

Data Analysis Project

AVOCADO PRICES:
HISTORICAL DATA ON
AVOCADO PRICES AND SALES
IN US MARKETS.

- BY DIVYANSH CHOUBISA

ABOUT DATA

THIS DATA WAS DOWNLOADED FROM KAGGLE

<https://www.kaggle.com/neuromusic/avocado-prices>

The data 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. 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.

Some relevant columns in the dataset:

- Date - The date of the observation
- AveragePrice - the average price of a single avocado
- type - conventional or organic
- year - the year
- Region - the city or region of the observation
- 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
- Small Bags
- Large Bags
- XLarge Bags

SUMMARY

In this data analysis project, I have analysed the data of AVOCADO PRICES (Historical Data on Avocado Prices and Sales in US Markets)

- I have made different graphs on the sales of two different kinds of avocado (conventional and organic) in different regions of United States where the mean average price of an avocado was high.
- I have analysed different patterns in sales of avocados in different regions of United States and concluded useful results.
- I have also done timeseries forecasting for the average prices of avocados.
- I have calculated the correlation between different attributes, performed statistical analysis and calculations.
- With different graphical visualization I have also created the maps showing the locations of the regions of United States on which I have performed data analysis.

Introduction

I have used Python programming language as my tool for my data analysis work. I have used different python libraries for this data analysis project. These libraries are:

1- Pandas

2- Matplotlib

3- numpy

4- Seaborn

5- Folium

6- fbprophet

The data from avocado contains 13 columns and I have attached photo of the dataset (after renaming of some columns so that they sound meaningful) and it exactly looks like the photo below -

| | Date | AveragePrice | Total Volume Sold | PLU_4046 | PLU_4225 | PLU_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 | 0.0 | conventional | 2015 | Albany |
| 1 | 2015-12-20 | 1.35 | 54876.98 | 674.28 | 44638.81 | 58.33 | 9505.56 | 9408.07 | 97.49 | 0.0 | conventional | 2015 | Albany |
| 2 | 2015-12-13 | 0.93 | 118220.22 | 794.70 | 109149.67 | 130.50 | 8145.35 | 8042.21 | 103.14 | 0.0 | conventional | 2015 | Albany |
| 3 | 2015-12-06 | 1.08 | 78992.15 | 1132.00 | 71976.41 | 72.58 | 5811.16 | 5677.40 | 133.76 | 0.0 | conventional | 2015 | Albany |
| 4 | 2015-11-29 | 1.28 | 51039.60 | 941.48 | 43838.39 | 75.78 | 6183.95 | 5986.26 | 197.69 | 0.0 | conventional | 2015 | Albany |

This dataset has 18249 rows and 13 columns. So, this dataset is significantly large.

For my data analysis and data manipulation I have mainly used the following columns-

1- Date

2- AveragePrice

3- Total Volume Sold

4- type

5- region

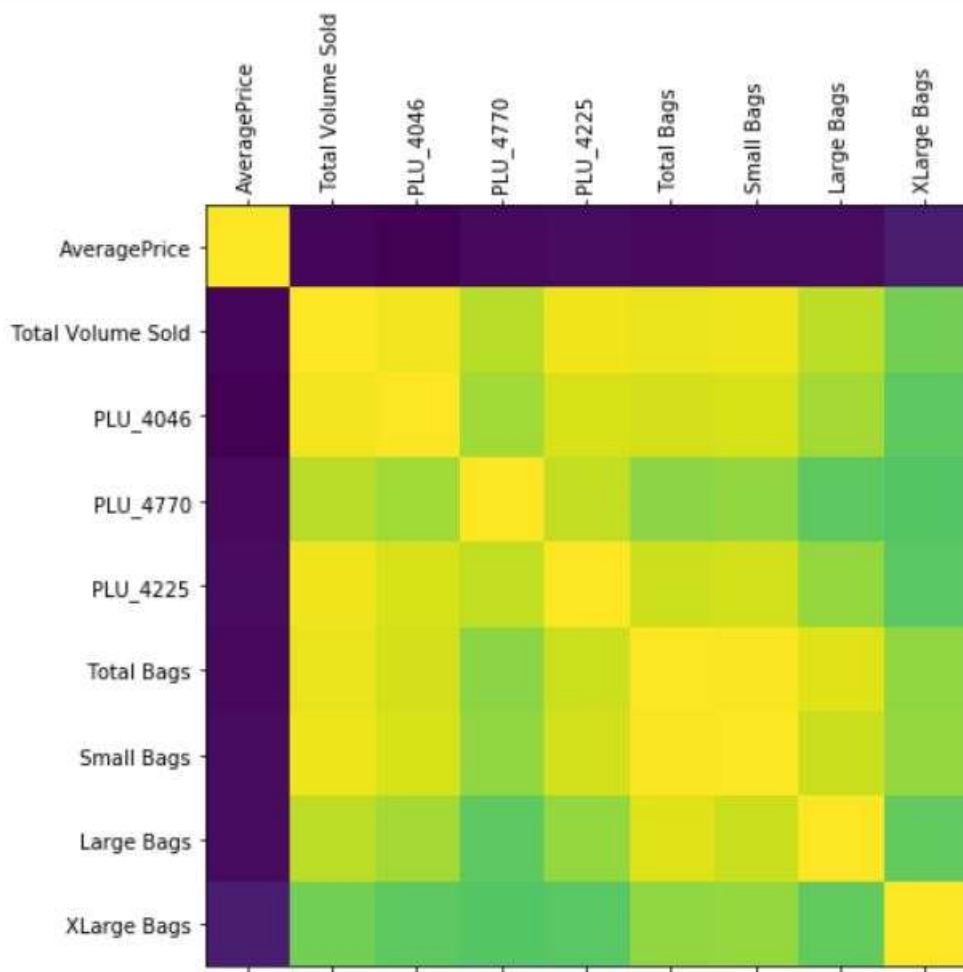
DATA ANALYSIS

To start with data analysis my very first step was to get some measure of relationship between the attribute of my data. So, I have calculated the correlation between the variables. Correlation is measure of the extent of independency between variables. This calculation of correlation between the variables has helped me to do my further analysis and also has given me some clues on to which attributes to focus on. I have created the table of pearson correlation coefficient between the variables. To understand this correlation data, I also have created the matshow map.

These values of correlation coefficients are calculated for whole dataset. The resulting coefficient is a value between -1 and 1 inclusive, where:

- **1:** Total positive linear correlation.
- **0:** No linear correlation, the two variables most likely do not affect each other.
- **-1:** Total negative linear correlation.

| | AveragePrice | Total Volume Sold | PLU_4046 | PLU_4770 | PLU_4225 | Total Bags | Small Bags | Large Bags | XLarge Bags |
|-------------------|--------------|-------------------|-----------|-----------|-----------|------------|------------|------------|-------------|
| AveragePrice | 1.000000 | -0.192752 | -0.208317 | -0.179446 | -0.172928 | -0.177088 | -0.174730 | -0.172940 | -0.117592 |
| Total Volume Sold | -0.192752 | 1.000000 | 0.977863 | 0.872202 | 0.974181 | 0.963047 | 0.967238 | 0.880640 | 0.747157 |
| PLU_4046 | -0.208317 | 0.977863 | 1.000000 | 0.833389 | 0.926110 | 0.920057 | 0.925280 | 0.838645 | 0.699377 |
| PLU_4770 | -0.179446 | 0.872202 | 0.833389 | 1.000000 | 0.887855 | 0.792314 | 0.802733 | 0.698471 | 0.679861 |
| PLU_4225 | -0.172928 | 0.974181 | 0.926110 | 0.887855 | 1.000000 | 0.905787 | 0.916031 | 0.810015 | 0.688809 |
| Total Bags | -0.177088 | 0.963047 | 0.920057 | 0.792314 | 0.905787 | 1.000000 | 0.994335 | 0.943009 | 0.804233 |
| Small Bags | -0.174730 | 0.967238 | 0.925280 | 0.802733 | 0.916031 | 0.994335 | 1.000000 | 0.902589 | 0.806845 |
| Large Bags | -0.172940 | 0.880640 | 0.838645 | 0.698471 | 0.810015 | 0.943009 | 0.902589 | 1.000000 | 0.710858 |
| XLarge Bags | -0.117592 | 0.747157 | 0.699377 | 0.679861 | 0.688809 | 0.804233 | 0.806845 | 0.710858 | 1.000000 |



I was also interested in knowing the count of conventional avocados and organic avocados in our main dataset and I found that the dataset contains 9126 entries of conventional avocados and 9123 entries of organic avocados. This means that throughout the United states conventional avocados are sold more than organic avocados by a count of 3 over entire timeline.

| Count | |
|--------------|------|
| Type | |
| conventional | 9126 |
| organic | 9123 |

But also, it doesn't necessarily mean that the sales of total volume of conventional avocados are more than organic avocados. Because for every different date there is difference in sales of Total Volume of Avocado. To this point we can't actually say that throughout the United states which type of avocados were sold highest in numbers of Total Volume over entire timeline.

So as my next goal was to compare the average price of conventional avocados and organic avocados in different regions of United States but the problem here was that our data was not grouped and for every different date, we have different average price of avocados. To solve this problem, I created a data frame which was grouped by regions and type of avocados and for every region for the entire timeline, I added the mean average price of the avocados for example for Albany region for every different date we have different average price of an avocado, so over the all dates of Albany region I have calculated mean average price of avocado. The data frame looks like below one.

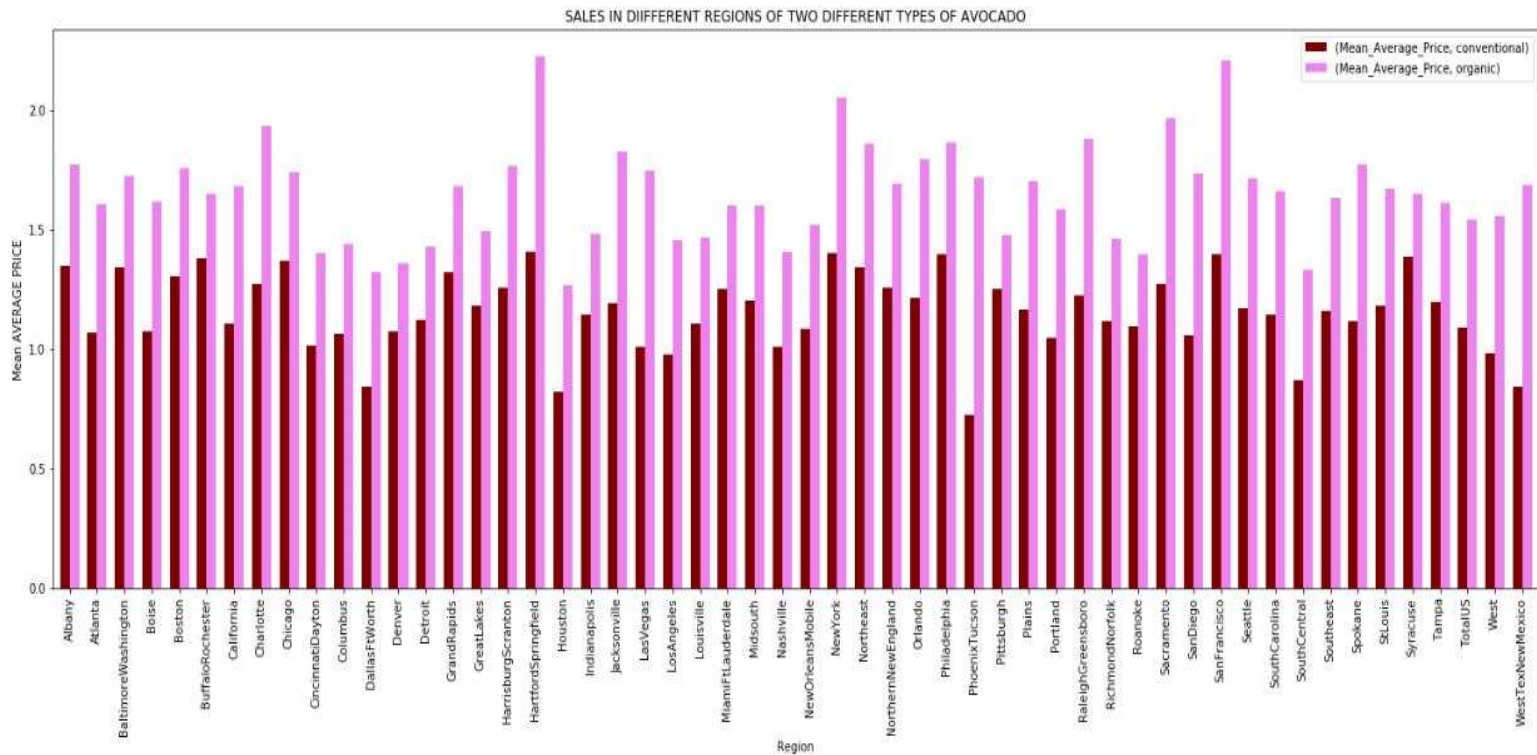
| | region | type | Mean_Average_Price |
|---|---------------------|--------------|--------------------|
| 0 | Albany | conventional | 1.348757 |
| 1 | Albany | organic | 1.773314 |
| 2 | Atlanta | conventional | 1.068817 |
| 3 | Atlanta | organic | 1.607101 |
| 4 | BaltimoreWashington | conventional | 1.344201 |
| 5 | BaltimoreWashington | organic | 1.724260 |
| 6 | Boise | conventional | 1.076036 |
| 7 | Boise | organic | 1.620237 |
| 8 | Boston | conventional | 1.304379 |
| 9 | Boston | organic | 1.757396 |

Then I further manipulated this data frame to make it more understandable and to plot a bar graph.

| type | Mean_Average_Price | | Mean_Average_Price | |
|---------------------|--------------------|----------|--------------------|---------------------|
| | conventional | organic | conventional | organic |
| region | | | | |
| Albany | 1.348757 | 1.773314 | count | 54.000000 54.000000 |
| Atlanta | 1.068817 | 1.607101 | mean | 1.158040 1.654010 |
| BaltimoreWashington | 1.344201 | 1.724260 | std | 0.162290 0.205752 |
| Boise | 1.076036 | 1.620237 | min | 0.728225 1.270769 |
| Boston | 1.304379 | 1.757396 | 25% | 1.070104 1.486154 |
| BuffaloRochester | 1.382308 | 1.651361 | 50% | 1.164260 1.657041 |
| California | 1.105444 | 1.685207 | 75% | 1.270533 1.755266 |
| Charlotte | 1.275089 | 1.936982 | max | 1.408047 2.229231 |
| Chicago | 1.369349 | 1.744201 | | |
| CincinnatiDayton | 1.015503 | 1.402899 | | |
| Columbus | 1.066272 | 1.439290 | | |

The statistics of the data frame.

So as a result, I created a bar graph which will compare mean average price of an avocado in different regions of United States. The bar graph is shown below-



Conclusion from the bar graph

-The major observation here is that the mean average price of an organic kind of avocado is higher than the conventional type of avocado in all the regions of United States.

-The second useful observation here is that Hartford Springfield and San Francisco have higher mean average price of a conventional as well as organic avocado than other regions of United States.

I further investigated my data frame to find the regions where

- 1- Mean Average Price of an organic avocado was minimum.
- 2- Mean Average Price of an organic avocado was maximum.
- 3- Mean Average Price of a conventional avocado was minimum.
- 4- Mean Average Price of a conventional avocado was maximum.

So, I started writing queries to my data frame that I had created earlier and I found the following results

- The minimum mean average price for an organic type of avocado (between dates: 2015-01-01 to 2018-03-25 available in our dataset) is 1.270769 in the region **Houston**

- The maximum mean average price for an organic type of avocado (between dates: 2015-01-01 to 2018-03-25 available in our dataset) is 2.229231 in the region **Hartford Springfield**
- The minimum mean average price for a conventional type of avocado (between dates: 2015-01-01 to 2018-03-25 available in our dataset) is 0.728225 in the region **Phoenix Tucson**
- The maximum mean average price for a conventional type of avocado (between dates: 2015-01-01 to 2018-03-25 available in our dataset) is 1.408047 in the region **Hartford Springfield**

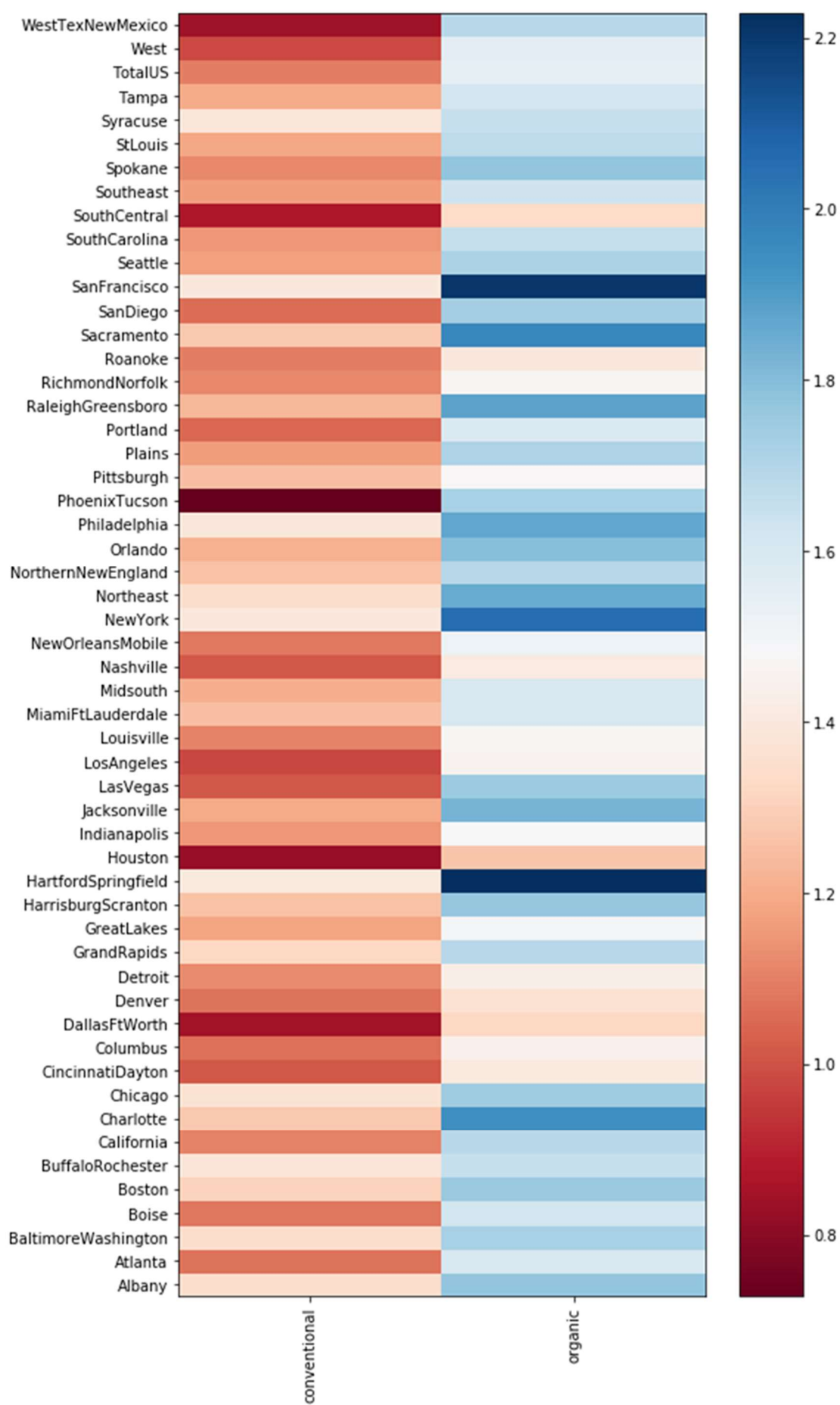
We can also verify these results from the bar graph.

I have also created a heatmap for more clear understanding of mean average price of an avocado in different regions.

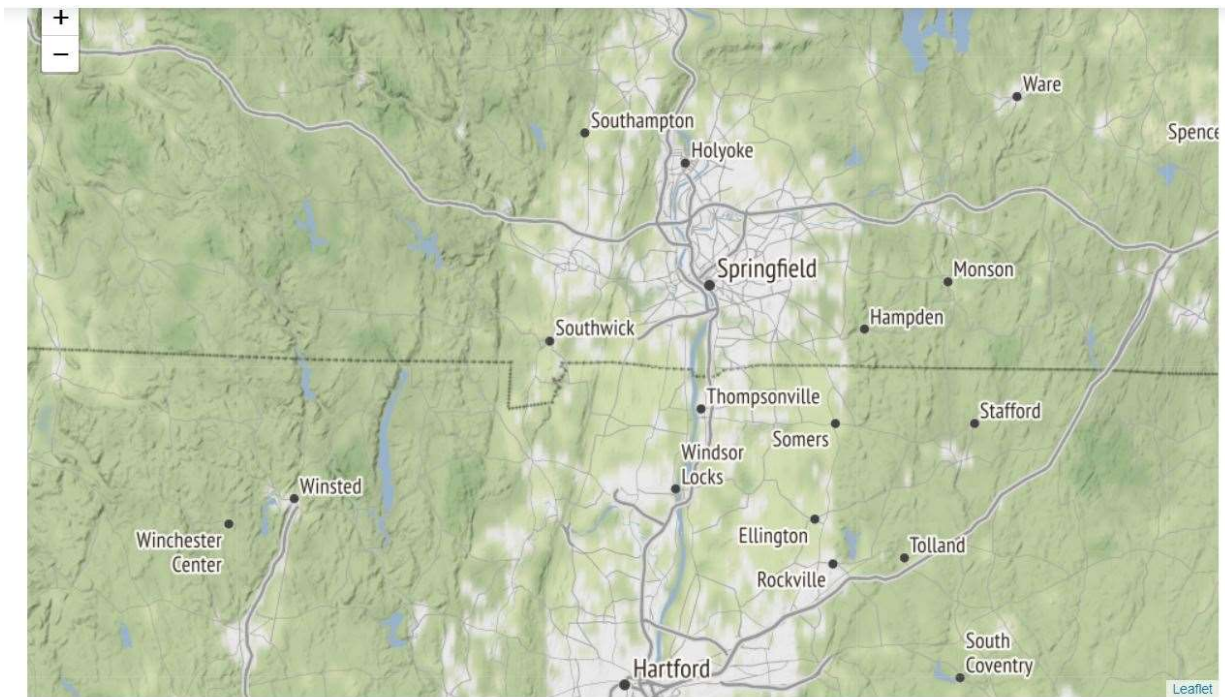
On analysing the heatmap below we can come on the conclusion that

The regions Hartford Springfield and San Francisco have on an average higher mean average price of avocados than other regions in United States.

The first stage of analysis is over and now my next step is to perform the data analysis part on the regions Hartford Springfield and San Francisco because they have on an average higher mean average price of avocados than other regions in United States.



DATA ANALYSIS: HARTFORD SPRINGFIELD



This is the location of Hartford Springfield.

From the main dataset I created a new data frame which contain the data of Hartford Springfield only and the data frame is shown below-

Out[34]:

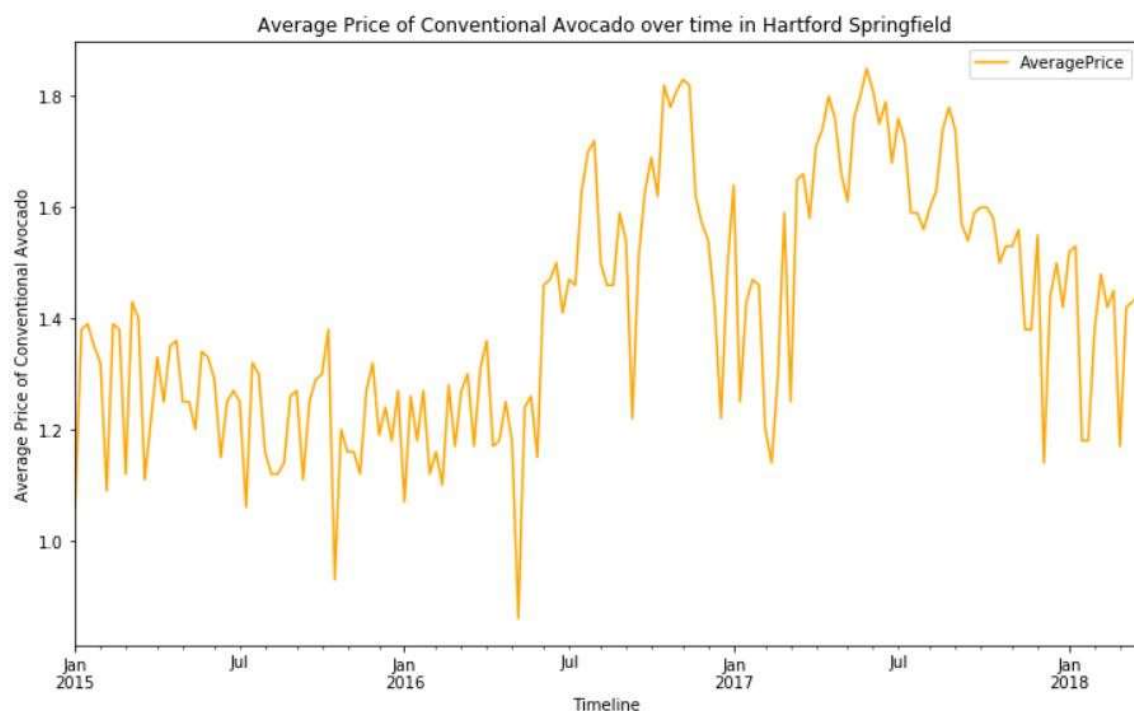
| | Date | AveragePrice | type | region |
|-----|------------|--------------|--------------|---------------------|
| 884 | 2015-12-27 | 1.27 | conventional | HartfordSpringfield |
| 885 | 2015-12-20 | 1.18 | conventional | HartfordSpringfield |
| 886 | 2015-12-13 | 1.24 | conventional | HartfordSpringfield |
| 887 | 2015-12-06 | 1.19 | conventional | HartfordSpringfield |
| 888 | 2015-11-29 | 1.32 | conventional | HartfordSpringfield |

From this data frame I further created a data frame which was having only conventional type of avocados. The data frame is shown below-

Out[37]:

| | Date | AveragePrice | type | region |
|-----|------------|--------------|--------------|---------------------|
| 935 | 2015-01-04 | 1.06 | conventional | HartfordSpringfield |
| 934 | 2015-01-11 | 1.38 | conventional | HartfordSpringfield |
| 933 | 2015-01-18 | 1.39 | conventional | HartfordSpringfield |
| 932 | 2015-01-25 | 1.35 | conventional | HartfordSpringfield |
| 931 | 2015-02-01 | 1.32 | conventional | HartfordSpringfield |

And I plot the curve to get the variation of average price of conventional type of avocados in Hartford Springfield over time. The curve is shown below-



Conclusion from this graph:

As we can see that from January 2015 graph starts and at that time the average price of conventional avocados was somewhere in between 1.0 to 1.2.

The average price of conventional avocados was at its lowest near the month of May 2016 but as soon as July 2016 came closer average price of conventional avocado starts to rise very quickly then again dropped to normal between 1.2 to 1.4 near September 2016 and December 2016 but as soon as the January 2017 came closer average price of conventional avocados raised again and highest this time was in between 1.6 to near 1.8.

The highest average price ever recorded was more than 1.8 near May 2017 and after September of 2017 it starts to drop down to the January of 2018.

I also queried my created data frame to get the exact dates of minimum and maximum average price ever recorded in this timeline.

The results are below-

| | Date | AveragePrice | type | region |
|------|------------|--------------|--------------|---------------------|
| 3725 | 2016-05-08 | 0.86 | conventional | HartfordSpringfield |

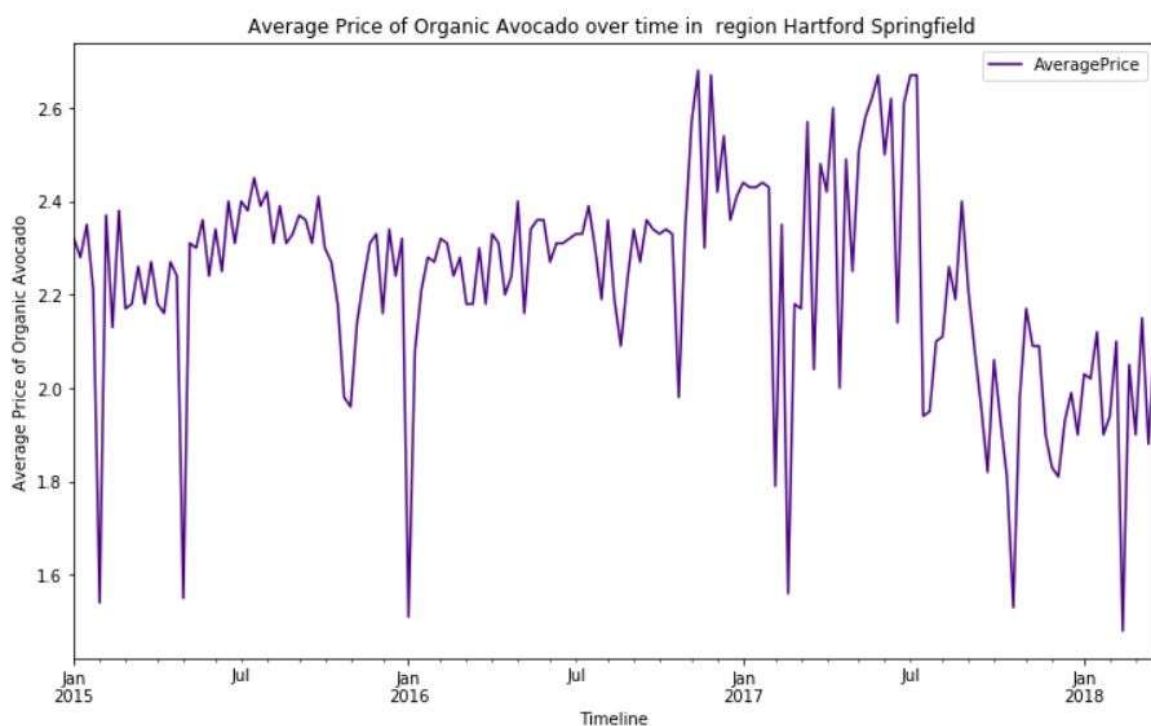
The minimum average price ever recorded was on 8th of May 2016

| | Date | AveragePrice | type | region |
|------|------------|--------------|--------------|---------------------|
| 6548 | 2017-05-28 | 1.85 | conventional | HartfordSpringfield |

The maximum average price ever recorded was on 28th of May 2017.

And we can verify these results from graph.

I also created a data frame for Average Price of organic avocado in the region Hartford Springfield the data frame is shown below-



Conclusion from this graph:

This analytical graph of variations of Average price of organic avocados over time in Hartford Springfield has a lot of variations.

The graph starts from January 2015 onwards and at that time the average price of organic avocado was between 2.2 to 2.4. This graph also tells us that the average price of organic avocado drops below 1.8, 6 times between January of 2015 to March of 2018.

The average price of organic avocados very much high near December 2016 and after starts to drop down and massively drop down below 1.8 around February and March 2017.

From the data frame I found that the minimum average price of organic avocado ever recorded was

| | Date | AveragePrice | type | region |
|-------|------------|--------------|---------|---------------------|
| 17810 | 2018-02-18 | 1.48 | organic | HartfordSpringfield |

The average price was minimum on 18th of February 2018.

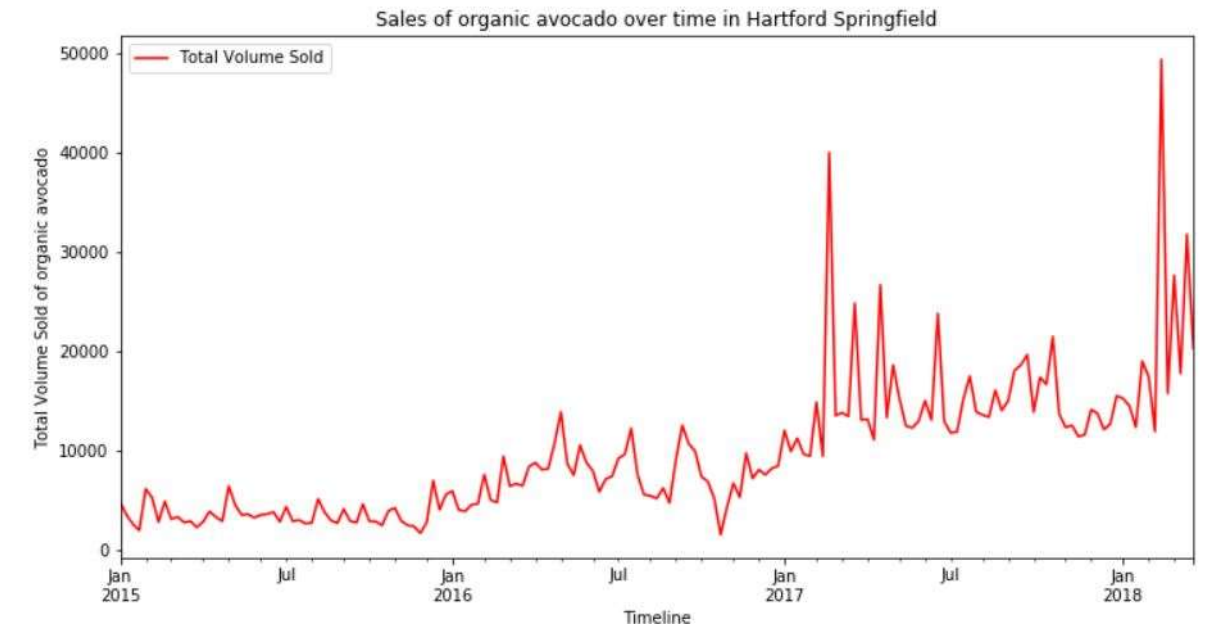
| | Date | AveragePrice | type | region |
|-------|------------|--------------|---------|---------------------|
| 12823 | 2016-11-13 | 2.68 | organic | HartfordSpringfield |

The average price was maximum on 13th of November 2016.

SALES ANALYSIS

I have also analysed the sales of organic and conventional avocados over time in Hartford Springfield.

Now lets first look at the sales of organic avocado in Hartford Springfield.



Conclusion from this graph:

This graph starts from January 2015 at that time the sales were between 0 to Ten Thousand.

Yes, in this graph sales are dropping sometimes but the graph is usually increasing in nature with respect to time.

It can be clearly seen from this graph that organic avocado sale is increasing after December 2016.

Based on analysis we have mainly two major results

The lowest sale of organic avocado in Hartford Springfield was observed on

| | Date | Total Volume Sold | type | region |
|--------------|------------|-------------------|---------|---------------------|
| 12826 | 2016-10-23 | 1494.63 | organic | HartfordSpringfield |

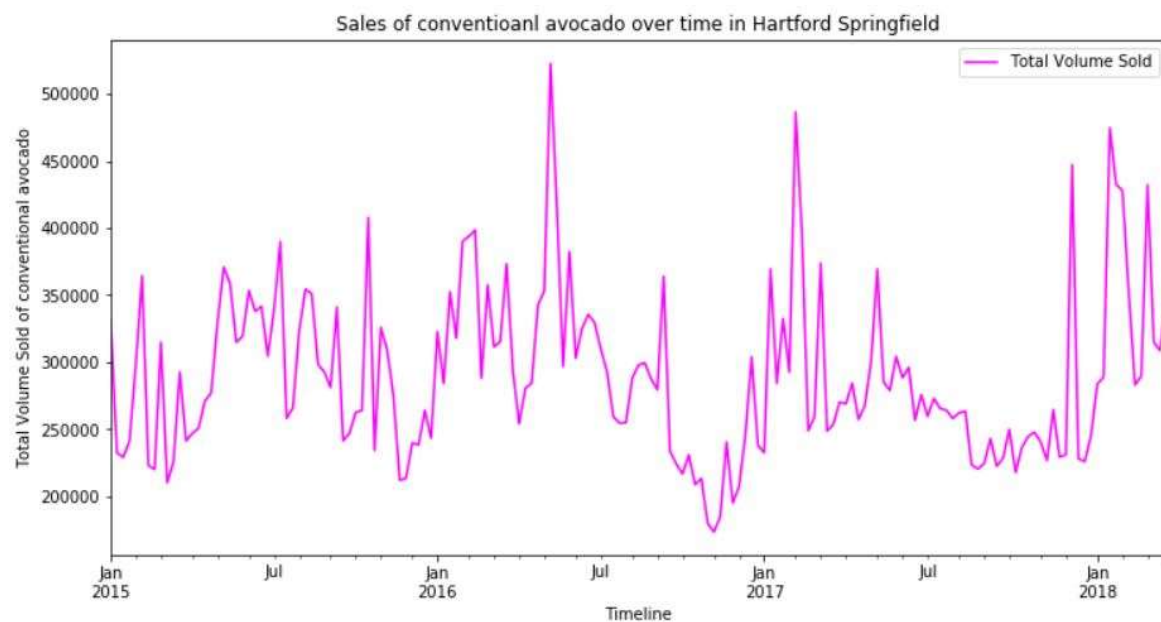
23rd of October 2016, only near about 1494 organic avocados were sold.

And the highest sale of organic avocado in Hartford Springfield was observed on

| | Date | Total Volume Sold | type | region |
|--------------|------------|-------------------|---------|---------------------|
| 17810 | 2018-02-18 | 49348.86 | organic | HartfordSpringfield |

18th of February 2018, near about 49348 organic avocados were sold.

Now let's view the sales of conventional avocados in Hartford Springfield



This graph has many variations it starts from January 2015 and at that time the sales were in between 30 thousand to 35 thousand. The sales drop down too much before December of 2016 and reach its maximum between June 2016 to July 2016.

The lowest sale of conventional avocado in Hartford Springfield was observed on

| | Date | Total Volume Sold | type | region |
|------|------------|-------------------|--------------|---------------------|
| 3699 | 2016-11-06 | 173620.97 | conventional | HartfordSpringfield |

6th of November 2016, only near about 173621 conventional avocados were sold.

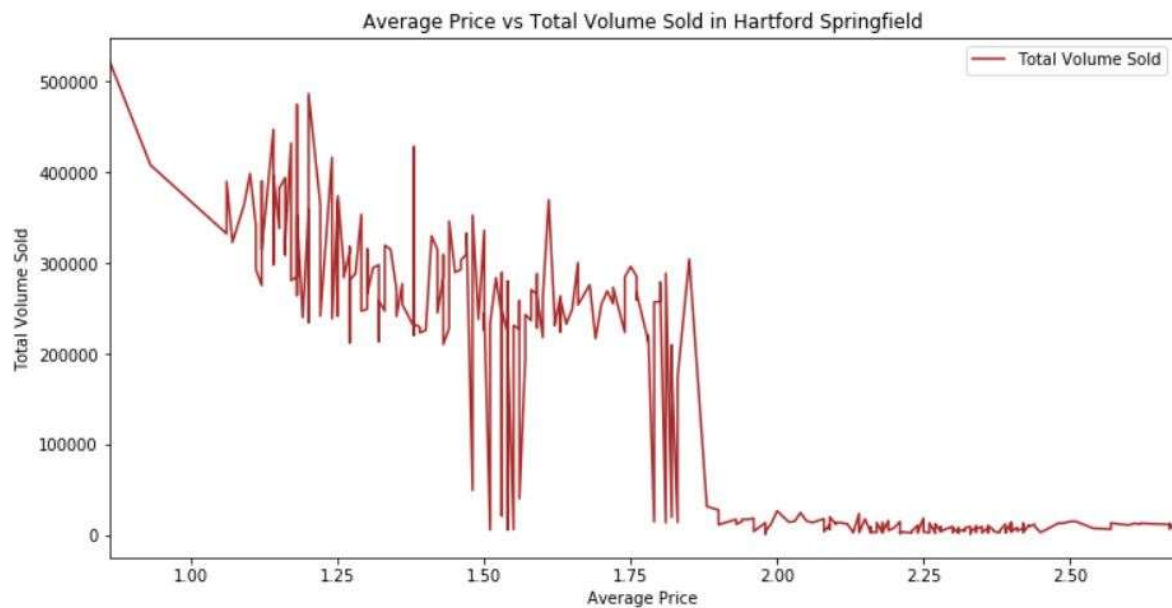
The highest sale of conventional avocado was observed on

| | Date | Total Volume Sold | type | region |
|------|------------|-------------------|--------------|---------------------|
| 3725 | 2016-05-08 | 522565.33 | conventional | HartfordSpringfield |

8th of May 2016, around 522565 conventional avocados were sold.

I have also analysed the average price vs total volume of avocados sales in Hartford Springfield.

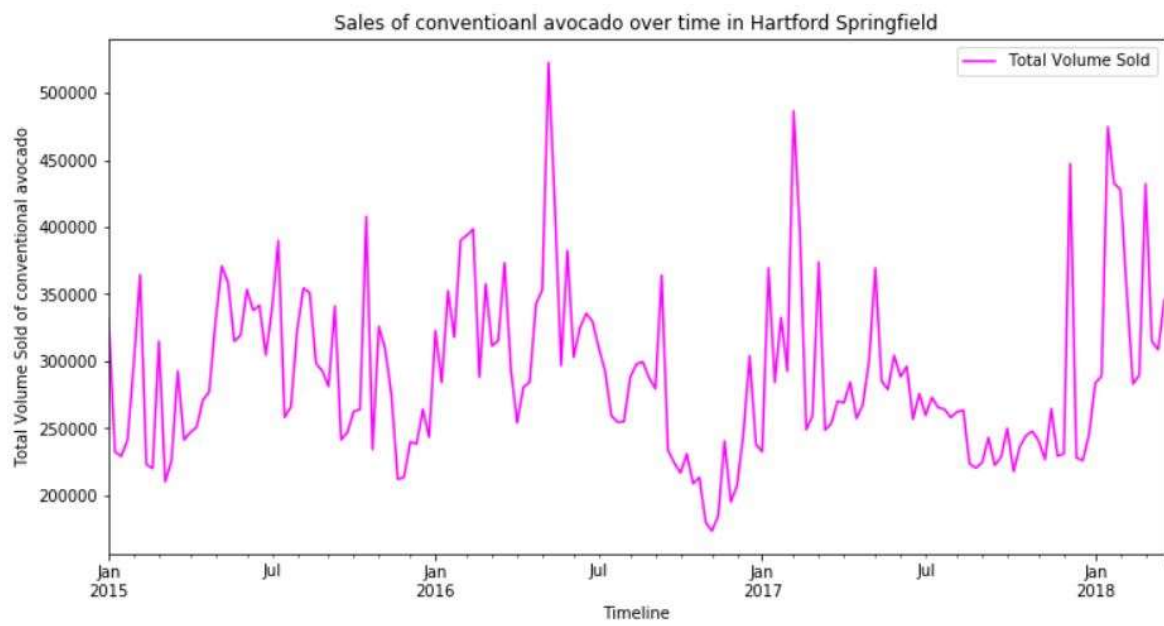
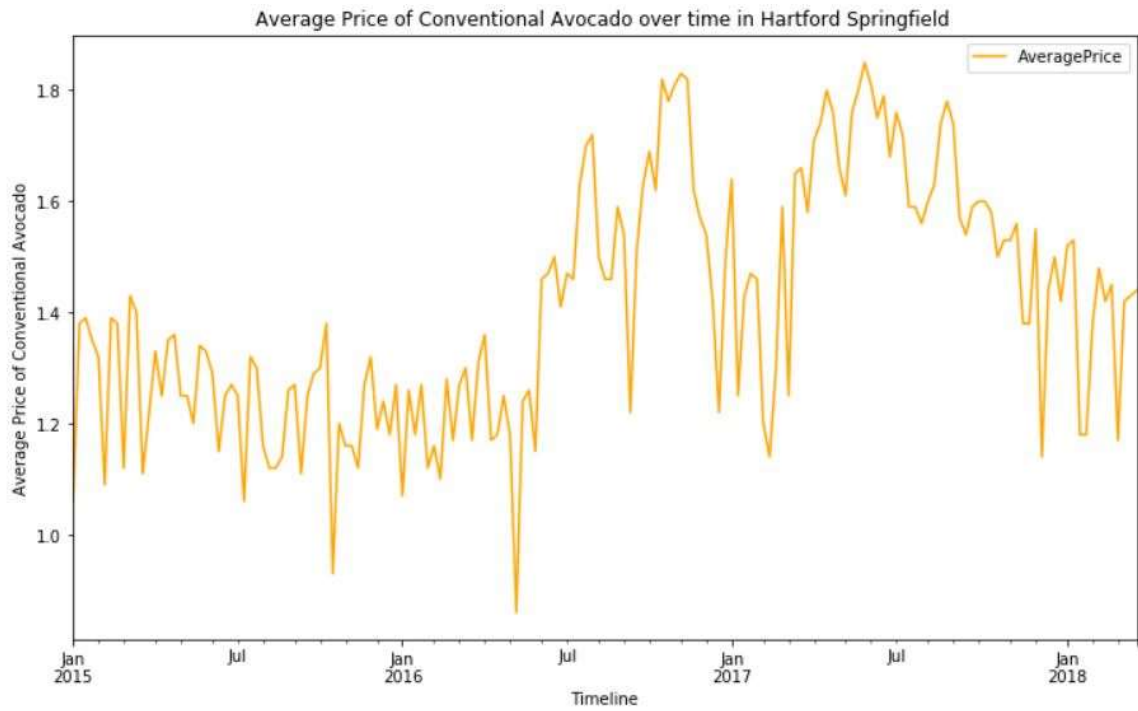
The graph is shown below-



Conclusion from this graph:

It is very clear from this graph that when the average price of avocados in Hartford Springfield is lesser than 1.00 the sale is Highest and as the average price keeps rising the sales witness downfall. This clearly indicates that if we increase average price too much avocados sales are going to be very less. But if the average price of the avocado is very less it might cause some problems in business. According to this graph the most appropriate range of average price where the sales are high and average price is also moderately high is between 1.6 to 1.75 where sales are in the range of 20 thousand to 40 thousand. The sales of avocado also depend on type of avocado customer wants to purchase and also the two types of avocados have different price range. If we slightly reduce the price to 1.50 customers behaviour changes as there is possibility that customers might want to buy avocado in the average price range of 1.25 to somewhere around 1.35. This is just an assumption.

GRAPHICAL COMPARISONS

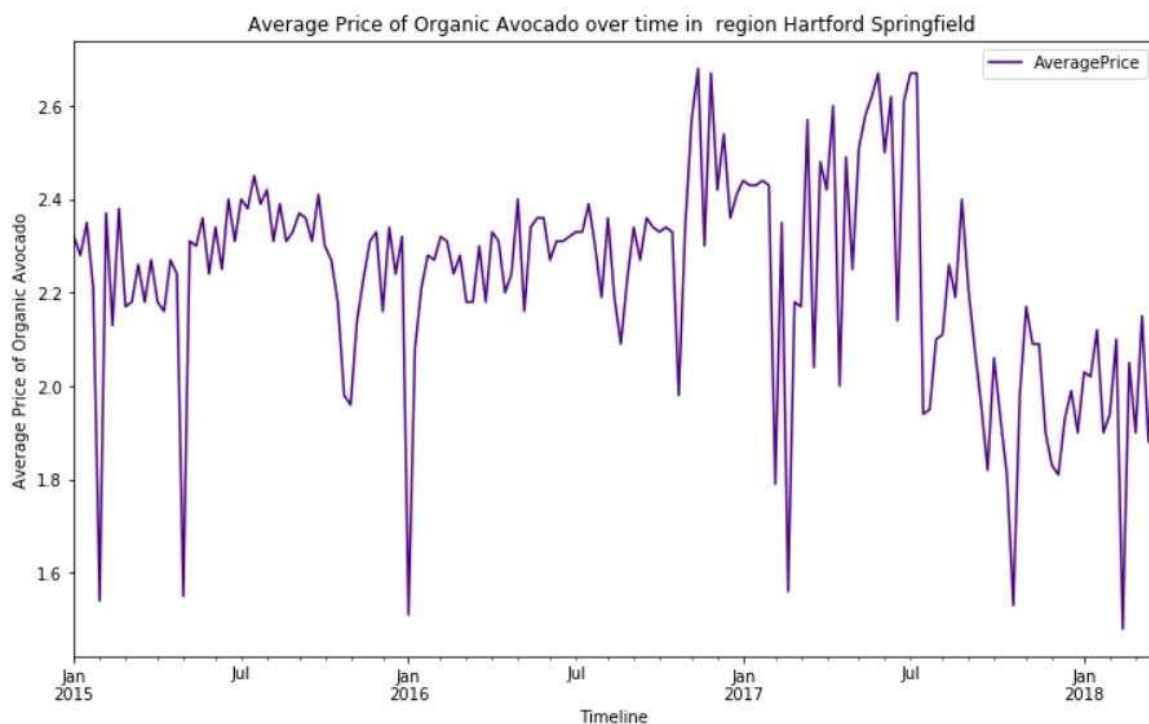


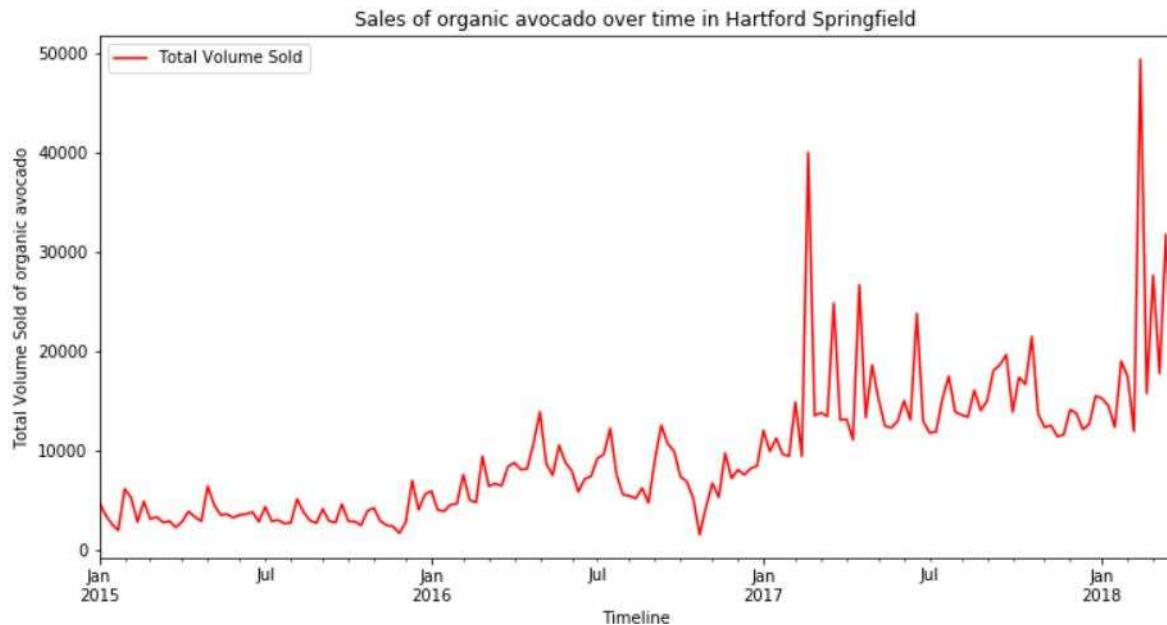
Here I have taken two graphs one of total volume sold of conventional avocado over time and other variation of average price of conventional avocado over time in the region Hartford Springfield.

The most important observation we can generate from these two graphs is that in the sales analysis graph the total volume sales is decreasing after September of 2016 and is lowest on 6th of November 2016, only near about 173621 conventional avocados were sold. At the same time in Average price variation over time graph of conventional avocados in Hartford Springfield from July 2016 graph initially drops a little bit but then it rises too quickly and after September 2016 the average price is again increasing very fast and so we can observe here the pattern that when average price of an avocado is increasing it is affecting the total volume sales of conventional avocado i.e. when the average price of an conventional avocado increases the sale of conventional avocados tends to decrease.

This is a very important observation as it is telling the behaviour of this market that if you increase average price too much the sales drops down.

Now let's compare the same graphs but for organic avocados.





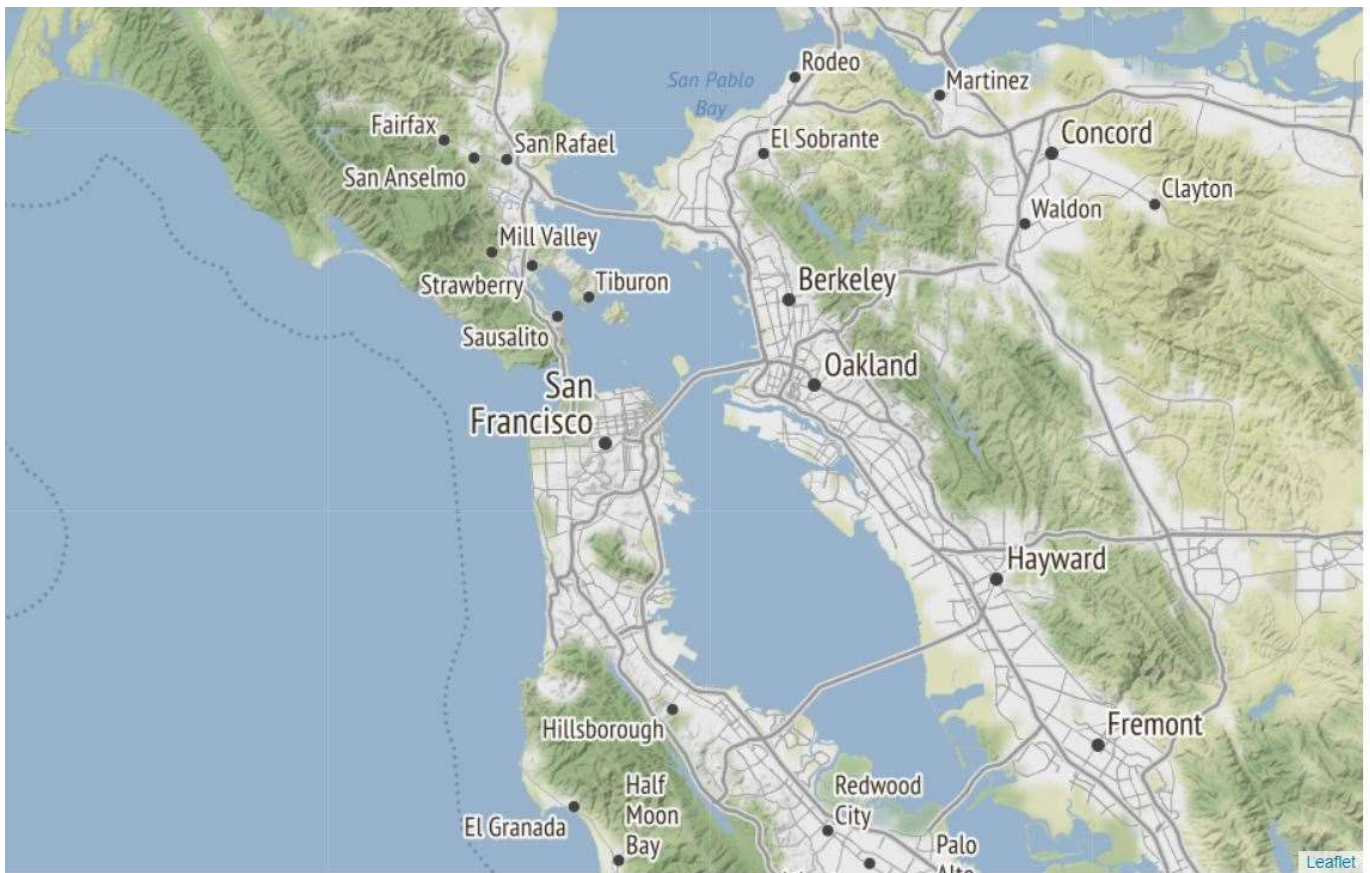
The total volume sales of organic avocado over time in Hartford Springfield is actually increasing in nature with respect to time with some drops in sales in between this is because as we seen the graph of variation of average price of organic avocado over time there are many massive drops and rise in average price over time but specifically if we look at that graph after December 2016 and compare it with total volume sales graph after December 2016 we observe that when average price is decreasing the sales have increased and the highest sale is observed after January 2018 and at the same moment the average price of organic avocado is down to 1.6 after January 2018 .

These both observations prove that average price affects sales of avocado inversely i.e. too much increase in average price brings the sales down.

At last I have also calculated the p-value and Pearson coefficient between average price of avocados and total volume of avocados sales for this region i.e. Hartford Springfield and I found that The Pearson correlation coefficient is -0.8997084706270533 with a p-value of P: 5.19813317092476e-123. And Since the p-value is < 0.001, the correlation between Average Price and Total Volume Sold is statistically significant, the coefficient of -0.899 shows that the relationship is negative and moderately strong.

With this our data analysis for Hartford Springfield Ends

DATA ANALYSIS: SAN FRANCISCO



This is the location of San Francisco.

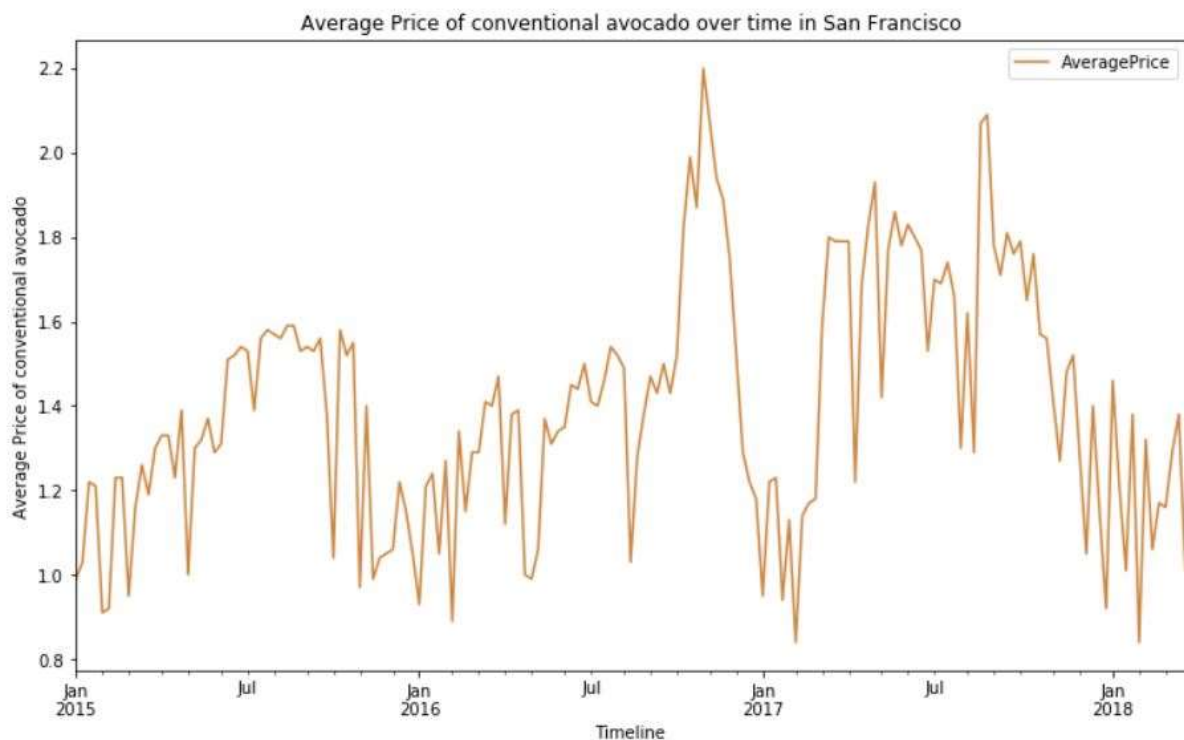
From the main dataset I created a new data frame which contain the data of San Francisco only and the data frame is shown below-

| | Date | AveragePrice | type | region |
|------|------------|--------------|--------------|--------------|
| 2184 | 2015-12-27 | 1.05 | conventional | SanFrancisco |
| 2185 | 2015-12-20 | 1.15 | conventional | SanFrancisco |
| 2186 | 2015-12-13 | 1.22 | conventional | SanFrancisco |
| 2187 | 2015-12-06 | 1.06 | conventional | SanFrancisco |
| 2188 | 2015-11-29 | 1.05 | conventional | SanFrancisco |

From this data frame I further created a data frame which was having only conventional type of avocados. The data frame is shown below-

| | Date | AveragePrice | type | region |
|------|------------|--------------|--------------|--------------|
| 2235 | 2015-01-04 | 0.99 | conventional | SanFrancisco |
| 2234 | 2015-01-11 | 1.03 | conventional | SanFrancisco |
| 2233 | 2015-01-18 | 1.22 | conventional | SanFrancisco |
| 2232 | 2015-01-25 | 1.21 | conventional | SanFrancisco |
| 2231 | 2015-02-01 | 0.91 | conventional | SanFrancisco |

And I plot the curve to get the variation of average price of conventional type of avocados in San Francisco over time. The curve is shown below-



Conclusion from the graph:

This graph starts from January 2015 , initially the average price of conventional avocados vary in between the range of 0.8 to 1.7 but as we reach August 2016 ,we see that after August 2016 we see that average price keeps on increasing and it crosses the value of 2.0 and from its peak value it starts to drop down .From October 2016 to February 2017 its drops below 1.0 again after September 2017 it starts to drop down and near the start of 2018 it drops down below 1.0.

Based on my analysis I have found that

The minimum average price of a conventional avocado was observed on

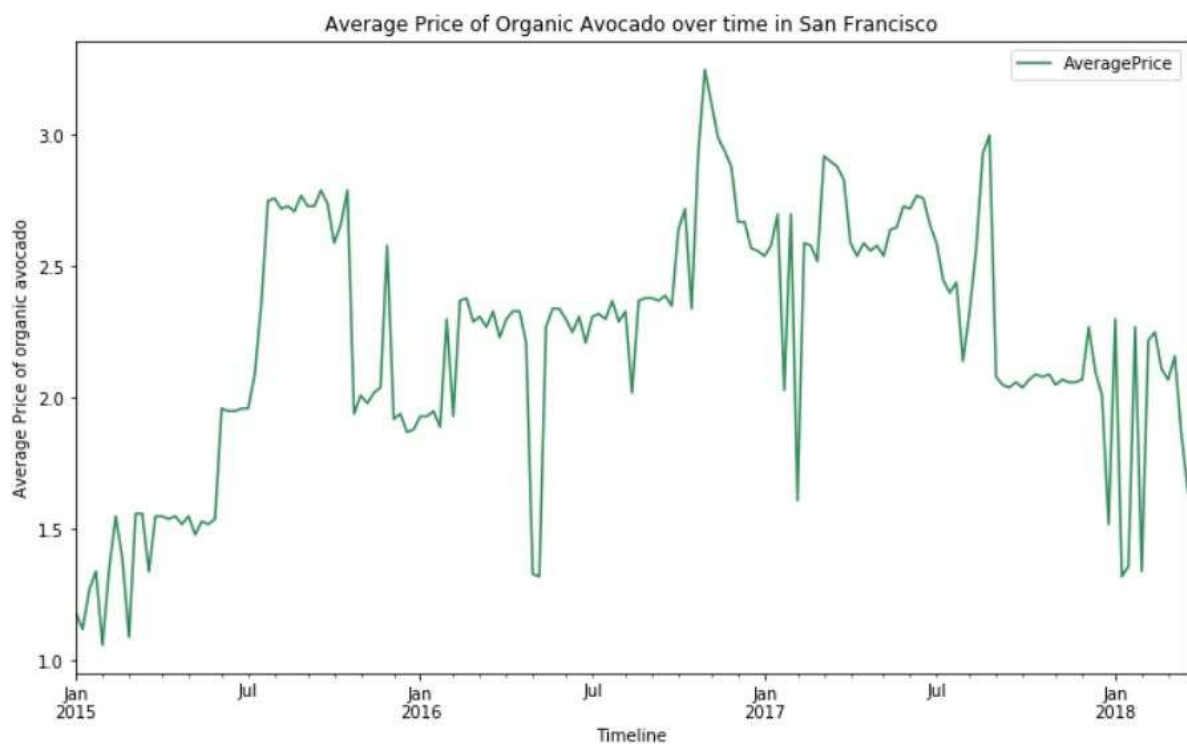
| | Date | AveragePrice | type | region |
|------|------------|--------------|--------------|--------------|
| 7889 | 2017-02-05 | 0.84 | conventional | SanFrancisco |
| 8989 | 2018-02-04 | 0.84 | conventional | SanFrancisco |

these two dates 5th of February 2017 and 4th of February 2018 with the value of 0.84

| | Date | AveragePrice | type | region |
|------|------------|--------------|--------------|--------------|
| 5000 | 2016-10-30 | 2.2 | conventional | SanFrancisco |

And the maximum average price of a conventional avocado was observed on 30th of October 2016 with the value of 2.2

I also created a data frame for Average Price of organic avocado in the region San Francisco the data frame is shown below-



Conclusion from this graph:

This graph starts from January 2015 and initially with time average price tends to increase with some drops in between. The main thing to notice from this graph is that after July 2016 average price starts to increase quickly and before the month of December or near around last months of 2016 it crosses the value of 3.0 and reaches its maximum.

Based on my analysis I found that

The minimum average price of organic avocado was observed on

| | Date | AveragePrice | type | region |
|-------|------------|--------------|---------|--------------|
| 11357 | 2015-02-01 | 1.06 | organic | SanFrancisco |

1st of February 2015 with a value of 1.06

And the maximum average price of organic avocado was observed on

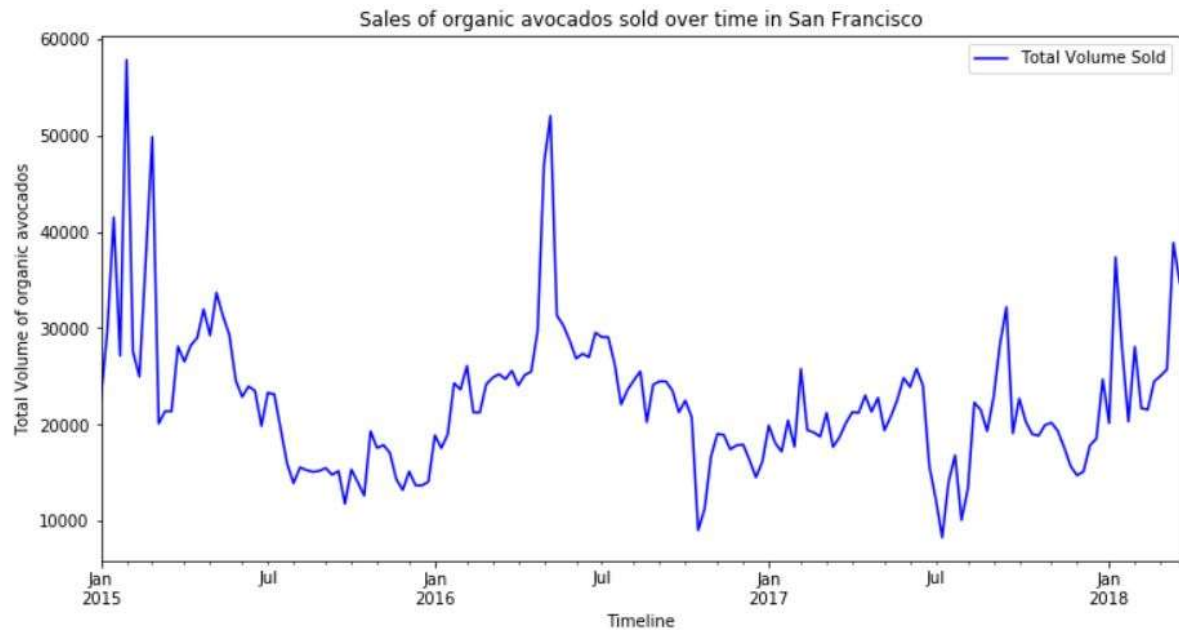
| | Date | AveragePrice | type | region |
|-------|------------|--------------|---------|--------------|
| 14125 | 2016-10-30 | 3.25 | organic | SanFrancisco |

30th of October 2016 with a value of 3.25.

SALES ANALYSIS

I have also analysed the sales of organic and conventional avocados over time in San Francisco.

Now let's first look at the sales of organic avocado in San Francisco.



Conclusion from graph:

This graph starts from January 2015 and in between January 2015 and before May 2015 the sales of organic avocado in San Francisco were high.

Also, in this graph the sales drop down to minimum in the intervals of July 2016 to December 2016 and March 2017 to September 2017.

Based on my data analysis I have found that

The minimum number of organic avocados sold in San Francisco was observed on

| | Date | Total Volume Sold | type | region |
|-------|------------|-------------------|---------|--------------|
| 16992 | 2017-07-09 | 8311.12 | organic | SanFrancisco |

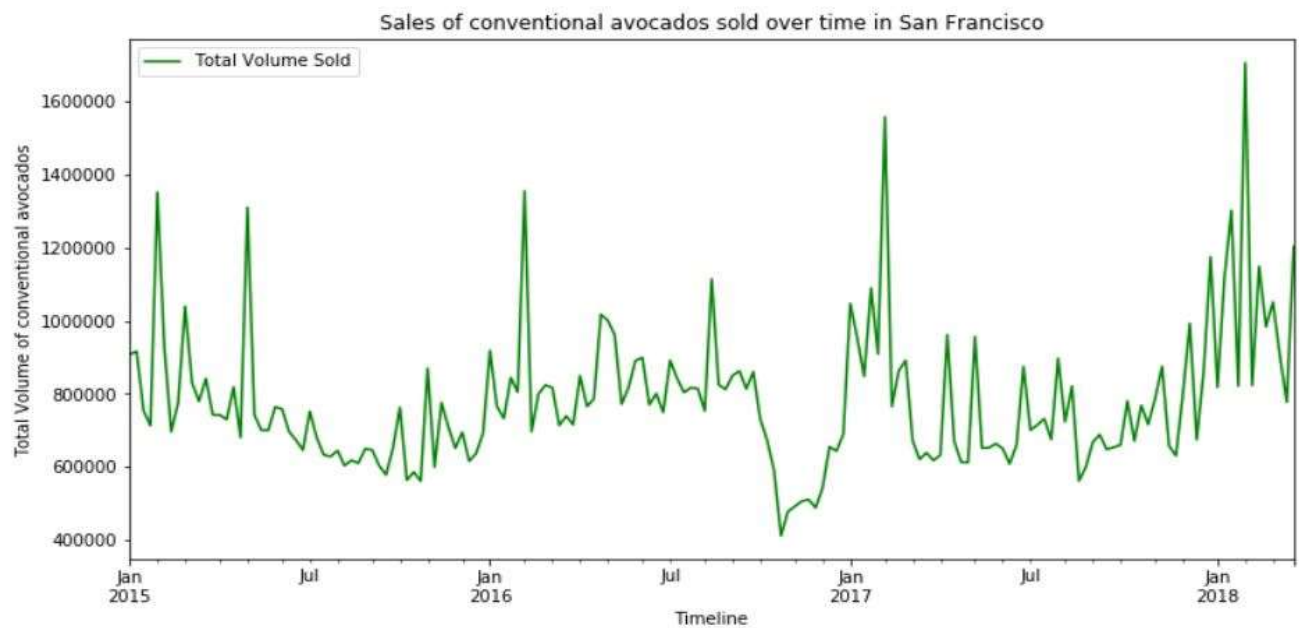
9th of July 2017 with a volume of 8311 avocados approx.

The maximum number of organic avocados sold in San Francisco was observed on

| | Date | Total Volume Sold | type | region |
|-------|------------|-------------------|---------|--------------|
| 11357 | 2015-02-01 | 57802.02 | organic | SanFrancisco |

1st of February 2015 with a volume of 57802 avocados approx.

I have also analysed the sales of conventional avocados in San Francisco



Conclusion from this graph:

This graph starts from January 2015. The major observation that can be taken out from this graph is that near the start of 2018 in between January to March the sales of avocados were rapidly increasing and it reaches its maximum in the start of 2018.

The second main observation is that the sales of avocados drop down to minimum after September 2016 and before December 2016.

Based on my analysis I found that

The minimum amount of sale of conventional avocados was observed on

| | Date | Total Volume Sold | type | region |
|-------------|------------|-------------------|--------------|--------------|
| 5001 | 2016-10-23 | 411873.66 | conventional | SanFrancisco |

23rd of October 2016 with the volume of 411873 avocados approx.

The maximum amount of sale of conventional avocados was observed on

| | Date | Total Volume Sold | type | region |
|-------------|------------|-------------------|--------------|--------------|
| 8989 | 2018-02-04 | 1706251.05 | conventional | SanFrancisco |

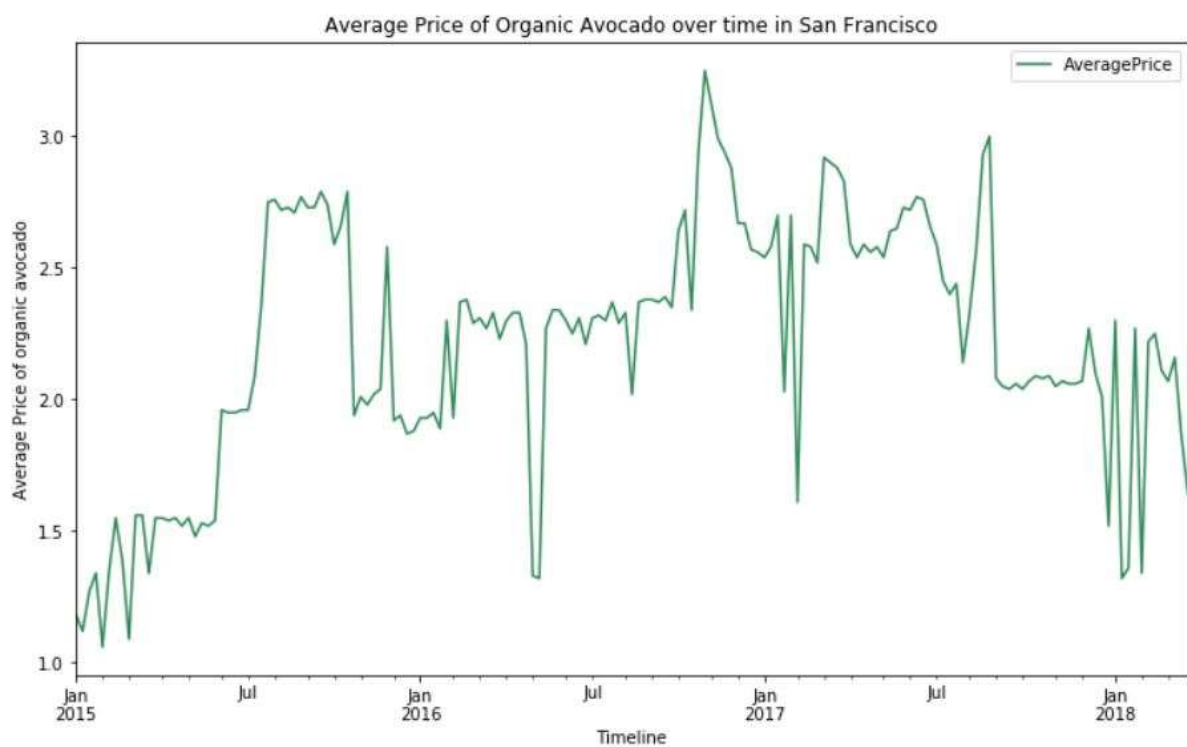
4th of February 2018 with the volume of 1706251 avocados approx.

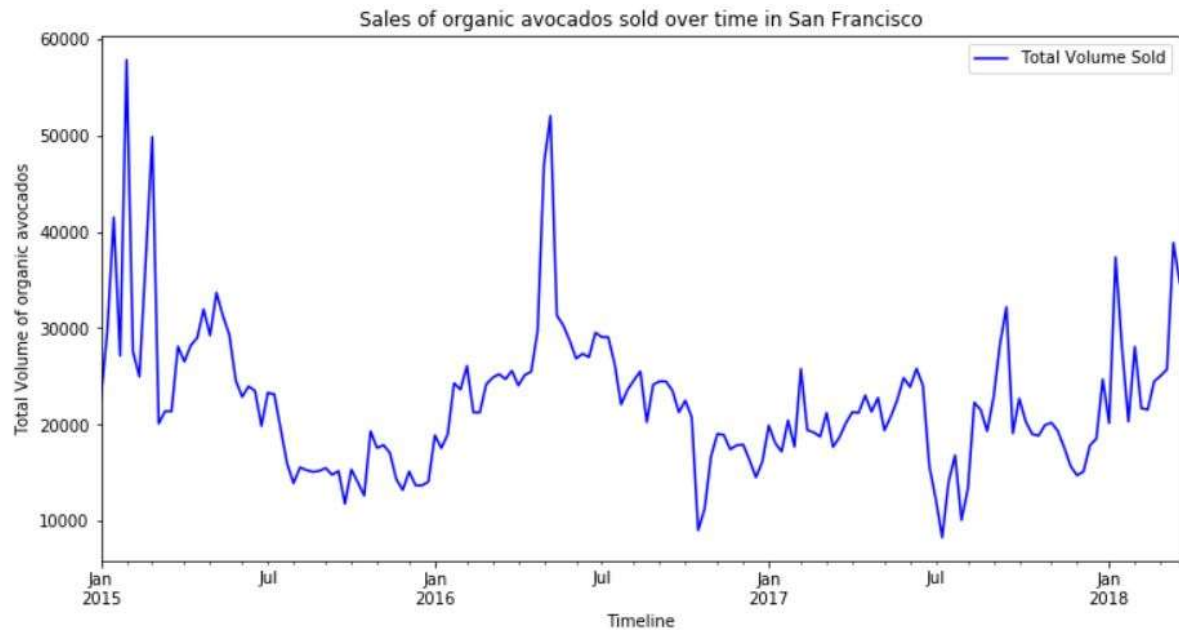
Average Price vs Total Volume of avocado sold analysis in San Francisco.



This graph shows that sales rapidly decreases as average price of avocado increases. (Type of avocado also important for the sales pattern).

GRAPHICAL COMPARISONS



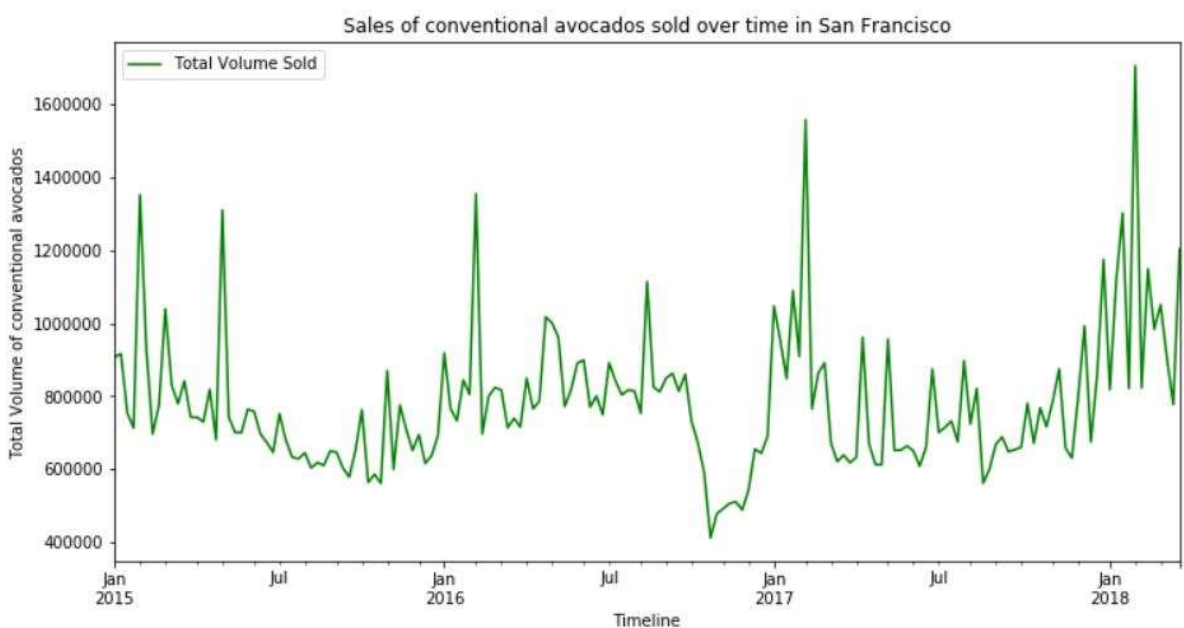
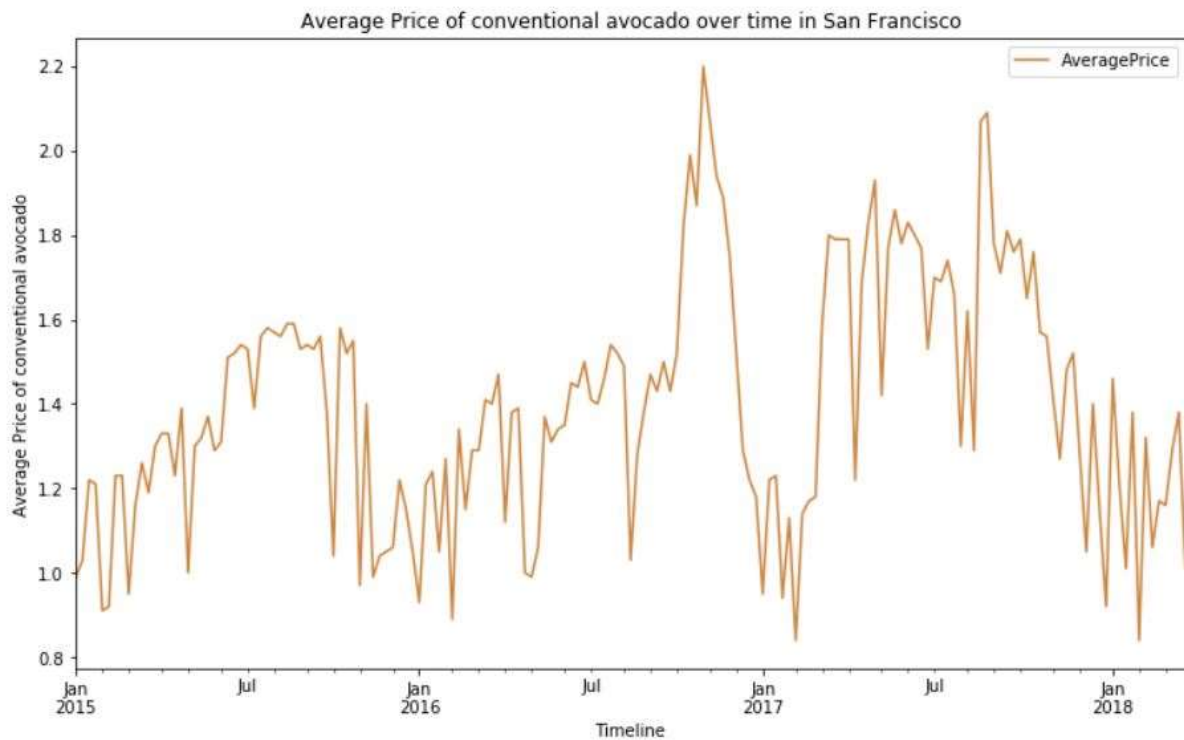


As we have observed the above graphs it is clear that in the beginning when the average price of organic avocados in San Francisco is low, the sale is at its peak somewhere between starting 3-4 months of 2015.

We know that sales of organic avocados drop down below 20 thousand between July 2016 to December 2016 and March 2017 to September 2017 because at the same time interval average price was increasing rapidly.

So, with the help of this conclusion we have found that with the increase in average price of organic avocado the sales drops down.

Similarly, when we compare the average price variation of conventional avocados over time with sales of conventional avocados over time



we get the similar trend that when the average price is running really low after December 2016 and before April 2017 sales of conventional avocados is running high. Average price is also running low in the end of 2017 and start of 2018 corresponding to which sales of conventional avocado is running high and achieving its maximum.

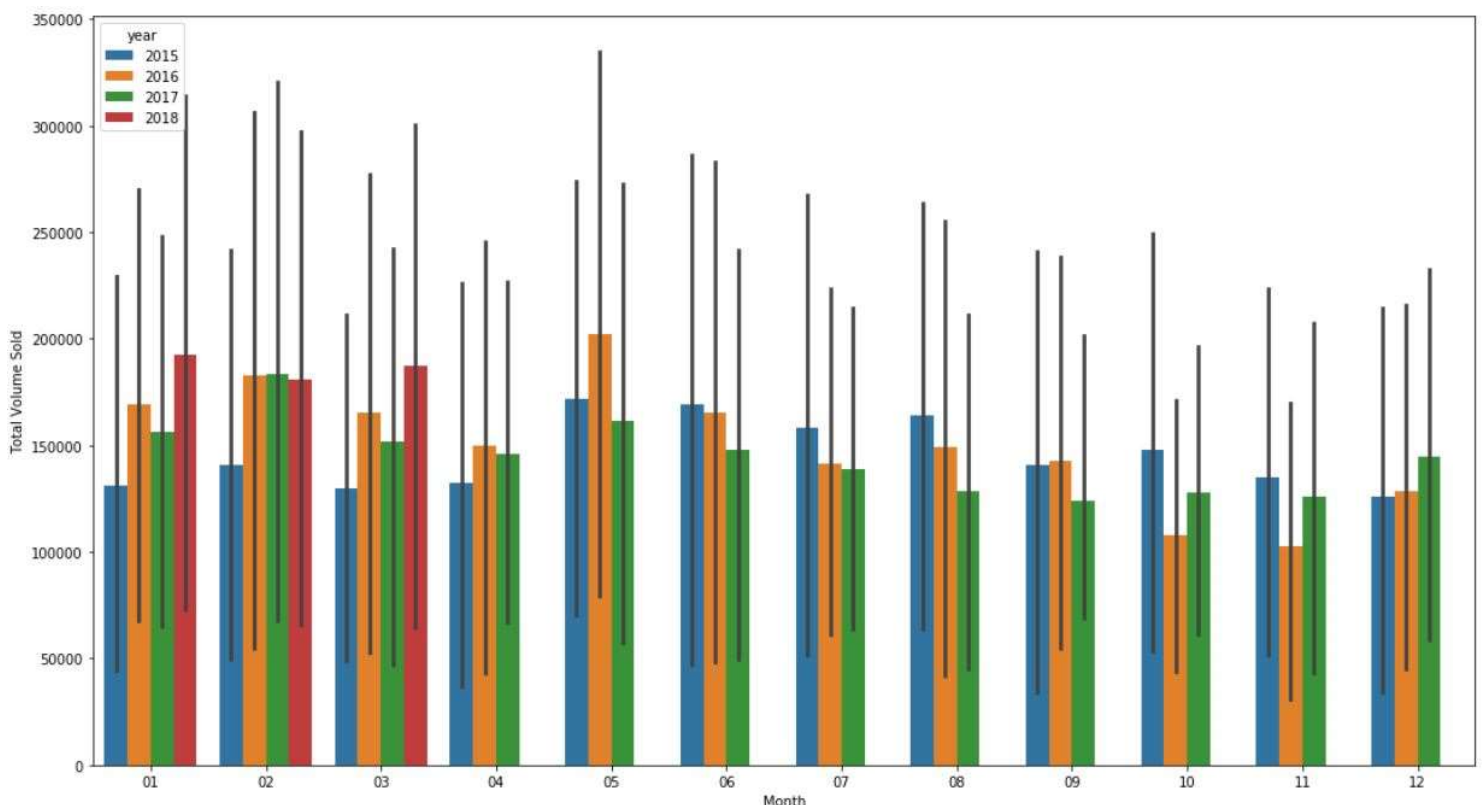
Again, these both observations prove that average price affects sales of avocado inversely i.e. too much increase in average price brings the sales down.

At last I have also calculated the p-value and Pearson coefficient between average price of avocados and total volume of avocados sales for this region i.e. San Francisco and I found that the Pearson Correlation coefficient is -0.7709556583470577 with a p-value of $P: 8.23994209244763e-68$

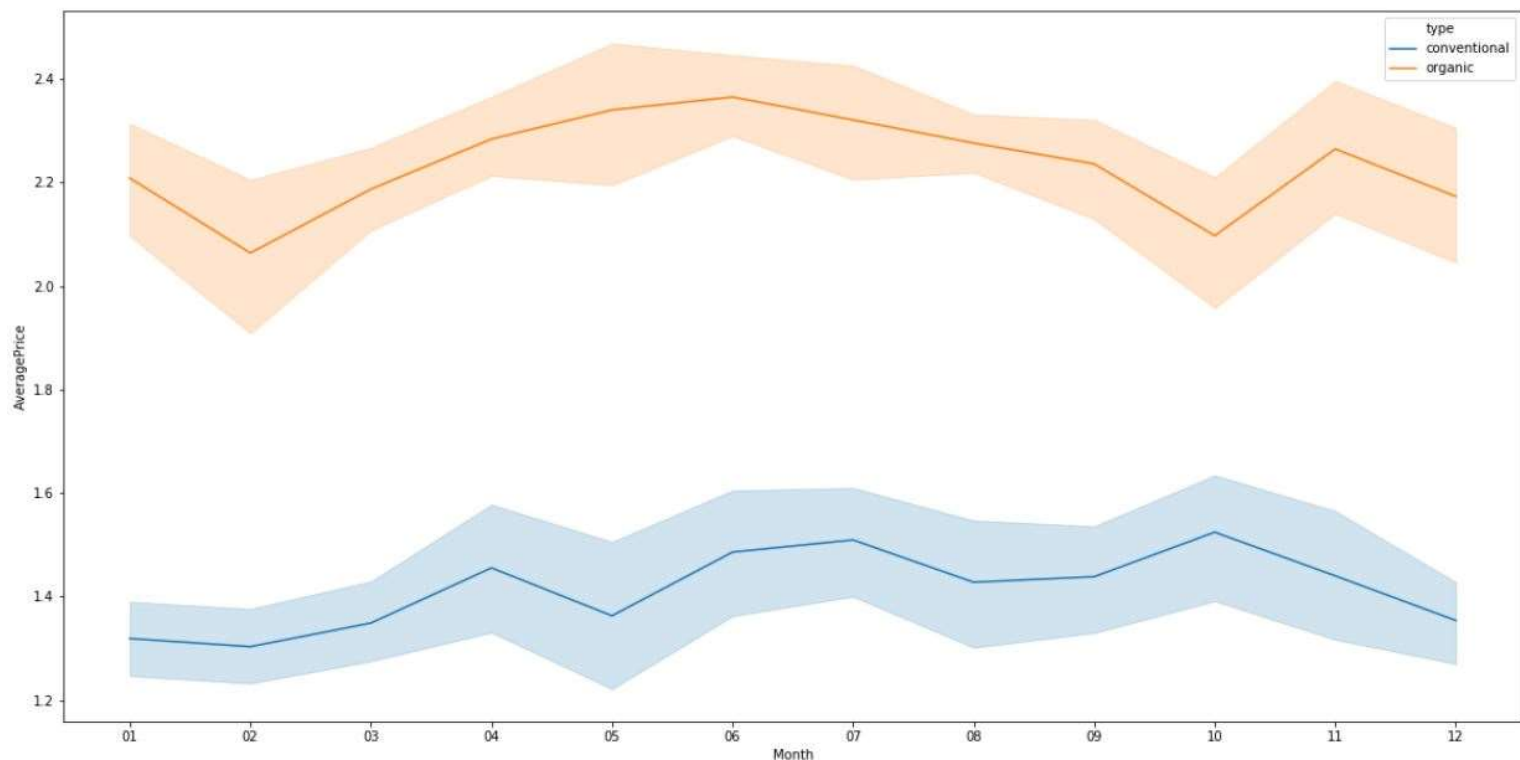
Since the p-value is < 0.001 , the correlation between average price and Total Volume Sold is statistically significant, the coefficient of -0.770 shows that the relationship is negative and moderately strong.

MONTH WISE GRAPHS

HARTFORD SPRINGFIELD

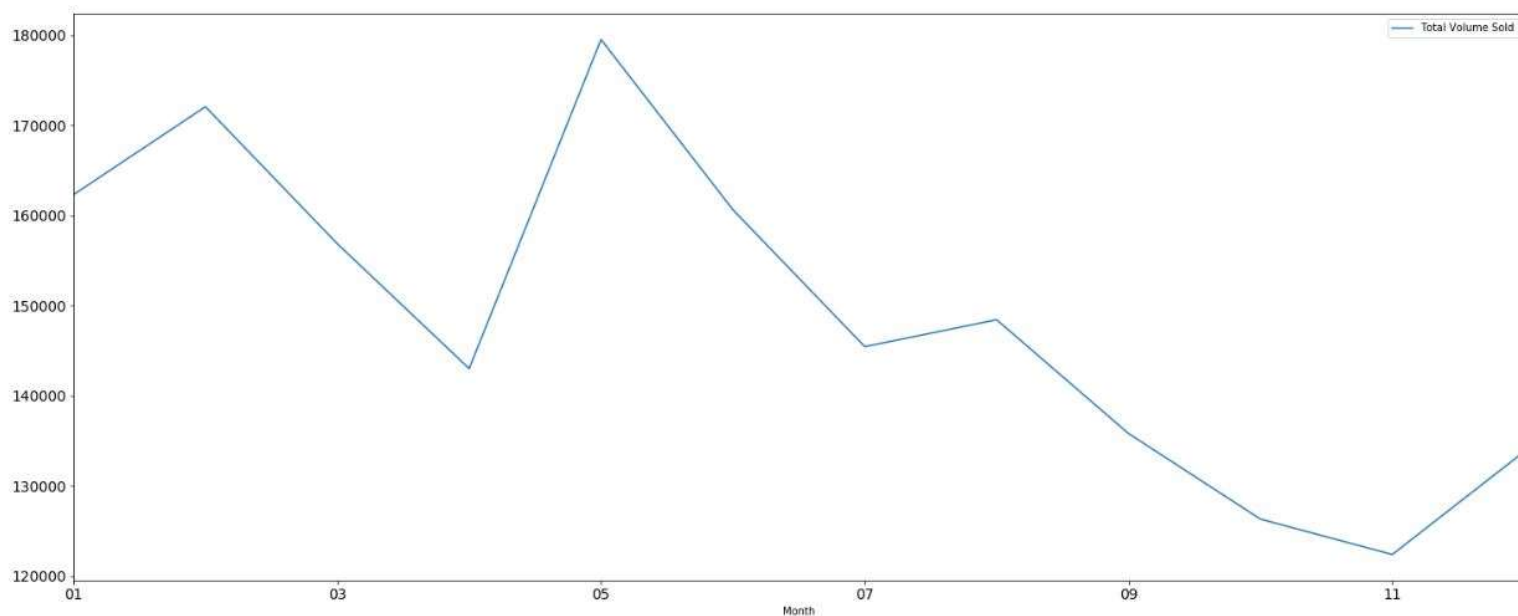


TOTAL VOLUME SOLD IN DIFFERENT MONTHS (2015-2018)



AVERAGE PRICE OF AVOCADOS DURING DIFFERENT MONTHS (2015-2018)

From this graph we can conclude that Average Prices of organic avocados in Hartford Springfield are very high during months of May, June and July. And for conventional June to July and September to October.



MEAN OF TOTAL VOLUME SALES DURING DIFFERENT MONTHS (2015-2018)

From this graph we can conclude that the sales of avocados in Hartford Springfield is very high in month of May as sales increases from April to May and from August to November the sales continuously decrease and reach its minimum in the month of November.

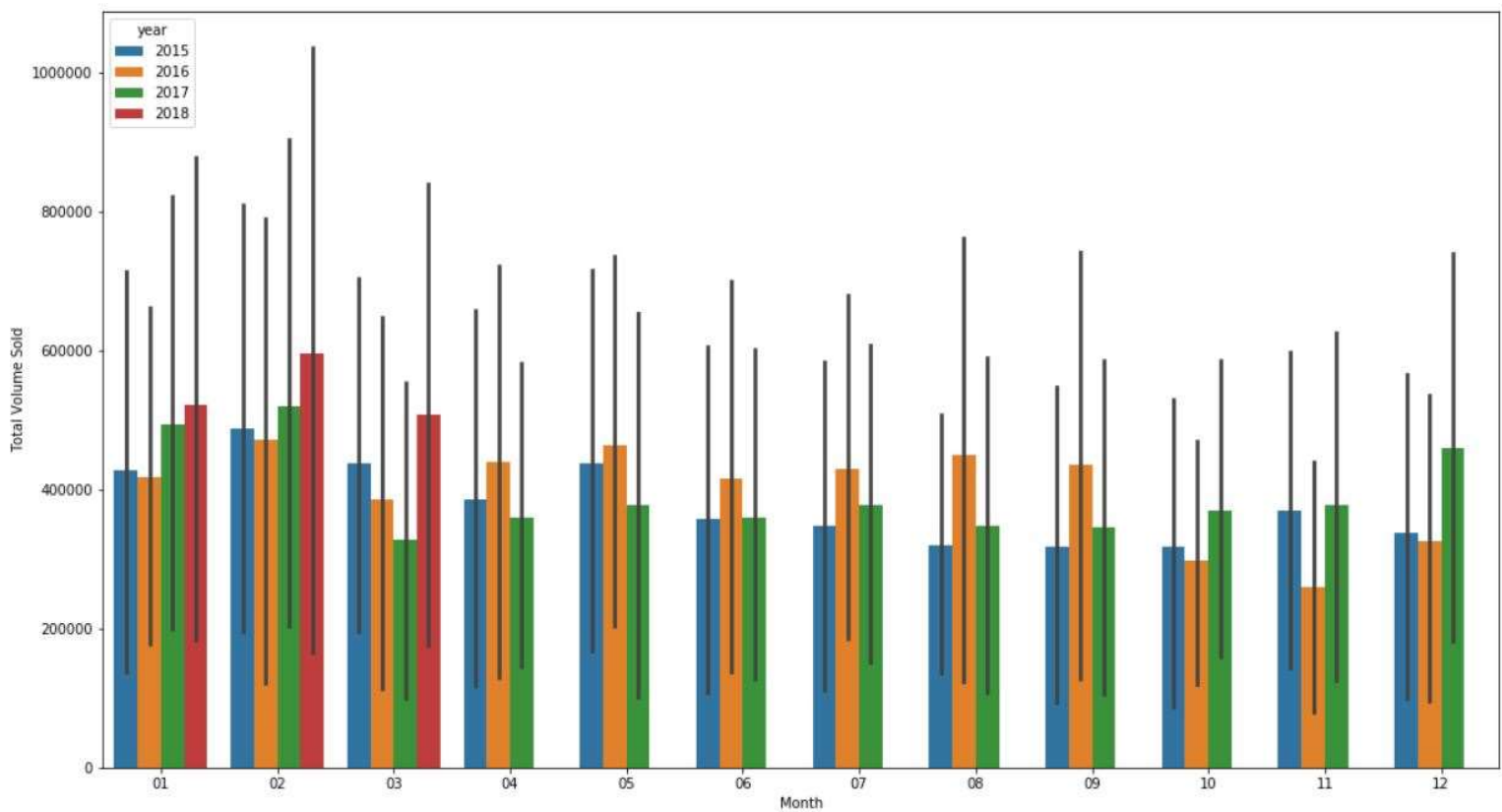
From my analysis I found that in Hartford Springfield the minimum sales occur in the month of November, the average(mean) of 122391 of total avocados sold from 2015 to 2018.

| Total Volume Sold | |
|-------------------|---------------|
| Month | |
| 11 | 122391.225769 |

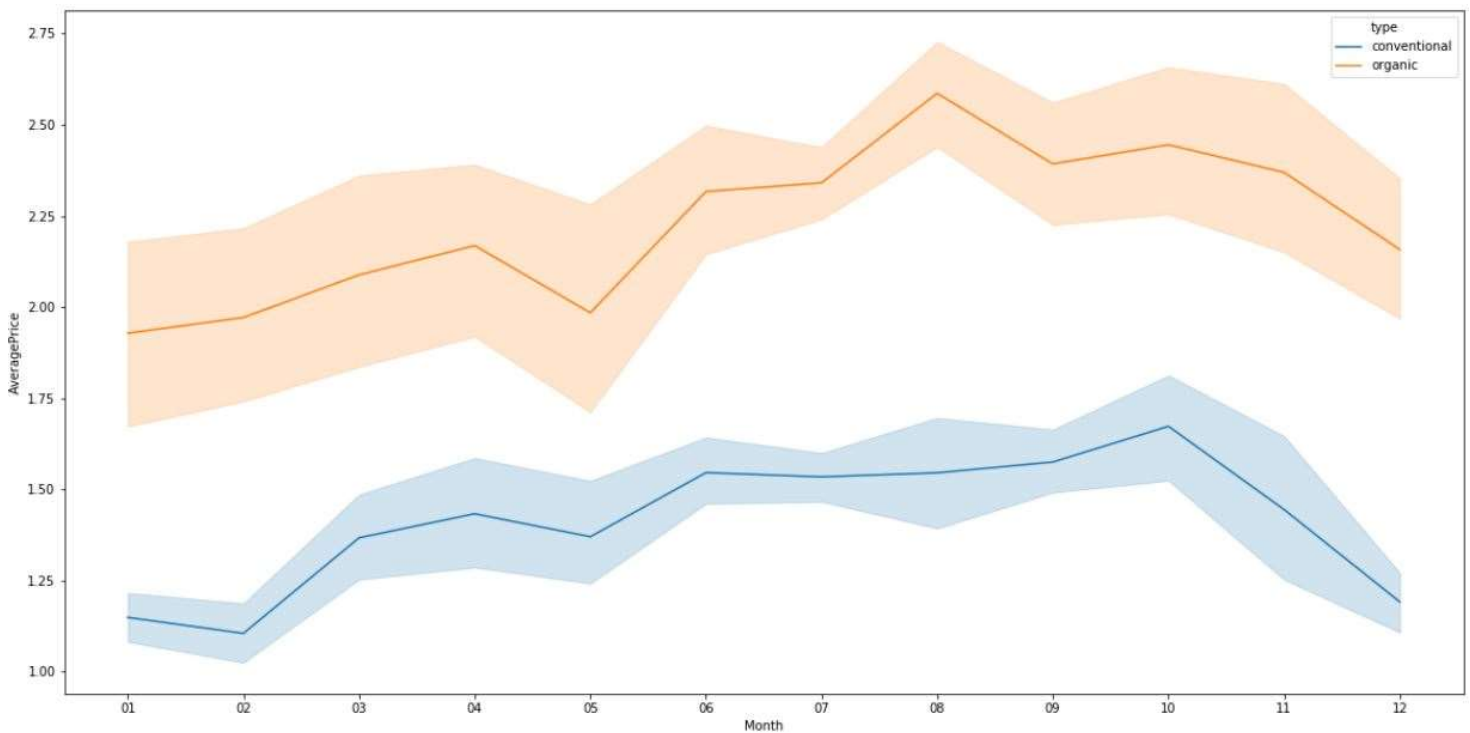
From my analysis I found that in Hartford Springfield the maximum sales occur in the month of May, the average(mean) of 179512 of total avocados sold from 2015 to 2018.

| Total Volume Sold | |
|-------------------|---------------|
| Month | |
| 05 | 179512.726429 |

SAN FRANCISCO

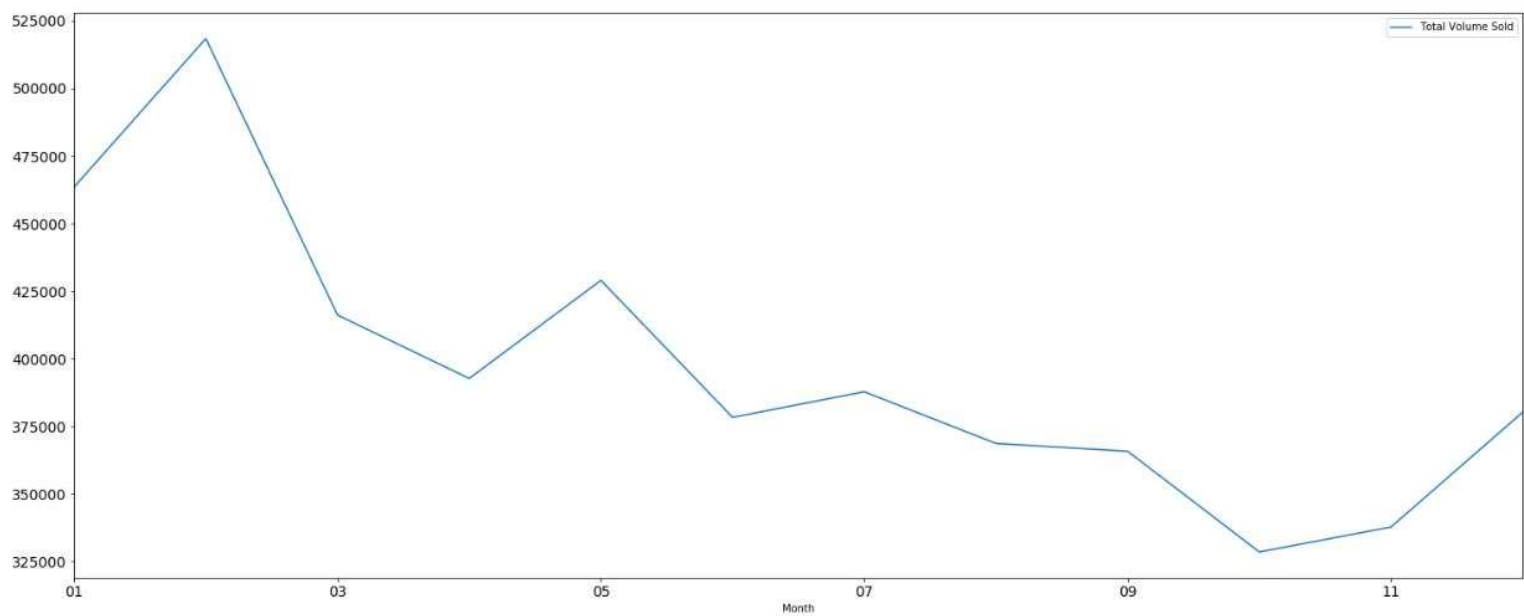


TOTAL VOLUME SOLD IN DIFFERENT MONTHS (2015-2018)



AVERAGE PRICE OF AVOCADOS DURING DIFFERENT MONTHS (2015-2018)

From this graph we can observe that in San Francisco the average price of organic avocados reaches its maximum in the month of August and minimum near May and for conventional it is maximum in the month of October and minimum in the months of January, February and December.



MEAN OF TOTAL VOLUME SALES DURING DIFFERENT MONTHS (2015-2018)

From this graph we can observe that in San Francisco the sales of avocados are highest in the month of February and lowest in the month of October.

From my analysis I found that in San Francisco the minimum sales occur in the month of October, the average(mean) of 328570 of total avocados sold from 2015 to 2018.

| Total Volume Sold | |
|-------------------|---------------|
| Month | |
| 10 | 328570.412857 |

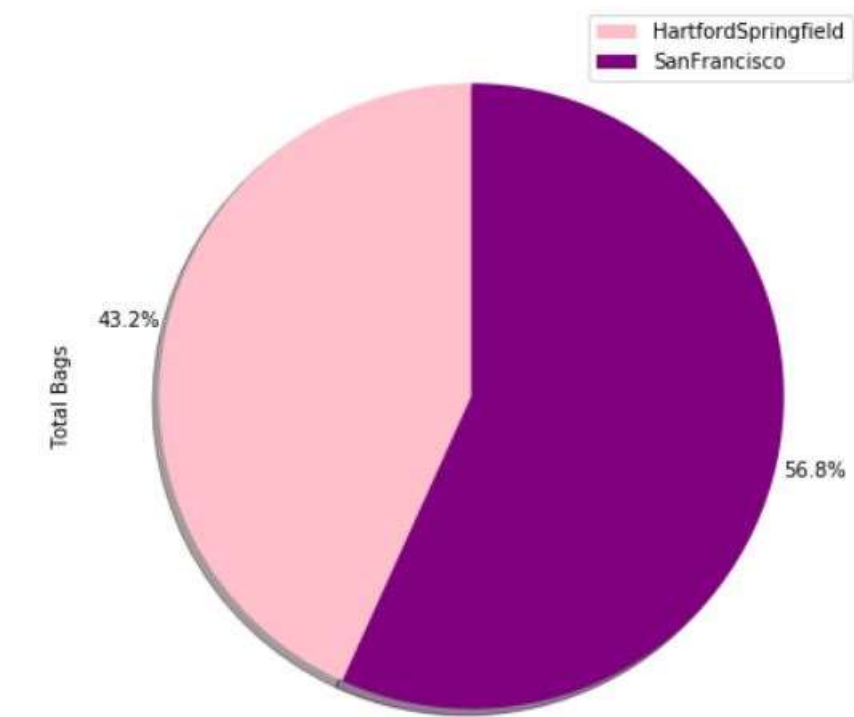
From my analysis I found that in San Francisco the maximum sales occur in the month of February, the average(mean) of 518460 of total avocados sold from 2015 to 2018.

| Total Volume Sold | |
|-------------------|---------------|
| Month | |
| 02 | 518460.020625 |

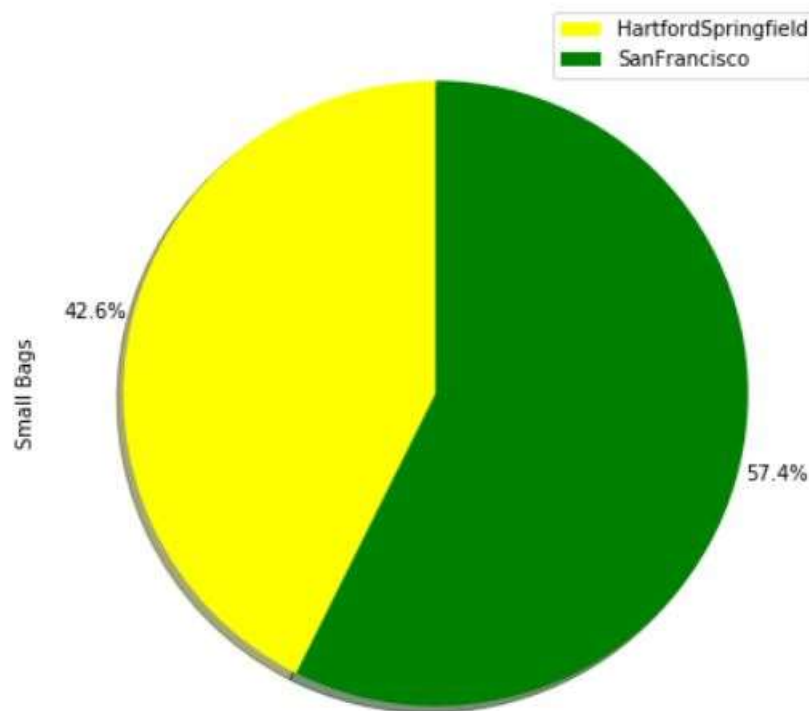
PIE CHARTS FOR THE SALES OF DIFFERENT SIZES OF BAGS OF AVOCADOS IN HARTFORD SPRINGFIELD AND SAN FRANCISCO

| | Total Bags | Small Bags | Large Bags | XLarge Bags |
|---------------------|-------------|-------------|------------|-------------|
| region | | | | |
| HartfordSpringfield | 11333012.63 | 10778560.68 | 476848.23 | 77601.72 |
| SanFrancisco | 14890924.97 | 14502351.14 | 176939.24 | 211634.59 |

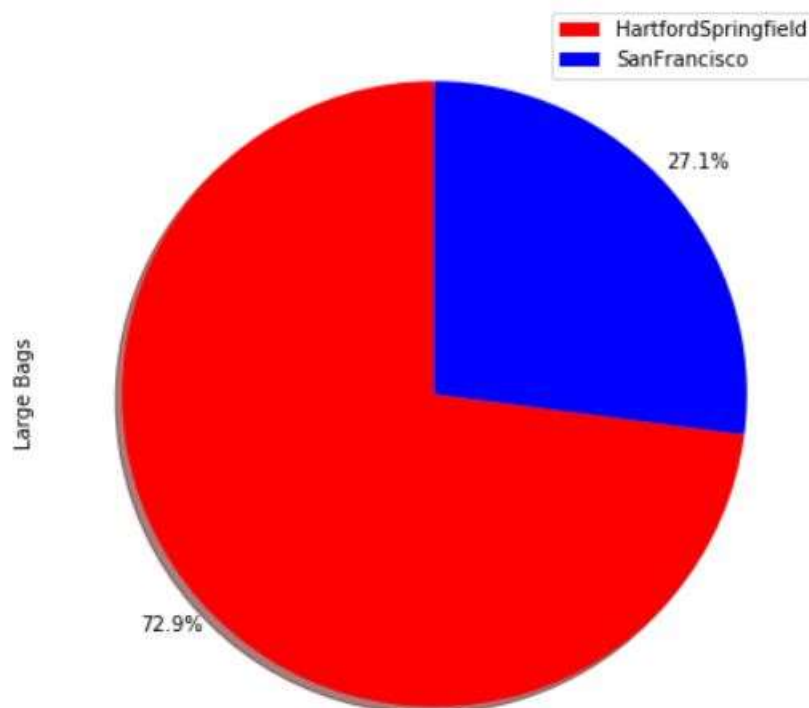
The above data shows the sales of different types of bags in the two regions from 2015 to 2018.



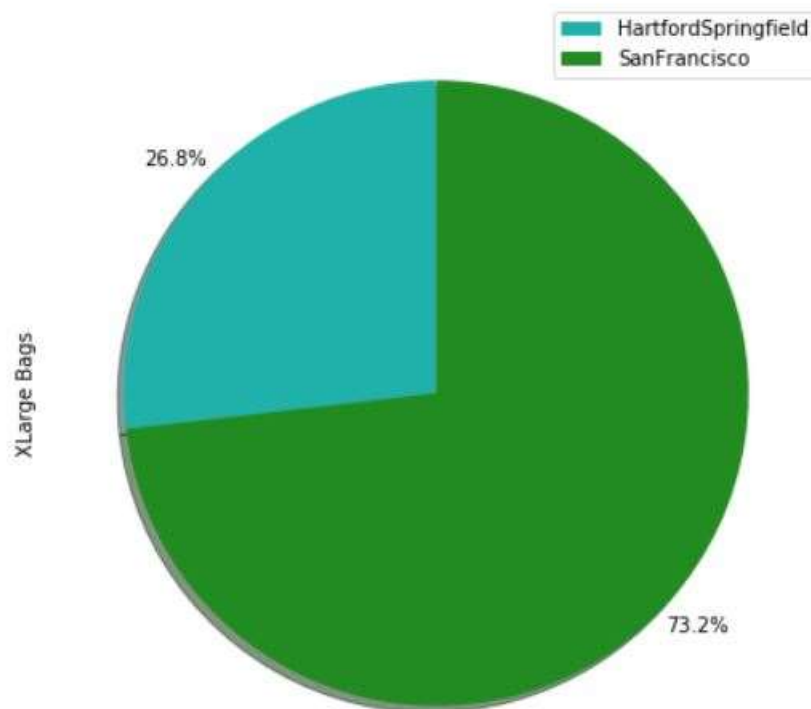
-From this pie chart we can conclude that in San Francisco there were more total number of bags throughout the timeline.



-From this pie chart we can conclude that in San Francisco there were more small bags throughout the timeline.



From this graph we can conclude that in Hartford Springfield there were more Large Bags throughout the timeline.

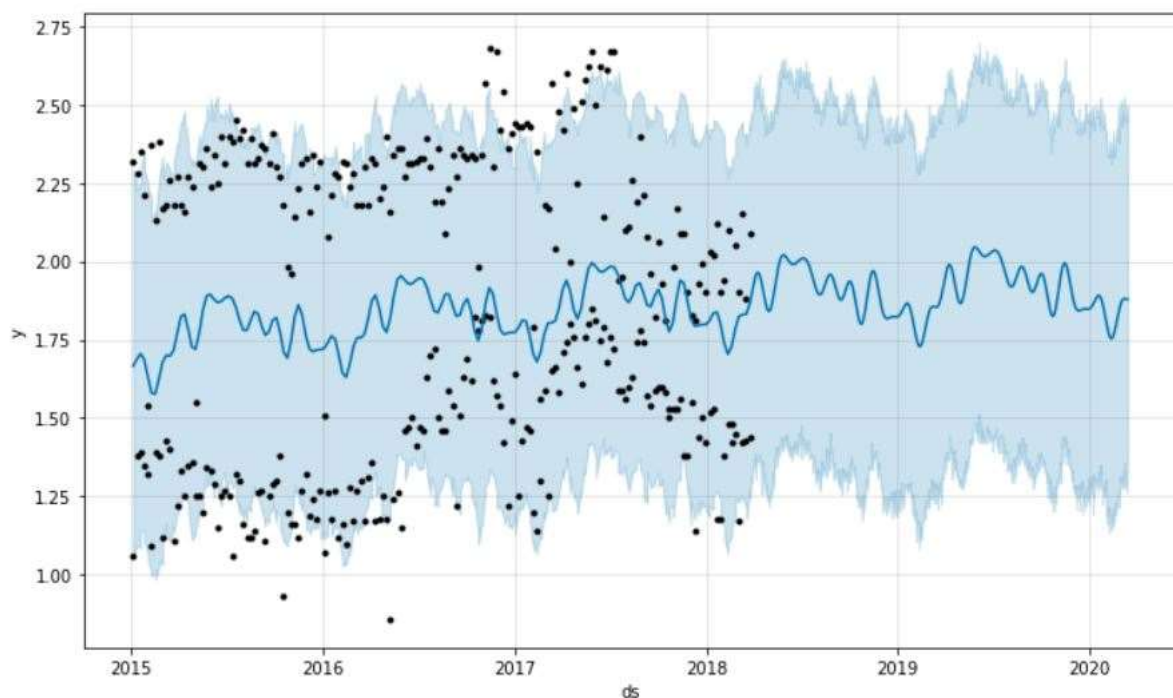


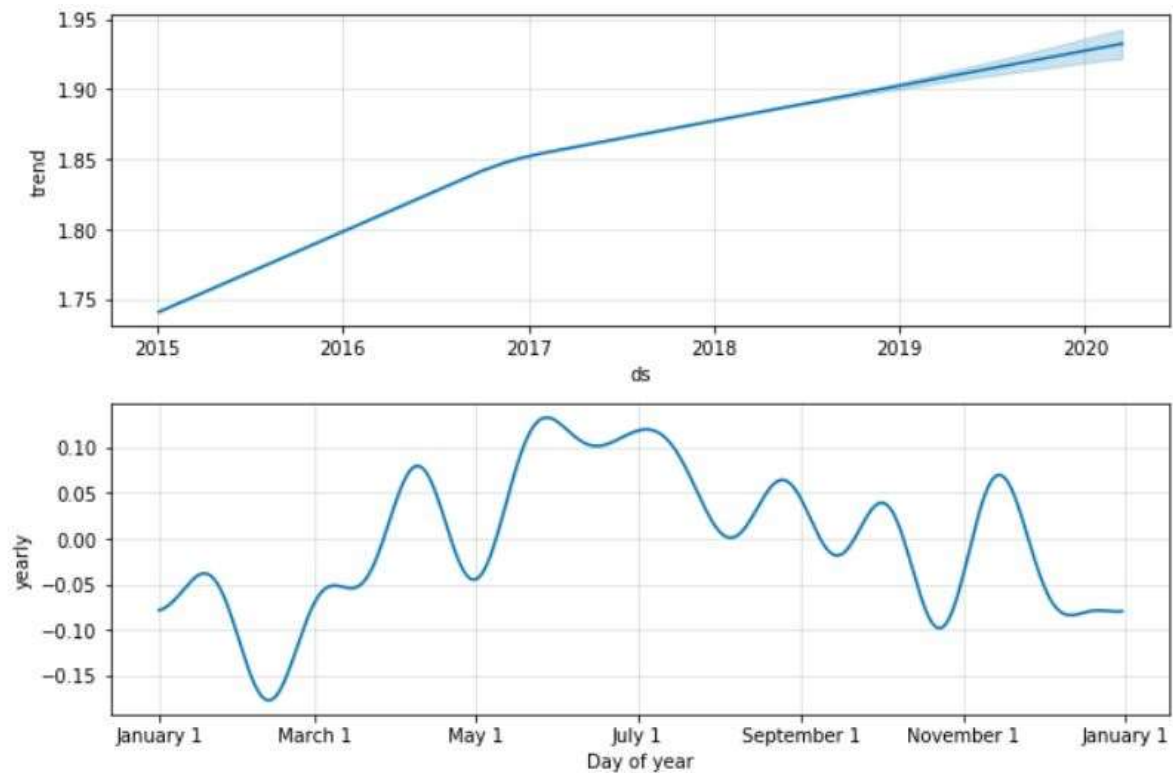
From this graph we can conclude that in San Francisco there were more XLarge bags throughout the timeline.

AVERAGE PRICE PREDICTIONS

In my final part of analysis, I have predicted the average prices for next two years near about for next 720 days.

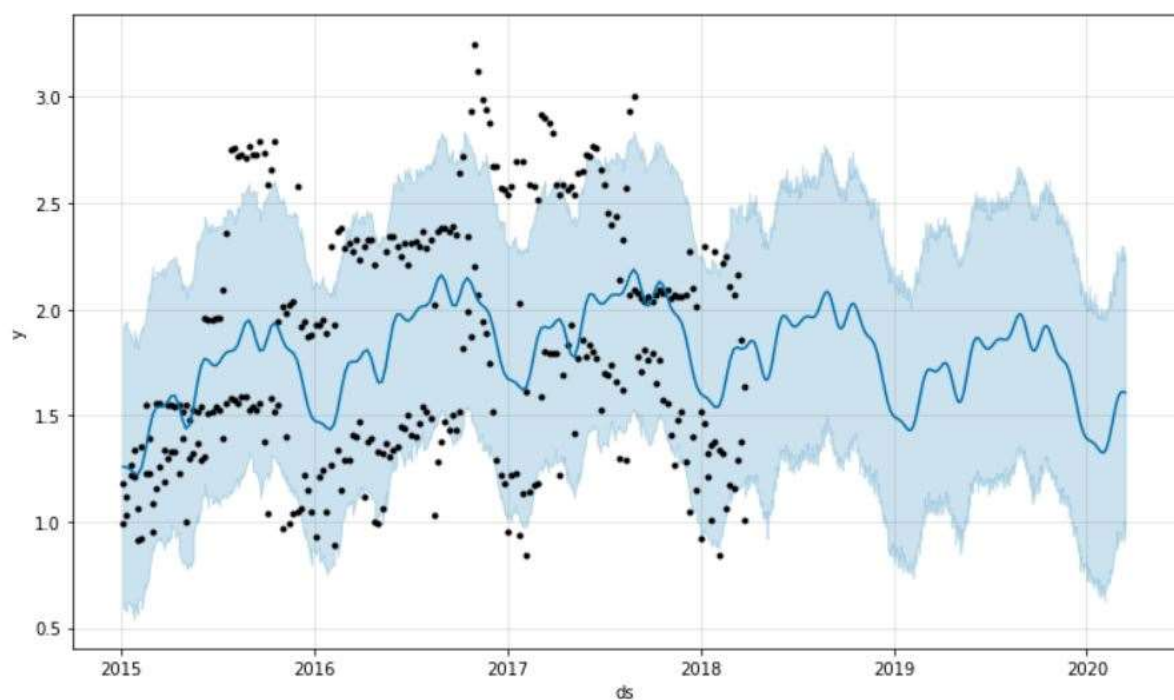
Hartford Springfield

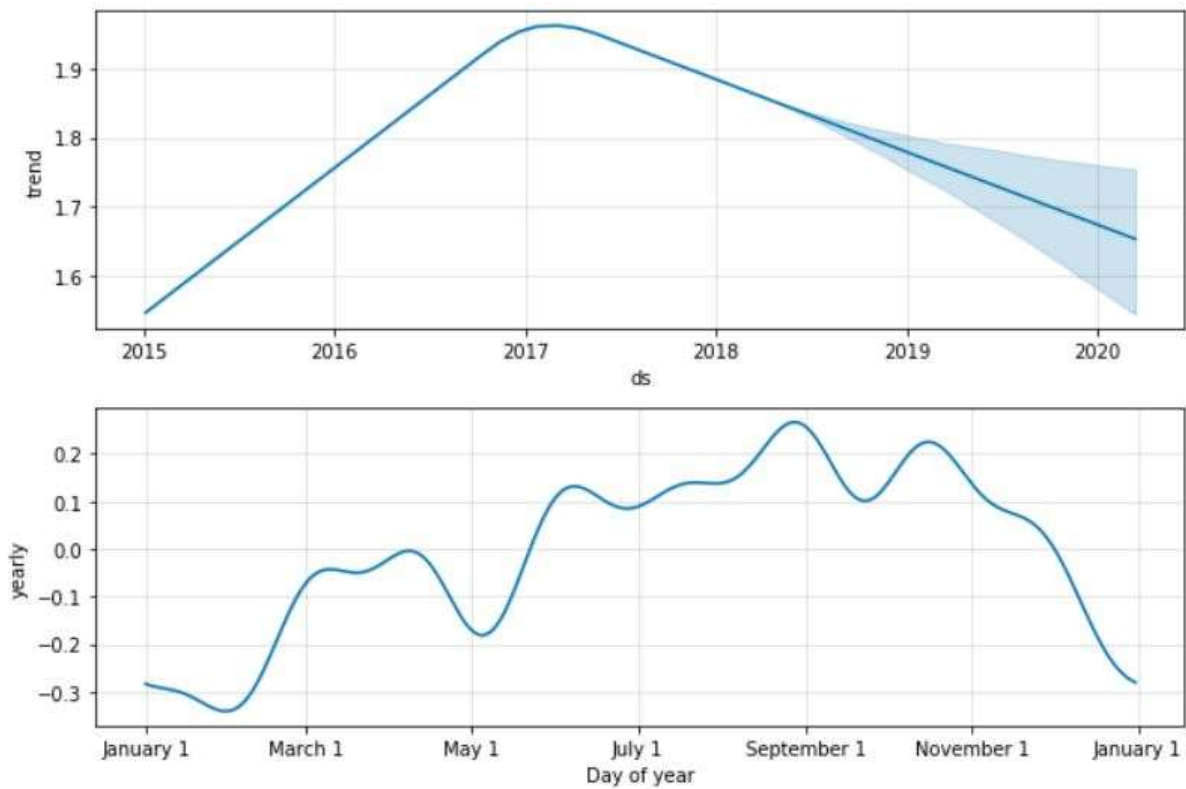




The first graph shows the predicted average prices of avocados with upper and lower bounds of prediction and the black dots representing actual values. The next two graphs show trends and yearly variations. Here we can see how components of the model affect our predictions.

San Francisco





The first graph shows the predicted average prices of avocados with upper and lower bounds of prediction and the black dots representing actual values. The next two graphs show trends and yearly variations. Here we can see how components of the model affect our predictions.