**AVOCADOCALYPSE**

Due to the rise of vegetarianism, I was wondering if the price of the avocado, depending on the varieties increased over time and if it has an impact on the total volume sold as well as the volume sold for three different PLU codes:

PLU4046: Hass small

PLU4225: Hass large

PLU4770: Hass extra large

I first imported, renamed some columns, inferred the schema and checked it out

**Questions to answer**

1. Does the price have an influence on the volume?
2. Which areas consumed avocados the most from 2015 to 2021?
3. The PLU4770 being the biggest Hass variety, is it also the one consumed the most?

**Hypothesis**

1. Yes the price has an influence on the volume and vice versa.
2. Mostly the highly developed areas because of the vegetarian trend
3. The consumer always going for the biggest, the cleanest, the most beautiful, is more likely to choose the biggest avocado variety

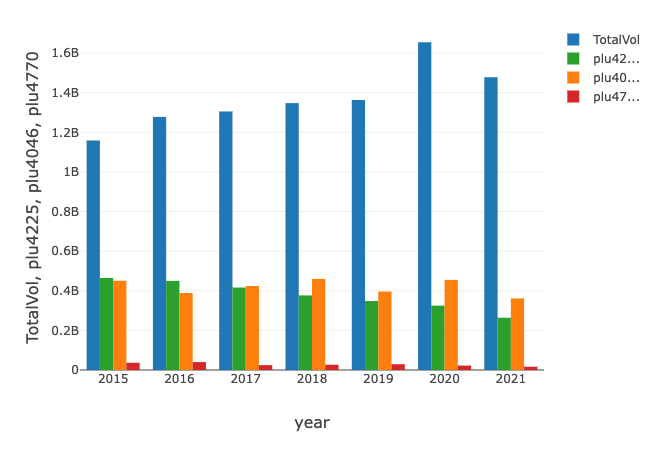
**Data Analysis**

I downloaded the data from the years 2015 to 2021 on the website <https://hassavocadoboard.com/category-data/>as a primary data source.

I then compiled them in one single sheet and cleaned the data, formatted the cells, and cleaned the region data as we had duplicates like “BaltimoreWashington” and “Baltimore/Washington” so I put all that in one single style.

I then decided to go for the metrics that interested me the most: the date, average price, the total volume, and the volume per PLU codes

Problem with the “region” field: there are not only cities but also general areas like IN ('TotalUS', 'West', 'California', 'SouthCentral', 'Northeast', 'Southeast', 'GreatLakes', 'Midsouth') so they had to be removed (or added) depending on what we wanted to know

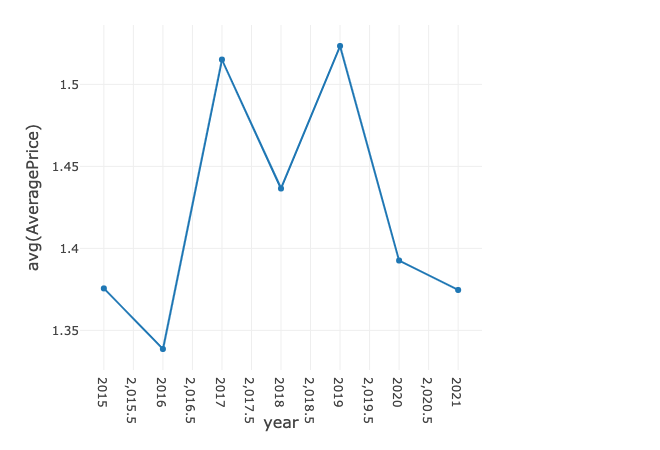


It seems that the PLU4225(Hass big size) and PLU4046 (Hass small size) are chasing each other until 2018 when plu 4046 (the small variety of Hass avocado) became the most sold.

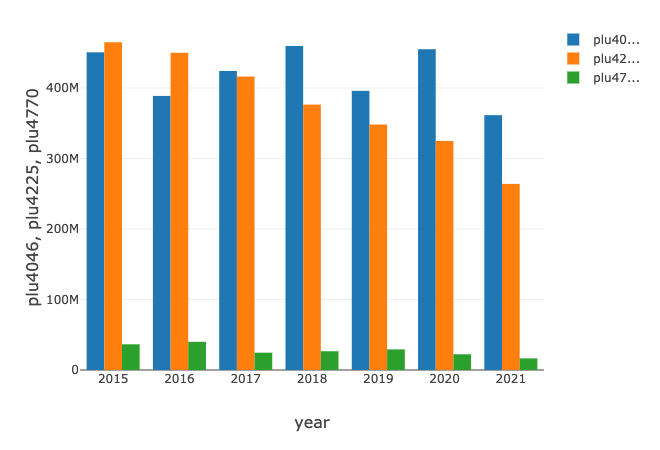
Fun fact: it takes 70 liters to grow ONE avocado so on average, 450 millions avocados are sold. So it took around 31,5 billion of liters of water.

Source:https://www.fairobserver.com/more/environment/hans-georg-betz-avocado-environment-water-footprint-production-consumption-europe-china-latin-america-news-13621/

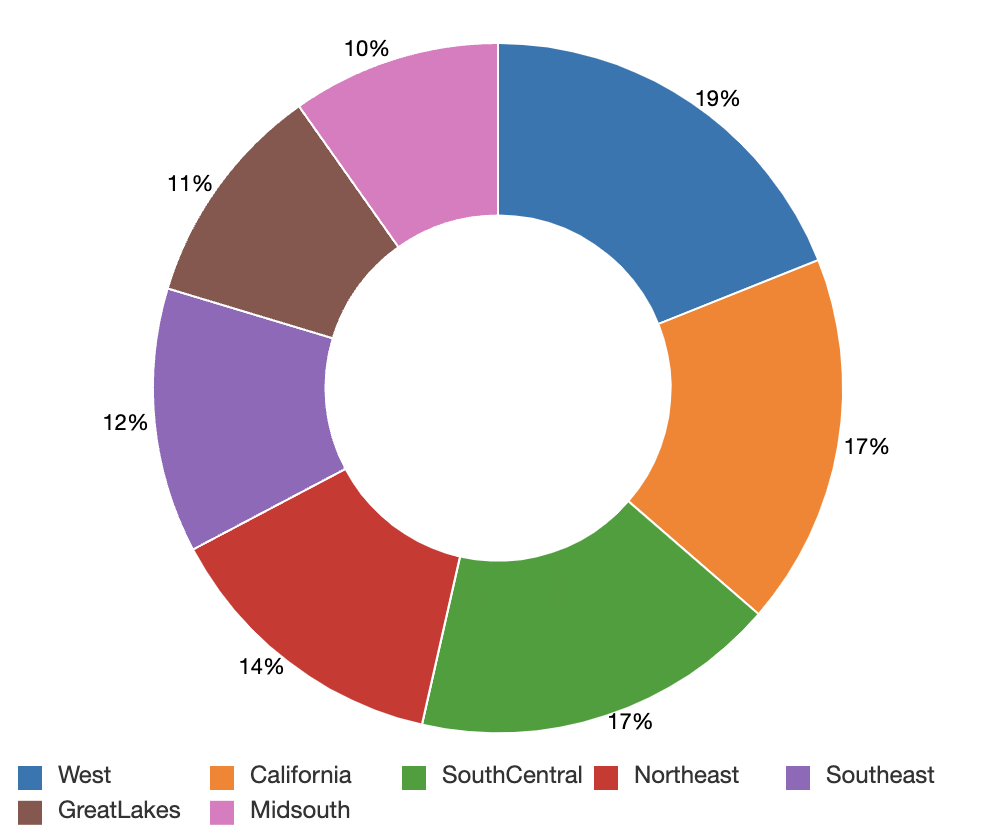
I then looked at the price, to see if it followed the trend of the volume sold:

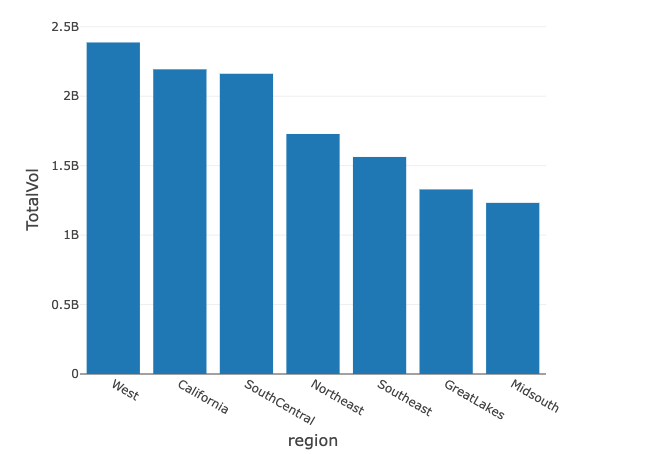


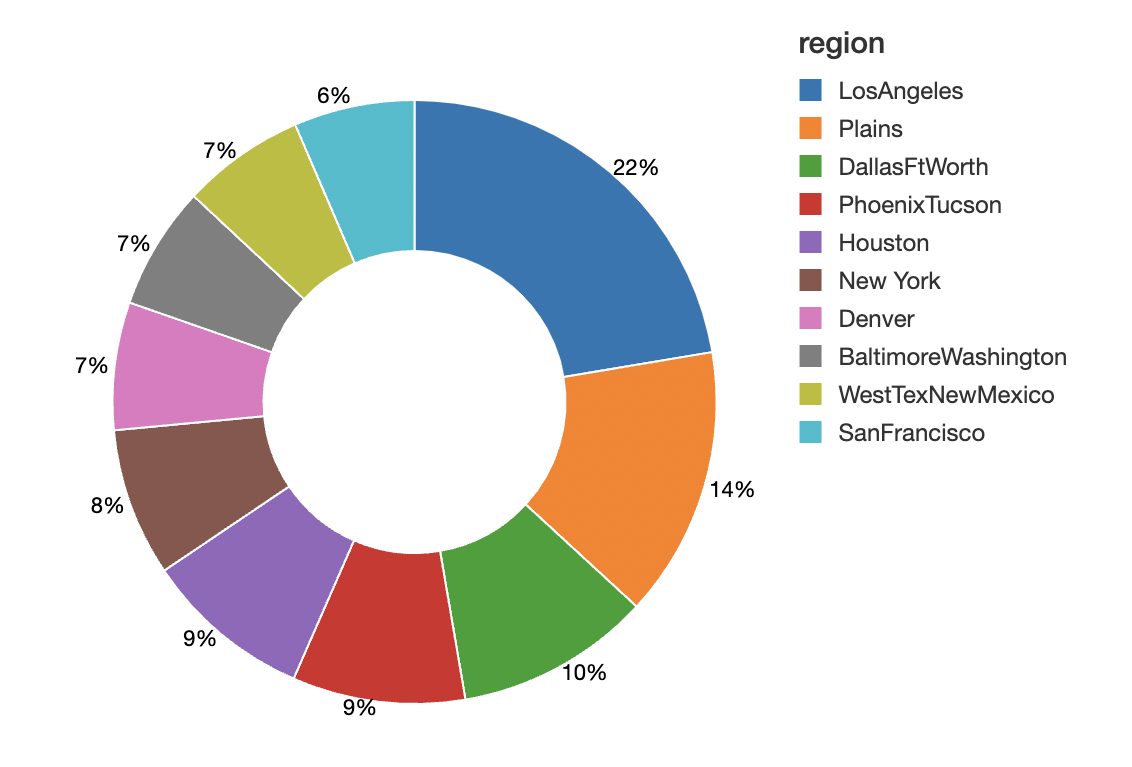
Let’s have a proper look at each variety of avocados . It’s clear that the plu4046 is ahead of the game now.



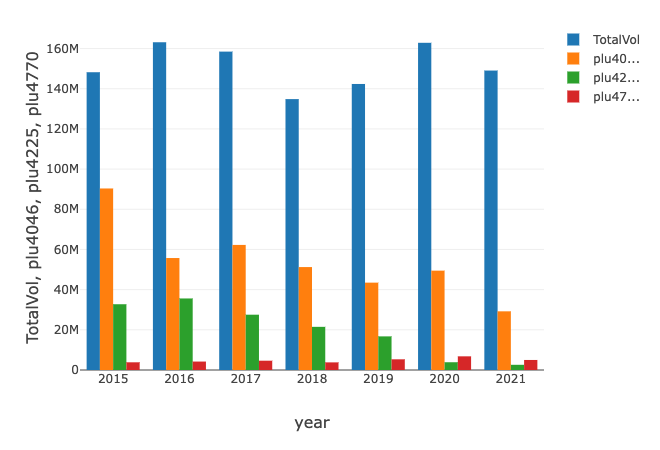
let’s have a look at each area





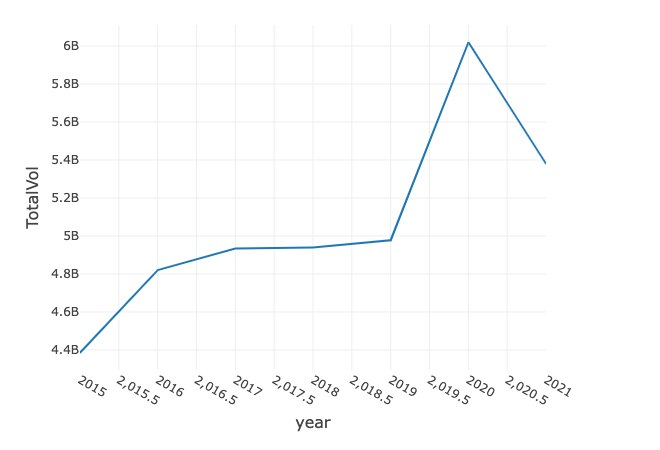


I want to see if the sales from the biggest avocado consuming city reflects the general sales of the avocado in the U.S and if one city is enough to see the general state of the US sales



when comparing it to the general sales, we see it’s not the case.

Lastly, I want to see if the price follows the sales trend.



What is the most amazing is from 2019 to 2020. the average price went from $1.5 to $1.3 but the sales went from 5 billions to 6 billions. Aside from that, the price alone is not sufficient enough to predict the sales, the quality of the product, the trends, the taste of people, in 2019 the COVID appeared so that surely played an effect on the sales of avocados.

**Questions to answer**

**1. Does the price have an influence on the volume?**

In a sense yes but it is not significant enough to have an influence, other variables come into place.

**2. Which areas consumed avocados the most from 2015 to 2021?**

West, California and South Central have consumed the most avocados from 2015 to 2021. Los Angeles is the city where the most avocados were sold during that period (1.05 billion avocados sold)

**3. The PLU4770 being the biggest Hass variety, is it also the one consumed the most?**

Quite the opposite, the PLU4770 is the least sold while the smaller one, the PLU 4046 is now the one sold the most.

However it would be important to compare the production units to the units sold to see if the fact that the PLU4770 being sold the least is because it is also the least produced.