Case Study Report: Avocado Prices — Market Insights

Aakif Altaf

2025-10-07

Contents

---

1 Introduction

Business context: A grocery retail client wants to understand historical avocado pricing and sales patterns across U.S. regions to improve pricing, promotions, and inventory planning.

Primary business question:

How can historical avocado pricing and sales trends inform retail decisions on pricing, promotion timing, and inventory allocation?

Sub-questions we answer:

1. How have average prices evolved over time (overall and by type)?
2. How do prices and sales vary across regions? Which regions are highest/lowest price?
3. Is there seasonality (monthly/weekly patterns) in price and volume?
4. What is the relationship between price and total volume (does demand fall as price rises)?

Data: Import & Quick Snapshot

```
## Rows: 18,249
## Columns: 14
## $ x1            <dbl> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16~
## $ date          <date> 2015-12-27, 2015-12-20, 2015-12-13, 2015-12-06, 2015-11~
## $ average_price <dbl> 1.33, 1.35, 0.93, 1.08, 1.28, 1.26, 0.99, 0.98, 1.02, 1.~
## $ total_volume  <dbl> 64236.62, 54876.98, 118220.22, 78992.15, 51039.60, 55979~
## $ x4046         <dbl> 1036.74, 674.28, 794.70, 1132.00, 941.48, 1184.27, 1368.~
## $ x4225         <dbl> 54454.85, 44638.81, 109149.67, 71976.41, 43838.39, 48067~
## $ x4770         <dbl> 48.16, 58.33, 130.50, 72.58, 75.78, 43.61, 93.26, 80.00,~
## $ total_bags    <dbl> 8696.87, 9505.56, 8145.35, 5811.16, 6183.95, 6683.91, 83~
## $ small_bags    <dbl> 8603.62, 9408.07, 8042.21, 5677.40, 5986.26, 6556.47, 81~
## $ large_bags    <dbl> 93.25, 97.49, 103.14, 133.76, 197.69, 127.44, 122.05, 56~
## $ x_large_bags  <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.~
## $ type          <chr> "conventional", "conventional", "conventional", "convent~
## $ year          <dbl> 2015, 2015, 2015, 2015, 2015, 2015, 2015, 2015, 2015, 20~
## $ region        <chr> "Albany", "Albany", "Albany", "Albany", "Albany", "Alban~
```

Table 1: Data summary

Name	av
Number of rows	18249
Number of columns	14
_____	
Column type frequency:	
character	2
Date	1
numeric	11
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
type	0	1	7	12	0	2	0
region	0	1	4	19	0	54	0

Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
date	0	1	2015-01-04	2018-03-25	2016-08-14	169

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
x1	0	1	24.23	15.48	0.00	10.00	24.00	38.00	52.00
average_price	0	1	1.41	0.40	0.44	1.10	1.37	1.66	3.25
total_volume	0	1	850644.01	3453545.36	84.56	10838.58	107376.76	432962.29	62505646.52
x4046	0	1	293008.42	1264989.08	0.00	854.07	8645.30	111020.20	22743616.17
x4225	0	1	295154.57	1204120.40	0.00	3008.78	29061.02	150206.86	20470572.61
x4770	0	1	22839.74	107464.07	0.00	0.00	184.99	6243.42	2546439.11
total_bags	0	1	239639.20	986242.40	0.00	5088.64	39743.83	110783.37	19373134.37
small_bags	0	1	182194.69	746178.51	0.00	2849.42	26362.82	83337.67	13384586.80
large_bags	0	1	54338.09	243965.96	0.00	127.47	2647.71	22029.25	5719096.61
x_large_bags	0	1	3106.43	17692.89	0.00	0.00	0.00	132.50	551693.65
year	0	1	2016.15	0.94	2015.00	2015.00	2016.00	2017.00	2018.00

Brief note on data quality & limitations:

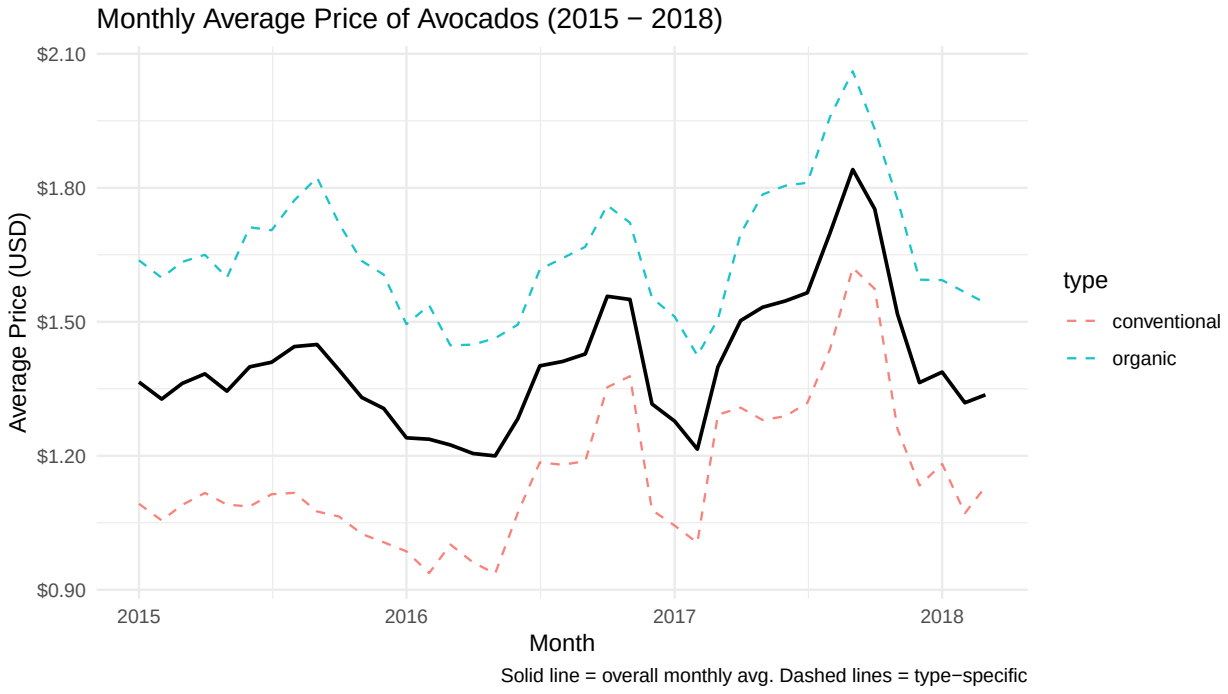
Data covers 2015–2018 across many U.S. regions; it's observational retail data.

Potential limitations: local promotions, missing marketing variables, no direct supply/wholesale price columns, and region aggregation hides within-region store variation. Treat correlations as directional signals, not causal proof.

2 Analysis & Insights

1) Time trends — Average price over time (overall and by type)

Question: Has the average retail price changed materially over time? Any divergence between organic vs conventional?



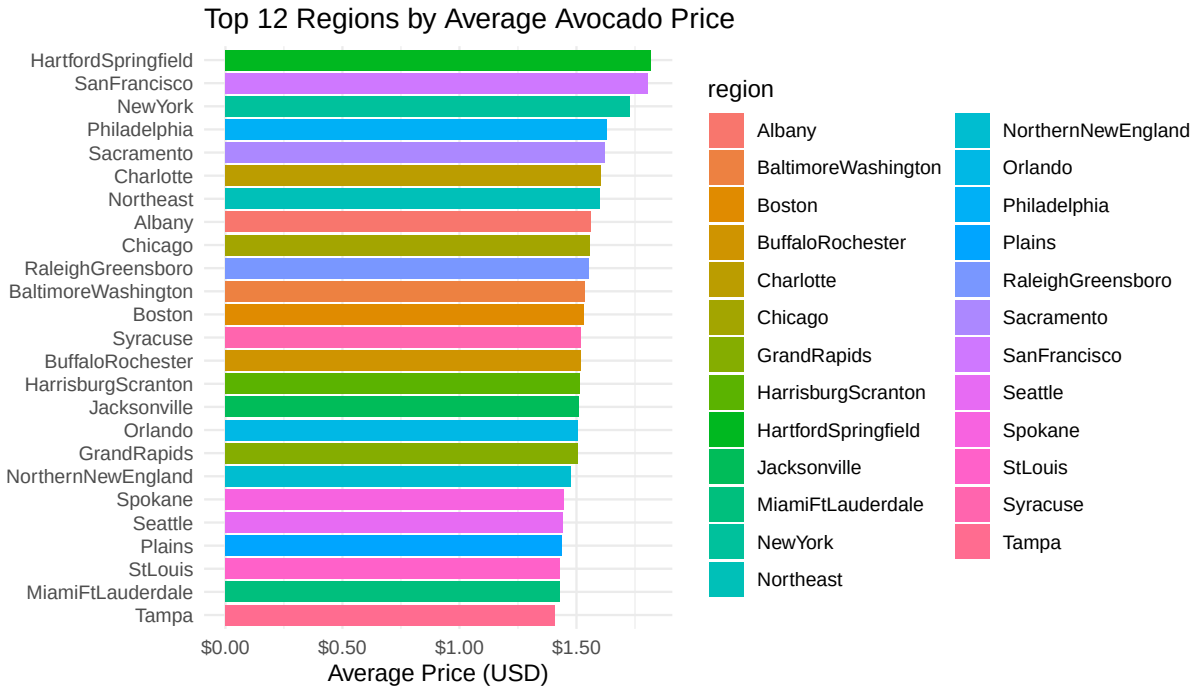
Insight:

Prices fluctuate seasonally and with supply conditions. Organic avocados consistently show a higher average price than conventional across the time span.

Business implication: organic commands a premium — target promotions and premium shelf space accordingly.

2) Regional comparison — Where are prices and volumes highest?

Question: Which regions consistently show higher prices and which show lower prices?



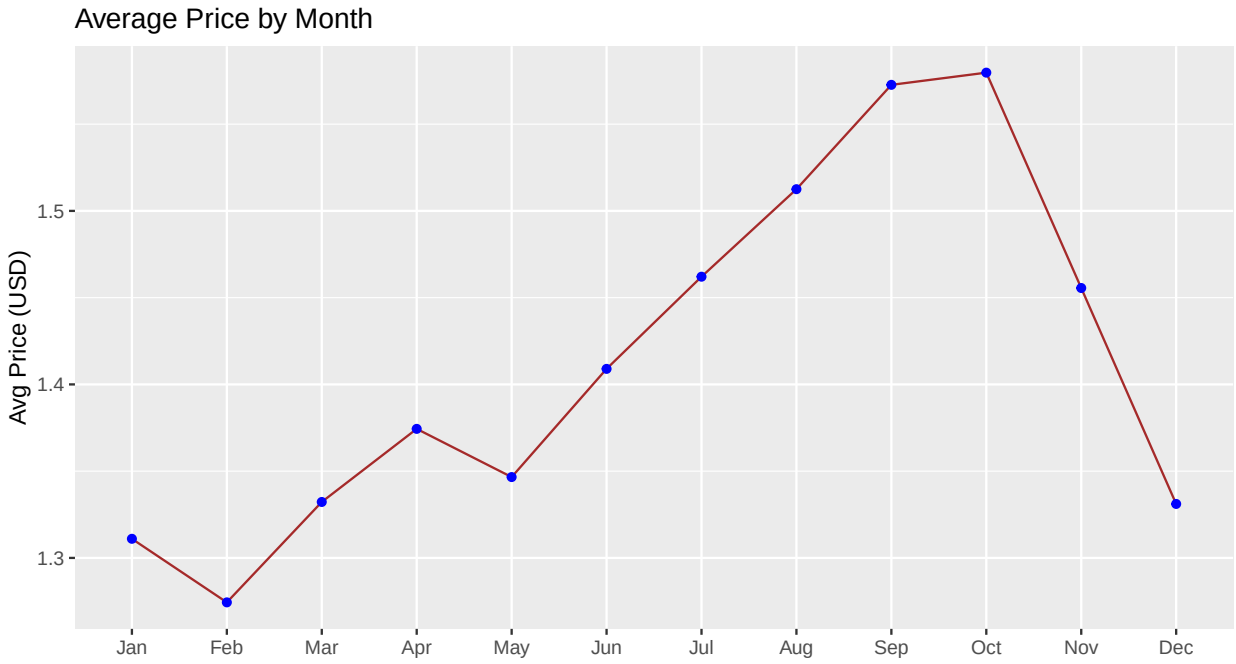
Insight:

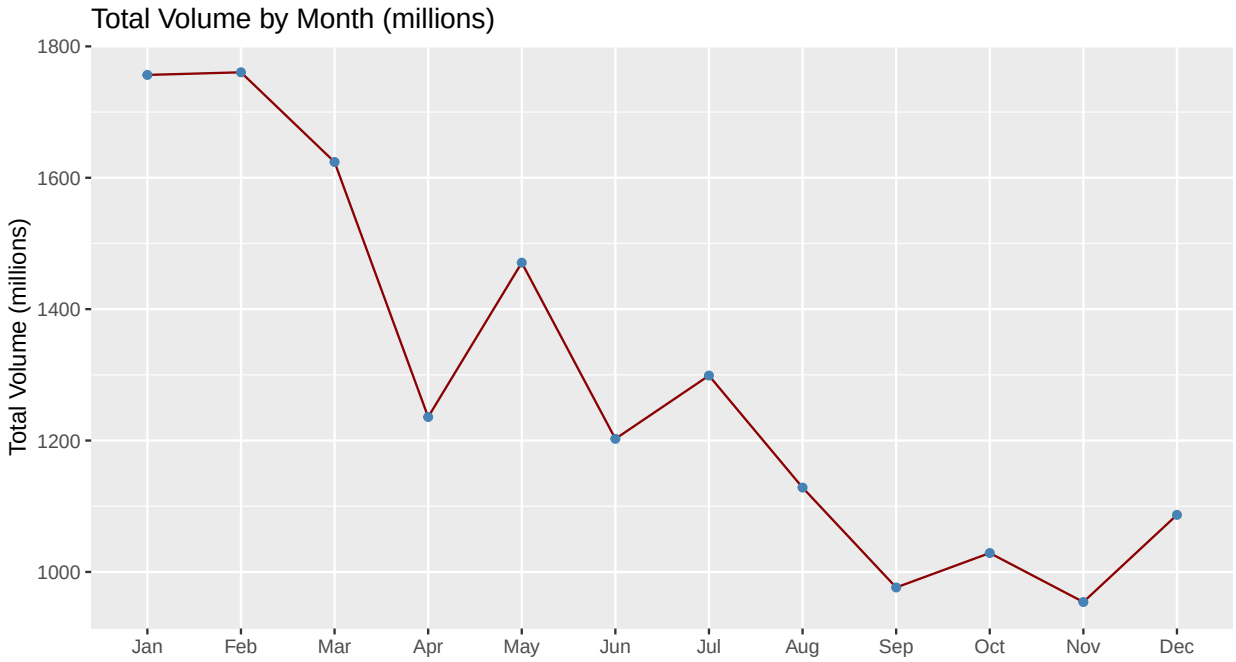
Some regions (e.g., **HartfordSpringfield** in this dataset) show notably higher average prices. These regional differences can be due to supply chain costs, demand, or retailer mix.

Business implication: allocate organic or premium stock to higher-price regions and consider price promotions in lower-price regions to increase volume.

3) Seasonality — monthly patterns (price & volume)

Question: Are there predictable monthly/seasonal patterns?





Insight:

Volume shows recurring peaks (seasonal demand) while prices also vary with season — often higher when supply tightens.

Business implication: increase inventory ahead of expected high-volume months and schedule promotions off-peak to smooth demand.

4) Price vs. Volume — correlation & regression

Question:* Is there an observable relationship between price and total volume?

```
## [1] -0.1927524
```

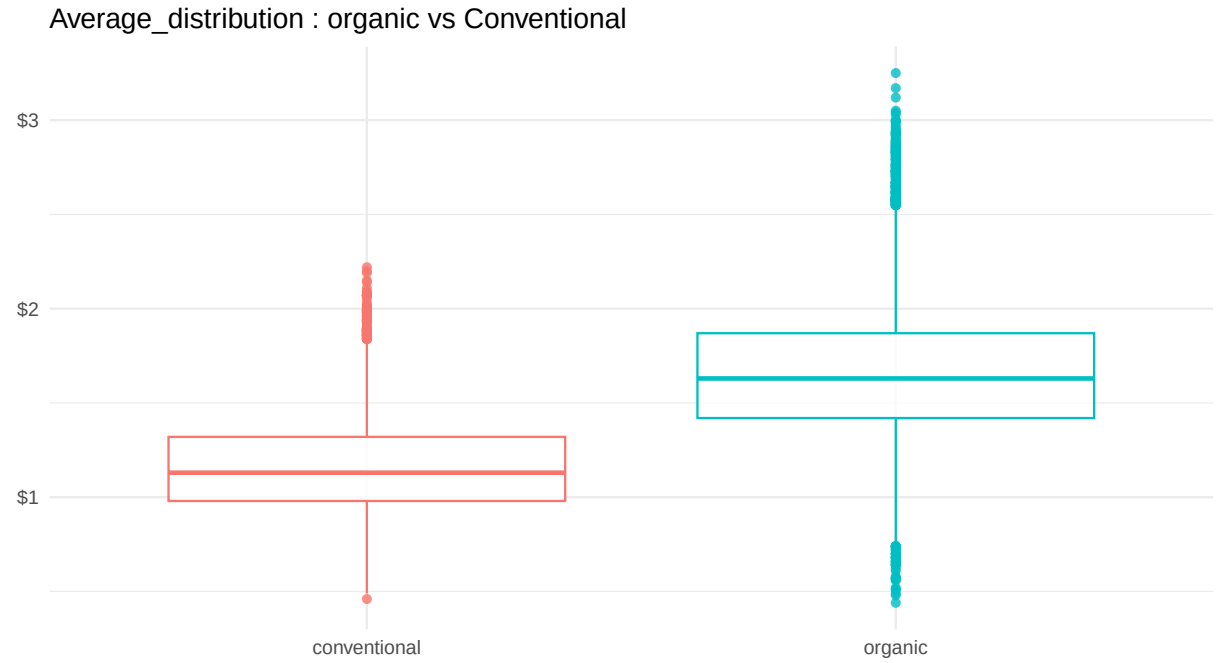
```
## # A tibble: 2 x 5
##   term           estimate std.error statistic   p.value
##   <chr>             <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)    3174918.    91115.      34.8 1.31e-257
## 2 average_price -1653136.    62301.     -26.5 2.94e-152
```

Insight:

Correlation is typically weak/negative (i.e., higher volumes often associated with lower prices), but because of strong seasonality and regional effects, a simple global correlation is noisy.

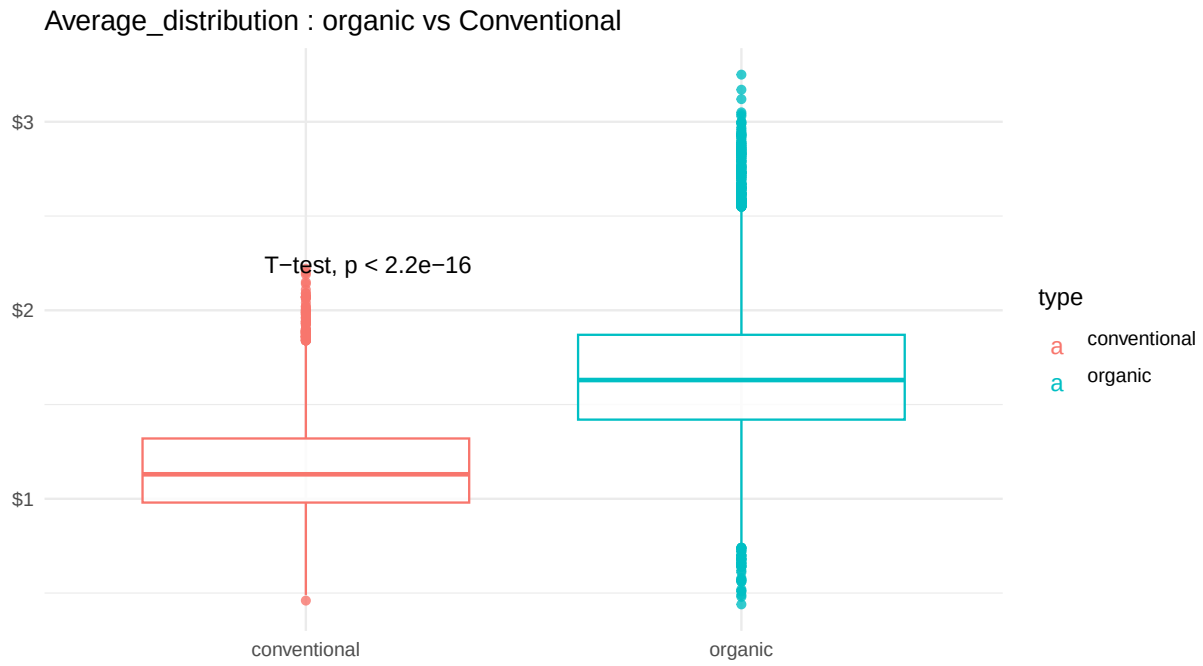
5) Conventional vs Organic — distribution and hypothesis test

Question: Are organic prices significantly higher than conventional?



Also checking the Statistically tests if the average prices of organic and conventional avocados are significantly different.

```
##
##  Welch Two Sample t-test
##
## data:  average_price by type
## t = -105.58, df = 16619, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group conventional and group organi
## 95 percent confidence interval:
##  -0.5051664 -0.4867517
## sample estimates:
## mean in group conventional      mean in group organic
##                   1.158040                   1.653999
```



Insight:

Organic avocados have a higher median/mean price than **conventional.** The t-test typically shows this difference is statistically significant.

Business implication: organic can be positioned as a premium product; profit margins and promotion strategies should reflect that.

#Answering the original business questions (concise)

What factors influence price? Seasonality, avocado type (organic > conventional), and regional market differences. Price-volume relations indicate higher supply => lower price in general, but region and promotion effects matter.

How do trends apply to retailers? Use seasonality to adjust inventory; push premium promotions for organic in higher-price regions.

How should this inform marketing? Time promotions to low-price months, advertise organic as premium in high-willingness regions, and use price/volume monitoring to avoid stockouts before peak demand.

3 Recommendations.

1. Region-specific assortment: Prioritize organic product allocation for high-price regions; add SKUs in high-volume regions.

2. Seasonal inventory planning: Build inventory for predicted high-volume months; negotiate supplier capacity ahead of time.

3. Targeted promotions: Offer discounts during lower-volume months to stimulate demand; avoid discounting right before major volume peaks.

4. Deeper analysis: Estimate price elasticity by region and type (regression with region & month fixed effects); test promotional lift with A/B experiments.

4 Reproducibility & Files

Source data: avocado.csv (Kaggle, original dataset).

5 Thank You

Thank you for reviewing this analysis. Contact: kmrtech99@gmail.com