

Predicting Per-Unit Prices of Hass Avocados

Yunyao Zhu

3/1/2021

Introduction

In 2017, an Australian property developer claimed that millennials were spending too much money on avocado toast instead of saving for their first home (CNNMoney 2017). While the millennial avocado toast stereotype is by no means convincing, avocados are indeed a favored fruit in the U.S. According to a market report (USAID 2014), the U.S. is the world's biggest importer of avocados. As more health benefits about avocados are discovered and promoted, demand surged even more and so did the price. The goal of this project is to investigate whether we can reasonably predict the price of avocados based on past prices and geographical data.

Existing Literature

A study (Evans 2009) investigated the various factors that might influence the avocado market. The researchers predicted that the avocado prices would likely decrease in the 2009-2010 season due to an increase in supplies and a probable decrease in demand caused by the financial crisis. Since the study was conducted over ten years ago, it seems be interesting to examine if the avocado market has changed over the years and if we can still predict the avocado prices given its supposed volatility in recent years.

Data

The dataset used in this analysis comes from the publicly accessible Kaggle Avocado Prices dataset, which credited the Hass Avocado Board for the collection and release of the data. This dataset contains 13 variables encompassing the per-unit prices, total volumes, regions, and sizes of Hass avocados, a cultivar of avocados. Data from the start of 2015 to the end of the first quarter of 2018 are available. Each entry represents one observation from one region in the U.S. during one week. There are over 18,000 entries in total.

There are 54 unique regions in the dataset. One of these regions, denoted by `TotalUS`, encompasses the avocado volumes in the country as a whole. Another 8 regions corresponds to large geographical divisions of the U.S. such as the `Northeast` region, `Southwest` region, etc. The remaining regions represent smaller geographical areas in a less consistent way, sometimes referring to a single city in a state (e.g. `Chicago`) while other times denoting cities across states (e.g. `BaltimoreWashington`). Some regions can also be ambiguous - cities such as `Albany` and `Jacksonville` can be found in more than one state. Thus, to manage the location data consistently and unambiguously, we choose to focus on the 8 geographical regions.

To enrich the current dataset with demographic data, the `df_state_demographics` dataset from the `choroplethr` package (Lamstein 2021) is aggregated and combined with the mean per-unit prices of Hass avocados (over the per-unit prices each week from 2015 to 2018) in each of the 8 geographical regions. Selected demographic data include the total population in the region, the average percentage of white population and black population, the average per capita income, and the average median age. (The averages are weighted by the population in the states in each geographical region).

Research Goals

Predicting the per-unit prices of Hass avocados Specifically, we want to use the data from 2015 to the first quarter of 2017 to predict the avocado prices in the following year. Doing so allows us to compare the predicted prices with the recorded prices and evaluate the accuracy of the predictions. Predicting one year of avocado prices enables us to examine any seasonal variations in pricing.

Investigating whether spatial relationships exist Does the region (where the observations were made) have any significant impact on the per-unit price of the avocados? If such spatial relationships are found, what might be some potential implications?

Exploratory Data Analysis

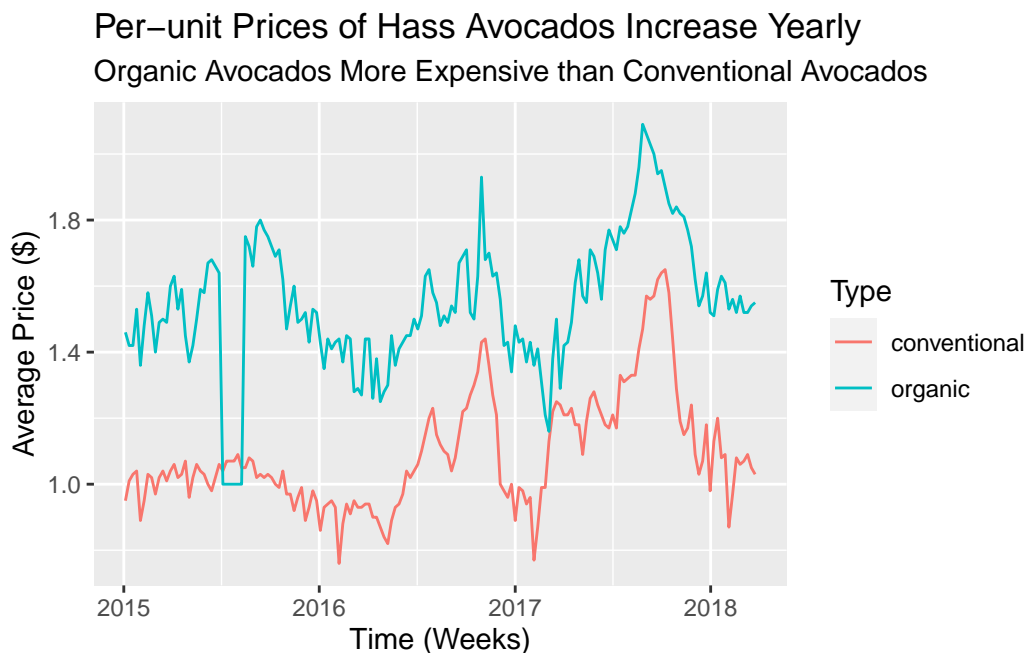


Figure 1: Average Prices of Hass Avocados Over Time in the U.S.

From the plot above, we see the the per-unit price of Hass avocados increases every year from 2015 to 2018. Seasonal trends within each year also seem to exist, with lower prices in first two quarters and higher prices in the last two quarters. Expectedly, the prices of organic avocados are almost always higher than those of conventional avocados, suggesting that the type of avocados can be a predictor in the price prediction model. Further analysis is required to investigate the sudden drop in price of organic avocados in mid-2015.

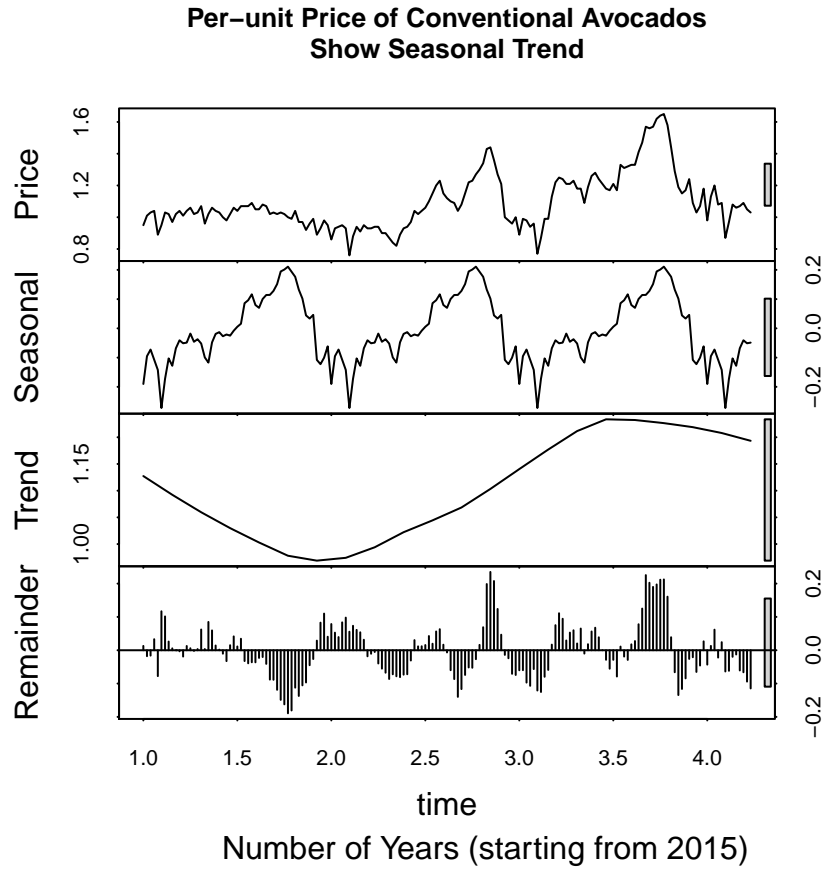


Figure 2: Time Series Plot of Per-unit Prices of Conventional Avocados in the U.S.

The plot above is the time series plot of the per-unit prices of conventional avocados in U.S. The plot of organic avocados shows very similar results. Compared to the Figure 1, this plot shows more directly the seasonal trends. We observe that the prices peak during the third quarter of each year and decrease to the lowest during the first quarter.

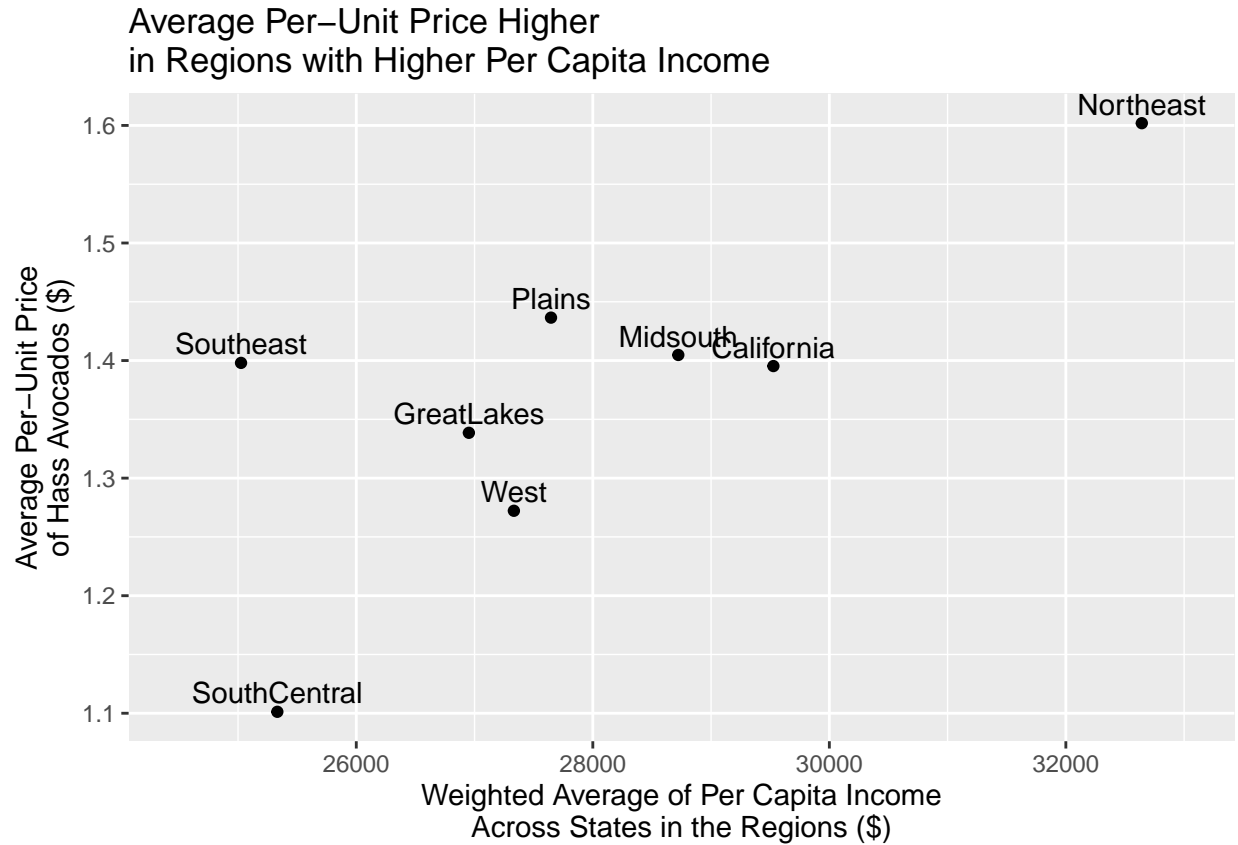


Figure 3: Scatter Plot of Per-unit Prices of Avocados v.s. Per Capita Income in Different Regions

In general, it seems that the regions with higher per capita income also have higher per-unit avocado prices. (Additional scatter plots show similar trends with higher median age and higher percentage of white population). An exception is the Southeast region, which has the lowest per capita income but a relatively high per-unit avocado price. Further analysis is required to better explain this observation.

Bibliography

CNNMoney. 2017. “Millionaire to Millennials: Lay Off the Avocado Toast If You Want a House.” *Cable News Network*.

Evans, Sikavas, Edward A. & Nalampang. 2009. “Forecasting Price Trends in the U.s. Avocado (Persea Americana Mill.) Market.” *Food Distribution Research Society, Vol. 40(2), Pages 1-10*. [https://doi.org/10.1016/S0140-6736\(17\)32812-X](https://doi.org/10.1016/S0140-6736(17)32812-X).

Lamstein, Ari. 2021. “Mapping Census Bureau Data in R with Choroplethr.” *United States Census Bureau*. <https://www.census.gov/data/academy/courses/choroplethr.html>.

USAID. 2014. “The Us Market for Avocado.” *ACCESO*. https://pdf.usaid.gov/pdf_docs/PA00KP28.pdf.