Avocado: what is expected the price and consumption of avocado in 2022?

Mireille P. Feudjio T.
Springboard School of Data

Outline

- Context and Problem statement
- Exploratory data analysis
- Forecasting methodology
- Results
- Conclusion

Context and Problem statement

- Avocado for diet, health, nutrition, and food
- Avocado is rich in B vitamins, vitamin E, vitamin K , folate
- U.S. consumption of avocados has generally trended upward since 1970
- 2.21 pounds per capita in 2000 to 7.81 pounds per capita in 2019 (Statista, 2020)

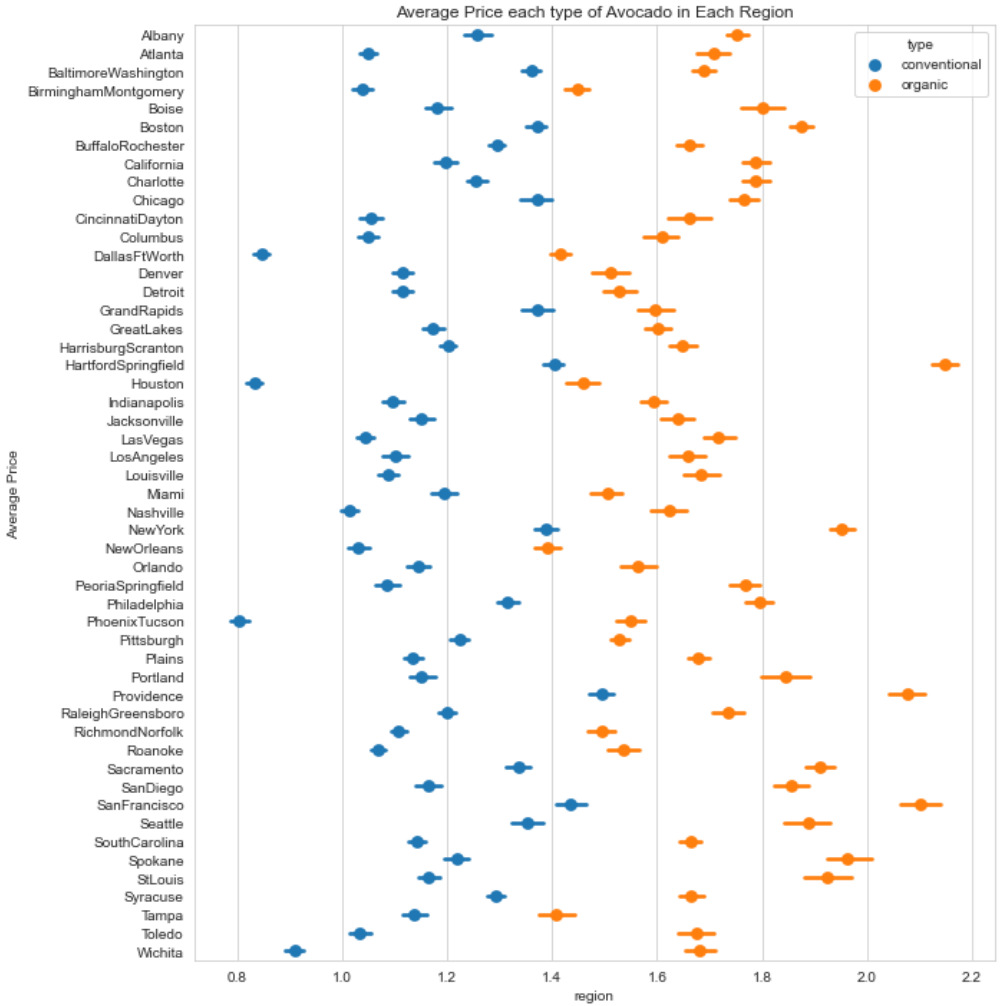
- United States produced 206,610 tons (NASS, 2020)

- Company Alpha based in the region of New York, wants to build a partnership with Avocado producers (farmers)

- Within this context, what are the expected avocado stock prices and quantities to be sold next year that can allow the Company Alpha base in New York to sign a contract with new producers and increase the price of avocado by 2%?

Exploratory data analysis (1)

Price behavior of avocados in all regions



Data wrangling

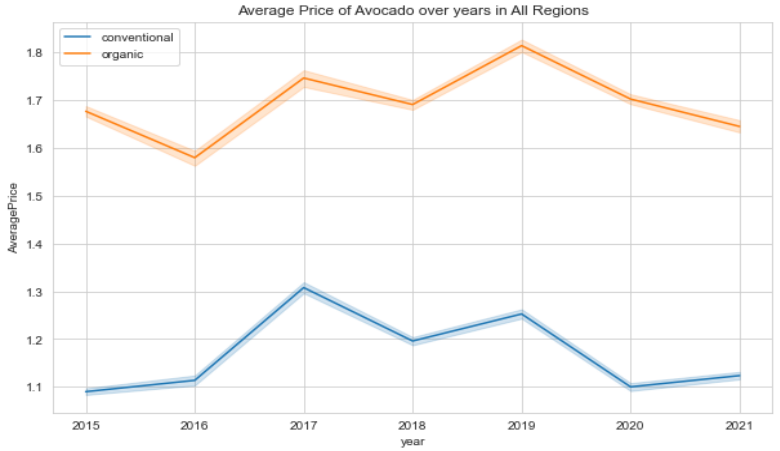
- renaming of incorrect string
- indexing date and convert to datetime,
- checking for duplicate, missing value and data type

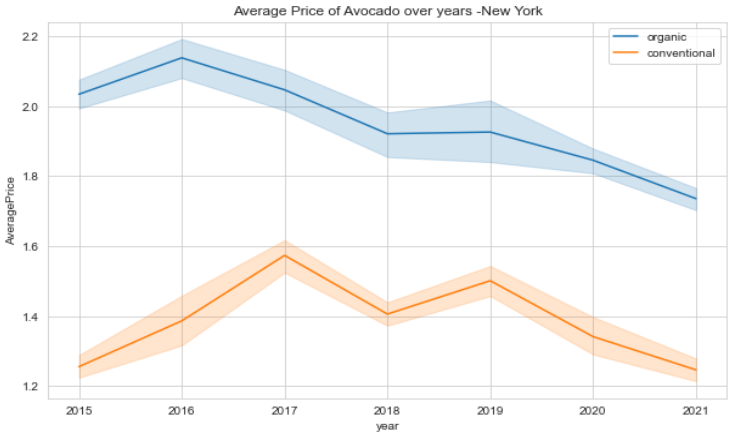
Dataset shape was 41045 rows and 13 columns.
There is no missing data, and no duplicate

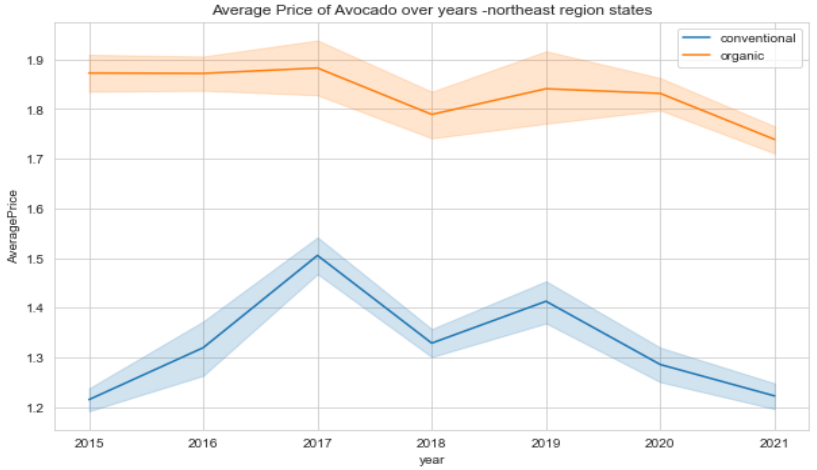
Exploratory data analysis (2)

Price trend of avocados since 2015

Top-down approach to analyzed and visualized the avocado price and consumption behave at different scales

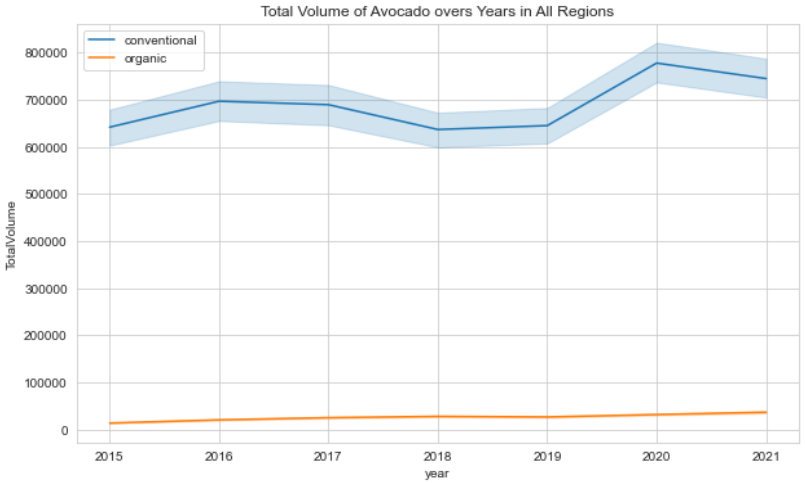


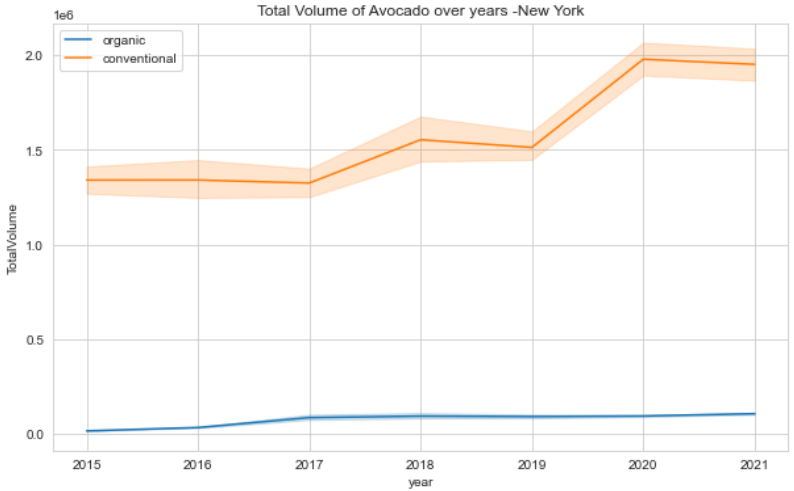


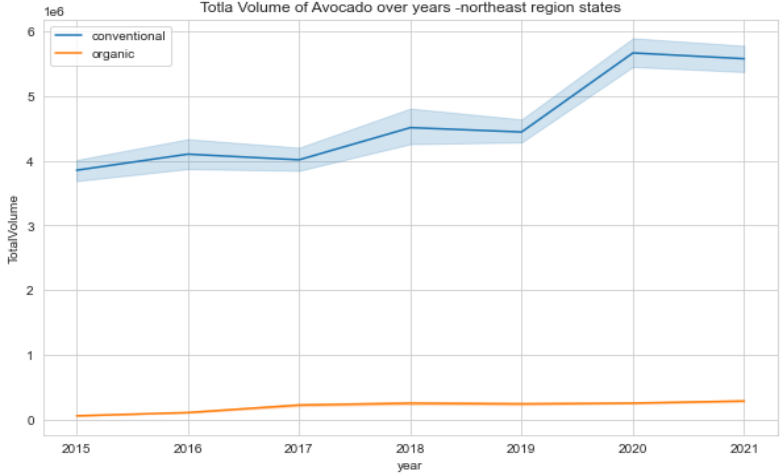


Exploratory data analysis (3)

Trend of avocados sold since 2015

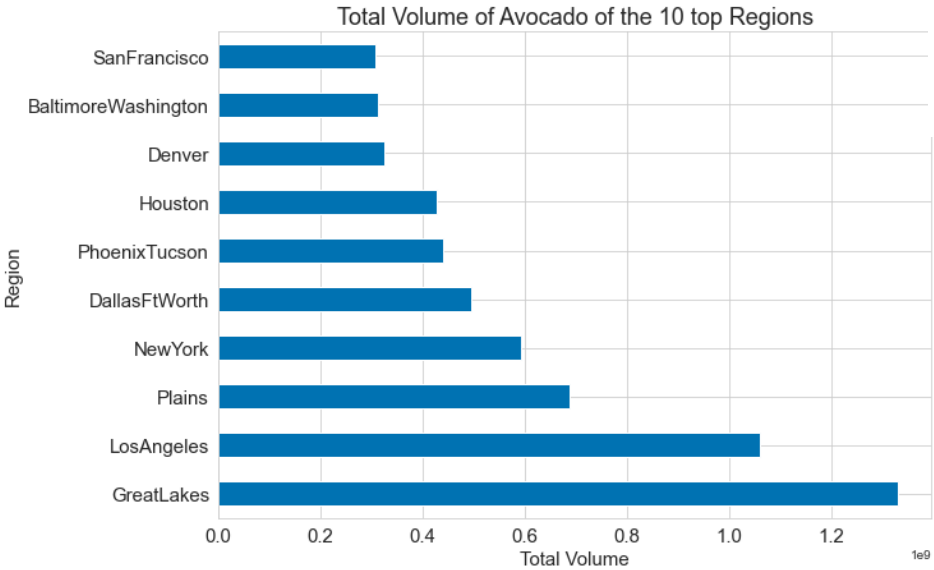


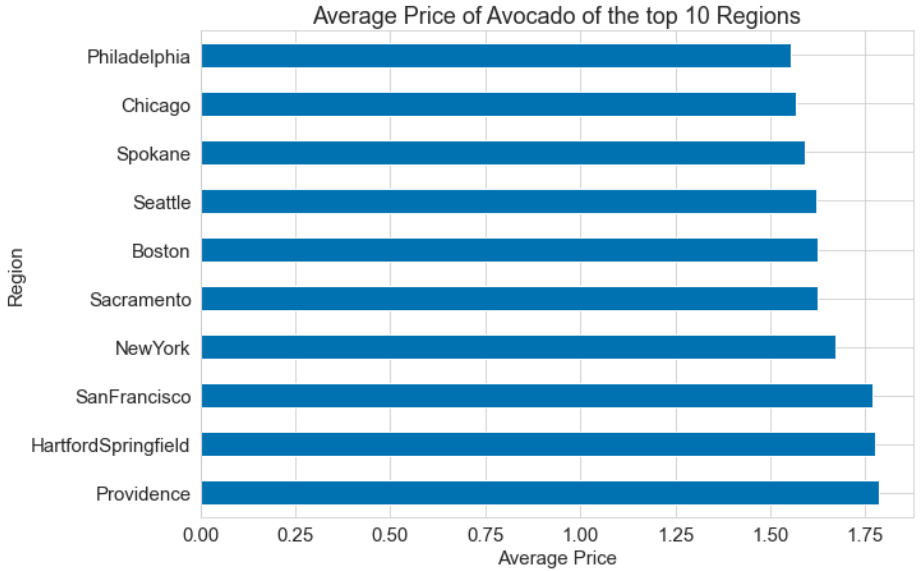




Exploratory data analysis (4)

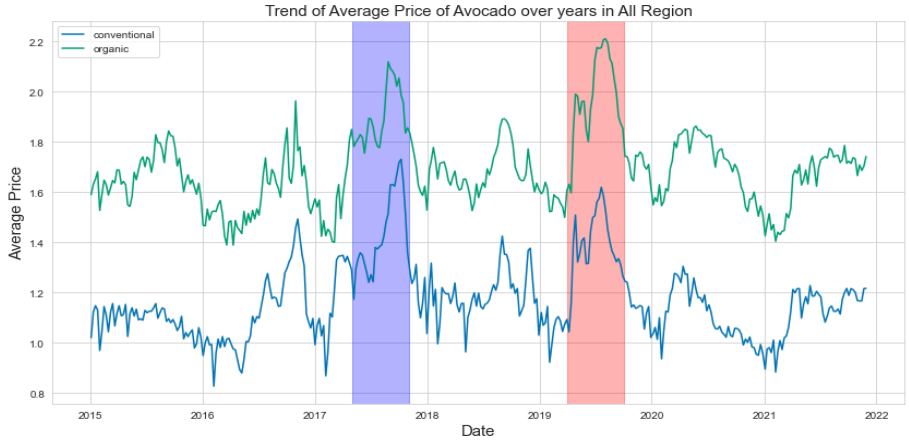
Top 10 regions with high price and high quantities sold of avocados (conventional and organic) since 2015?

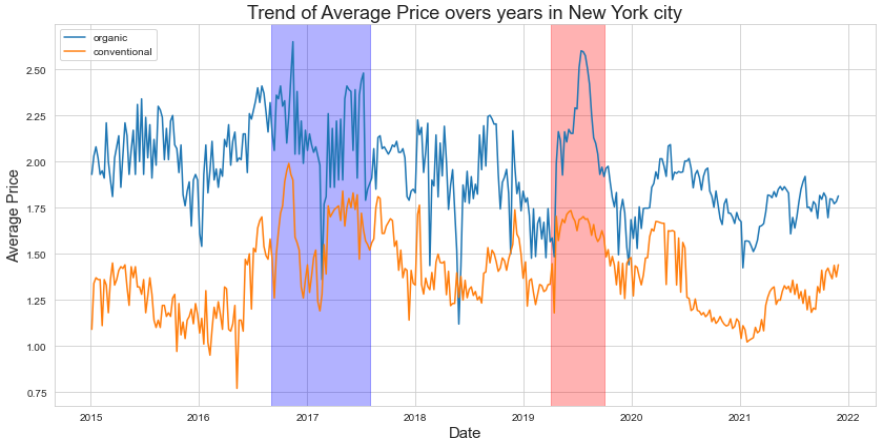


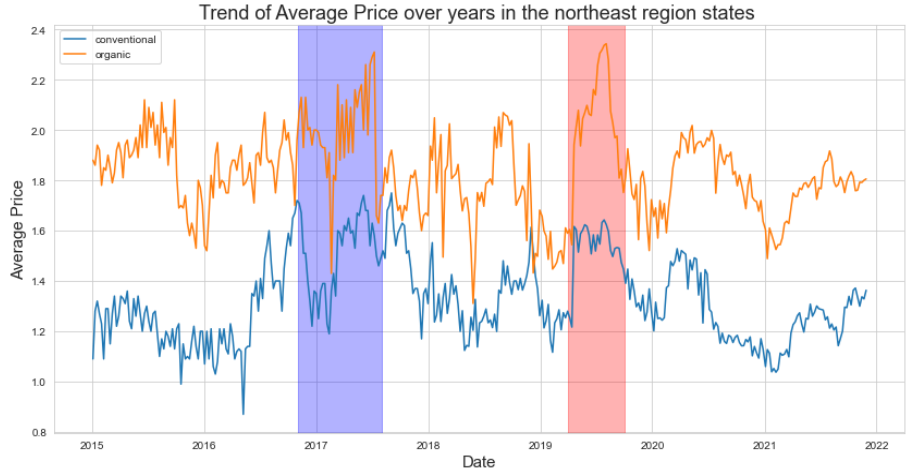


Exploratory data analysis (4)

Weekly trend of avocado prices between years

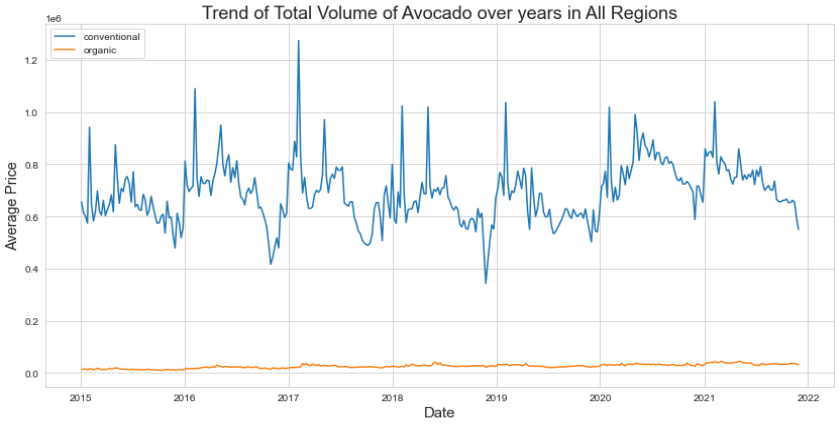


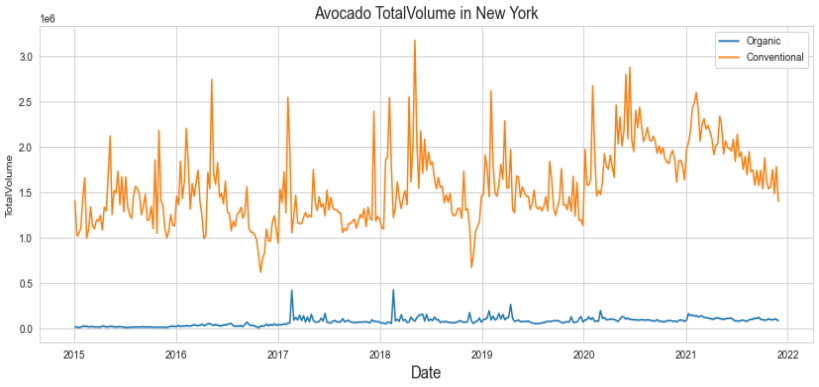


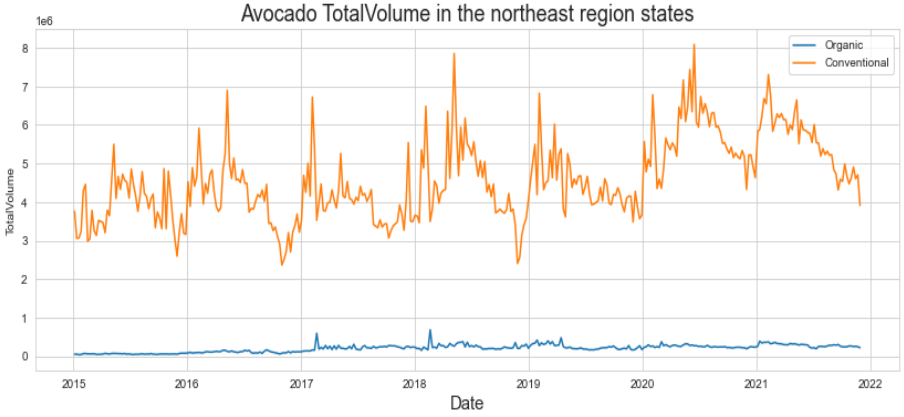


Exploratory data analysis (4)

weekly trend of avocado sold (volume) between years

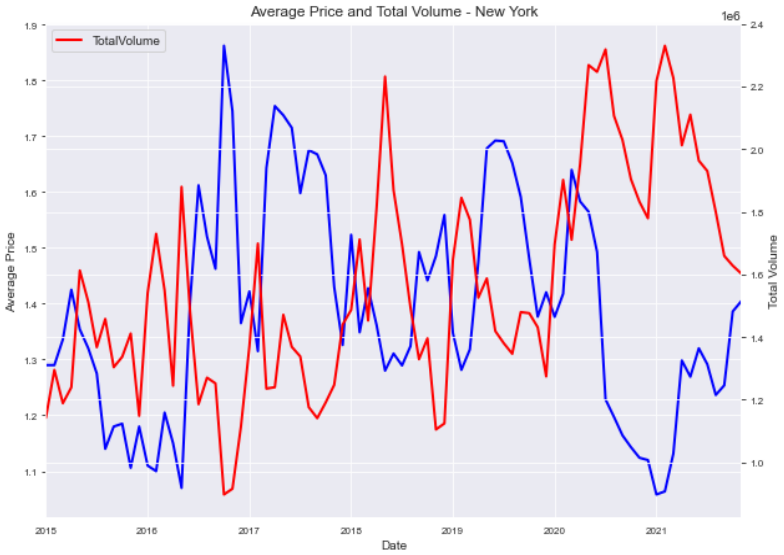


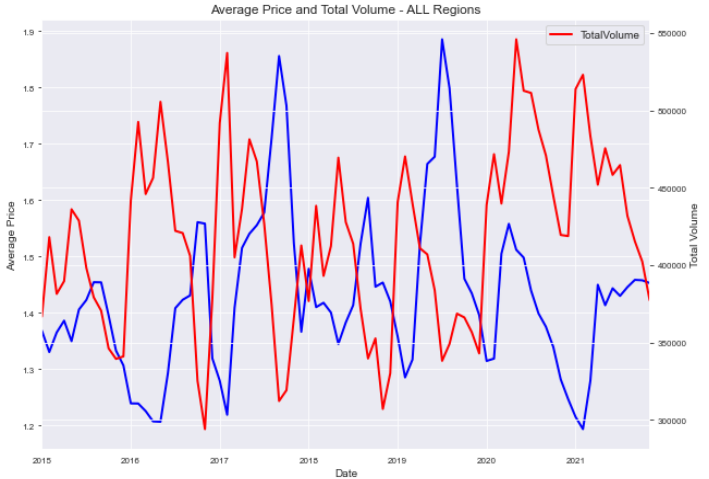


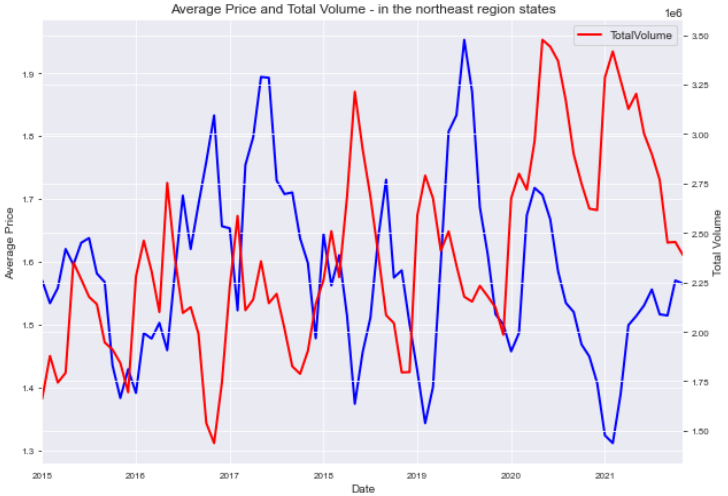


Exploratory data analysis (5)

Trend of prices and volume trend of avocado over years







Exploratory data analysis (6)

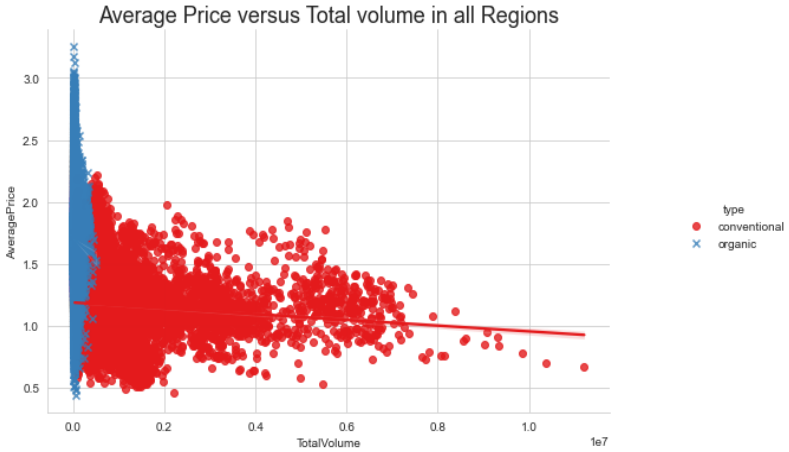
Type of avocado (organic and conventional) impact the price and the consumption

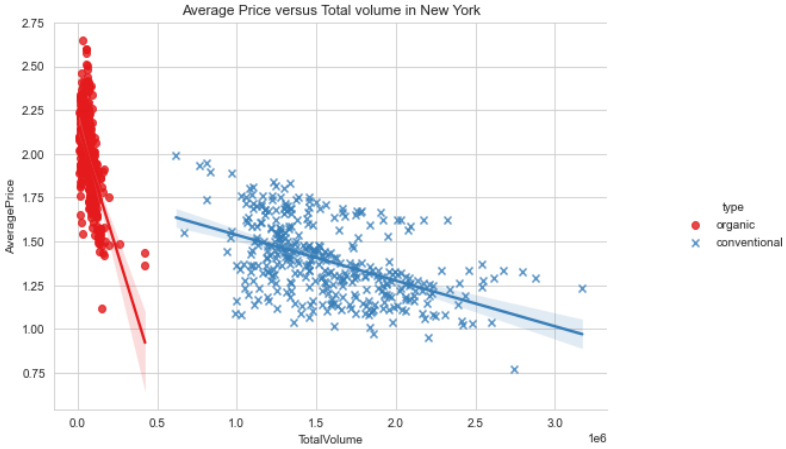
Table. Summary statistic of Avocado prices and volume in New York

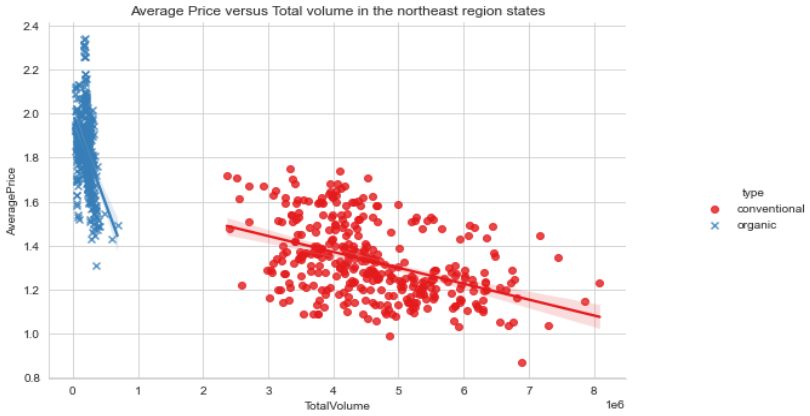
	conventional Avocado		Organic Avocado	
	Average Price	Total Volume	Average Price	Total Volume
count	361.000	361.000	361.000	361.000
mean	1.389	1566999.169	1.952	75484.313
std	0.213	414175.263	0.242	47249.091
min	0.770	618279.770	1.119	8442.790
25%	1.230	1255552.680	1.798	40280.960
50%	1.360	1481997.420	1.931	75479.880
75%	1.540	1838475.070	2.110	97048.630
max	1.990	3172572.870	2.650	424186.840

Exploratory data analysis (7)

Relation between the average price, total volume, and avocados type





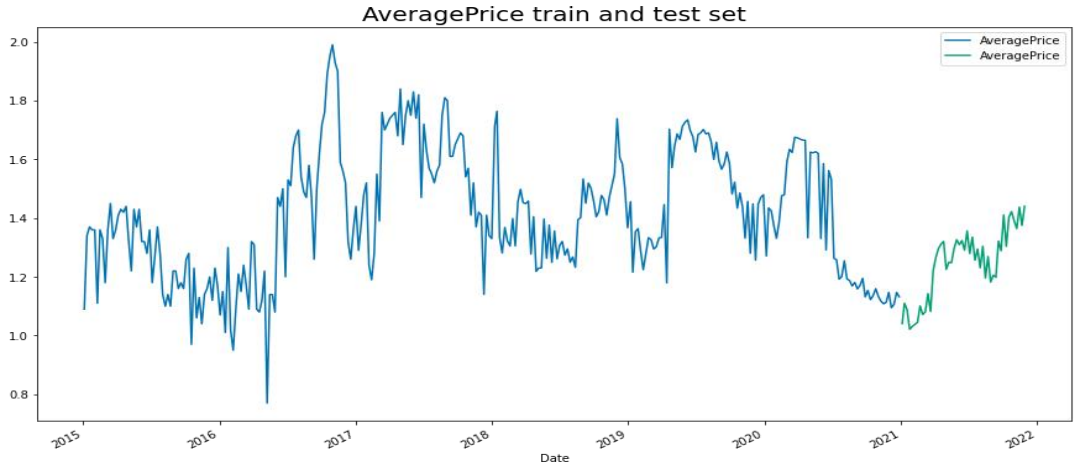


Forecasting methodology (1)

1. Check the stationarity of the time series and define the order of the differencing (d)

Augmented Dickey-Fuller Test (ADF test) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests.

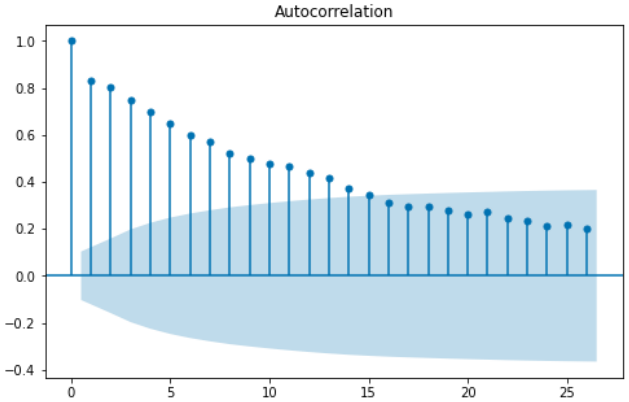
2. Get the train and test set

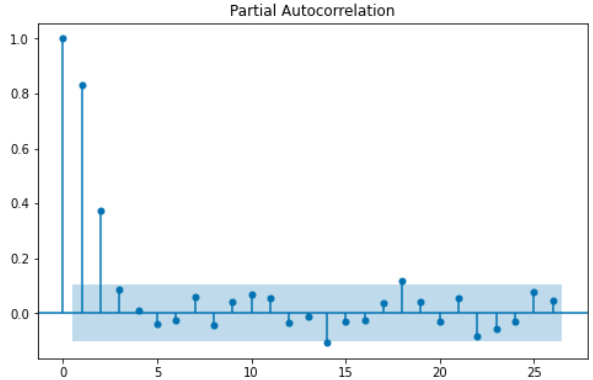


Forecasting methodology (2)

3. identification of AR and MA orders

AR term (p) with partial autocorrelation function(pacf)
MA term (q) with autocorrelation function (acf)





From the plot of ACF and PACF above, it suggested that model is a AR(2) , with p =2 and q = 0

Evaluate different p, d, and q values - choice of best model according to AIC and BIC

	p	q	d	AIC	BIC
1st	1	1	0	-438.909479	-427.670869
2nd	3	1	0	-438.894161	-420.163145

Explored the seasonal with auto_arima function to identify the P,Q,D

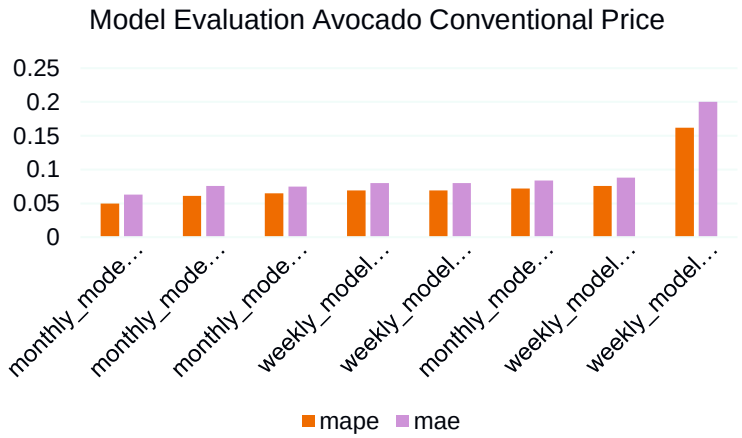
Results

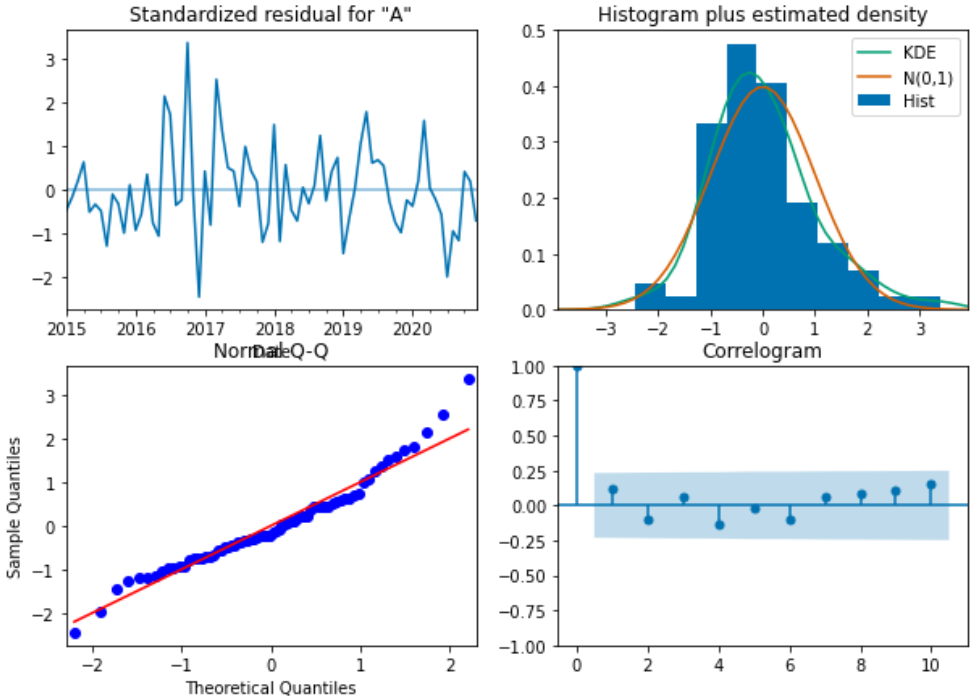
Table. New York conventional avocado price models descriptions

Model name	p,d,q non seasonal part	P,Q,D seasonal part	Used of exogenous variable or not	Name of results metric
results_1	3,0,1	-		monthly_model_sarimax
results_2	1,0,1	-		weekly_model_sarimax
results_3	1,0,1	-		weekly_model_arima
results_4	8,0,1	-		monthly_model_arima
results_s	3,0,0	0,0,0,52		weekly_model_saisonal
results_s2	1,0,0	0,0,1,52		monthly_model_saisonal
results_ex	1,0,1	0,0,1,52	with volume as exogenous	weekly_model_exog
results_ex2	1,0,1	0,0,0,52	with volume as exogenous	monthly_model_exog

Results (1)

case: conventional avocado price

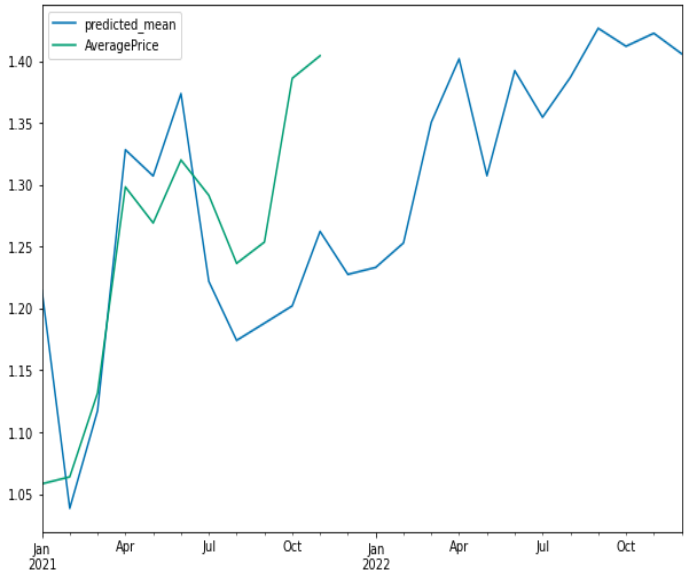


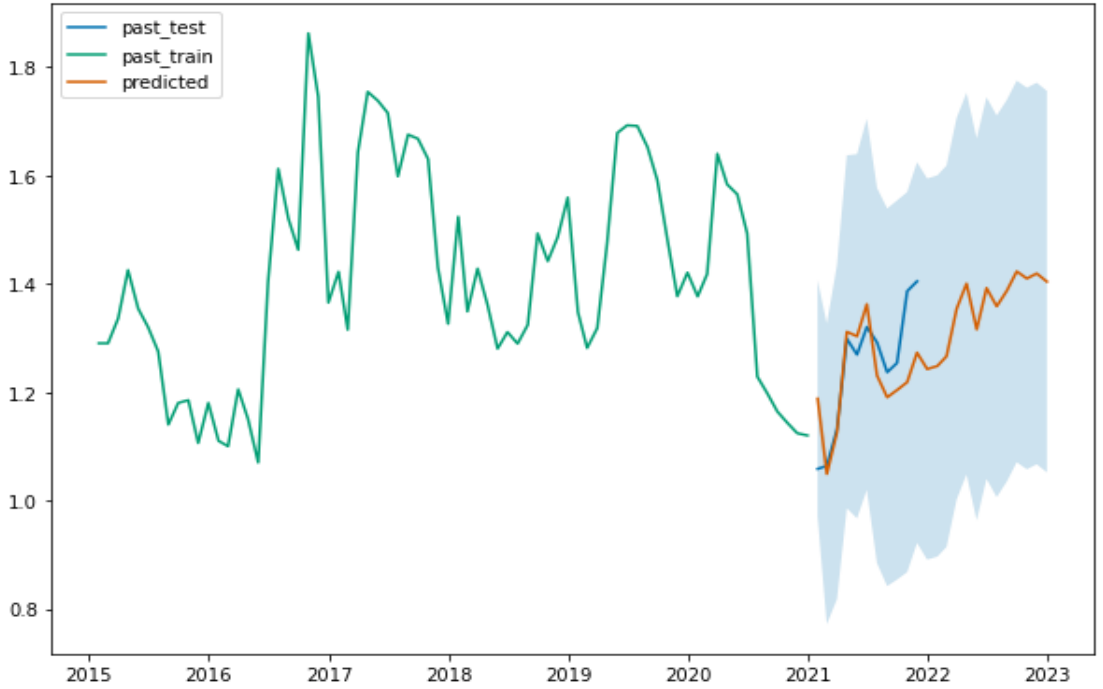


Model specification (1,0,0,0,0,1),52

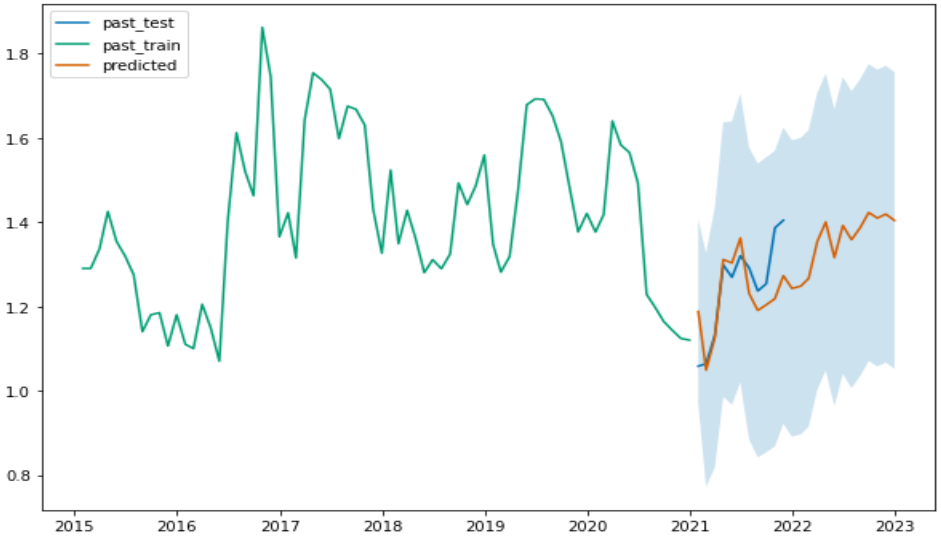
Results (2)

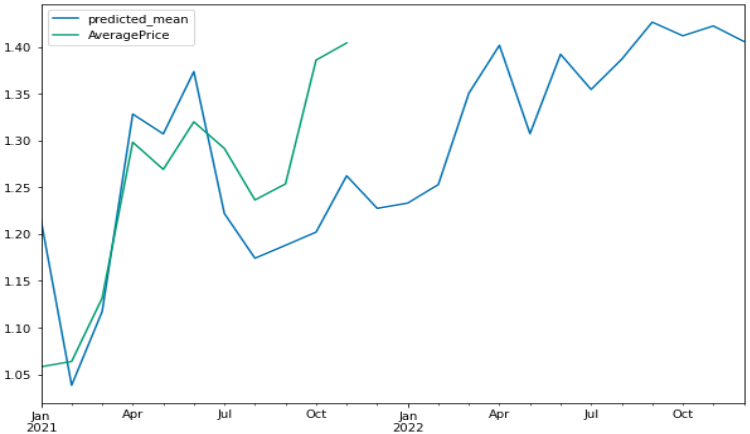
case: conventional avocado price





case: conventional avocado price





	lower AveragePrice	upper AveragePrice	pred_mean_AveragePrice	AveragePrice	Diff (pred – actual)
2021-01-31 00:00:00	0.970	1.406	1.188	1.058	0.130
2021-02-28 00:00:00	0.772	1.326	1.049	1.064	-0.015
2021-03-31 00:00:00	0.818	1.434	1.126	1.132	-0.006
2021-04-30 00:00:00	0.986	1.637	1.311	1.298	0.013
2021-05-31 00:00:00	0.967	1.639	1.303	1.269	0.034
2021-06-30 00:00:00	1.020	1.704	1.362	1.320	0.042
2021-07-31 00:00:00	0.885	1.577	1.231	1.292	-0.061
2021-08-31 00:00:00	0.842	1.539	1.190	1.237	-0.046
2021-09-30 00:00:00	0.854	1.553	1.204	1.254	-0.050
2021-10-31 00:00:00	0.868	1.569	1.218	1.386	-0.168
2021-11-30 00:00:00	0.922	1.624	1.273	1.404	-0.132
2021-12-31 00:00:00	0.891	1.594	1.242		
2022-01-31 00:00:00	0.896	1.600	1.248		
2022-02-28 00:00:00	0.914	1.618	1.266		
2022-03-31 00:00:00	1.002	1.706	1.354		
2022-04-30 00:00:00	1.048	1.752	1.400		
2022-05-31 00:00:00	0.964	1.668	1.316		
2022-06-30 00:00:00	1.040	1.744	1.392		
2022-07-31 00:00:00	1.006	1.710	1.358		
2022-08-31 00:00:00	1.035	1.739	1.387		
2022-09-30 00:00:00	1.071	1.775	1.423		
2022-10-31 00:00:00	1.058	1.761	1.409		
2022-11-30 00:00:00	1.067	1.771	1.419		
2022-12-31 00:00:00	1.052	1.756	1.404		

Conclusion

- Analysis of the commercialization of avocado showed a clear distinction between organic and conventional avocados,
- the price of the organic avocado is higher than that of the conventional avocado,
- the total volume of organic avocados sold is less than that of conventional avocados.
- New York is among the top 5 regions with high price (and high volume) in organic and conventional avocado (occupied 4th or the 5^{th} rang).
- when total volume increases, price decreases and vice versa.

- 8 models have been developed for each type of avocado (conventional price, conventional volume, organic price, organic volume).
- Our models did good. The predicted prices of avocados (conventional and organic) are close to real values. the average of the difference between this value (pred – actual) is -0,024, the MAE is 0.063 and the MAPE is 0,050. More, the Pearson correlation is 0,712.

Conclusion

Before covid 19, the price and volume of avocado was constantly increasing. the marketing of avocado was affected between 2019 and 2021. With time series analysis, this had a considerable impact on forecasting the future value of 2022. If considering that everything is back to normal (less impact of covid 19), we can suggest to the company to revise the prices of avocado considering the local trend, and the values of the upper limit of the price and volume.

How can we improve the current model? maybe add exogenous variables indicating if it is a covid period or not, indicating the population, the percentage of people affected by covid for each observation.

THANK You