

The Future of Avocados

analyzing the avocado industry

Xiangzhen He | Katherine Hurtado-da Silva | Jazmin Logroño

# TITLE PAGE

California, the leading avocado producing state in America, is expected to increase its overall avocado production in the future. However, fluctuating precipitation levels has shown it is expected to adversely affect the Santa Barbara and Ventura avocado production, leaving most of the farming pressure to San Luis Obispo, San Diego, and San Bernardino. Increasing wildfires and limited water supply both serve as rationale for farmers to explore sustainable irrigation methods. Specifically, drip irrigation is recommended to maintain avocado growth and as a wildfire preventative measure. Since regional temperatures are not varying significantly, this recommendation is expected to increase production and reduce its vulnerability to fires, therefore allowing the California avocado industry to meet the ever-increasing consumer demands. Furthermore, the stabilization of the avocado production industry is expected to alleviate the fluctuating average avocado prices for customers, particularly for Connecticut, North Carolina, Illinois, and Massachusetts residents, who are known for having historically high retail costs for these goods.

# INTRODUCTION

People are becoming more aware of the health impacts caused by poor eating habits. As awareness increases, people seek out healthier foods. According to the Food Network, avocados are good for the heart and eyes, they help in growth and development, help lower blood pressure, and can help consumers lose weight. The health benefits offered by avocados have made them a popular commodity in the United States. Since 2000, United States avocado per capita consumption has significantly increased from around 2 pounds in 2001 to more than 8 pounds in 2017 (Statista, 2020). Though most avocados are imported from other countries like Mexico and Peru, California is the main avocado producer out of all the states in America. Unfortunately, California has been vulnerable to recent environmental changes and requires an evaluation of these factors in order to sustain its agricultural industry.

# SPECIFICATION

## PROBLEM

California, the state with the largest avocado production, has been suffering from an increase in droughts and wildfires in recent years. Historical production, pricing, and climate changes were analyzed to determine how the avocado industry is expected to change in the future in order to meet demand and mitigate significant fluctuations in average avocado prices across the United States.

## HYPOTHESIS

We predicted average temperatures, precipitation, and total wildfires to have significant impacts on avocado production forecasts. Given recent global warming concerns, we expected to see rising temperatures, declining precipitation levels, and increasing wildfires. Furthermore, these anticipated trends were expected to have adverse effects on avocado production, resulting in rising consumer prices.

## About the Data

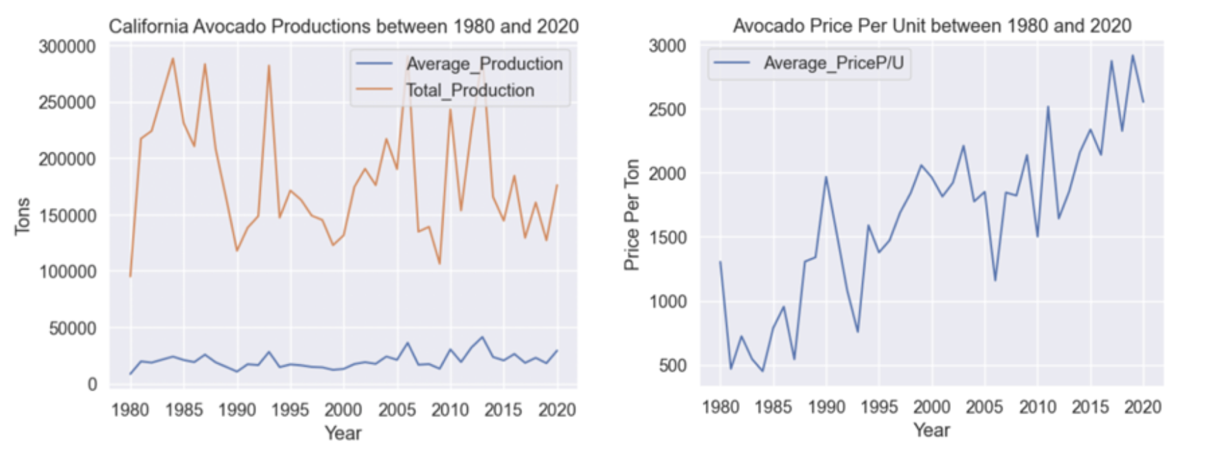
The main dataset, California Avocado Production[[1]](#footnote-2), is hosted by Kaggle and consists of 420 rows with yearly values from 1980-2020 for twenty-seven California counties. The data includes information about Harvested Acres, Yield, Production, Price per Unit, and Value in USD. In this report, Year, County, Production, and Price P/U variables will be primarily used to perform the analysis. During the data cleaning process, extra spaces were removed from column names and value cells, records that have missing values were excluded, and the data types of some variables were corrected.

In addition, weather data was obtained from the United States National Weather Service to help analyze temperature and precipitation impacts on avocado production and prices. California wildfire data was obtained from the Department of Forestry and Fire Protection to provide information about how many wildfires have occurred each year in the state of California. Moreover, a second Kaggle dataset[[2]](#footnote-3), was used to evaluate avocado prices in different States throughout 2015-2018.

# OBSERVATIONS

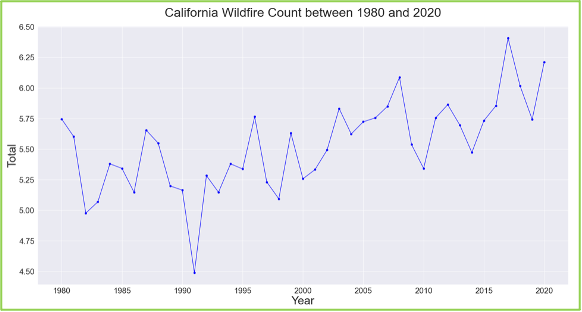
## STATE-WIDE TRENDS

The avocado total production and unit price was first reviewed at the state level. Two time series plots were created to show how total production and average price per pound have changed between 1980 and 2020.

**Figure 1**: Time series plots for historical California Avocado Production and Unit Price

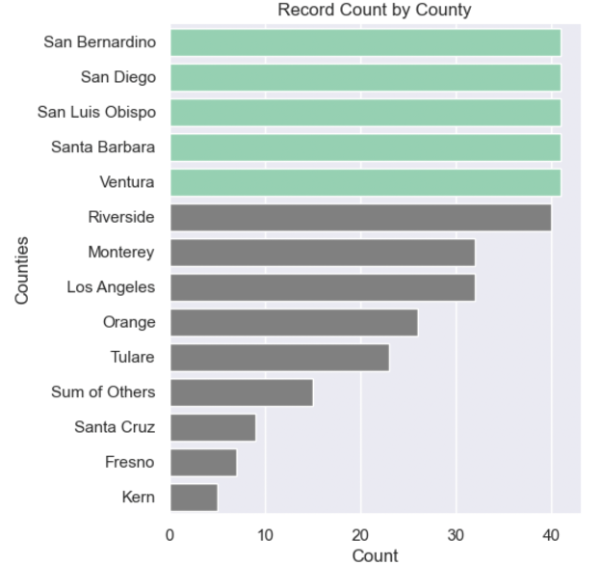
The graphs in Figure 1 show that there is a cyclical pattern in total production and unit price, and the intervals between each cycle have been getting shorter and shorter in the last decade. Generally, the total avocado production has decreased while the cost has increased in recent years. Given the rising demand of avocados over the years in the United States, but decreasing production of these goods in California, the rising prices were expected and were substantiated by these figures.

Figure 2 shows wildfires in California have been increasing over time. Despite the data being representative of the entire state, a recent article reported Ventura County was less vulnerable to this phenomenon than San Diego (The Sacramento Bee, 2022). Specifically, Ventura had 44% of its properties at moderate risk or higher, while San Diego had more than 80% of its area subject to wildfires. This data may explain why Ventura replaced San Diego as the largest avocado producer in California in recent years. The increasing trend in total wildfire occurrences suggests it is correlated to the declining avocado production over the years.

**Figure 2**: California Wildfire History

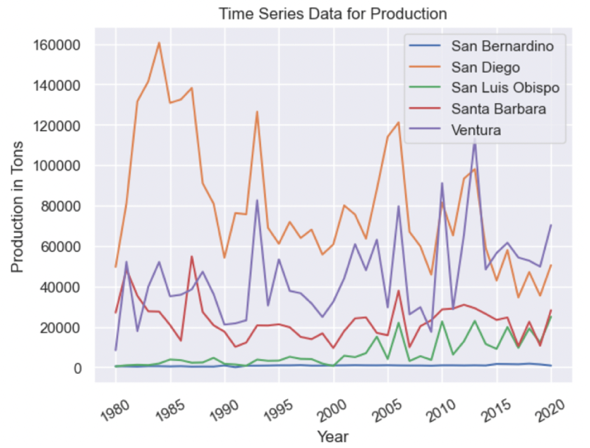
## COUNTY-WIDE TRENDS

When reviewing values at the county level, the bar plot below summarizes the number of records in each county, which suggests that only five counties, San Bernardino, San Diego, San Luis Obispo, Santa Barbara, and Ventura County, have all 41 years of records between 1980 and 2020.

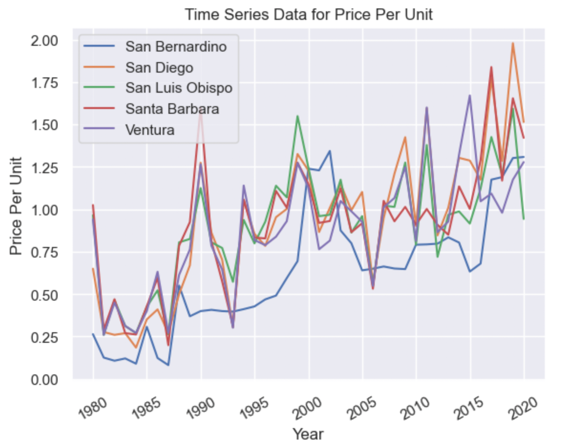


**Figure 3**: Number of Records in Each County

Two time series plots were created to show how total production and average unit price have changed over the years in these five counties. As shown in Figure 4 below, San Diego and Ventura County produced the highest amount of avocado in California. Historically, San Diego County was the leading avocado producer in California. However, Ventura County has surpassed them since 2015 and has been able to offer its avocados at a low price compared to the remaining counties.

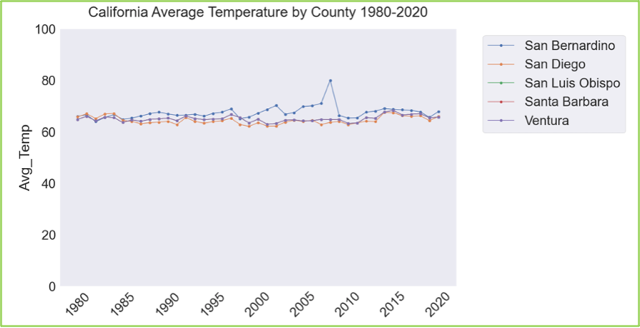


**Figure 4:** Time Series Plot of Total Avocado Production at the County Level

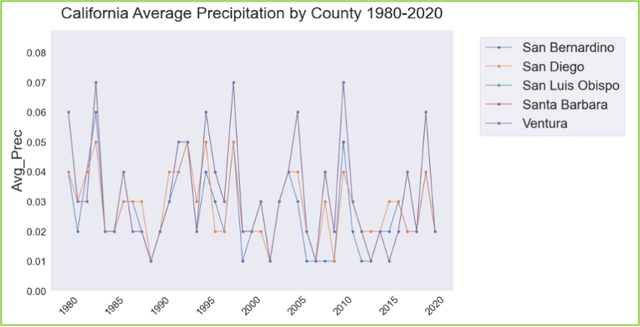


**Figure 5**: Time Series Plot of Average Unit Price at the County Level

Annual mean temperatures and precipitation levels for each California county with complete records are plotted below. Temperature variations between these five counties do not appear until 1985, where San Bernardino began and remained having the highest temperatures. On the other hand, San Diego’s temperature slightly decreased in 1985, and continued to have the lowest temperatures out of all the counties in the plot. Ventura, San Luis Obispo, and Santa Barbara are all in the Long Beach region and share the same temperature and precipitation levels, despite having varying production outputs. Over time, these counties did not have significant fluctuations in temperature. It is worth noting that as of 2015, all five counties have temperatures that do not vary by more than a few degrees Fahrenheit and that there is no significant positive or negative trend. Lastly, though precipitation levels also fail to demonstrate an observable and strong relationship, Ventura, San Luis Obispo, and Santa Barbara County do appear to have all the highest peaks of these levels.



**Figure 6**: Time Series Plot of Annual Mean Temperature



**Figure 7**: Time Series Plot of Annual Mean Precipitation Levels

## Avocado Prices in the US

The Avocado Prices data consists of avocado prices in different cities for 2015-2018. Prices are aggregated at a state level for twenty-eight States to create an average for avocado prices. Exploring the data provided insights on States with the highest and lowest average avocado prices. States with the highest average avocado prices include Connecticut, North Carolina, Illinois, Massachusetts, and Maryland. States with the lowest average avocado prices include Texas, Colorado, and Ohio. Arizona, New Mexico, and Colorado are relatively close to California compared to New York or North Carolina. The average avocado prices are relatively low compared to States with higher average prices that are further away from California. Figure 8 is a map using color to denote average avocado prices in each State. States with no data are colored gray.

A picture containing map

Description automatically generated**Figure 8**: Map representing Average Avocado Prices from 2015-2018

# ANALYSIS

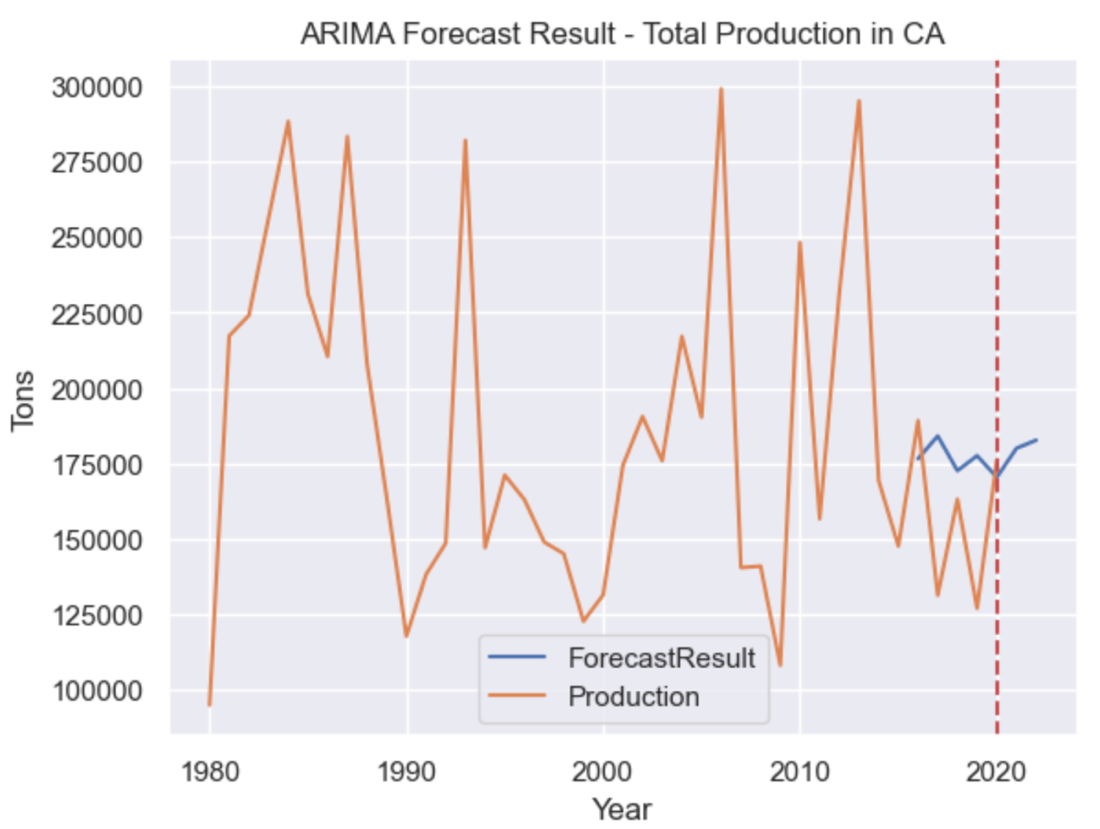
Forecasting models use historical time series data to help us understand the future of avocado production and prices.

## Time Series Analysis Using ARIMA

Time series analysis was first performed using the Autoregressive Integrated Moving Average or ARIMA models. The ARIMA model is trained from past values and then used to predict future values. Additionally, models need to be trained from stationary time series data. Therefore, the Augmented Dickey-Fuller (ADF) test needs to be performed on the time series data to check stationarity. If the data is non-stationary, differencing will be performed to make the time series stationary. Moreover, the autocorrelation plots and partial autocorrelation plots are created to determine the parameters for ARIMA models. When building the models, 90% of the historical values are used as the training data, and the remainder 10% of the historical data is used for model evaluation.

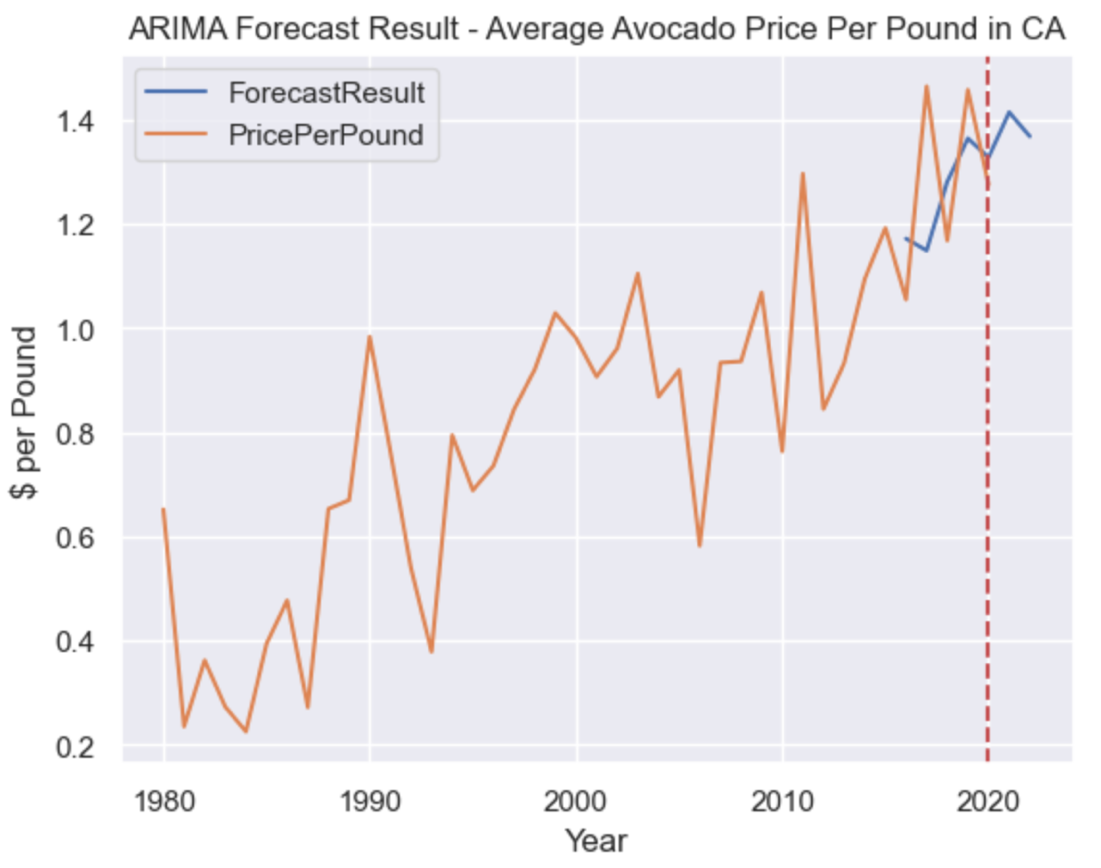
**ARIMA ANALYSIS: STATE-WIDE PRODUCTION AND UNIT PRICE ANALYSIS**

In this section, the ARIMA model was used to predict statewide avocado total production and average unit price in 2021 and 2022. First order differencing was performed using the diff() function to make both time series data stationary. In the end, we predict total avocado production will continue to grow in 2021 and 2022, as shown in the plot below. The predicted total production in 2021 is 180,044 tons and 182,708 tons in 2022.



**Figure 9**: ARIMA Forecasting for Statewide Avocado Production

In addition, we predict unit prices will slightly go up in 2021 and decrease in 2022, as shown in Figure 10. The predicted average price per pound in 2021 is $1.41 and $1.37 in 2022.



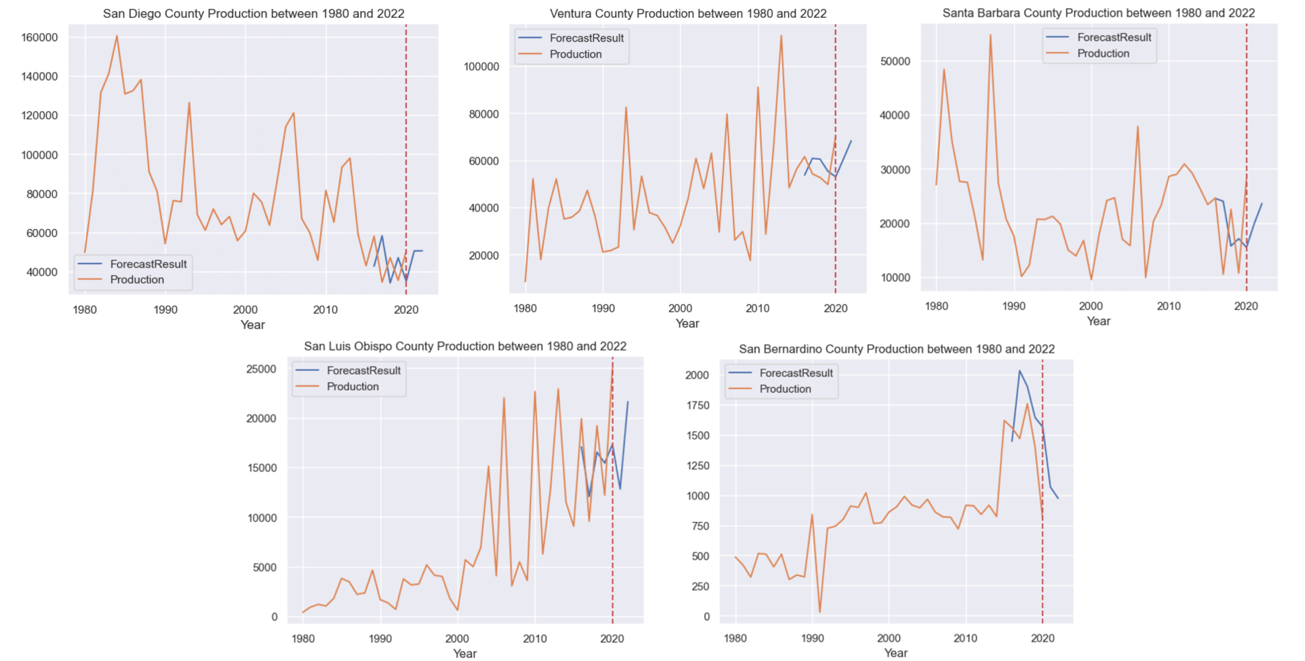
**Figure 10**: ARIMA Forecasting for Statewide Avocado Unit Price

The root mean squared errors (RMSE) for each model was calculated using the evaluation or test data. The RMSE for the model used to predict total production is around 33,510, and the RMSE for the unit price prediction model is approximately 0.166.

**ARIMA ANALYSIS: COUNTY-WIDE PRODUCTION AND UNIT PRICE ANALYSIS**

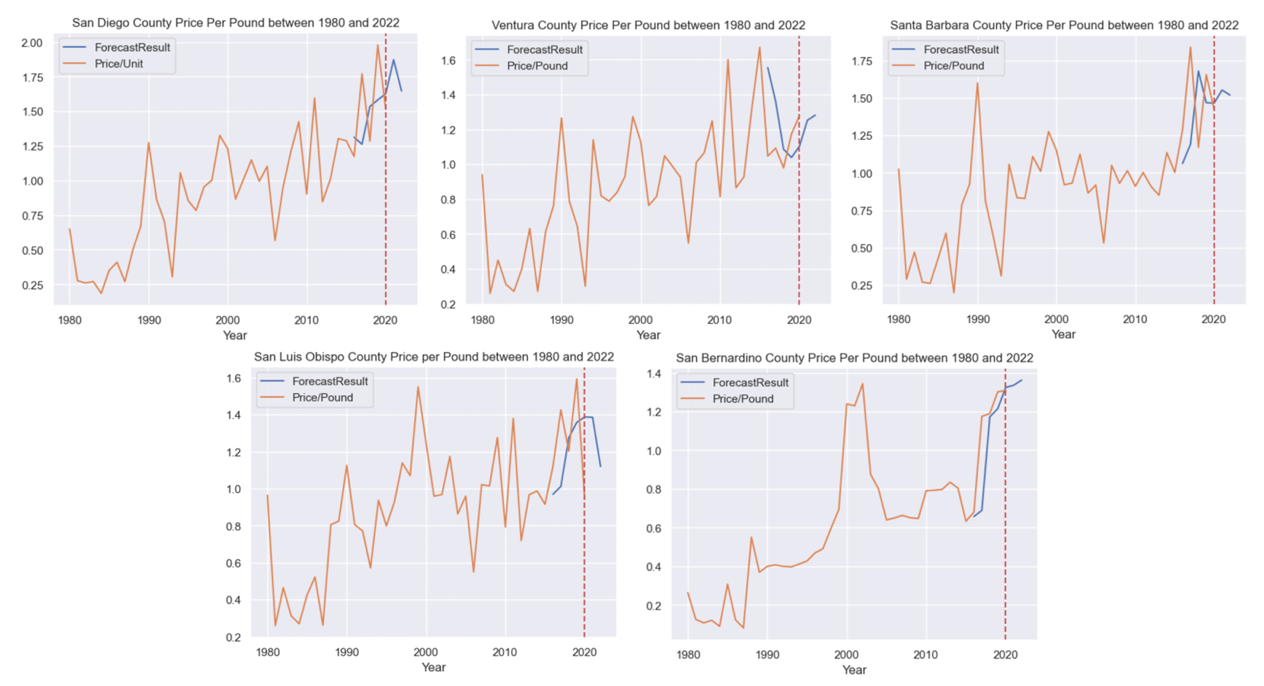
In this section, ARIMA models were built to predict total avocado production and the average unit price at the county level for five counties that have all 41 years of records, still using 90% of data to train the models.

The graphs below present the predicted total avocado productions for each county.

**Figure 11:** ARIMA Forecasting Results for Total Production at the County Level

As shown in the graphs above, the total avocado production in San Diego, Ventura, and Santa Barbara Counties will continue to grow in 2021 and 2022, and Ventura County will remain as the biggest avocado producer within the state of California. For San Luis Obispo County, production will most likely decrease in 2021 and increase again in 2022, similar to the patterns found in the previous cycles. Unlike other counties that all show an overall growing trend, the production amounts in San Bernardino County will continue to go down in 2021 and 2022.

In addition, the graphs below present the predicted average avocado price per pound in each county.

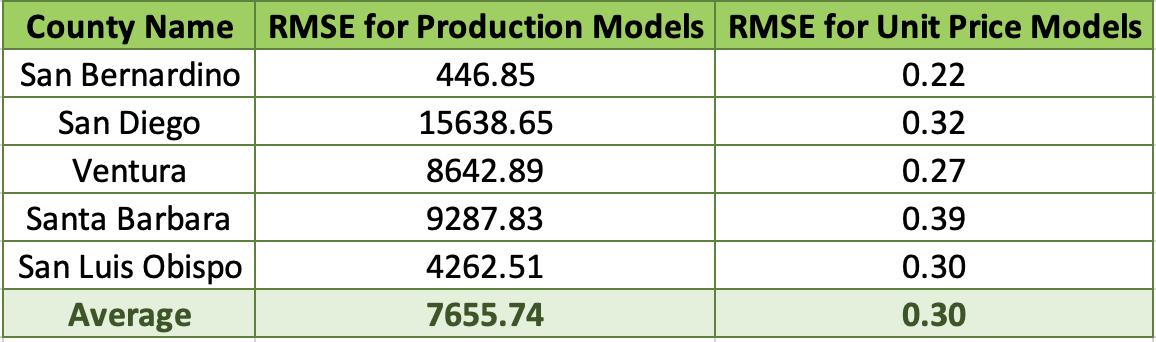
**Figure 12:** ARIMA Forecasting Results for Average Price Per Pound at the County Level

As shown in Figure 12, the average avocado unit price in Ventura and San Bernardino Counties will grow in both the years 2021 and 2022. The average unit prices in San Diego and Santa Barbara Counties are likely to increase in 2021 but decrease slightly in 2022. In San Luis Obispo County, the unit price is likely going to remain unchanged in 2021 and then decrease substantially in 2021.

Based on the prediction results, San Diego County will probably to continuedly provide avocados at the highest price in 2021 and 2022 at above $1.6 per pound, which is much higher than the predicted state average price of around $1.4 per pound. Ventura County, the current largest avocado producer in California, will most likely sell its avocados at around $1.3-$1.4 per pound in 2021 and 2022. In 2022, San Luis Obispo County is predicted to be reaching the peak of its avocado production cycle and providing avocados at the lowest unit price among all five counties at around $1.0 per pound.

Therefore, both Ventura County and San Luis Obispo County will be good options for companies who want to buy large amounts of avocados at lower prices. However, smaller businesses that wish to buy at the lowest cost should probably choose to purchase from San Luis Obispo County.

Lastly, the Root Mean Square Error (RMSE) for each model was calculated using the evaluation or test data, and the results are summarized in the table below.



**Table 1**: Evaluation results for ARIMA Models built using data from each County

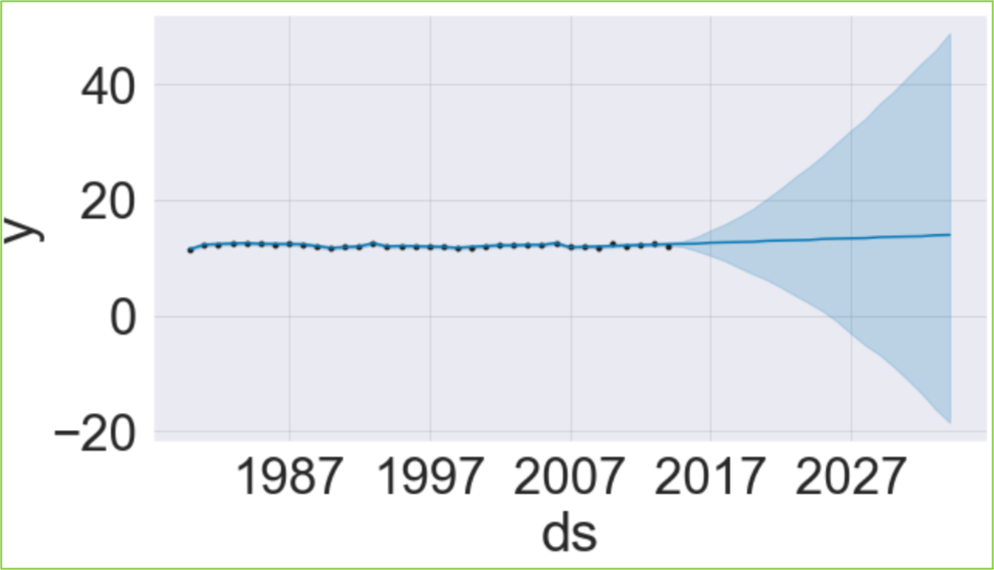
The average RMSE for the models used to predict total production in each county is approximately 7,656, and the average RMSE for the unit price prediction models is about 0.30.

## Time Series Analysis Using PROPHET

Facebook’s Prophet is a Python library that uses an additive regression model with a piecewise linear or logistic growth curve trend. Data between 1980 and 2014 inclusive, was saved as the train set, while the remaining six records of the data was saved as the test set, otherwise known as an out-of-sample. Since production trends do not appear to be linear, the multiplicative seasonality mode was used. An uncertainty interval of 95% was selected over the default of 80% to ensure that 95% of the sample fit within the boundaries that the upper and lower predicted bands establish, represented by yhat\_upper and yhat\_lower respectively. Since Prophet models are optimally trained from stationary time series data, the Augmented Dickey-Fuller (ADF) test was also performed on the time series data to check for stationarity. Log transformation was applied to production data, wildfire, and temperature data to reduce non-stationarity. The inverse of the log transformation was applied to the forecasted production that was normalized to effectively calculate the RMSE and evaluate the model’s performance.

**PROPHET ANALYSIS: STATE-WIDE PRODUCTION AND WILDFIRE REGRESSOR**

At the state level, the total production for all records was aggregated by year to forecast how much California is expected to produce. Since wildfire data is reflective of the entire state and not by county, this is the only model where it was applied as a regressor. The figure below shows the baseline results of California’s production forecast. Despite forecasting increasing avocado production as well, its associated RMSE value of 144,444.84 tons means the ARIMA model had higher fidelity by having the significantly lower error of 33,510 tons. When wildfire data was added as a regressor, it magnified the forecasting error to 6.15 x 1069, ultimately concluding it was not helpful in forecasting state-wide avocado production.



**Figure 13**: Prophet Forecasting Results for Average Production at the State Level

**PROPHET ANALYSIS: COUNTY-WIDE PRODUCTION AND WEATHER REGRESSORS**

Three versions of the Prophet forecasts were generated for each county, one serving as the baseline with avocado production data alone, the second with annual mean temperature included as a regressor, and the third with the annual average precipitation levels replacing the temperature regressor. The two regressors were never added together to form a fourth model due to the adverse impact one or both had on the baseline results. Since temperature and precipitation shape the optimal conditions for avocado growth, they were used as regressors for production forecasts alone. This data was also available by county, therefore enabling a focused analysis for each of these regions’ production.

In all five county analyses, the addition of temperature as a regressor increased the RMSE exponentially and was determined to be unhelpful in forecasting avocado production. Using precipitation levels as a regressor reduced the RMSE values for all models except for San Bernardino. For all county-level production forecasts, the prophet models with the smallest RMSE values outperformed the ARIMA forecasts, indicating precipitation as a regressor was effective in forecasting county-wide avocado production in four of the five counties.

The figures below show the Ventura and Santa Barbara baseline models forecast increasing avocado production. However, when precipitation was added as a regressor to both models, their performances improved by reducing the RMSE from 4,355.55 to 281.70 and 12,025.95 to 4,509.12 tons, respectively. Precipitation had an adverse effect on both production trends, indicating that Ventura and Santa Barbara are expected to have declining avocado production in the future.



**Figure 14**: Baseline Prophet Forecasting Results for Ventura Production

Chart

Description automatically generated

**Figure 15**: Baseline Prophet Forecasting Results for Santa Barbara Production

San Luis Obispo and San Diego also had their RSME values improve with the addition of precipitation as a regressor. Though San Luis Obispo originally showed a declining forecast, precipitation as regressor forecasted it would have increased avocado production. San Diego production maintained the slight increase in production when precipitation was added to its model, making it the only model without a trend reversal due to precipitation. However, it is worth noting that despite the RMSE improvements of 12,930.17 to 1,409.11 and 51,369.82 to 9,739.99 tons, San Diego’s error is more significant, making its subtle projected increases questionable.

Chart

Description automatically generated

**Figure 16**: Baseline Prophet Forecasting Results for San Luis Obispo Production

Chart, line chart

Description automatically generated

**Figure 17**: Baseline Prophet Forecasting Results for San Diego Production

The addition of each regressor widened the error for San Bernardino’s forecast, making the baseline model with an RMSE of 436.46 tons the best. The figure below shows the optimal baseline model projects this county will have increased avocado production. Table 2 summarizes the evaluative RMSE values for all county forecasts.

Chart, line chart

Description automatically generated

**Figure 18**: Baseline Prophet Forecasting Results for San Bernardino Production

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **County** | **Production RMSE on Test Set** | **Production & Temperature RMSE on Test Set** | **Production & Precipitation RMSE on Test Set** | **Production & Wildfire RMSE on Test Set** |
| Ventura | 4,355.55 | 2.2189758173940270406 e+1305 | **281.70** | N/A |
| San Bernardino | **436.46** | 1.7409024065682025788 e+699 | 10,828.81 | N/A |
| Santa Barbara | 12,025.95 | 21,130.35 | **4,509.12** | N/A |
| San Luis Obispo | 12,930.17 | 16,893.49 | **1,409.11** | N/A |
| San Diego | 51,369.82 | 2.7546451867950760886e+1783 | **9,739.99** | N/A |

**Table 2**: Evaluation results for Prophet Models by County

## forecasting state average avocado prices with prophet

A Prophet model was trained for each of the twenty-eight States in the Avocado Prices dataset to understand average Avocado prices for the next four years. States with the highest average Avocado prices like Connecticut, North Carolina, Illinois, and Massachusetts trends are reviewed. The Prophet model predicts average Avocado prices in Connecticut decline from 2019-2020, slightly increase in 2021 and continue to decrease after 2021 as shown in Figure 19.

Chart, line chart

Description automatically generated**Figure 19**: Prophet Forecasting Model for Connecticut

North Carolina and Massachusetts share the same trend for average Avocado prices. Average Avocado prices will continue to rise as shown in Figures 20 and 21. However, average prices in Massachusetts will increase at a more rapid rate.

Chart, scatter chart

Description automatically generated

**Figure 20**: Prophet Forecasting Model for North Carolina

Chart, line chart

Description automatically generated

**Figure 21**: Prophet Forecasting Model for Massachusetts

The trend is unique for Illinois average Avocado prices. As shown in Figure 22, average Avocado prices will remain the same until 2020. Average prices will increase after 2020 and then remain constant for the next few years.

Chart, line chart

Description automatically generated

**Figure 22**: Prophet Forecasting Model for Illinois

The gif image below in Figure 23 reveals interesting changes for other States. Idaho, for example, has a relatively low average Avocado price value before 2019. According to the Prophet model, Idaho will have one of the highest average Avocado prices in 2019, decrease in 2020 and then continue to increase after 2021 to once again having one of the highest average State values of Avocado prices.

**Figure 23:** Maps of Prophet Model Forecast of Average Avocado Prices for 2019-2022

# RECOMMENDATION

The ARIMA results indicate that avocado total production in California will continue to grow in 2021 and 2022, but its long way from past production highs is reason for concern. When considering the effect of precipitation in the forecasts, San Luis Obispo, San Diego, and San Bernardino are expected to have increased production, showing they are being spared from the adverse effects precipitation levels had on Ventura, and Santa Barbara farmland. To mitigate seasons with low precipitation levels, it is recommended California farmers create basins to collect water from rain and use it to form a specialized irrigation schedule for their avocado farms. Though wildfires proved to be unhelpful in the forecasts, its growing trend suggests sustainable methods like drip irrigation are necessary to not only cultivate avocados, but to prevent fires from forming and spreading as well.

The avocado state average per-unit price will increase in 2021 and then decrease in 2022. Ventura and San Luis Obispo Counties are good options for avocado vendors seeking large amounts of avocado quantities for lower prices. San Luis Obispo County holds one of the lowest costs for avocados and might be of interest to smaller businesses. Evaluating demand at a state level can help optimize avocado prices for each state.

# REFERENCES

1. Statista. (2022, November). *U.S. per capita consumption of fresh avocados 2000-2021*. Retrieved December 19, 2022, from <https://www.statista.com/statistics/257192/per-capita-consumption-of-fresh-avocados-in-the-us/>
2. *5 surprising health benefits of avocados*. Food Network. (n.d.). Retrieved December 18, 2022, from <https://www.foodnetwork.com/healthyeats/recipes/2014/12/5-surprising-health-benefits-of-avocados>
3. The Sacramento Bee. (2022, May). *Is my California home at risk for wildfires? New tool lets you check your ZIP code*.Retrieved November 26, 2022, from

<https://www.sacbee.com/news/california/fires/article261495002.html>

1. *Avocado PLU's*. California Avocados. (n.d.). Retrieved December 18, 2022, from <https://californiaavocado.com/retail/avocado-plus/>

1. https://www.kaggle.com/datasets/jarredpriester/california-avocado-production-19802020 [↑](#footnote-ref-2)
2. https://www.kaggle.com/datasets/neuromusic/avocado-prices [↑](#footnote-ref-3)