

# AVOCADO PRICE FORECASTING

Kang, Andini Wulandari





# About the Avocado Price Dataset

The dataset represents weekly 2018 retail scan data for National retail volume (units) and price. Retail scan data comes directly from retailers' cash registers based on actual retail sales of Hass avocados. Starting in 2013, the table below reflects an expanded, multi-outlet retail data set. Multi-outlet reporting includes an aggregation of the following channels: grocery, mass, club, drug, dollar and military. The Average Price (of avocados) in the table reflects a per unit (per avocado) cost, even when multiple units (avocados) are sold in bags. The Product Lookup codes (PLU's) in the table are only for Hass avocados. Other varieties of avocados (e.g. greenskins) are not included in this table.

**Date** :The date of the observation  
**AveragePrice** : the average price of a single avocado  
**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  
**Total Bags** : Total number of bags  
**Small Bags** : Total number of small bags  
**Large Bags** : Total number of large bags  
**XLarge Bags** : Extra Large Bags  
**type** : conventional or organic  
**year** : year of the date  
**region** : the city or region of the observation

Data Information

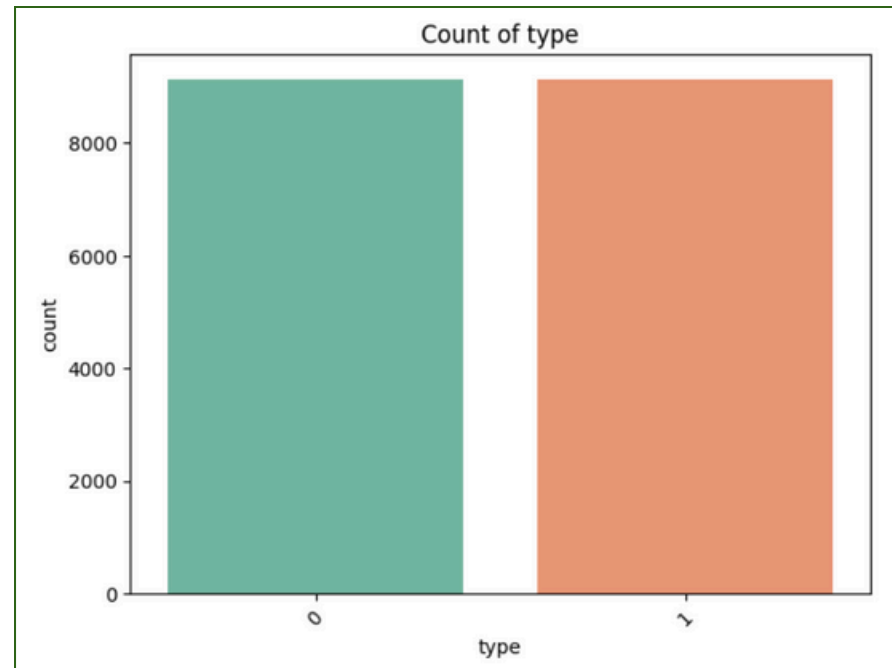
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18249 entries, 0 to 18248
Data columns (total 13 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Date            18249 non-null  object
1   AveragePrice    18249 non-null  float64
2   Total Volume    18249 non-null  float64
3   4046            18249 non-null  float64
4   4225            18249 non-null  float64
5   4770            18249 non-null  float64
6   Total Bags      18249 non-null  float64
7   Small Bags      18249 non-null  float64
8   Large Bags      18249 non-null  float64
9   XLarge Bags     18249 non-null  float64
10  type            18249 non-null  object
11  year            18249 non-null  int64
12  region          18249 non-null  object
dtypes: float64(9), int64(1), object(3)
memory usage: 1.8+ MB
```

Data Attributes

Top 5 rows of Data

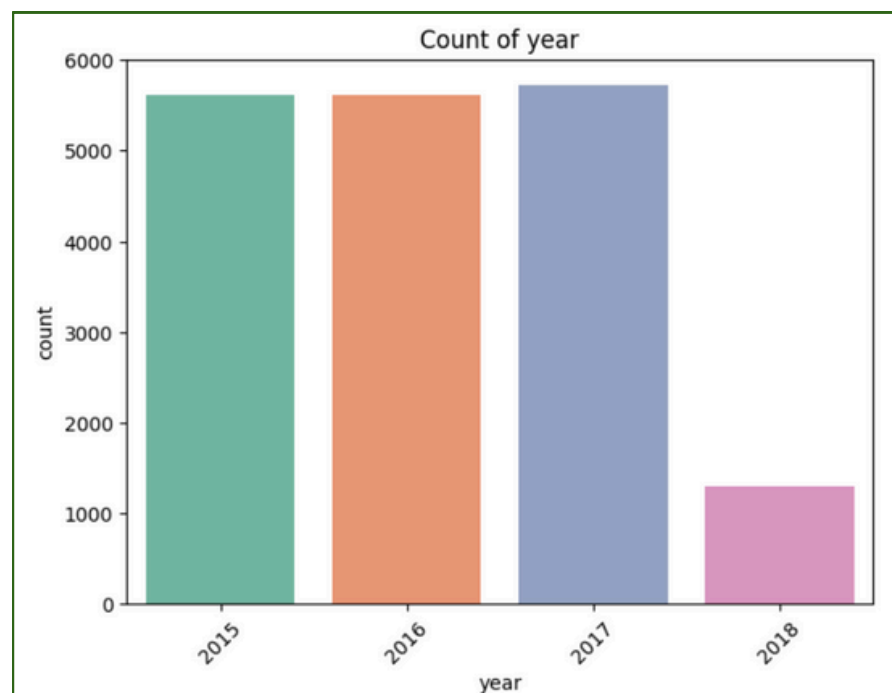
	Date	AveragePrice	Total Volume	4046	4225	4770	Total Bags	Small Bags	Large Bags	XLarge Bags	type	year	region
0	2015-12-27	1.33	64236.62	1036.74	54454.85	48.16	8696.87	8603.62	93.25	0.00	conventional	2015	Albany
1	2015-12-20	1.35	54876.98	674.28	44638.81	58.33	9505.56	9408.07	97.49	0.00	conventional	2015	Albany
2	2015-12-13	0.93	118220.22	794.70	109149.67	130.50	8145.35	8042.21	103.14	0.00	conventional	2015	Albany
3	2015-12-06	1.08	78992.15	1132.00	71976.41	72.58	5811.16	5677.40	133.76	0.00	conventional	2015	Albany
4	2015-11-29	1.28	51039.60	941.48	43838.39	75.78	6183.95	5986.26	197.69	0.00	conventional	2015	Albany

# Data Visualization



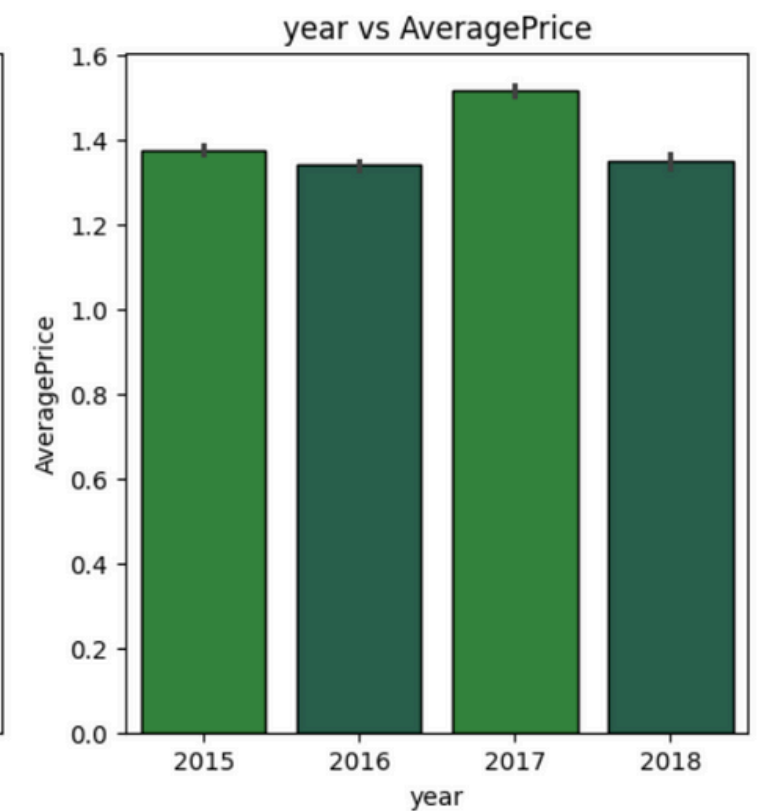
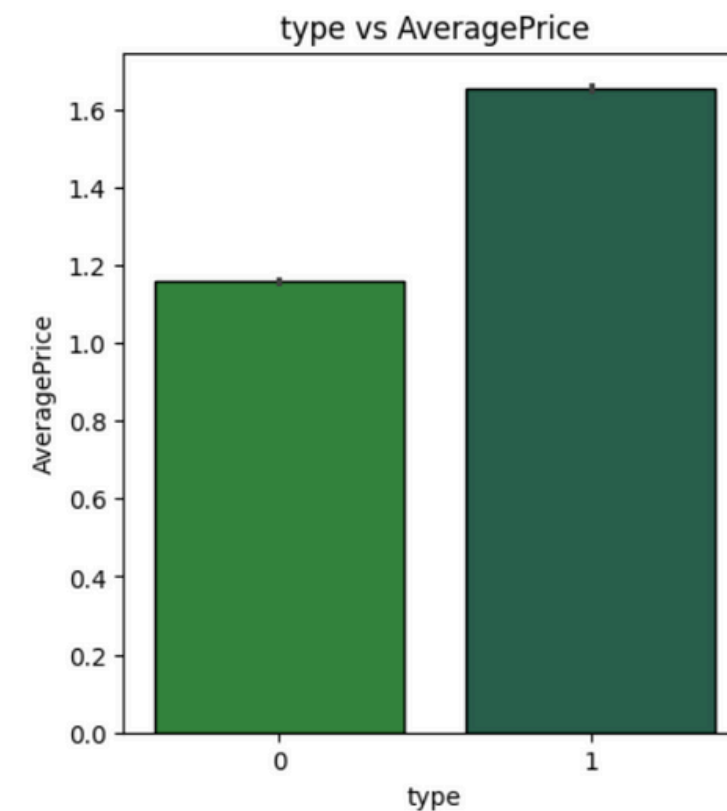
## Type Distribution

- Conventional and Organic avocados are present in almost equal numbers.



## Yearly Data Distribution

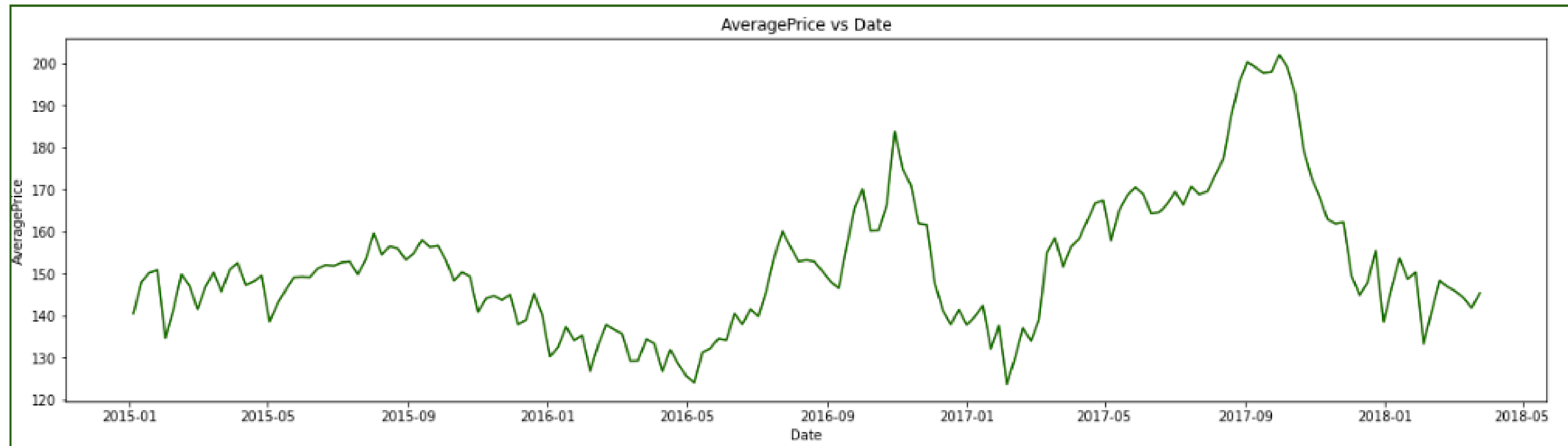
- Data points are consistent for the years 2015, 2016, and 2017.
- There is a sharp drop in 2018, likely due to limited or missing data.



## Average Price Analysis of Avocados

- By Type
  - Conventional (0) avocados have a lower average price compared to Organic (1) avocados.
- By Year
  - The average price remains relatively stable during 2015, 2016, and 2018.
  - A significant price increase is observed in 2017.

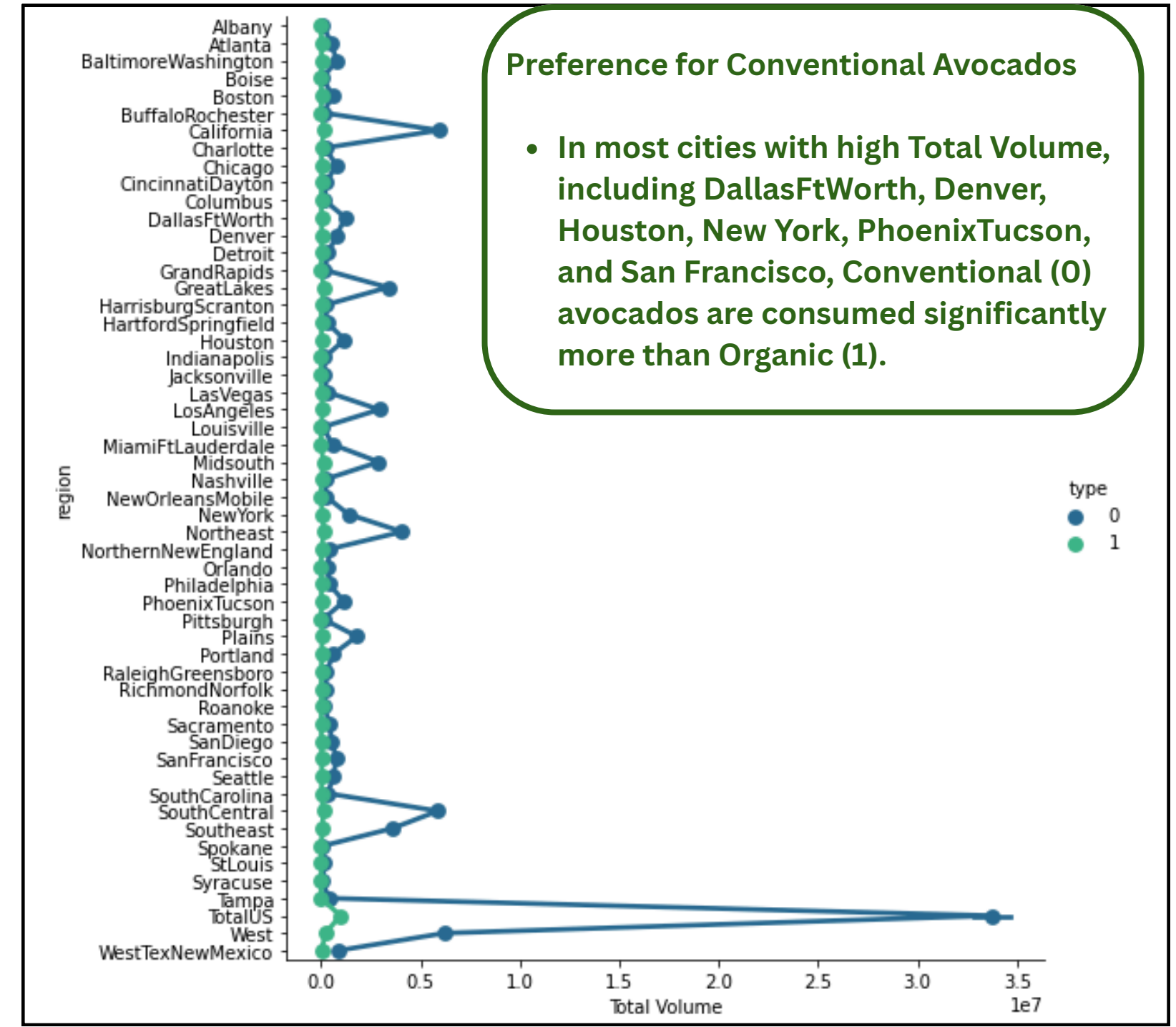
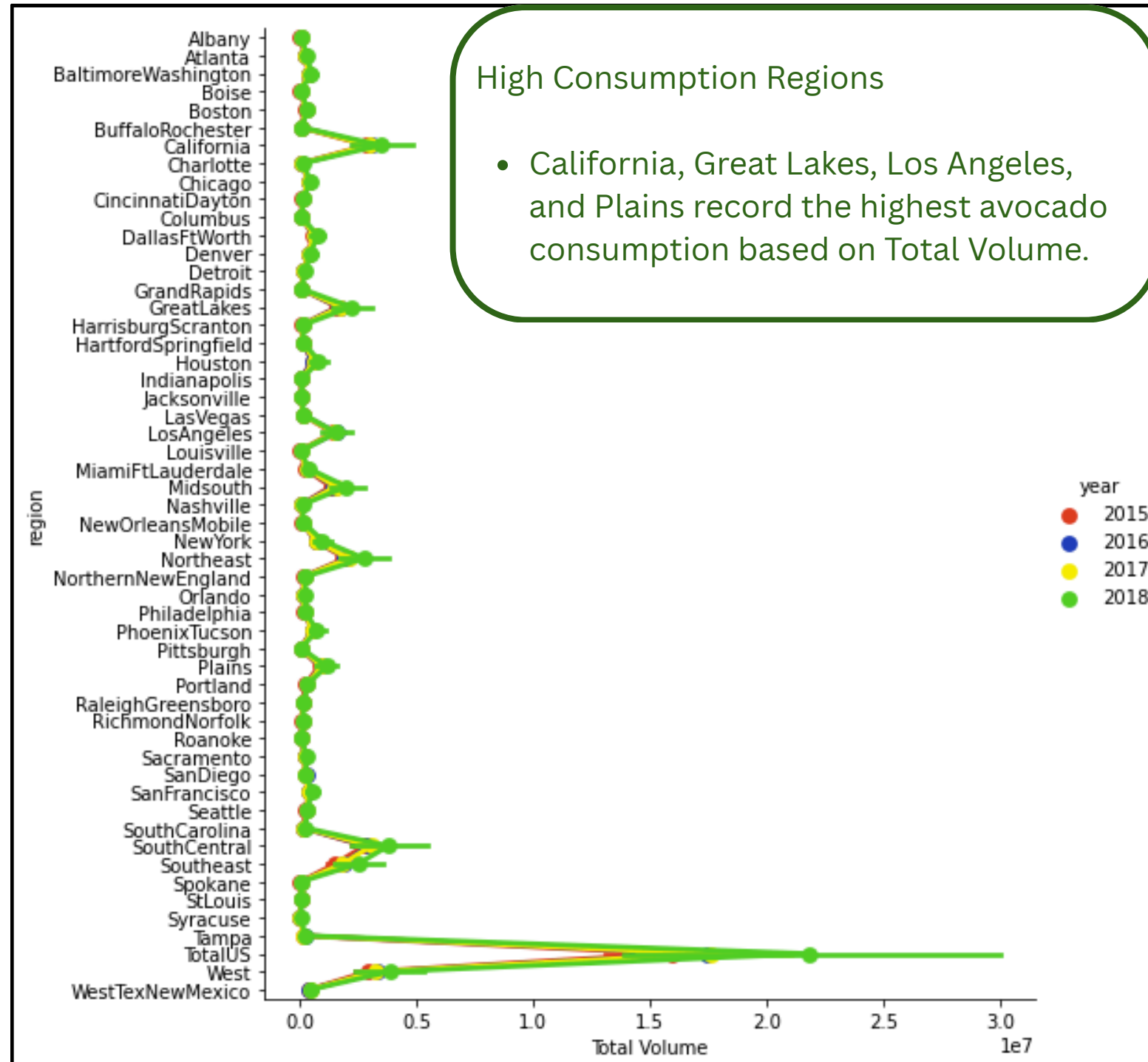
# Data Visualization



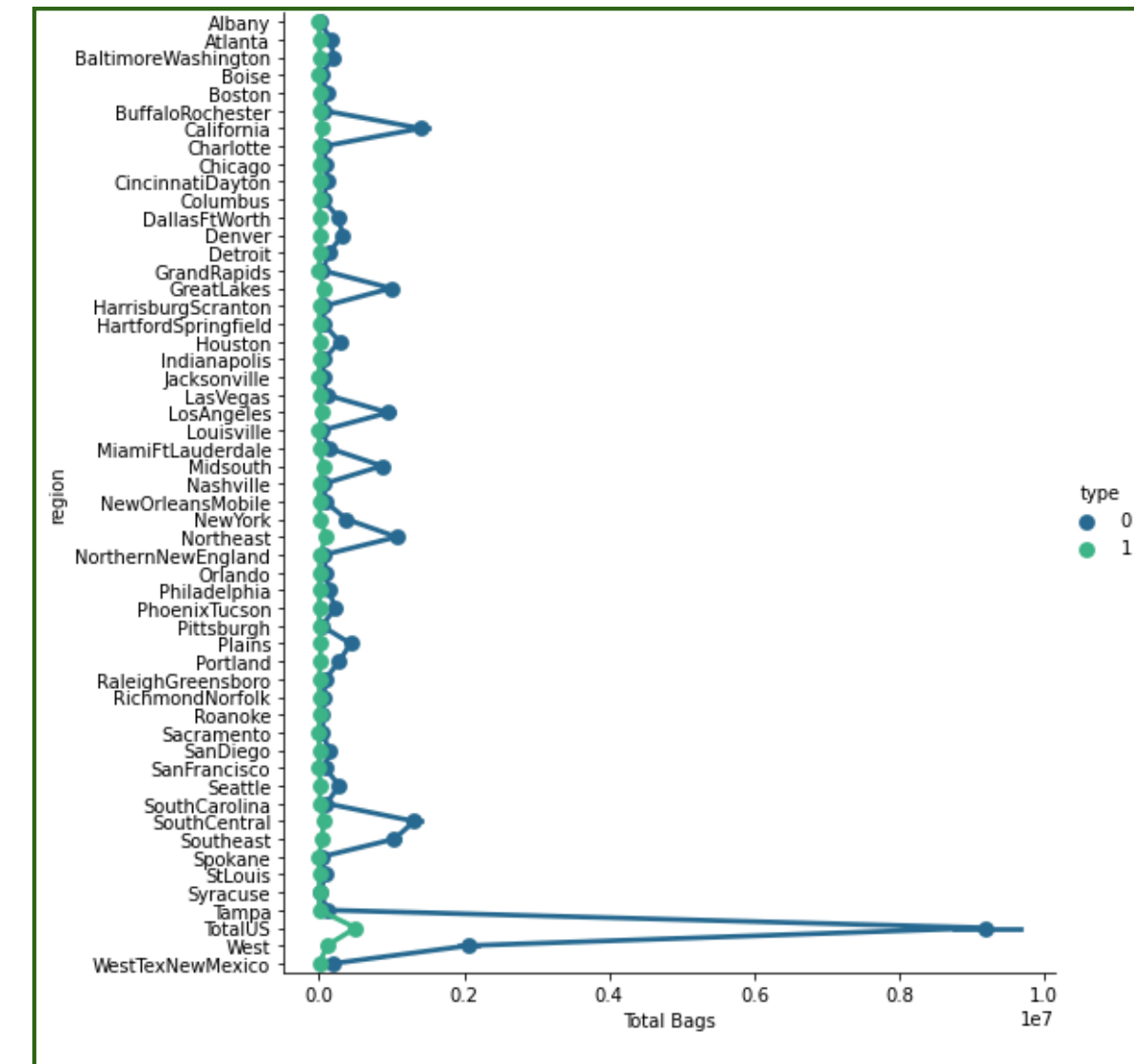
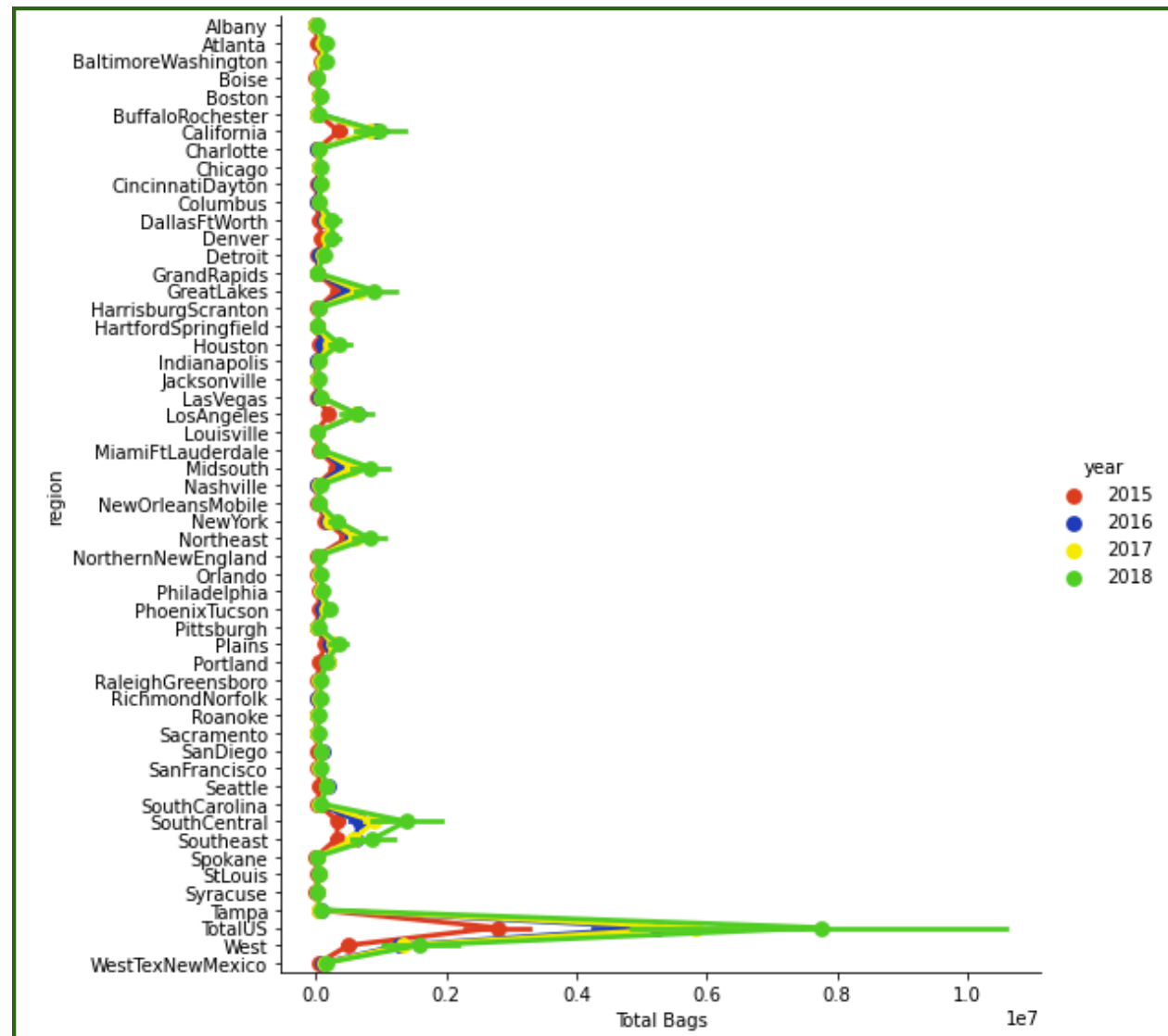
## Average Price Trend Over Time

- A positive upward trend is observed in Average Price over time.
- Three repetitive peaks appear at consistent intervals.
- Average Price drops around December / January.
- Highest prices are recorded during the months September to November.

# Data Visualization



# Data Visualization



## Correlation Between Bags & Total Volume

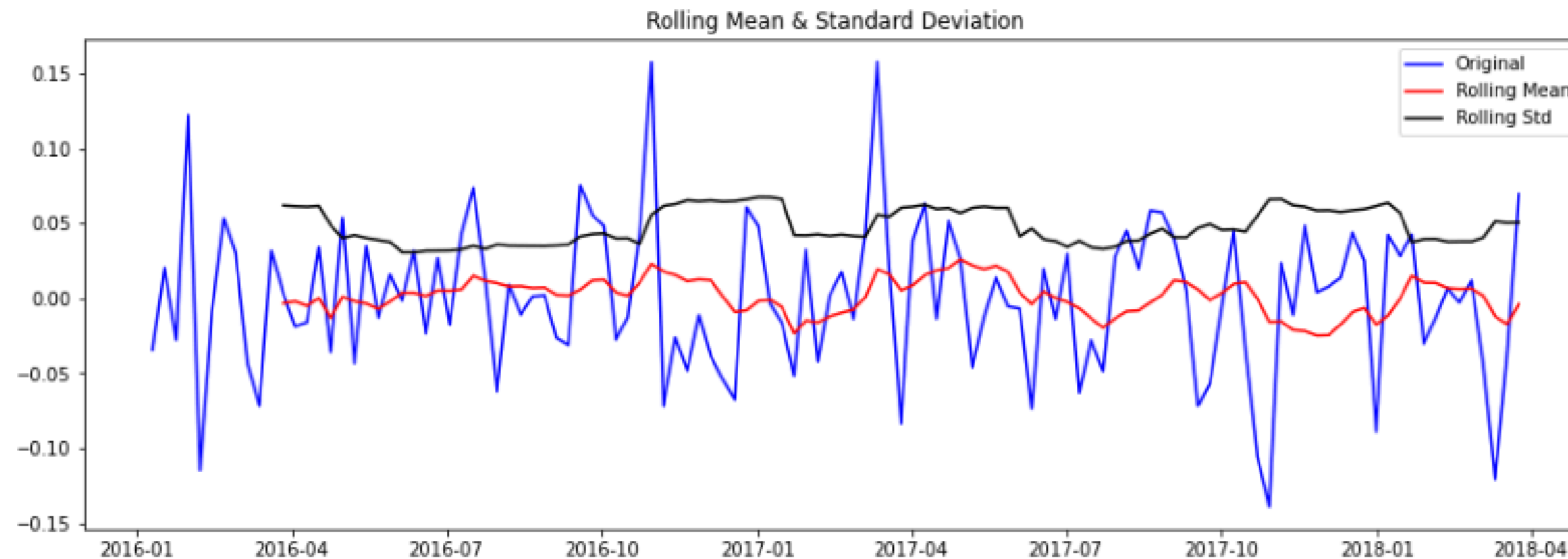
Bag usage shows a strong positive correlation with Total Volume, as evidenced by TotalUS data. There is a clear and consistent increase in bag usage in direct proportion to the rise in avocado demand, indicating a cause-and-effect relationship.



# Dickey-Fuller Test

To check whether the time series is stationary or not using the Augmented Dickey-Fuller (ADF) Test.

Objective



## Results of Dickey-Fuller Test:

Test Statistic	-4.60
p-value	0.00
#Lags Used	11.00
Number of Observations Used	104.00
Critical Value (1%)	-3.49
Critical Value (5%)	-2.89
Critical Value (10%)	-2.58
dtype:	float64

## Interpretation:

- Since  $p\text{-value} < 0.05$  and  $\text{Test Statistic} < \text{Critical Values} \rightarrow \text{Reject } H_0 \rightarrow \text{Data is stationary}$

## Conclusion:

- The dataset is stationary at a 99% confidence level, making it suitable for SARIMA modeling.

# SARIMA Overview: What & Why

## What is SARIMA?

SARIMA (Seasonal Auto-Regressive Integrated Moving Average) is an extension of ARIMA that captures seasonality in time series data.

Model order:

$(p, d, q) \times (P, D, Q, M)$

- $(p, d, q)$ : same as ARIMA parameters
- $(P, D, Q)$ : seasonal components
- $M$ : seasonal period (here 52 weeks  $\rightarrow$  1 year)

## Why SARIMA?

SARIMA is used to forecast weekly avocado prices by capturing both trend and seasonality patterns.



# SARIMA Model Configuration

## Model Setup

- **Model: SARIMAX(1, 1, 2) × (0, 1, 0, 52)**
- **Dependent Variable: Log\_AveragePrice**
- **Sample Period: Jan 2015 – Mar 2018**
- **Observations: 169**

## Model Fit Quality

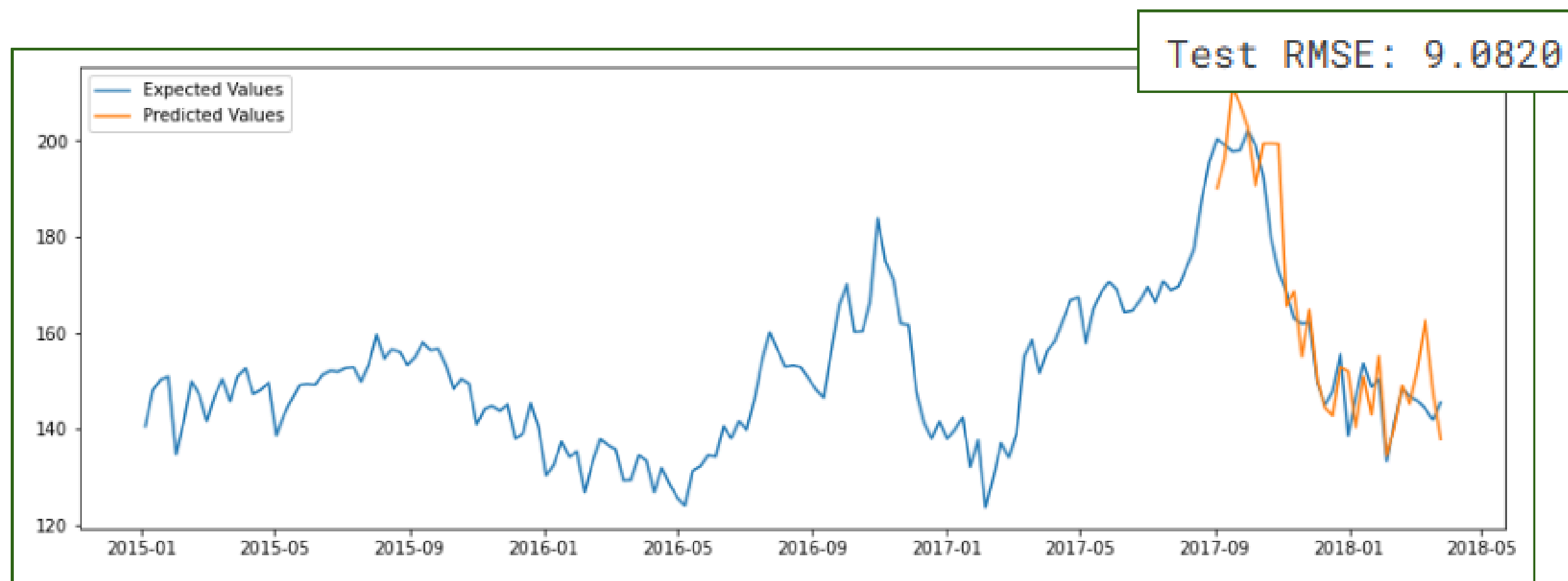
- AIC: -361.85 → Low → Good fit
- BIC: -350.83 → Supports model efficiency
- HQIC: -357.37

## Conclusion

- Model has a good overall fit based on AIC/BIC.
- Significant AR(1) and MA(1) terms indicate strong temporal dependence.
- Residuals show slight non-normality but acceptable for forecasting.
- SARIMA is suitable for weekly avocado price forecasting.

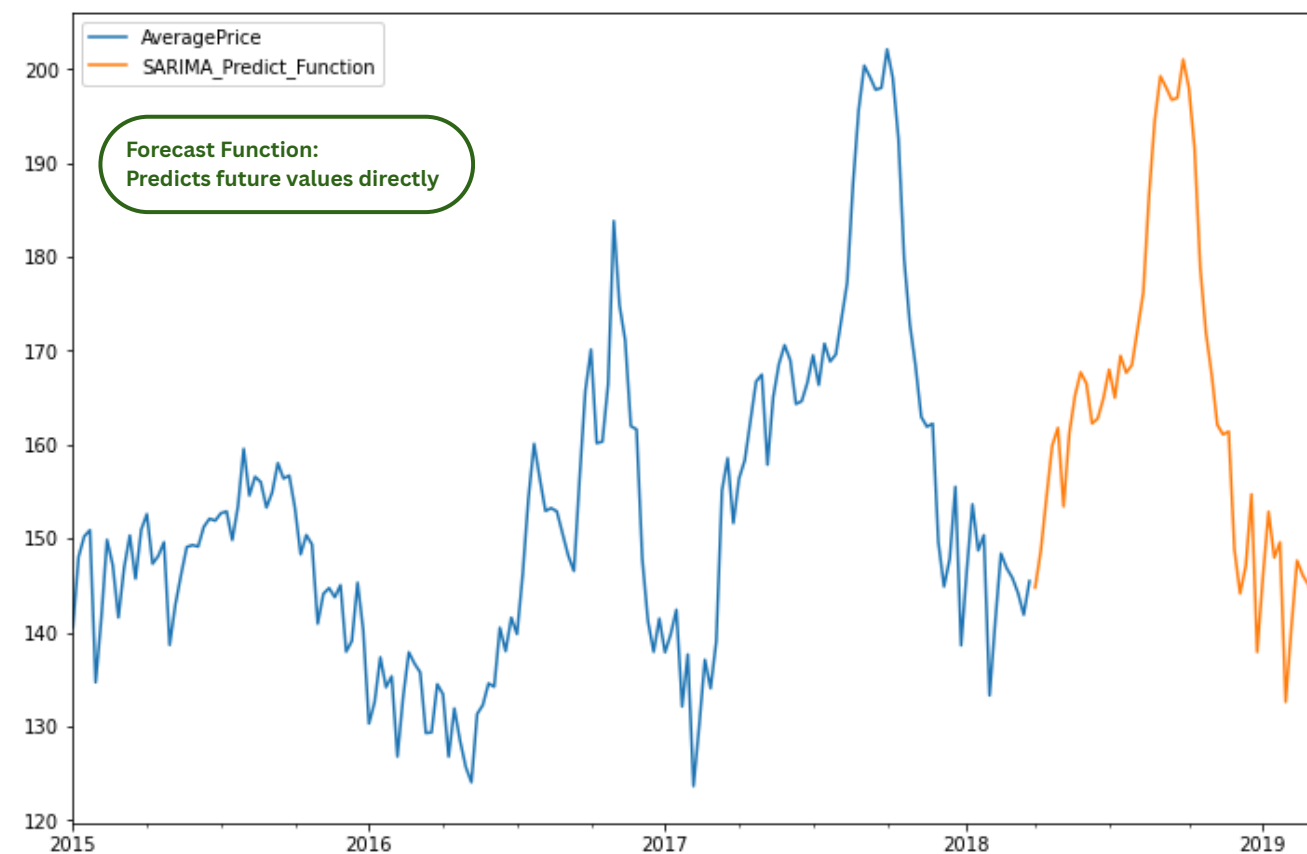
SARIMAX Results						
=====						
Dep. Variable:	Log_AveragePrice	No. Observations:	169			
Model:	SARIMAX(1, 1, 2)x(0, 1, [], 52)	Log Likelihood	184.923			
Date:	Mon, 12 Sep 2022	AIC	-361.845			
Time:	14:14:49	BIC	-350.831			
Sample:	01-04-2015	HQIC	-357.374			
	- 03-25-2018					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
ar.L1	0.7727	0.169	4.566	0.000	0.441	1.104
ma.L1	-0.7458	0.184	-4.047	0.000	-1.107	-0.385
ma.L2	-0.1546	0.109	-1.416	0.157	-0.369	0.059
sigma2	0.0024	0.000	9.163	0.000	0.002	0.003
=====						
Ljung-Box (Q):	77.68	Jarque-Bera (JB):	7.83			
Prob(Q):	0.00	Prob(JB):	0.02			
Heteroskedasticity (H):	1.16	Skew:	0.10			
Prob(H) (two-sided):	0.64	Kurtosis:	4.26			
=====						

# In-Sample Forecasting Results



The SARIMA model was applied using a rolling forecast on the last 30 data points to evaluate its predictive performance. The results show that the predicted values closely follow the actual values, indicating a good model fit. Although slight overfitting is observed in certain peaks and drops, the overall accuracy remains acceptable, with a test RMSE of 9.082. These findings confirm that the model performs well on historical data and is reliable enough to proceed with out-of-sample forecasting.

# Out-Sample Forecasting Results



The results obtained from both methods are almost identical, with only minor differences in decimal values. This indicates that the model produces consistent and reliable predictions.

Moreover, the model successfully captures the seasonal patterns present in the data and shows a slight upward trend in the projected period. These findings confirm that the SARIMA model is highly suitable for generating short- to medium-term forecasts of avocado prices, providing a solid foundation for future decision-making.



# Key Insights & Conclusion

## SARIMA Forecasting

### Key Insight

- SARIMA model effectively models weekly avocado prices.
- Seasonality is well captured after applying seasonal differencing.
- High accuracy demonstrated by low RMSE and consistent forecasts.
- Forecast Function and Predict Function produce almost identical results.
- Suitable for short- to medium-term forecasting.

### Conclusion

- The dataset is well-suited for learning time series analysis and exploring various forecasting techniques.
- Through this project, we gained insights into stationarity, data transformations, and seasonal modeling.
- The SARIMA model provides reliable projections of future avocado prices, making it valuable for business decision-making.

thank you

---

---

---

---