



<https://cado-rice-data.herokuapp.com/>

CADO Creations

COOL AVOCADO DATA OBSERVATIONS

Could This Be You?

SEARCHING FOR MS. RIPE

SCA (Single California Avocado) is searching for a FHNWP (Fun Health Nut with Personality) in Houston who shares his California attitude and fun-loving lifestyle.



Background

Avocados are the darling of the produce section. They're the go-to ingredient for guacamole dips at parties. And they're also turning up in everything from salads and wraps to smoothies and even brownies.

What a better fruit to analyze?!?

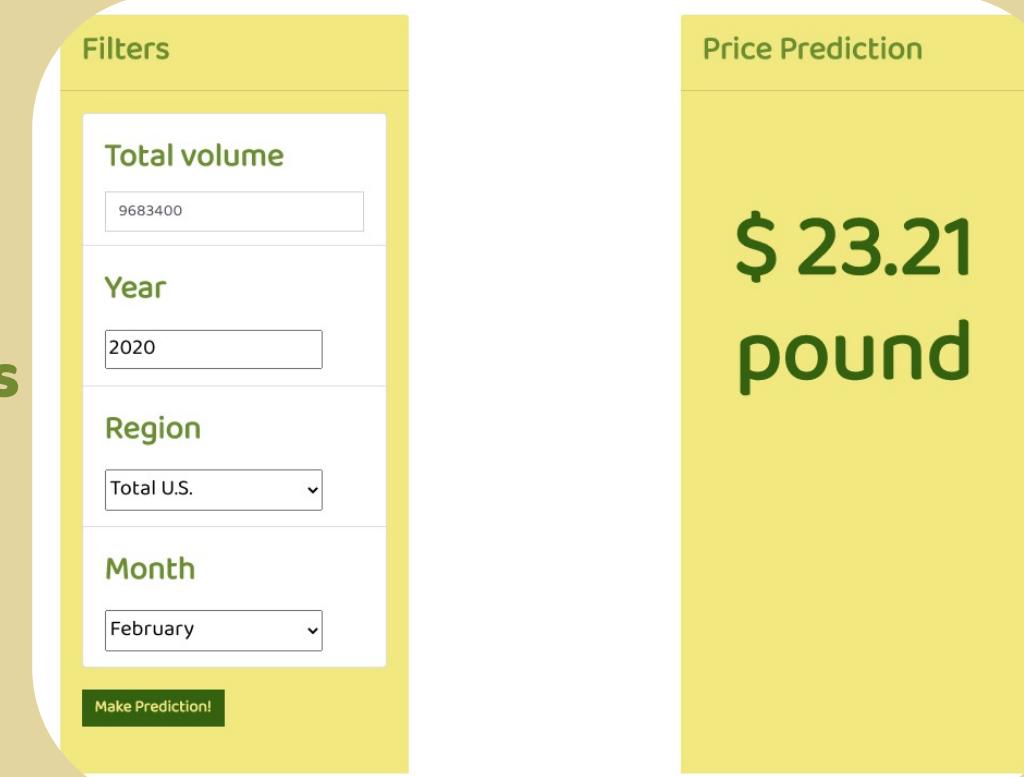
Objectives

Create an application to analyze and explore data analysis and price predictions for Hass Avocados in the United States



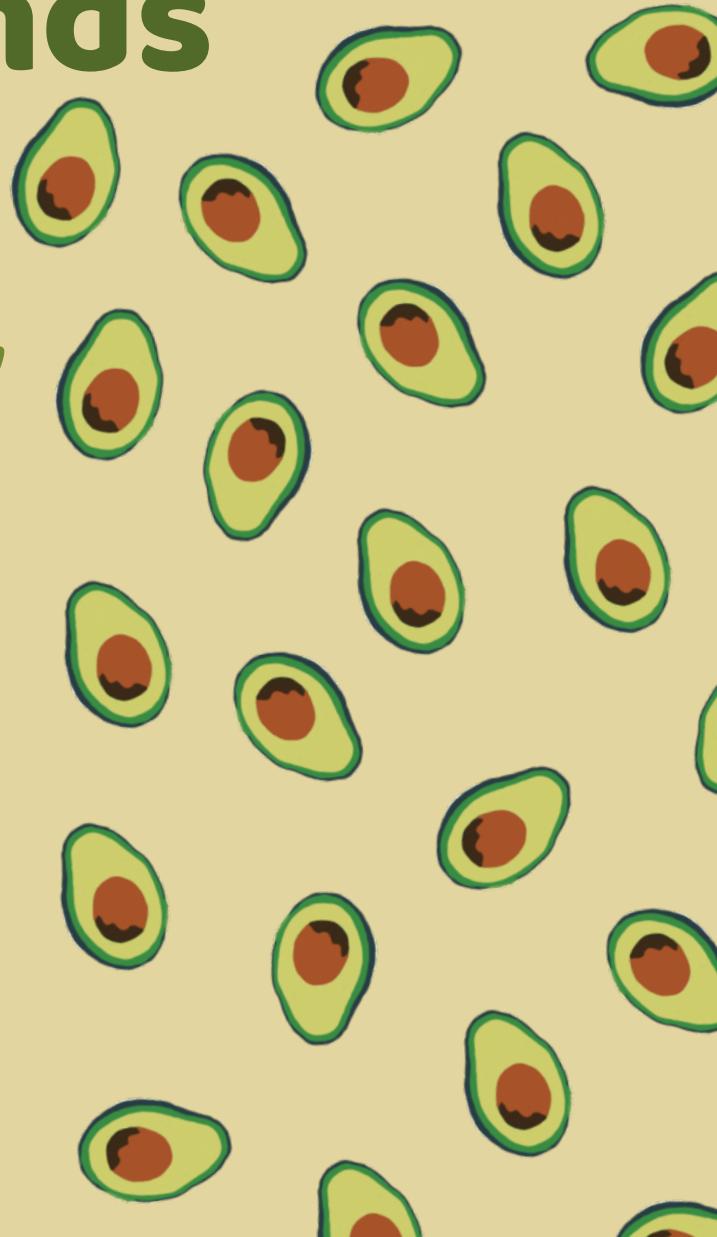
Technologies Used

Pandas, Matplotlib, Seaborn, HTML, CSS, JavaScript, Tableau Public, Heroku, Python, D3, Flask, Scikit-learn, Github



Basic Hypothesis, Trends

- 1. Avocado sales will vary directly with the season (hotter months in the countries of production will result in an increase in quantity and quality of the world's avocados.)**
- 2. Avocado prices will correlate with avocado sales volume.**



Inspirations

“...a magnet for
money-hungry
cartels.”

IMDb.com

Full Cast and Crew | Trivia | User Reviews | IMDbPro | More ▾

SHARE

Rotten (2018–2019)

The Avocado War

57min | Documentary | Episode aired 4 October 2019

7.6 /10 169 Rate This

Season 2 | Episode 1 | < Previous | All Episodes (12) | Next >



The avocado's rise from culinary fad to must-have superfood has made it a lucrative crop - and a magnet for money-hungry cartels.

Director: Lucy Kennedy
Stars: Latif Nasser, Zoila Quiroz, Noel Stehly | See full cast & crew »

Watch on Netflix with subscription

+ Add to Watchlist

Reviews
2 user | 1 critic

<https://www.imdb.com/title/tt11064620/>



CODO Creations

Baloo Tamma Regular - Font



“This avocado
armchair could be
the future of AI.”

Extract the DATA

Kaggle - <https://www.kaggle.com/timmate/avocado-prices-2020.csv>

Data used “is an updated version of the avocado dataset originally compiled from the Hass Avocado Board (or HAB, for short) data and published on Kaggle by Justin Kiggins in 2018. The dataset features historical data on avocado prices and sales volume in multiple cities, states, and regions of the USA.”

This updated version contains data from 4 January 2015 up to 17 May 2020



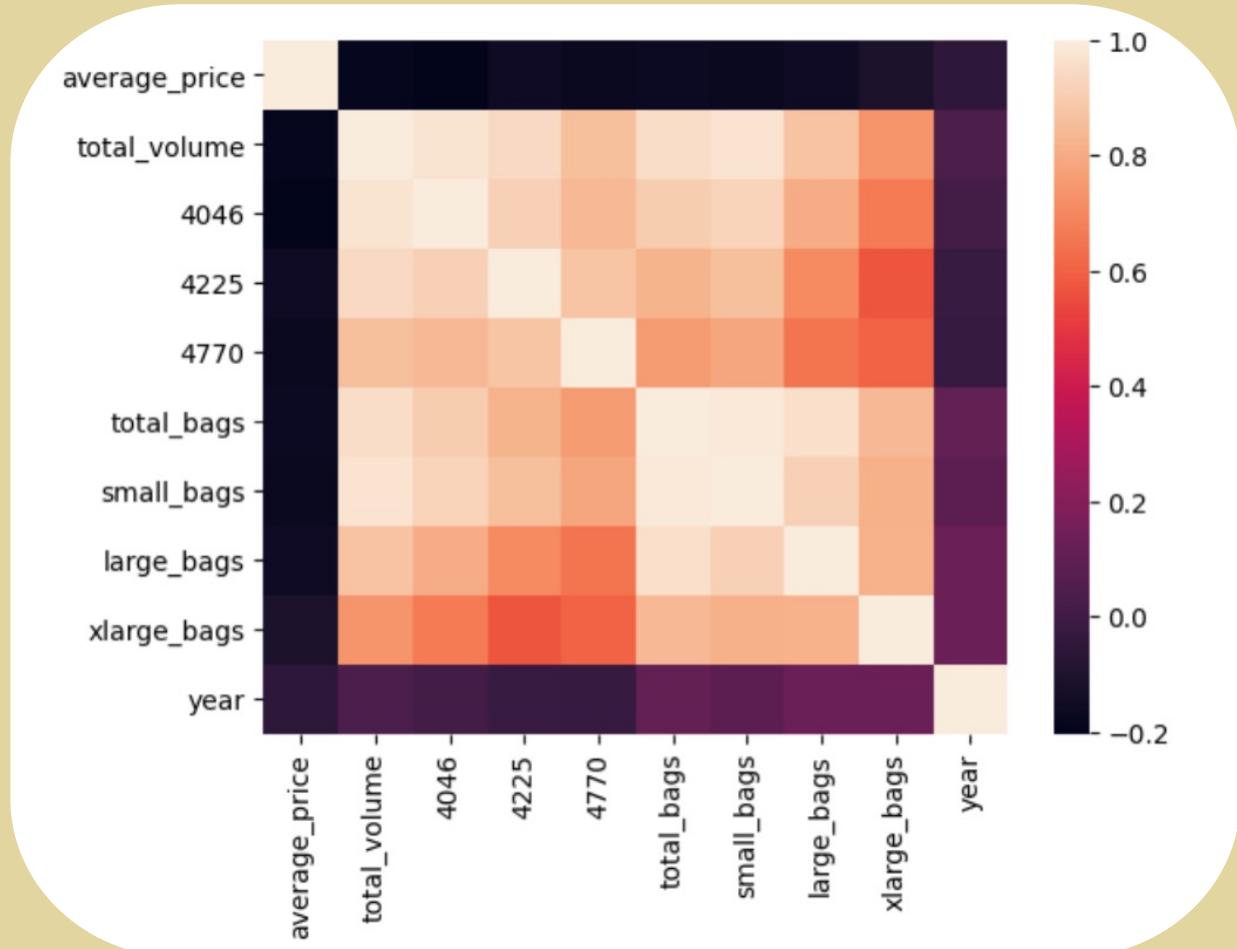
Some relevant columns in the dataset:

- **Date** - The date of the observation
- **Average Price** - 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



Investigating Predictions / Correlations

- Price Predictions
- Volume vs. Price
- Affects of Seasonal changes
- Differences between
Organic vs. Conventional



Transforming the Data

- Type column (conventional/organic)
 - Using One-Hot Encoding for Categorical Data
- Month column (1-11) – Using Label Encoding from 1 to 11 for each month respectively, excluding December per null values found in December.



Machine Learning

The machine-learning portion of the avocado project is based on multiple linear regression, the oldest and most tried-and-true form of linear regression. It functions by minimizing the squared distance between predicted and actual values, where predicted values are produced by multiplying a set of input variables by a set of coefficients.

These coefficients are changed until the most accurate picture is formed (i.e., the least error-prone model is created).



Machine Learning

In the avocado example, this most-accurate picture turned out to be only slightly accurate. It seems that even though the price of a good is predictable (avocado prices have followed steady trends over the years), the *short-term movement* in that price is still highly unpredictable. We found that the most significant variables in the regression were volume of avocados. Besides that, each region and year seems to have a significant effect on the model. Prices do not just follow a trend, but a geographical pattern as well.



Machine Learning Deployment

- Our process for Machine Learning included deploying our model using a Flask app.
- We utilized pickle to load and save the data to filter a user's input to display the results
- The predictions were returned as a numeric type to a jinja template in html.



Avocado Average Price by Region

Discover what the single avocado price is in each region of the United States through the years 2015-2020

Regions like Hartford/Springfield and San Francisco have unusually high average prices ranging from \$1.62 - \$1.97.

Average Price per Region

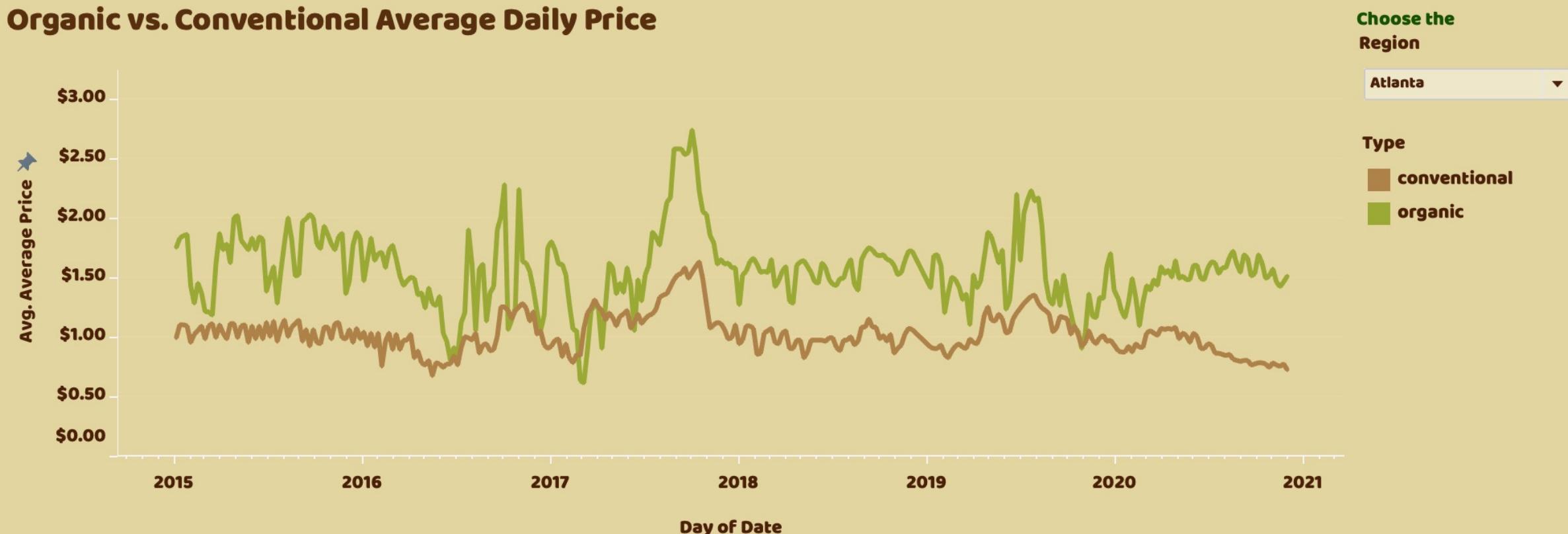


Organic vs. Conventional Average Price Comparison

Here we can see that the organic avocados, have always been priced higher through the years.

Undoubtedly, the price of production for organic avocados have always kept the price higher than the conventional type.

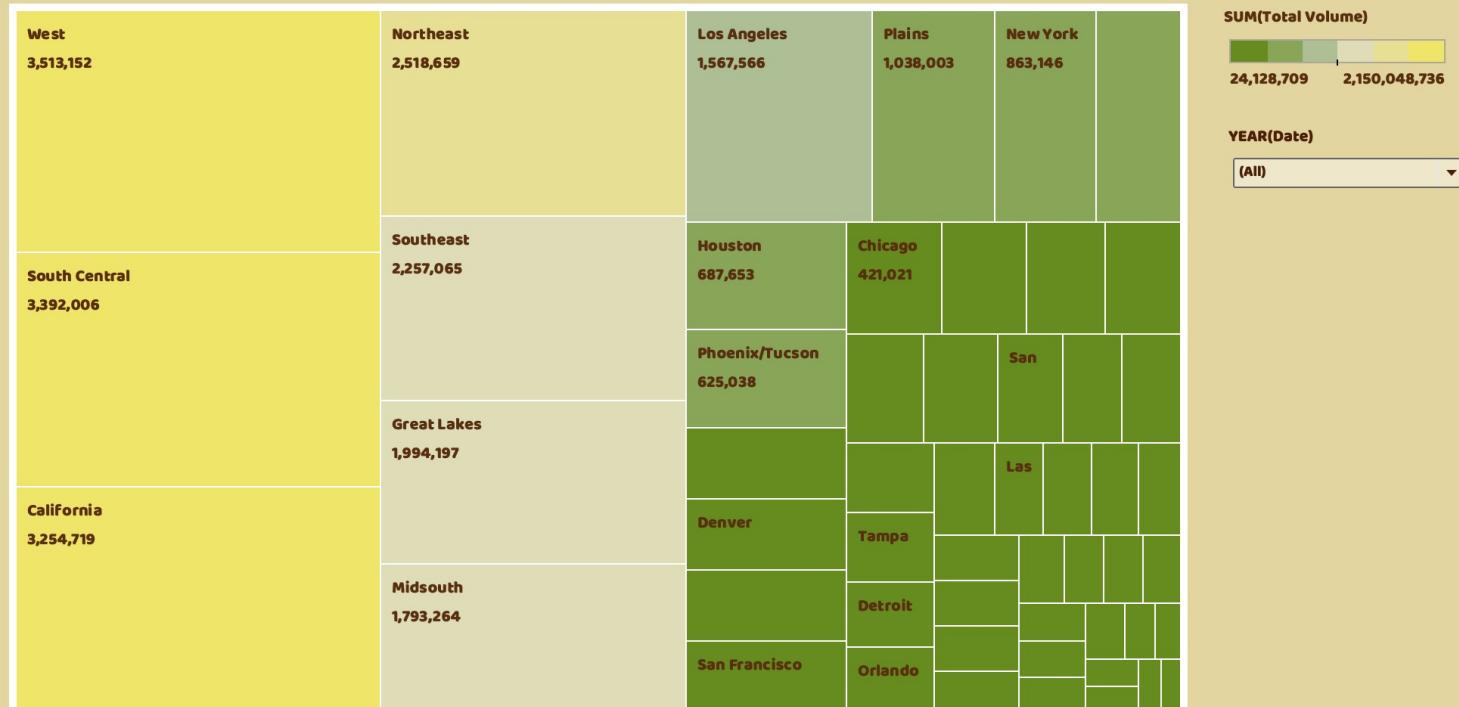
Organic vs. Conventional Average Daily Price



Total Volume of Avocados Sold

1. Winter holidays and Super Bowl parties have affected the quantity sold.
2. Because the avocado fruit is not necessarily an American recipe staple for winter holidays, such as Christmas or Hanukkah, we can expect that the 4th quarter to have lower sales.
3. Through each region and year span, the 3rd and 4th quarters generally have lower volume sales, and the 1st and 2nd quarters have higher quantities of avocados sold.

Total Volume (lb) of Avocados Sold



Avg. Total Volume

Total Avocados Sold by Region

Region	2018 Q2	2018 Q3	2018 Q4	2019 Q1	2019 Q2	2019 Q3	2019 Q4	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4
West	3,513,152	3,445,270	2,901,740	4,150,534	3,050,253	3,150,424	3,354,747	3,350,861	4,707,773	4,348,490	3,419,007	4,392,222	4,729,570	4,070,234	3,020,604
South Centr.	73,563	3,605,365	3,181,247	4,165,511	3,666,078	3,327,656	3,354,249	4,143,369	4,719,619	4,375,905	3,461,874	4,434,778	4,908,107	4,401,893	3,723,657
California	07,224	3,289,808	2,809,420	3,772,394	3,348,646	3,079,699	3,048,807	3,717,627	4,186,101	4,148,459	3,199,669	4,218,429	4,225,503	3,899,857	3,199,669
Northeast	98,732	2,831,034	2,228,655	3,138,673	2,868,001	2,572,020	2,422,740	3,158,843	3,830,681	3,583,028	2,831,939	3,763,022	4,223,451	3,849,603	3,147,470
Southeast	98,803	2,481,895	2,172,386	3,138,106	2,633,523	2,409,138	2,561,001	3,248,106	3,303,018	3,037,084	2,739,026	3,401,247	3,508,801	3,243,459	3,025,543
Great Lakes	09,363	2,044,306	1,791,779	2,398,133	2,130,145	1,947,798	1,915,420	2,524,111	2,964,902	2,723,682	2,031,174	2,679,259	2,901,773	2,578,667	2,187,206
Midsouth	90,578	1,918,994	1,618,657	2,248,029	2,116,142	1,876,199	1,684,782	2,371,899	2,814,842	2,554,727	1,997,528	2,815,524	3,069,526	2,853,134	2,258,487
Los Angeles	79,443	1,500,549	1,329,864	1,824,435	1,588,578	1,425,211	1,437,262	1,725,780	1,947,026	1,958,677	1,507,140	1,931,752	1,972,267	1,843,823	1,507,140

Avocado Price Compared to Volume



Here we can see the growing price of avocados from 2019 to 2020

Seasonality is always a contributing factor with avocados, but overall, the price is on the increase.

Conclusions

1. In 2019 we see that prices override volume. Particularly from April to December
2. In 2019, March had the highest volume compared to other months within that year
3. In 2020, The highest peak months were May and August exceeding the avg. price mark for both months
4. In 2020, we see that avg price takes a dive in November at \$1.22 compared to a high \$1.66 in April

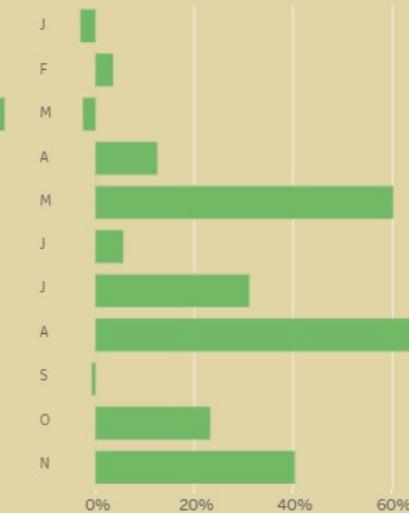
Are Avocados Expensive??



Difference in Average Price from 2019 to 2020



2019-2020 % Volume Difference



Avocado Price compared to Volume



Recommendations

- 1. Limit productivity of large bags and xlarge bags of avocados.**
- 2. Increase production of smaller bags.**
Smaller bags are more popular for consumers to purchase. This maybe due to the avocados short shelf life of 3-4 days.
- 3. Pay attention to the peak months of April, May, July, and August when sales of avocados increase.**



Limitations

- 1. Dataset is limiting...**
 - It does not include weather data - conditions that would've affected avocado production
 - It does not include Manufacturing cost and Production Cycle data
- 2. The use of a mathematical model with future dates and past data doesn't give a perfect prediction to depend on.**





Future Work

Including Weather Data for
each respective Region





CODO Creations About Us



My name is Phillecia Qualls, and I'm an IT Analyst for Decision Information Resources. I hold a Bachelor's degree in Business Management, a Masters' degree in Business Administration, as well as a Masters' in Management Information Systems. People find me to be an upbeat, self-motivated team player with excellent communication skills. My experience includes designing, analyzing, and implementing efficient IT systems, planning and overseeing projects (e.g., upgrades, hardware/software installations), provide advice and technical training, and much more.



Passionate Data Analyst. Team worker with great ability to learn and able to quickly adapt to a new environment, global mindset and fluent in English and Spanish. Interested in KPI's, Dashboard, ETL, Data Analysis, Data Engineering, Databases and Data maps.



Industrious, multifaceted Data Analyst with skills in Graphic Design and Education. Having a dedication for collecting and rearranging data into a clear, understandable narrative. Tackles projects within collaborative teams and independently to enable the data to be simply broken down, displayed and interpreted by all. Passionate for the details and producing solutions to problems.



I am a self-taught programmer with an intense interest in classical Economics. I am currently pursuing a major in Economics at Tulane University. I love numbers and solving problems, and I hope to one day start my own small business.