Hass Avocado Board Dataset Analysis

Scenario

In this case study, I analyzed U.S. retail sales of Hass avocados to understand how price and volume vary over time, across regions, and between conventional vs. organic products. The goal is to discover trends that could guide decisions such as regional promotions, seasonal inventory planning, and pricing strategy. I applied the full data analysis lifecycle (Ask → Prepare → Process → Analyze → Share → Act) and demonstrated the same analysis across spreadsheets, SQL, and R.

The analysis demonstrates the use of three tools: Google Sheets, SQL, and R, for data exploration, visualization, and insight generation.

Act

Business Task:
Identify differences in pricing and sales volume between conventional and organic Hass avocados across time and regions, and generate insights to guide business benefits.

Key Questions:

- How do sales volumes differ between conventional and organic avocados?
- What is the price premium for organic compared to conventional?
- How do prices and sales vary by region and over time?

Dataset Overview

The dataset contains weekly retail data derived from sale systems.

- AveragePrice reflects the per-unit (per avocado) price, even for bagged sales.
- PLUs (4046, 4225, 4770) identify Hass avocado items
- The data covers 54 U.S. regions/markets (Albany, Atlanta, BaltimoreWashington, Boise, Boston, BuffaloRochester, California, Charlotte, Chicago, CincinnatiDayton, Columbus, DallasFtWorth, Denver, Detroit, GrandRapids, GreatLakes, HarrisburgScranton, HartfordSpringfield, Houston, Indianapolis, Jacksonville, LasVegas, LosAngeles, Louisville, MiamiFtLauderdale, Midsouth, Nashville, NewOrleansMobile, NewYork, Northeast, NorthernNewEngland, Orlando, Philadelphia, PhoenixTucson, Pittsburgh, Plains, Portland, RaleighGreensboro, RichmondNorfolk, Roanoke, Sacramento, SanDiego, SanFrancisco, Seattle, SouthCarolina, SouthCentral, Southeast, Spokane, StLouis, Syracuse, Tampa, TotalUS, West, and WestTexNewMexico.) from 2015-2018.

Prepare

- Data Source: Hass Avocado Board, weekly retail scan data (2015–2018).
- Scope: 18,249 rows covering 54 U.S. regions.

Key Columns in Dataset

Date - The date of the observation
AveragePrice - the average price of a single avocado
type - conventional or organic
year - the year
Region - the city or region of the observation
Total Volume - Total number of avocados sold
4046 - Total number of avocados with PLU 4046 sold
4225 - Total number of avocados with PLU 4225 sold
4770 - Total number of avocados with PLU 4770 sold

Process

Google Sheets

- Created pivot tables for yearly totals, average volumes, and min/max values.
- Generated charts showing sales trends and regional comparisons.

SQL (BigQuery)

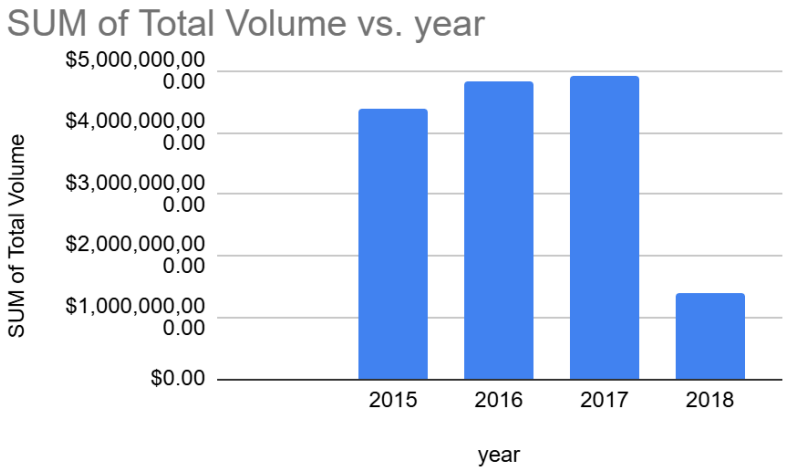
- Row counts and distinct values confirmed dataset coverage.
- Summarized total volume by type (conventional ≈ 97% of sales).
- Computed average price by region & type.
- Analyzed monthly price trends and regional winners.

R (tidyverse)

- Cleaned and structured CSV file.
- Calculated descriptive statistics (overall averages, organic premium).
- Produced visualizations:
    - Price over time by type
    - Regional average prices (bar chart)
    - Market share (pie/bar)
    - Seasonality (line chart by month)

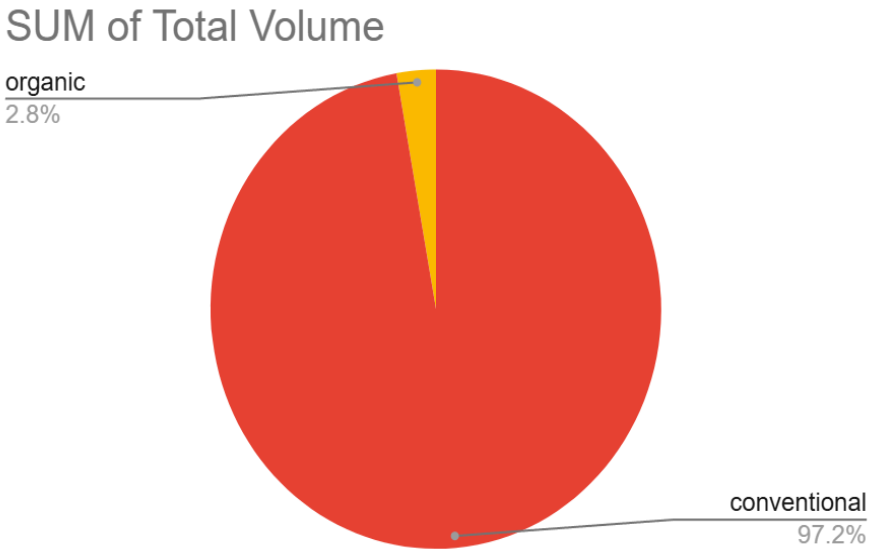
Spreadsheet Analysis

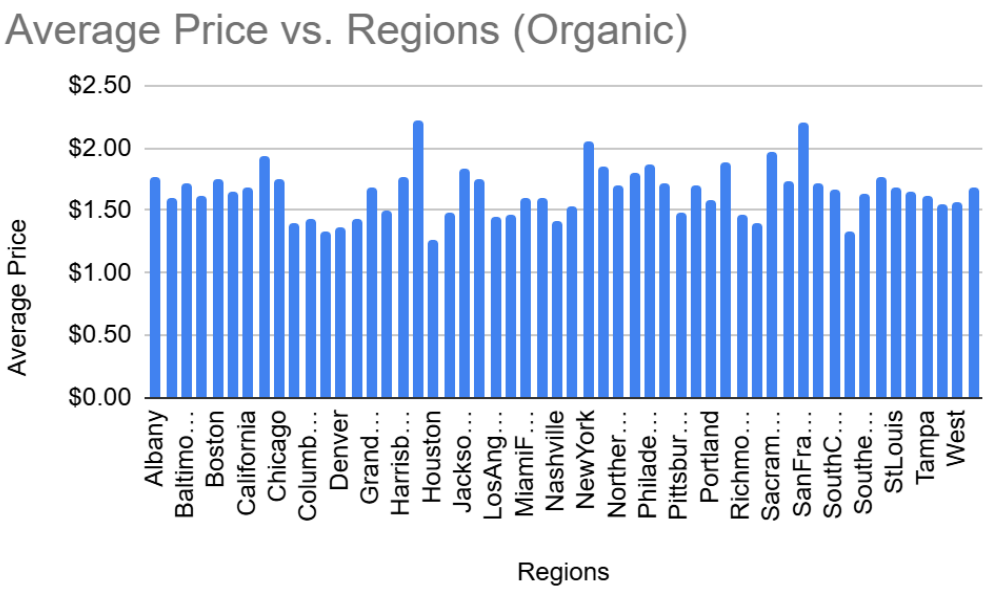
I used Google Sheets to create pivot tables and charts. Google Sheets was especially helpful for quick year-to-year comparisons and for visualizing averages across multiple regions. The following charts highlight total volumes, market share, and regional price differences.



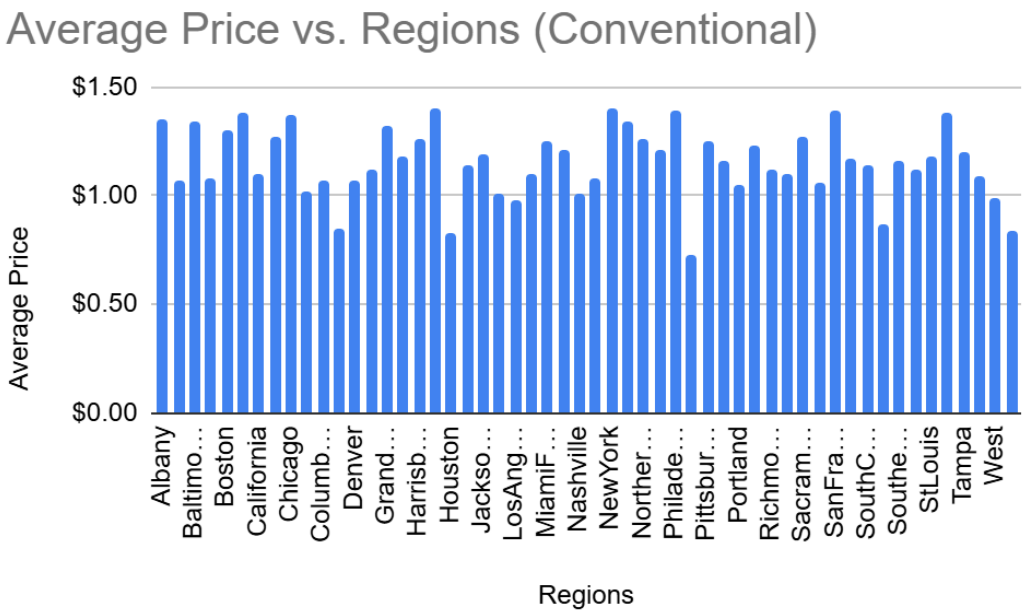
This bar chart shows the total volume of avocados sold each year from 2015 to 2018. Sales increased steadily from 2015 through 2017, with 2017 representing the peak at nearly 5 billion units sold. In 2018, sales appear significantly lower, but this is likely because the dataset only includes a partial year. Overall, the trend shows a growing avocado market over this period, which may be due to higher consumer demand and expanded distribution. For businesses, this indicates a consistent upward trajectory and the importance of planning for sustained growth in demand.

This pie chart visualizes the proportion of sales between conventional and organic avocados. Conventional avocados dominate, representing 97.2% of the total market, while organics account for only 2.8%. This emphasizes that although organics carry a price premium, they remain an unpopular product compared to conventional avocados. For consumers, this reflects broader affordability and accessibility of conventional avocados. For businesses, this means that, conventional must remain the primary focus for scale, while organic can be leveraged as a product in markets that support premium pricing.





This bar chart highlights average organic avocado prices across regions. HartfordSpringfield, San Francisco, and New York stand out with the highest averages, all exceeding $2.00. In contrast, regions like Houston and SouthCentral show prices closer to $1.25-$1.40. This variation suggests regional differences in consumer willingness to pay, influenced by factors such as cost of living, transportation expenses, and consumer preference for organic produce.



This bar chart compares average conventional avocado prices across regions. Prices remain lower overall, rarely exceeding $1.50. Houston, PhoenixTucson, and SouthCentral regions

consistently show the lowest prices, with averages near $0.80-$0.90, while HartfordSpringfield and New York top the list with averages near $1.40. This suggests conventional avocado pricing is more stable across regions compared to organic.

SQL Analysis

To explore the dataset, I used Google BigQuery with SQL. SQL allowed me to confirm dataset coverage, summarize sales volumes, analyze prices across regions, and compare performance between conventional and organic avocados. The following queries and results highlight key findings.

1.) Row count, time span, coverage

```
SELECT
  COUNT(*) AS total_rows,
  COUNT(DISTINCT year) AS distinct_years,
  COUNT(DISTINCT region) AS distinct_regions,
  COUNT(DISTINCT type) AS distinct_types
FROM `avocado-analysis-sql.avocado.avocado_dataset`;
```

Query results

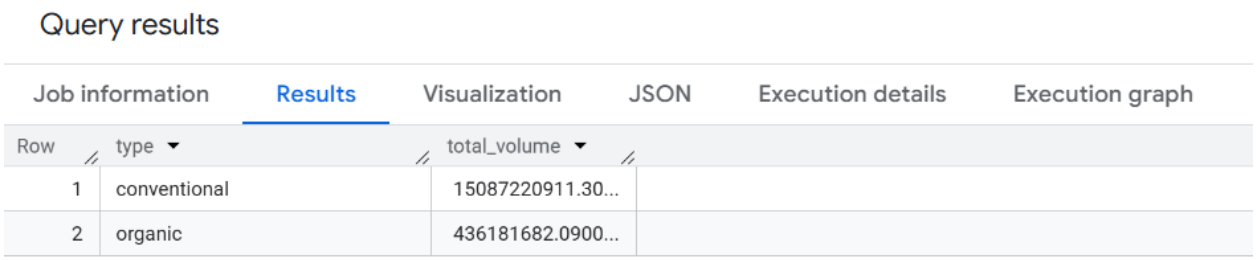
Job information | Results | Visualization | JSON | Execution details | Execution graph

Row	total_rows	distinct_years	distinct_regions	distinct_types
1	18249	4	54	2

The initial query confirmed that the dataset contains 18,249 rows spanning four years (2015-2018) across 54 distinct regions and two types of avocados (conventional and organic). This demonstrates that the dataset is both comprehensive and structured, making it reliable for trend and comparison analysis.

2.) Which type (conventional vs organic) sells more overall?

```
SELECT
  type,
  SUM(`Total Volume`) AS total_volume
FROM `avocado-analysis-sql.avocado.avocado_dataset`
GROUP BY type
ORDER BY total_volume DESC;
```



When calculating total volume by type, the results show that conventional avocados dominate with over 15 billion units sold, compared to just 436 million units for organic. This means that conventional avocados make up roughly 97% of all sales, a value consistent with what was seen in the spreadsheet analyses. The result highlights that while organics are growing, conventional avocados remain the foundation of the market.

3.) Average price by region and type

```
SELECT
  region,
  type,
  ROUND(AVG(AveragePrice)) As average_price
FROM `avacado-analysis-sql.avacado.avacado_dataset`
GROUP BY region, type
ORDER BY region, type
```

region	type	average_price	
Albany	conventional	1	
Albany	organic	2	
Atlanta	conventional	1	
Atlanta	organic	2	
BaltimoreWashington	conventional	1	
BaltimoreWashington	organic	2	
Boise	conventional	1	

Boise	organic	2	
Boston	conventional	1	
Boston	organic	2	
BuffaloRochester	conventional	1	
BuffaloRochester	organic	2	

(the rest of the sorted sql is linked)

By averaging prices across all regions, a clear pattern emerges: organic avocados consistently cost about $2 per unit, while conventional averages closer to $1. While some regions like CincinnatiDayton, Columbus, and Pittsburgh report slightly lower prices for both types, the organic premium is present everywhere. This shows that organics are universally positioned as a premium product.

4.) Monthly average price trend

```
SELECT
  DATE_TRUNC(Date, MONTH) AS month,
  type,
  ROUND(AVG(AveragePrice)) as average_price
FROM `avacado-analysis-sql.avacado.avacado_dataset`
GROUP BY month, type
ORDER BY month, type;
```

month	type	average_price
2015-01-01	conventional	1
2015-01-01	organic	2
2015-02-01	conventional	1
2015-02-01	organic	2
2015-03-01	conventional	1
2015-03-01	organic	2
2015-04-01	conventional	1
2015-04-01	organic	2
2015-05-01	conventional	1
2015-05-01	organic	2
2015-06-01	conventional	1

2015-06-01	organic	2
2015-07-01	conventional	1
2015-07-01	organic	2
2015-08-01	conventional	1
2015-08-01	organic	2
2015-09-01	conventional	1
2015-09-01	organic	2
2015-10-01	conventional	1

(the rest of the sorted sql is linked)

When aggregating prices by month, the results show consistent differences between conventional and organic. Conventional averages around $1, while organics average $2, with fluctuations across the years. Notably, both types experience synchronized price increases in certain months, especially during 2017, indicating seasonal supply and demand cycles that affect the market as a whole.

5.) Which type sells the most by region (all years combined)

```
WITH by_region AS (
  SELECT
    region,
    SUM(CASE WHEN type = 'organic' THEN `Total Volume` ELSE 0 END)  AS
org_vol,
    SUM(CASE WHEN type = 'conventional' THEN `Total Volume` ELSE 0 END) AS
conv_vol,
    SUM(`Total Volume`) AS total_vol
  FROM `avacado-analysis-sql.avacado.avacado_dataset`
  GROUP BY region
)
SELECT
  region,
  org_vol,
  conv_vol,
  total_vol,
  CASE
    WHEN conv_vol > org_vol THEN 'conventional'
    WHEN org_vol > conv_vol THEN 'organic'
    ELSE 'tie'
  END AS winner_type,
  ROUND(100 * GREATEST(org_vol, conv_vol) / NULLIF(total_vol, 0), 2) AS
```

```
winner_share_pct,
  ROUND(GREATEST(org_vol, conv_vol) - LEAST(org_vol, conv_vol), 2) AS
volume_margin
FROM by_region
ORDER BY volume_margin DESC;
```

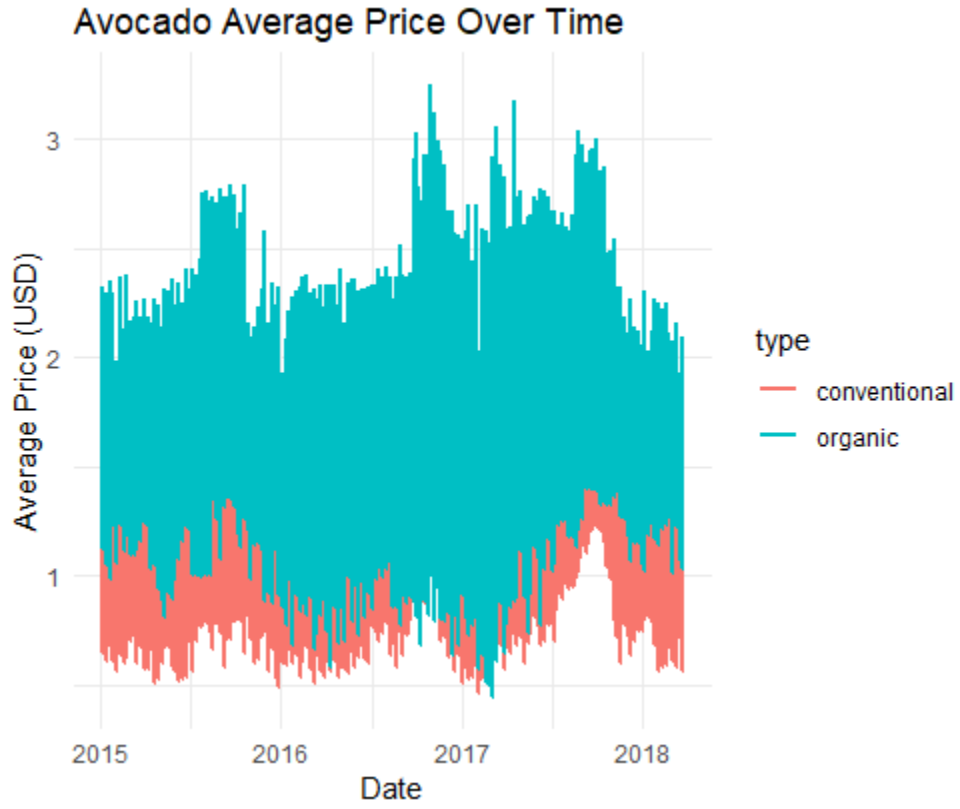
region	org_vol	conv_vol	total_vol	winner_type	winner_share_pct	volume_margin
TotalUS	163,518,595.44	5,701,221,586.36	5,864,740,181.80	conventional	97.21	5,537,702,990.92
West	37,418,566.21	1,049,360,589.54	1,086,779,155.75	conventional	96.56	1,011,942,023.33
California	25,384,812.11	1,003,596,841.06	1,028,981,653.17	conventional	97.53	978,212,028.95
SouthCentral	18,118,464.27	993,161,154.56	1,011,279,618.83	conventional	98.21	975,042,690.29
Northeast	24,226,096.09	689,054,813.47	713,280,909.56	conventional	96.6	664,828,717.38
Southeast	10,039,655.90	605,198,753.21	615,238,409.11	conventional	98.37	595,159,097.31
GreatLakes	22,074,687.29	567,567,862	589,642,549.29	conventional	96.26	545,493,174.71
LosAngeles	13,386,275.15	494,510,272.48	507,896,547.63	conventional	97.36	481,123,997.33
Midsouth	17,980,440.88	490,368,916.32	508,349,357.20	conventional	96.46	472,388,475.44
Plains	8,275,873.26	302,912,655.85	311,188,529.11	conventional	97.34	294,636,782.59
NewYork	8,990,571.40	231,743,556.13	240,734,127.53	conventional	96.27	222,752,984.73

(the rest of the sorted sql is linked)

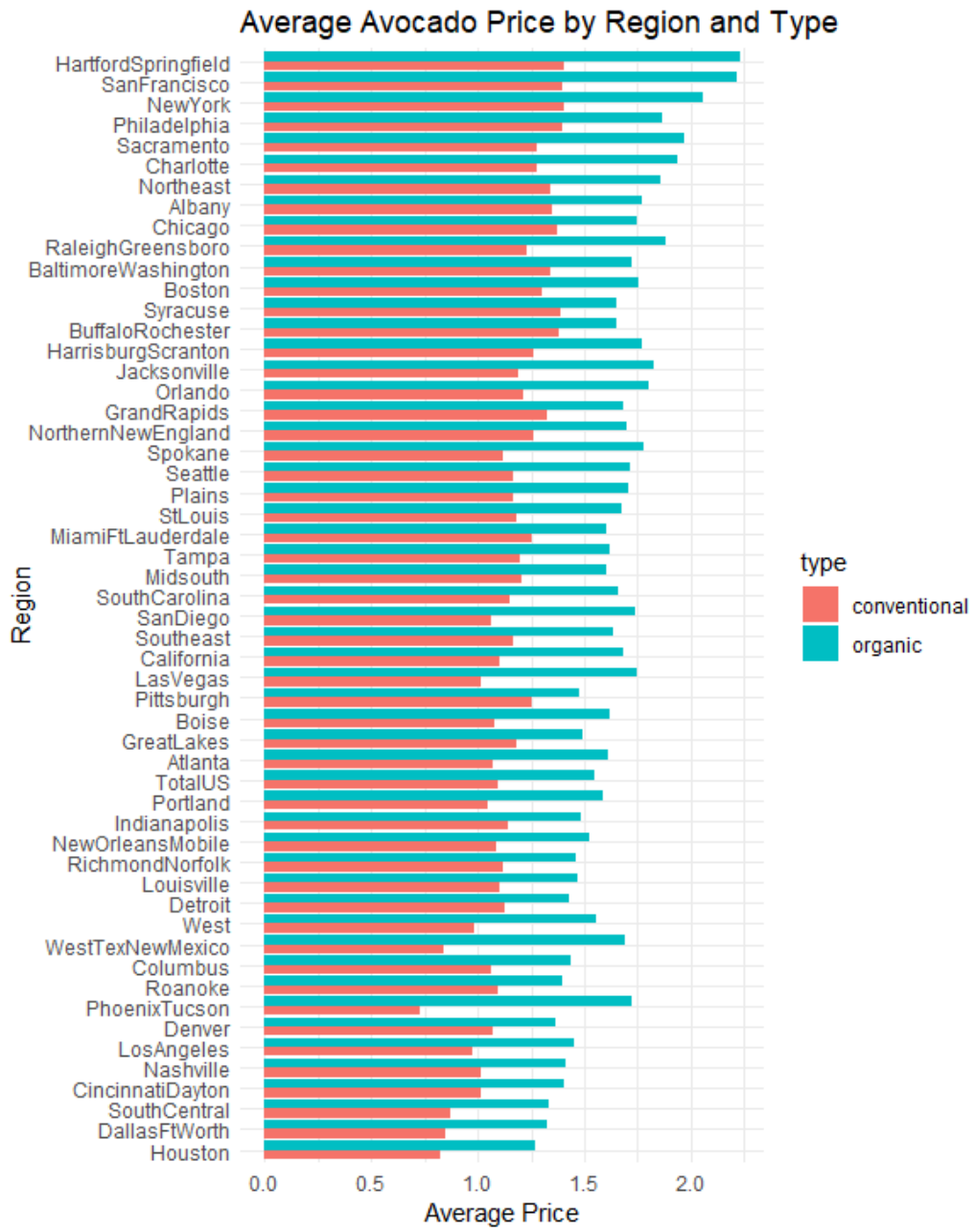
Breaking down sales volumes by region shows that in every single region, conventional avocados outsell organic by a wide margin. For example, in California, conventional sales reached over 1 billion units compared to just 25 million units for organic, while in the Total U.S. view, conventional claimed 97.21% of the share. Even in markets like San Francisco and New York, which support higher organic prices, the conventional category remains dominant. This emphasizes that while organics can provide high margins in select markets, conventional avocados are the key driver of total sales across the country.

R Analysis

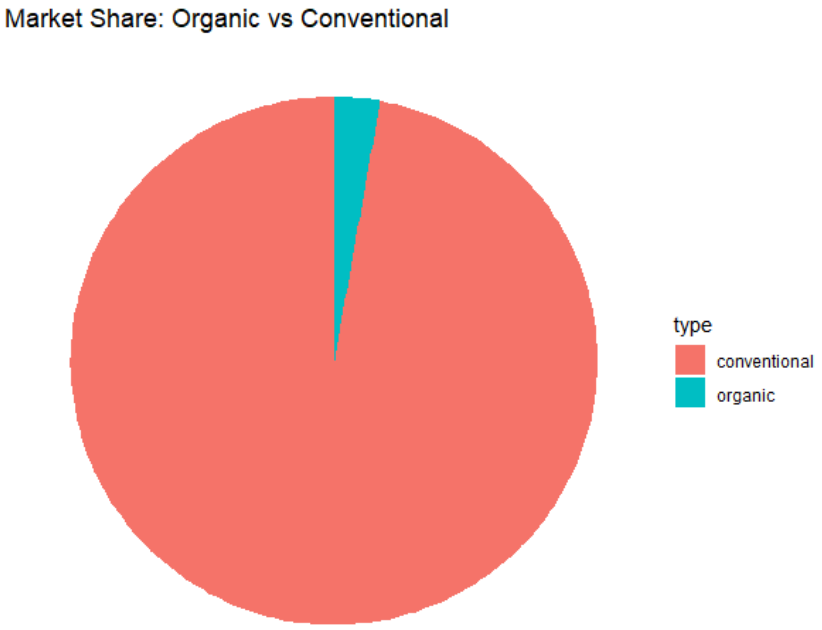
To further explore the dataset, I used R and the tidyverse library to clean the data, perform descriptive analysis, and generate visualizations. R allowed me to uncover patterns and trends in avocado pricing and sales that are easier to interpret through charts than raw tables.



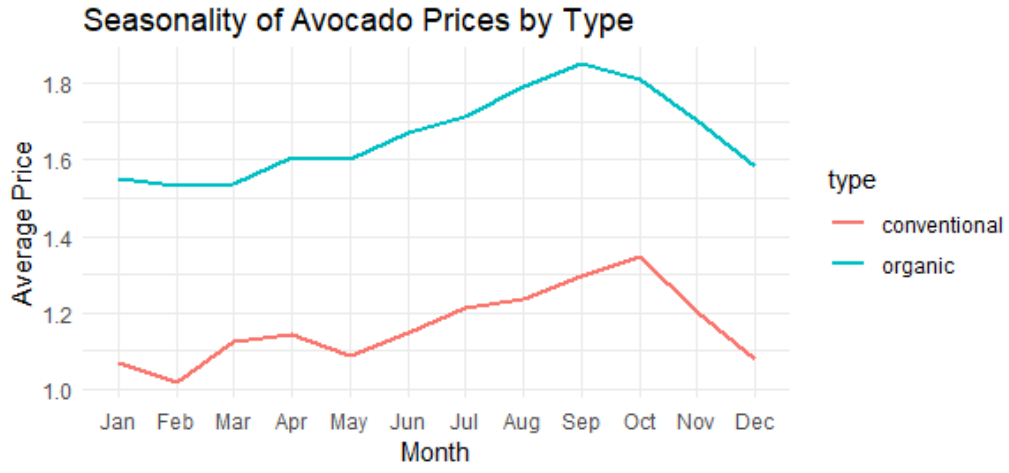
This line chart represents the average price of conventional and organic avocados from 2015 to 2018. Organic avocados consistently remain higher in price compared to conventional, with averages ranging from about $1.60 to $2.00. Conventional avocados, on the other hand, generally stay between $1.00 and $1.30. A noticeable spike occurs in 2017, where both organic and conventional prices increase sharply, with the rise being more significant for organic avocados. This indicates that both categories are affected by external market or seasonal pressures, though the premium for organic remains stable. From a consumer perspective, this suggests that organic buyers are willing to continue paying more even during price increases, while for businesses, this highlights the need to monitor supply chain dynamics closely, especially during seasonal peaks.



This visual represents the functionalization of avocado prices across fifty-four U.S. regions. Organic avocados are consistently more expensive than conventional in every region, with HartfordSpringfield, San Francisco, and New York leading with average organic prices above $2.00. On the lower end, cities like Houston, SouthCentral, and PhoenixTucson report the cheapest conventional prices, often below $1.00. The graph highlights clear regional differences where urban and coastal regions sustain higher premiums compared to inland areas. There could be many reasons for these differences, including transportation costs, consumer preferences, and local market conditions. From a consumer standpoint, location influences whether it is worth paying for organic avocados. From a business perspective, this chart helps identify where profit margins are highest and where organic demand could be further cultivated.



This pie chart demonstrates the significant gap in sales between conventional and organic avocados. Conventional avocados dominate the market, accounting for roughly 97% of total units sold. Organic avocados, though consistently priced higher, make up only about 3% of sales volume. This shows that while organics attract a smaller niche audience, conventional avocados are the true market driver in terms of scale. For businesses, this highlights the importance of keeping conventional avocados as the main supply focus while promoting organic avocados in regions where consumers are more willing to pay the premium.



This line chart demonstrates how avocado prices fluctuate seasonally throughout the year. Organic avocados show a steady increase from early spring into late summer, reaching their highest point in September and October at around $1.80 before declining toward December. Conventional avocados follow a similar seasonal pattern but remain consistently lower by $0.40 to $0.50 compared to organic. These synchronized shifts suggest shared supply and demand cycles, such as seasonal harvest patterns or consumer demand surges in the fall. For consumers, this means prices are predictably higher in late summer and early fall, while businesses can use this seasonal knowledge to prepare for price increases and manage inventory more effectively.

Analyzation Summary

Key Findings:

- Conventional dominance: Conventional avocados account for ~97% of units sold (15.1B vs 0.44B organic).
- Organic premium: Organic averages $1.65 vs conventional $1.16 → premium ≈ $0.49
- Regional variation:
    - Highest organic prices: SanFrancisco ($2.21), HartfordSpringfield ($2.23), NewYork ($2.05).
    - Lowest conventional prices: Houston ($0.83), SouthCentral ($0.87), PhoenixTucson ($0.73).
- Seasonality: Prices rise in late summer/fall, peaking around September–October 2017.
- In every region, conventional outsells organic.

Act

Recommendations according to the data provided:

1. Pricing & Promotions
    - Maintain conventional avocados as the primary volume driver.
    - Use targeted organic promotions in high-income regions such as San Francisco, the Northeast, and New York.

2. Inventory Planning
    - Prepare for seasonal price increases in late summer and fall.
    - Adjust supply chain and stock management to ensure availability during these peak months.

3. Assortment Strategy
    - Tailor pack sizes and PLU assortments by market. For example, emphasize larger PLUs in club stores vs. smaller ones in grocery

Conclusion

This analysis demonstrates how the Hass Avocado Board dataset can be explored through spreadsheets, SQL, and R to uncover pricing and sales insights. Conventional avocados dominate the U.S. market, while organics consistently maintain a premium across regions. Seasonal and regional patterns highlight opportunities for targeted promotions, smarter inventory management, and product assortment adjustments. Together, these findings can help retailers and suppliers align strategies with consumer demand while maximizing profitability.

Credits & References

- Image Credit
    - Hass Avocado Tree product image: Simply Trees
- Data Sources
    - Hass Avocado retail sales dataset: Kaggle – Avocado Prices
    - Hass Avocado Board: Hass Avocado Board website