Dianna Sinicrope  
DSC530 Final Project  
8/11/19  
Analysis of Avocado Sales

The dataset I found represented data collected on avocado prices from 2015 to 2018. The Haas Avocado Board collected the data.

My questions about the data were reasonably straightforward. I was curious as to whether organic avocados are more expensive than conventionally grown avocados (and if so, by how much?). I was also curious as to whether the price of organic avocados correlates with the price of conventionally grown avocados. As a starting point, I was interested in learning about the dataset in general and was open to other questions that might arise along my EDA journey.

While the actual EDA process for me ended up being filled with a lot of wrong turns and going down rabbit holes (which was mostly expected), I was able to answer my questions after some digging. For my first question, I found that, yes, organic avocados do cost more than conventional avocados. This wasn’t a surprise, and on average, an organic avocado costs $0.50 more than a conventional avocado. The thing that surprised me here, though, was that while organic avocado prices have increased over time, conventional avocado prices have risen more drastically. Narrowing down on the data, I found that both types saw an increase in 2017 over 2016, but conventional avocado prices increase by +17% that year, and organic prices only increased by +12% YOY.

2017’s success in the avocado prices world piqued my curiosity too since, in 2018, conventional prices and organic prices both dropped an equal -80%. My first hypothesis was that the 2017 vs. 2018 differences were a response to supply and demand. A quick Google search brings up a 2017 CNN article citing, “Two key drivers are behind the summer price surge: A weak harvest this year and insatiable demand for avocados” (Gillespie, 2017). They go on to say that the average American eats 7 lbs. of avocado a year, up from 1989’s 1.1lbs. It’s great that through my EDA I was able to pinpoint this data and then conduct additional research on why it happened.

Organic and conventional avocado prices also have a strong correlation. 76% of the changes in organic pricing are also seen in conventional avocado pricing. I would hypothesize that this is also due to the same supply and demand issue that emerged in my last question. I assume that conventional and organic avocado consumers are two different target markets, perhaps those who buy organic vs. those who don’t. However, while those two types of consumers are willing to pay separate prices for organic and non-organic produce, they both share an interest in buying avocados, and that shared interest seems to wax and wane similarly with seasonality.

The insights from my research could be applied to broader industries. The grocery store industry could use this study to provide insights on what they should and shouldn’t be providing for their customers. The media could see that they were correct in reporting on the “avocado craze” (or at least that’s what I’m calling it, since I feel like I saw plenty of headlines referring to avocados, avocado toast, and Millennials and avocado toast in 2017). They could also use my time series findings about the volume of avocados sold to wonder what other seemingly mundane purchases they can highlight as a cultural phenomenon. Avocado growers could use my volume time series analysis to realize that the best time of year for avocado purchases is in February (although, I do assume that they already know this and I assume that this is peak season for avocados). Farmers and other food industry companies could use this to their advantage. Perhaps Tostitos needs to make sure their supply chains are in functioning well in February due to the large number of avocados being purchased. Perhaps bread companies should advertise avocado toast more heavily every February. The possibilities are many.

In my EDA, I also found another interesting piece of information, even though it wasn’t within the scope of my original question. I found that in general, the number of avocados sold and the price of avocados have a negative correlation. The number of avocados sold tends to increase when the price decreases (and when the price increases, the quantity sold drops). This is also something I would have probably assumed ahead of the EDA, but it was reassuring to see the data validated by a correlation test.

I felt confident with the variables that I had access to. However, looking back, I think I would have liked to have dived more deeply into some of the region-specific sales data. While the code and chart didn’t make the cut for my PowerPoint presentation, I did find some avocado prices in specific places that were seemingly high compared to most other places. It wasn’t quite within the scope of this project, but perhaps for another project, I could address why prices in certain cities are so much higher, considering all else equal. And as I hinted before, it might be interesting to pair this data with data from other food brands (like Tostitos, for example) to see if changes in sales for other foods are caused by the seasonality of avocados.

I was careful not to make any assumptions about the data ahead of time, but I did hit some roadblocks along the way. The thing I had the most trouble with was segmenting my data. The “type” variable (for organic vs. conventional avocados) was in one column, and even though I used the pd.pivot\_table() method to get my data arranged the way I wanted it, I still ran into some issues, particularly with the regression analysis part of my project. I tried both the Think Stats method for linear regression and a Matplotlib method and wasn’t able to get them to work. In the end, I calculated it in Excel since I felt like the data would be beneficial to my overall analysis. I would have preferred to be able to handle the entire report in Python.

References

Gillespie, P. (2017, September 9). Avocado prices have soared 125% this year. Retrieved from https://money.cnn.com/2017/09/09/news/economy/avocado-prices-soar/index.html

*Volume and Price Data*[CSV]. (2018). Haas Avocado Board. Retrieved from https://www.kaggle.com/neuromusic/avocado-prices