standards of excellence for food retailers."[6] Because of this they have to strive for the highest quality products that they can offer. They need to squeeze every bit of margin that they can out of each product that they sell. While this is less successful than Walmart's approach, it is still working for them as they netted about $16 billion dollars' worth of revenue across only 500 stores.[7]

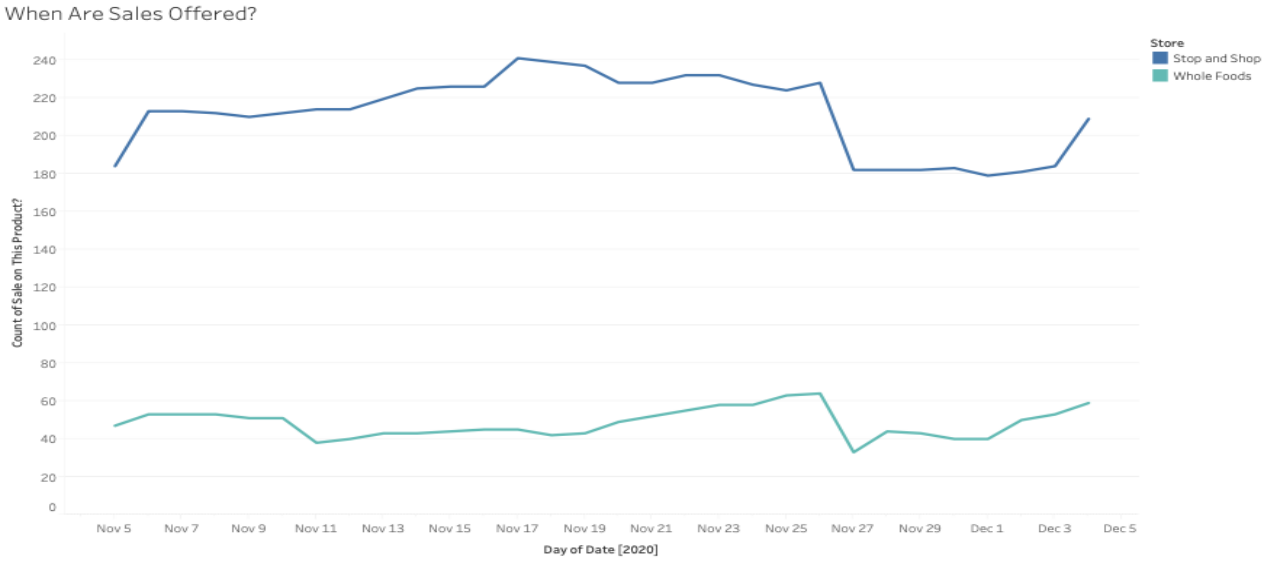
[6] https://www.wholefoodsmarket.com/mission-values
[7] https://www.foodindustry.com/articles/top-10-grocers-in-the-united-states-2019/



Among 4 stores, only Stop and Shop and Whole Foods had their sales information on their product (information shows below). We see that stop and shop has more products on sale than Whole Foods in all product types. Dairy products in Stop and Shop have most sales items, over 6 times of Whole Foods'.



Our project also provides information on what was the best time to buy groceries, before or after holidays. This year's Thanksgiving recess was on November 26-November 29, 2020. From the chart below, both stores had a relatively high count of sales on the first day of Thanksgiving. The number of products on sale had a sudden drop on the second day of the Thanksgiving recess. On days after Thanksgiving recess, the counts of sales were lower than the count on Thanksgiving night.



Insight & Suggestion

Our grocery store price comparison project is a good guide for setting up pricing strategy and sales arrangement. We recognize that pricing strategy depends on multiple dimensions, such as food standard, freshness and season. Generally, we suggest that local small grocery stores should set their grocery prices between Walmart and Whole Foods. Our data showed that Walmart has the lowest price on average among almost all sections.[8] We think that Walmart is a comprehensive retail chain, whose grocery prices are very attractive to customers, but what makes them different and hard to emulate is that they carry so many other items. These items allow Walmart to not go chasing after the margins in all of their sessions that somewhere like Whole Foods or another small grocery store would have to. Their prices are clearly in lack of profit margin for smaller, local companies. Whole Foods has a name for its *Food Ingredient*

[8] The only one exception is their meat section, which was over the prices of Stop and Shop and Whole Foods by about $1.

Quality Standards, targeting shoppers that are willing to pay more for healthier food. Therefore, Whole Foods's prices set a standard for the amount that customers are willing to pay for healthier food.

In regard to the four stores that we looked at, they have different focuses. Wegmans and Stop and Shop are more regionally focused. They are normally found across the eastern seaboard. Because of this, their reach is limited. That being said, they do a great job supplying the towns/communities that are sometimes focused on them. A perfect example of this is Pittsford which is where the largest Wegmans is located. I mention this because, even though Pittsford is the largest location, it pales in comparison to your average Walmart in both size and price point. So, with this in mind, we would have to suggest going either the Walmart route, or follow the Whole Foods blueprint. To further back up this point, both Walmart and Whole Foods were ranked in the Top 10 Grocers by revenue for 2019 (1 and 10 respectively) so they must be doing something right.