**Millennials vs. Avocados**

I. Questions

We began our research for this project by visiting the resources presented to us. One of those sources, Kaggle, caught our attention with the diverse selection of datasets. On the website was featured a dataset named, “The Price and Sales of Avocado”. Intrigued by this data and our appreciation for avocado toast, we began asking the question: Do millennials affect the prices of avocados?

By asking this simple question, we began to narrow down our questions according the data available, which resulted in the following:

* Does the millennial population percentage affect the price of avocados?
* Does the average household income affect the price of avocados?
* Does the weather forecast affect the price of avocados?

II. Sources & Data Cleanup

*The Price and Sales of Avocado* - Kaggle.com

The dataset of ‘The Price and Sales of Avocado’ is the primary data of all avocado information. We first began with cleaning up this dataset in order to locate usable information. Our team removed unnecessary and duplicated data from this set and narrowed it down to date, average price, total volume, type, year, and region. Much of this data needed to be formatted due to the variation syntax in selected cities.

Census Data - Census.gov

The United States Census Bureau provided with data pertaining to people and economy. Our team began collecting this data by doing an API pull. This pull was then cleaned up by removing much of unnecessary data and only left with data that pertained to year, city, state, age range, total numbers, and income.

Weather - OpenWeather.com

In order to obtain any sort of weather information pertaining to our list of cities, we were required to do an API pull for such data. Due to the limited information provided for education purposes, information was limited. Our team worked with the weather data available.

Locations - Gmaps

Once we finalized the list of cities that our cleaned up avocado dataset included, we did an API pull on Gmaps for these cities. Since the previous datasets included information regarding cities, there was no cleanup necessary.

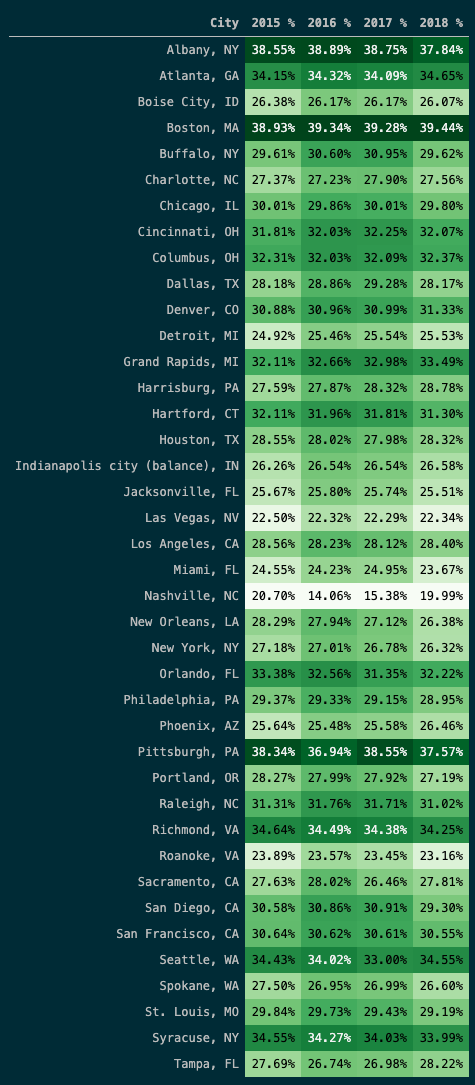
III. Data Analysis

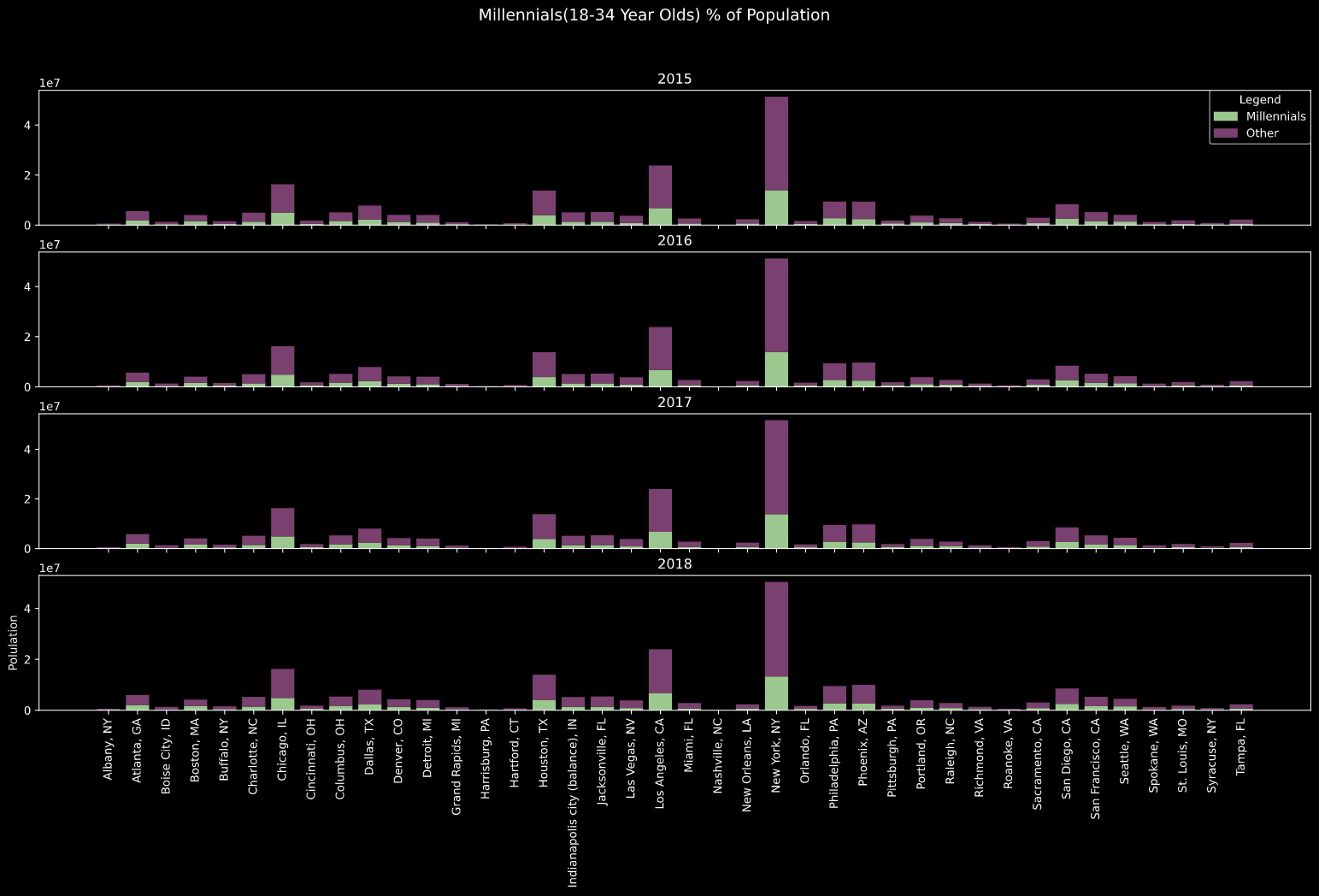
The following questions are answered with their findings:

Does the millennial population percentage affect the price of avocados?

Millennial Population vs. Average Avocado Price

First, we began by obtaining the millennial population numbers, which included the average percent in the overall population by city and year.





A picture containing text, green, scoreboard

Description automatically generatedBox and whisker chart

Description automatically generated

Once we had the necessary information regarding millennial population and the average price of avocados, we were able to create the following table that answers our question.

A picture containing graphical user interface

Description automatically generated

The line equation for Millennial is y = 181922.13x + -93868.77

The r-Squared for Millennial is 0.021035153654571443

The line equation for Other is y = 111806.21x + -68659.15

The r-Squared for Other is 0.02217732375560323

Does the average household income affect the price of avocados?

Average Household Income vs. Average Price Analysis

Since our initial avocado data included data on the average price on avocados, we decided to further use that information and obtain the following results.

A picture containing text, road, outdoor, street

Description automatically generatedChart, box and whisker chart

Description automatically generated

From there we used the census data to obtain the average household income by city and by year.

Chart

Description automatically generated

With all the necessary collected data, our team to pull both avocado and income datasets together to create the following table that shows a variety of results by year.

Chart

Description automatically generated

The line equation for 2015 is y = 6076.81x + 41442.05

The r-Squared for 2015 is 0.004458147154015406

The line equation for 2016 is y = 25900.41x + 17284.68

The r-Squared for 2016 is 0.10472518139494967

The line equation for 2017 is y = 28068.11x + 11416.46

The r-Squared for 2017 is 0.09341235056496099

The line equation for 2018 is y = 76348.44x + -48057.92

The r-Squared for 2018 is 0.4123616254754726

Does the weather forecast affect the price of avocados?

City Weather vs. Avocado Prices

With the limited weather data available for educational purposes, we were able to use that data and compare the average avocado prices to attempt to find a correlation between city weather temperature and avocado prices.

By completing an API pull from OpenWeather and using the cleaned avocado data with city and prices information, our team was able to create the following data.

Chart, scatter chart

Description automatically generated

The line equation is y = -0.0x + 1.56

The r-Squared is 0.040818288930768776

IV. Conclusion

The results that we received from our data does not conclusively state that millennials are a contributing factor on the average prices of avocados. In an overall comparison of total millennial population vs. all other total population the end data (r-values) is too low to be able to confirm.

Median income is another factor that we must take in consideration as our data shows that the average household income may have help shape the average price of avocados. Each year does vary in how related this data is. At least in 2018, we can see that it may have been a contributing factor yet overall data does not reflect this.

When it came to reviewing our city weather data and avocado prices, our team came to the outcome that there is no correlation that reflects that weather temperature has an effect on the prices on avocados.

Overall, the r-values from our completed data does not show any correlation as it is low, closer to 0 rather than 1. As a team, we have concluded that further data and research is needed in order to fully be able to make a correlation between millennials and other factors to the prices of avocados. There are many factors that may contribute to the pricing of avocados that we did not have data to include, such as the availability of avocados during peak vs non-peak season, harvest yield, or distance between the harvest location to the purchase location.