Project

AVOCADO SALES AND PRICE ANALYSIS – KAGGLE DATA

Team Members

- Seongkyoung Ryu (0725164)
- Yash Mecwan (0734905)
- Ronak Prajapati (0734911)
- Zerong Yang (0734120)
- Yi ZHou (0730368)
- Vanita Patel (0734890)

Data Overview

- Avocado sales data was downloaded from the Hass Avocado Board website in May 2018.
- The data represents daily retail scan data from Jan 2015 to Mar 2018 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 data 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.

Data Profiling

Columns in the Dataset

- **Date** – the date of the observation
- **Average Price** – the average price of the single avocado
- **Total Volume** – total number of Avocado sold
- **Type** – conventional or organic
- **Year** – the year of the date
- **Region** – the city or region of the Avocado sold
- **4046** – total number of avocados sold with PLU 4046 (small hass)
- **4225** – total number of avocados sold with PLU 4225 (large hass)
- **4770** – total number of avocados sold with PLU 4770 (extra large hass)
- **Total Bags** – total number of units Avocado sold in bags
- **Small Bags** - total number of units Avocado sold in small bags
- **Large Bags** - total number of units Avocado sold in large bags
- **Xlarge Bags** - total number of units Avocado sold in extra large bags

Data Cleaning

- Original data separated into three datasets: City, Region, Total US
- Created two more datasets for mapping City and Region to Tableau Map

Original Data

A	B	C	D	E	F	G	H	I	J	K	L	M	N
	Date	AveragePr	Total Volu	4046	4225	4770	Total Bags	Small Bags	Large Bags	XLarge Ba	type	year	region
0	27-12-2015	1.33	64236.62	1036.74	54454.85	48.16	8696.87	8603.62	93.25	0	conventio	2015	Albany
1	20-12-2015	1.35	54876.98	674.28	44638.81	58.33	9505.56	9408.07	97.49	0	conventio	2015	Albany
2	13-12-2015	0.93	118220.2	794.7	109149.7	130.5	8145.35	8042.21	103.14	0	conventio	2015	Albany
3	06-12-2015	1.08	78992.15	1132	71976.41	72.58	5811.16	5677.4	133.76	0	conventio	2015	Albany
4	29-11-2015	1.28	51039.6	941.48	43838.39	75.78	6183.95	5986.26	197.69	0	conventio	2015	Albany
5	22-11-2015	1.26	55979.78	1184.27	48067.99	43.61	6683.91	6556.47	127.44	0	conventio	2015	Albany
6	15-11-2015	0.99	83453.76	1368.92	73672.72	93.26	8318.86	8196.81	122.05	0	conventio	2015	Albany
7	08-11-2015	0.98	109428.3	703.75	101815.4	80	6829.22	6266.85	562.37	0	conventio	2015	Albany
8	01-11-2015	1.02	99811.42	1022.15	87315.57	85.34	11388.36	11104.53	283.83	0	conventio	2015	Albany
9	25-10-2015	1.07	74338.76	842.4	64757.44	113	8625.92	8061.47	564.45	0	conventio	2015	Albany
10	18-10-2015	1.12	84843.44	924.86	75595.85	117.07	8205.66	7877.86	327.8	0	conventio	2015	Albany
11	11-10-2015	1.28	64489.17	1582.03	52677.92	105.32	10123.9	9866.27	257.63	0	conventio	2015	Albany
12	04-10-2015	1.31	61007.1	2268.32	49880.67	101.36	8756.75	8379.98	376.77	0	conventio	2015	Albany
13	27-09-2015	0.99	106803.4	1204.88	99409.21	154.84	6034.46	5888.87	145.59	0	conventio	2015	Albany
14	20-09-2015	1.33	69759.01	1028.03	59313.12	150.5	9267.36	8489.1	778.26	0	conventio	2015	Albany

Data Cleaning

Cleaned Data

1. Avocado Sales by City

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Date	AveragePr	Total Volu	4046	4225	4770	Total Bags	Small Bags	Large Bags	XLarge Bag	type	year	city
2	27-12-2015	1.33	64236.62	1036.74	54454.85	48.16	8696.87	8603.62	93.25	0	conventio	2015	Albany
3	20-12-2015	1.35	54876.98	674.28	44638.81	58.33	9505.56	9408.07	97.49	0	conventio	2015	Albany
4	13-12-2015	0.93	118220.2	794.7	109149.7	130.5	8145.35	8042.21	103.14	0	conventio	2015	Albany
5	06-12-2015	1.08	78992.15	1132	71976.41	72.58	5811.16	5677.4	133.76	0	conventio	2015	Albany

2. Avocado Sales by Region

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Date	AveragePr	Total Volu	4046	4225	4770	Total Bags	Small Bags	Large Bags	XLarge Bag	type	year	region
2	27-12-2015	0.9	5040365	1833947	1760956	232755.9	1212707	1090140	110737.4	11829.59	conventio	2015	California
3	20-12-2015	0.94	4695737	1676601	1543281	266689.8	1209165	1061704	136747.1	10714.52	conventio	2015	California
4	13-12-2015	0.87	5259354	1806690	1627241	232985.1	1592438	1404012	180150.4	8275.96	conventio	2015	California
5	06-12-2015	0.78	5775536	1943066	2100246	221957.3	1510267	1376641	126664.4	6962.06	conventio	2015	California

3. Avocado Sales by US Total

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Date	AveragePr	Total Volu	4046	4225	4770	Total Bags	Small Bags	Large Bags	XLarge Bag	type	year	category
2	27-12-2015	0.95	27297984	9626901	10197890	1184340	6288852	4850404	1252626	185822	conventio	2015	TotalUS
3	20-12-2015	0.98	25083647	8710022	9329862	1201020	5842744	4618390	1025049	199305.1	conventio	2015	TotalUS
4	13-12-2015	0.93	28041335	9855054	10805839	1016163	6364280	4964462	1371440	28377.23	conventio	2015	TotalUS
5	06-12-2015	0.89	28800397	9405464	12160839	931830.6	6302263	5005077	1233956	63229.39	conventio	2015	TotalUS
6	29-11-2015	0.99	22617999	8094804	9003178	731008.4	4789009	3901953	856560.3	30495.62	conventio	2015	TotalUS

4. Region mapping data

	A	B
1	region	state
2	California	CA
3	GreatLakes	IL
4	GreatLakes	IN
5	GreatLakes	MI

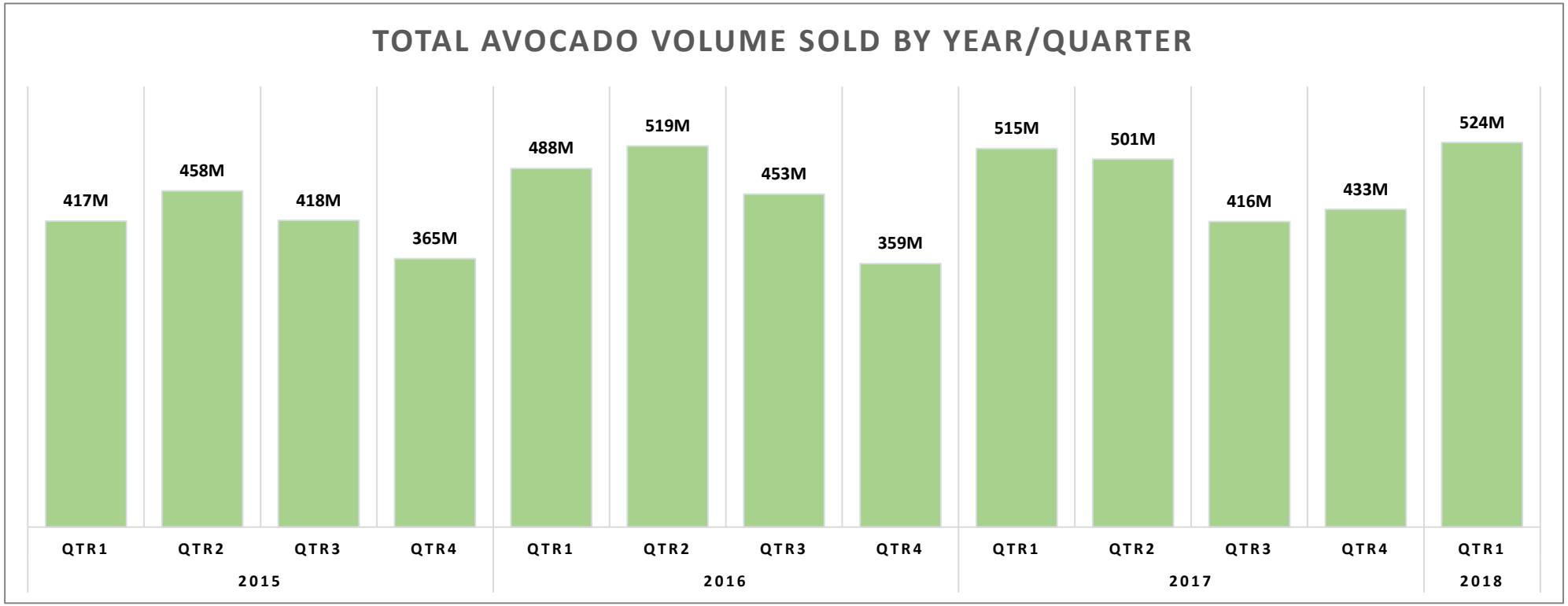
5. City mapping data

	A	B	C
1	region	state	city
2	Albany	New York	Albany
3	Atlanta	Georgia	Atlanta
4	BaltimoreWashington	Maryland	Baltimore
5	Boise	Idaho	Boise

Which is the peak(highest) period by quarter and a year?

Descriptive Analysis (Excel)

Row Labels	Sum of Total Volume
2015	
Qtr1	417M
Qtr2	458M
Qtr3	418M
Qtr4	365M
2016	
Qtr1	488M
Qtr2	519M
Qtr3	453M
Qtr4	359M
2017	
Qtr1	515M
Qtr2	501M
Qtr3	416M
Qtr4	433M
2018	
Qtr1	524M
Grand Total	5865M



The peak quarter for all the three years can be seen in Quarter2 in 2016 with total volume 519M. The trend can be seen that the sales are increasing in Quarter1 and Quarter2 each year and are gradually decreasing in Quarter3 and Quarter4. There is another pattern observable from the given data that each year for every quarter the sales are incremental as compared to same quarter of previous year on average of 50M.

Which is the peak(highest) period by quarter and a year?

SQL Queries (SQL Server Management Studio)

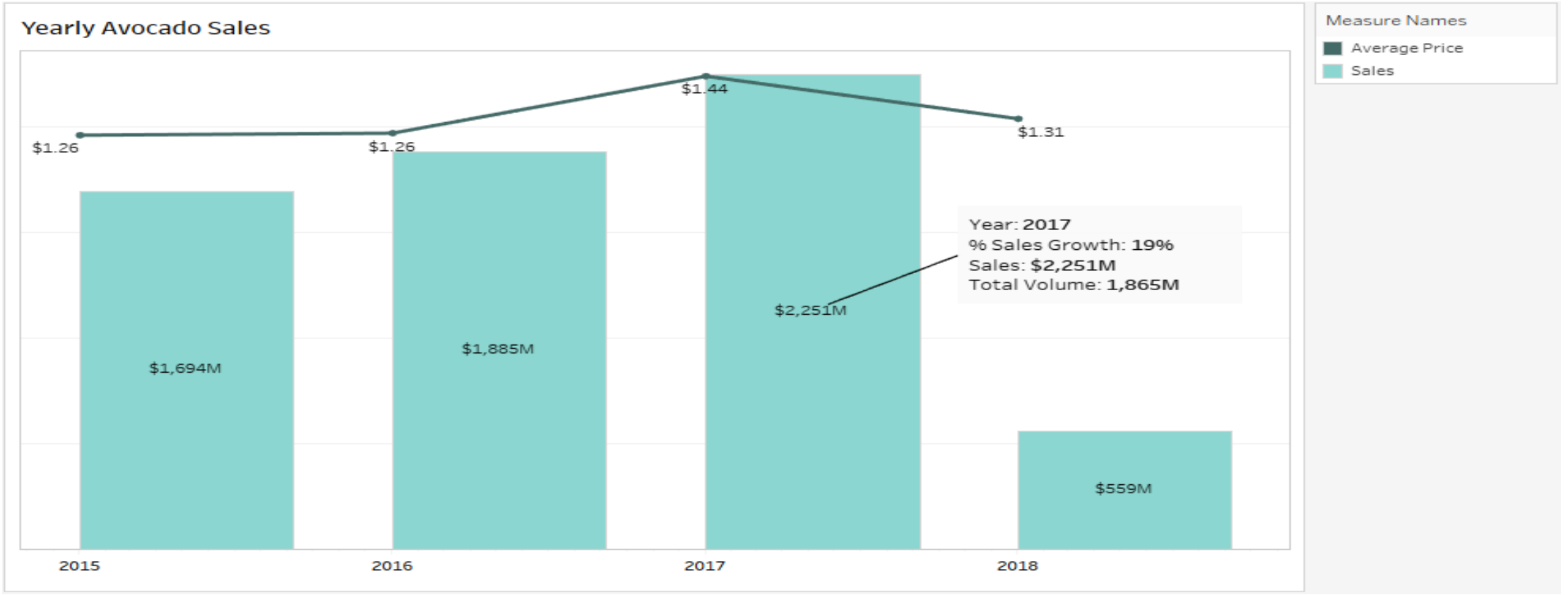
```
SELECT
	Year,
	Quarter,
	sum(Volume)  AS Total_Volume
FROM
	(
	Select
		avocado_total.year AS Year,
		avocado_total.[Total Volume] AS Volume,
		DATEPART(QQ,avocado_total.Date) AS Quarter
	FROM
		avocado_total
	) Total
GROUP BY
	[year],
	[Quarter]
```

Results | Messages

	Year	Quarter	Total_Volume
1	2015	1	396129185.8125
2	2016	1	513411875.4375
3	2017	1	467076642.75
4	2018	1	292547143.875
5	2015	2	470879703.03125
6	2016	2	448314925.3125
7	2017	2	505837030
8	2018	2	106265978.125
9	2015	3	411633280.25
10	2016	3	432537758
11	2017	3	437114868
12	2018	3	38220488.75
13	2015	4	378613052.34375
14	2016	4	424894570.8125
15	2017	4	454665015.3125
16	2018	4	86598647.375

Which is the peak(highest) period by quarter and a year?

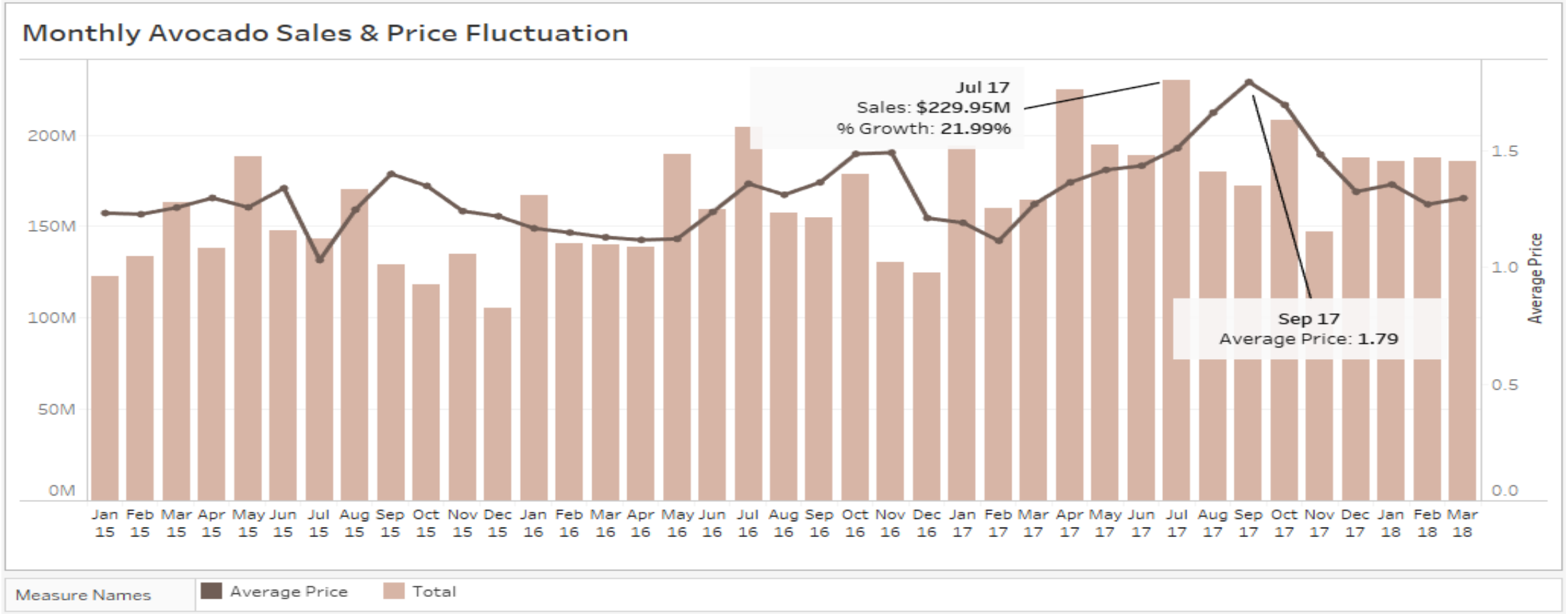
Data Visualization (Tableau)



The volume can be seen increasing every year though the growth in 2016 was by 11% and that of 2017 was 20%. Yet, the price seems disruptive ranging from $1.26 to $1.44. The average price recorded at $ 1.26 in 2015 and in 2016. It increase to 1.44 In 2018, the average price started from $ 1.26.

Which is the peak(highest) period by quarter and a year?

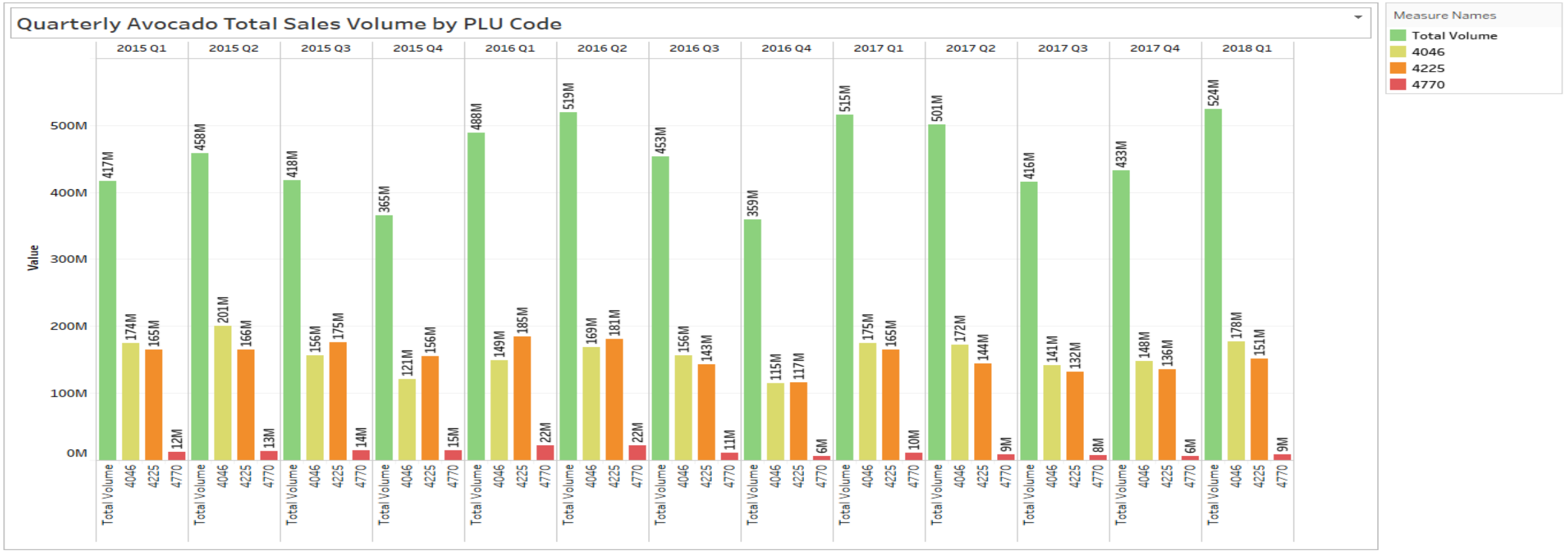
Data Visualization (Tableau)



The monthly avocado sales seem completely volatile in nature. The prices are quite stable from Jan 2015 to Jun 2016 (except June 2015) post that period it shows fluctuation as well. The highest sales were registered in July 2017 and April 2017 of $223M with 21% growth from previous month but the highest price was registered in September 2017 of $1.79.

Which is the peak(highest) period by quarter and a year?

Data Visualization (Tableau)

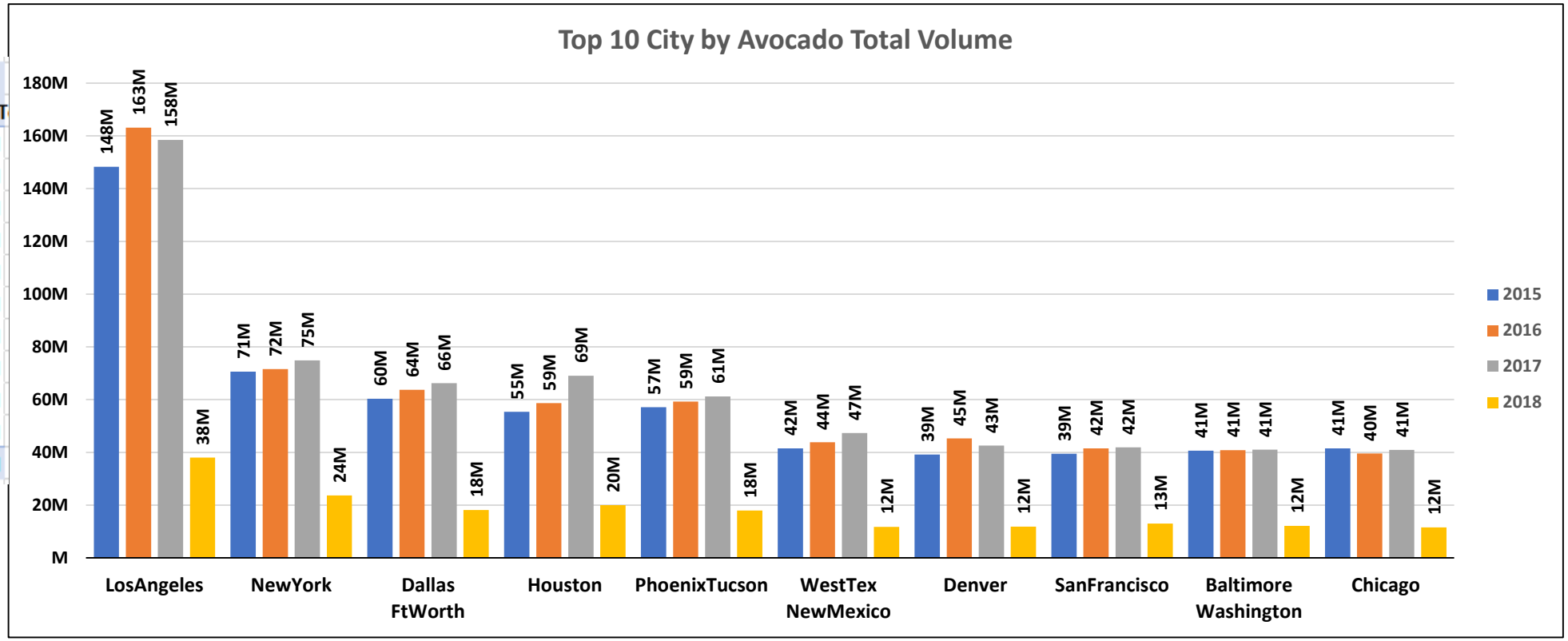


Here in bar graph we can see sales volume by PLU code which are indicating size of avocados: 4046(small/medium avocado),4225(large avocado),4770(extra large avocado). Among total volume of Avocado sold, average 49% from total were sold by PLU level 4046 and 4225 and same pattern is found in all quarter. 4773 PLU code only contribute 2%.

Best market for Avocado sale

Descriptive Analysis (Excel)

Sum of Total Volu	Colur				
Row Labels	2015	2016	2017	2018	Grand T
LosAngeles	148M	163M	158M	38M	508M
NewYork	71M	72M	75M	24M	241M
DallasFtWorth	60M	64M	66M	18M	208M
Houston	55M	59M	69M	20M	203M
PhoenixTucson	57M	59M	61M	18M	196M
WestTexNewMexi	42M	44M	47M	12M	145M
Denver	39M	45M	43M	12M	139M
SanFrancisco	39M	42M	42M	13M	136M
BaltimoreWashing	41M	41M	41M	12M	135M
Chicago	41M	40M	41M	12M	134M
Grand Total	594M	628M	644M	178M	2044M



The chart shows the top 10 best markets for avocado sales. The sales of LosAngeles are incomparably higher then the other cities. Among the other cities the sales are quite stable with very less fluctuation in amounts. The sales in each city are stable for every year.

Best market for Avocado sale

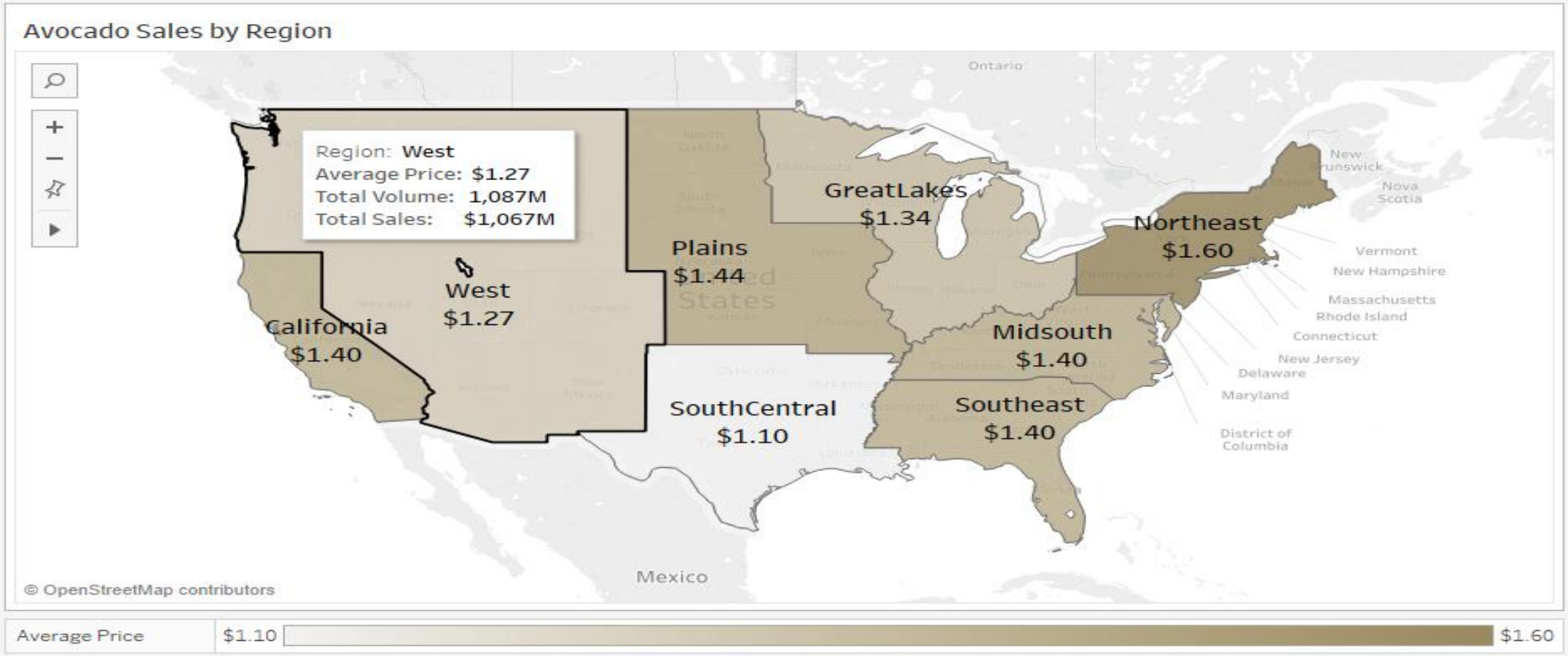
SQL Queries (SQL Server Management Studio)

```
SELECT Top 10
	region AS City,
	year,
	sum([Total Volume])  AS Total_Volume
FROM
	avocado_city
	Group BY
	region,
	year
order by
	Total_Volume DESC
```

1	LosAngeles	2016	163142302.589844
2	LosAngeles	2017	158481174.789063
3	LosAngeles	2015	148220693.265625
4	NewYork	2017	74850673.5625
5	NewYork	2016	71546562.9951172
6	NewYork	2015	70607636.0771484
7	Houston	2017	69093092.7763672
8	DallasFtWorth	2017	66240982.7529297
9	DallasFtWorth	2016	63695249.8535156
10	PhoenixTucson	2017	61170534.1743164

Best market for Avocado sale

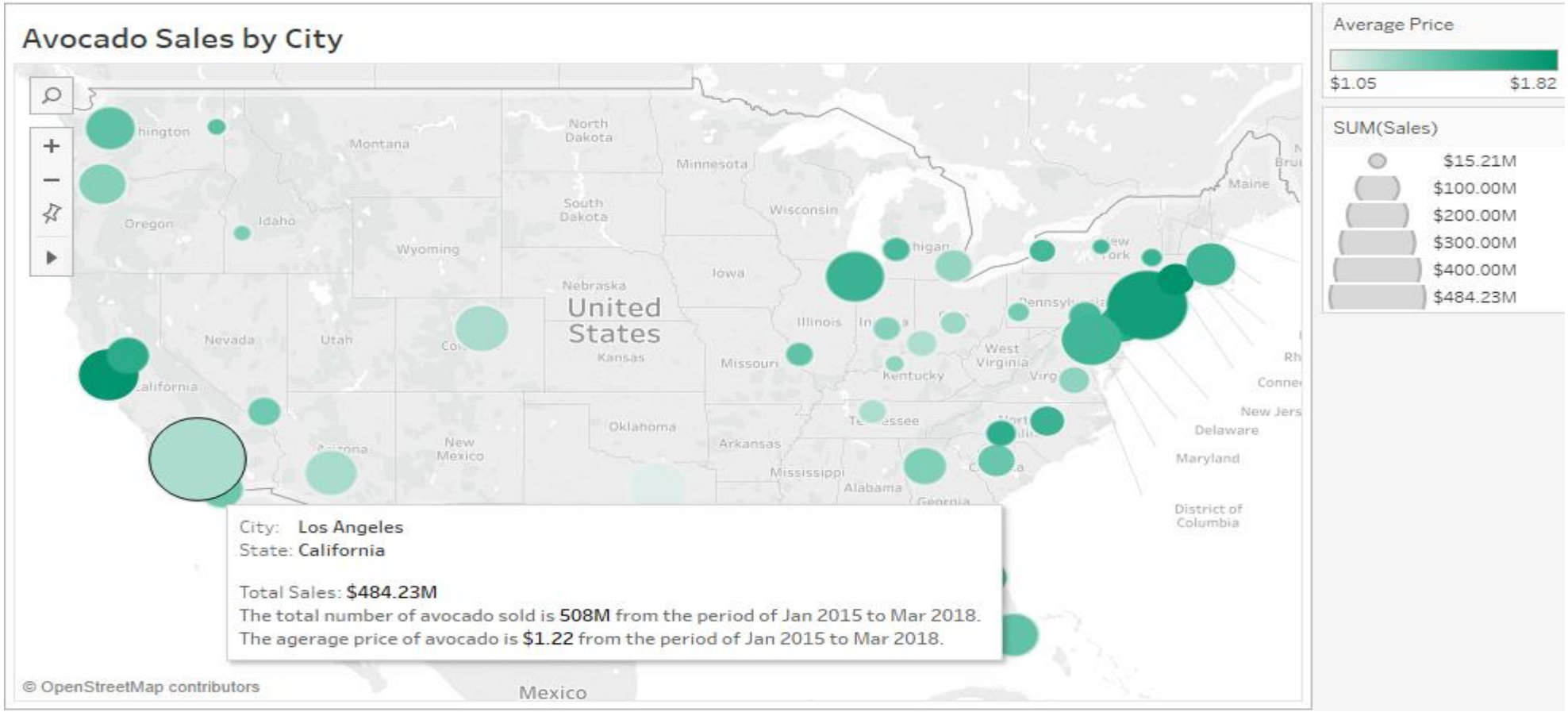
Data Visualization (Tableau)



When compared regionally, the sales of the West region amounts to the highest of $10,67M followed by NorthEast region with $960M. The price in SouthCentral region is the lowest that is $1.10 and the highest is NorthEast region with price of $1.60 . The prices of all the other regions range in between.

Best market for Avocado sale

Data Visualization (Tableau)

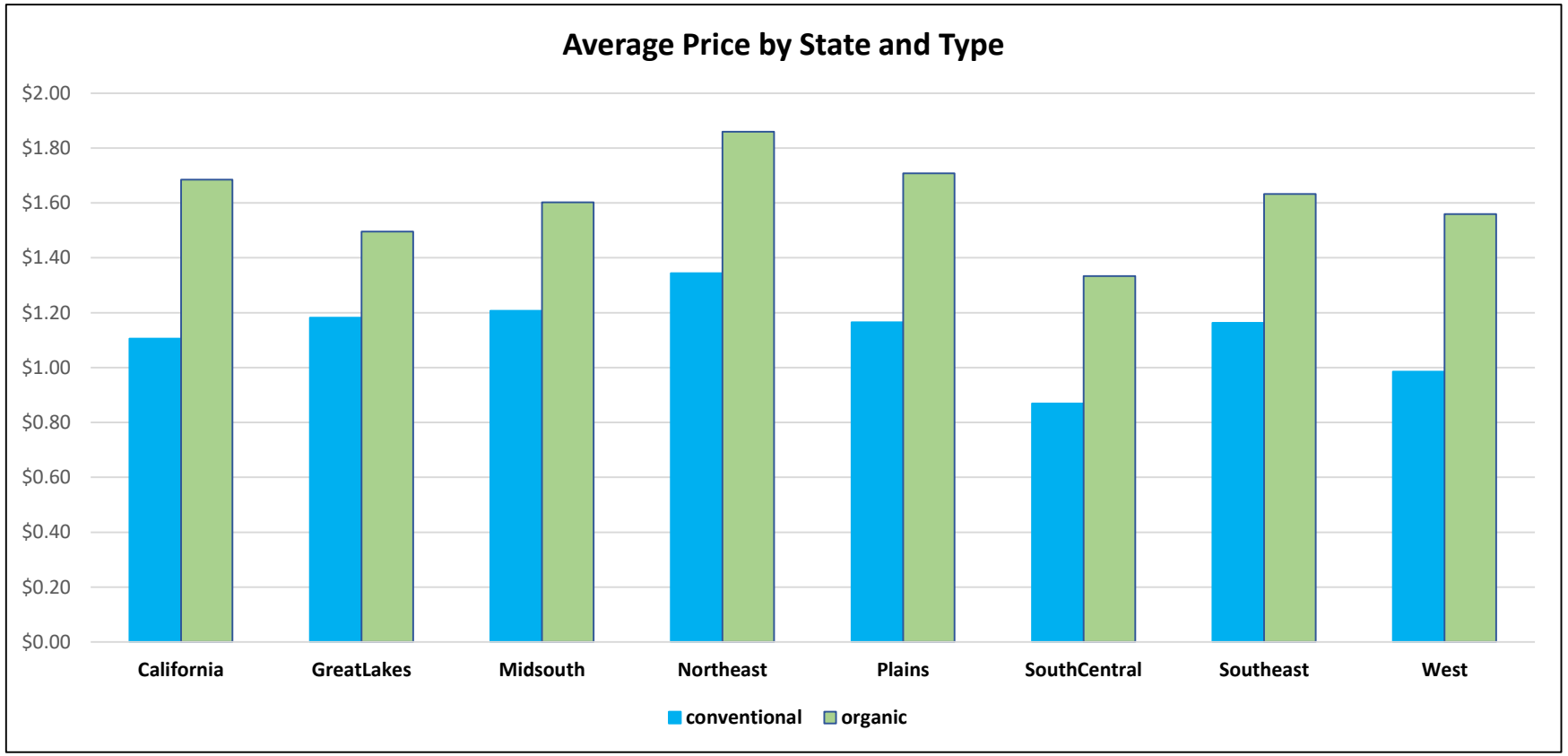


The darker shade of colour shows cities with the highest prices of Avocado. The highest prices are noted in HartFord, San Francisco and New York of $1.82, $1.80 and $1.73 respectively. The size of the circle shows the amount of sales. The highest sales are notes in Los Angeles, New York and Baltimore amounting to $485M, $335M and $180M respectively.

What is the difference of price between organic and conventional avocado?

Descriptive Analysis (Excel)

Average of AveragePrice region	type conventional	organic
California	$1.11	$1.69
GreatLakes	$1.18	$1.50
Midsouth	$1.21	$1.60
Northeast	$1.34	$1.86
Plains	$1.17	$1.71
SouthCentral	$0.87	$1.33
Southeast	$1.16	$1.63
West	$0.99	$1.56
Grand Total	**$1.13**	**$1.61**



The chart shows how the average price of Avocado by state and type. As compared to conventional Avocado the price for organic avocados is higher. The price range for conventional Avocado is $0.87 to $1.34 and that of organic Avocado is $1.33 to $1.86 for all the regions. As

What is the difference of price between organic and conventional avocado?

SQL Queries (SQL Server Management Studio)

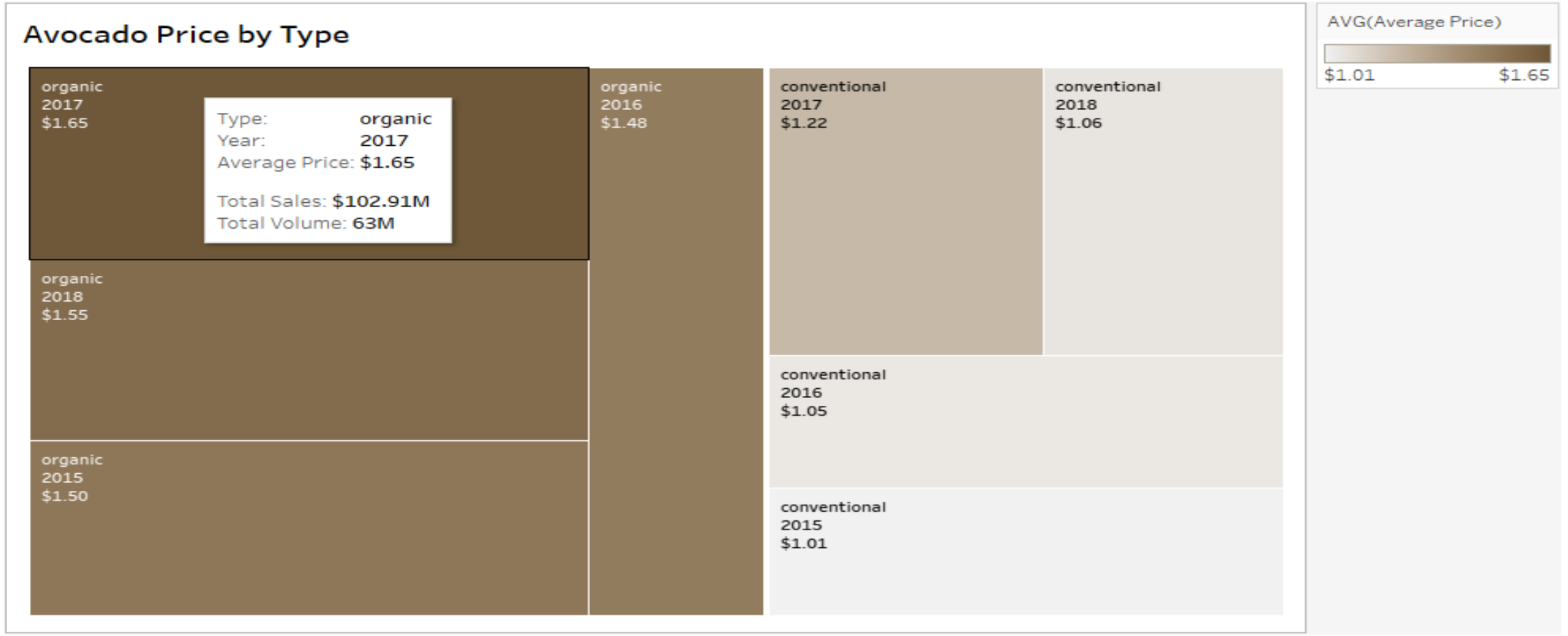
```
SELECT
	region AS Region,
	AVG(AveragePrice) AS Average_Price,
	type AS Avacado_Type
FROM
	avocado_state
GROUP BY
	region,
	type
Order BY
	Region ASC
```

Results | Messages

	Region	Average_Price	Avacado_Type
1	California	1.10544378658724	conventional
2	California	1.68520709960418	organic
3	GreatLakes	1.49520710304644	organic
4	GreatLakes	1.18189348941724	conventional
5	Midsouth	1.60236686715008	organic
6	Midsouth	1.20715975867221	conventional
7	Northeast	1.34443787357511	conventional
8	Northeast	1.85940827773168	organic
9	Plains	1.70751478403983	organic
10	Plains	1.16550295522227	conventional
11	SouthCen...	1.33307692073506	organic
12	SouthCen...	0.86940828230254	conventional
13	Southeast	1.16301775190252	conventional
14	Southeast	1.63301775300291	organic
15	West	1.55934911507827	organic
16	West	0.985088753277...	conventional

What is the difference of price between organic and conventional avocado?

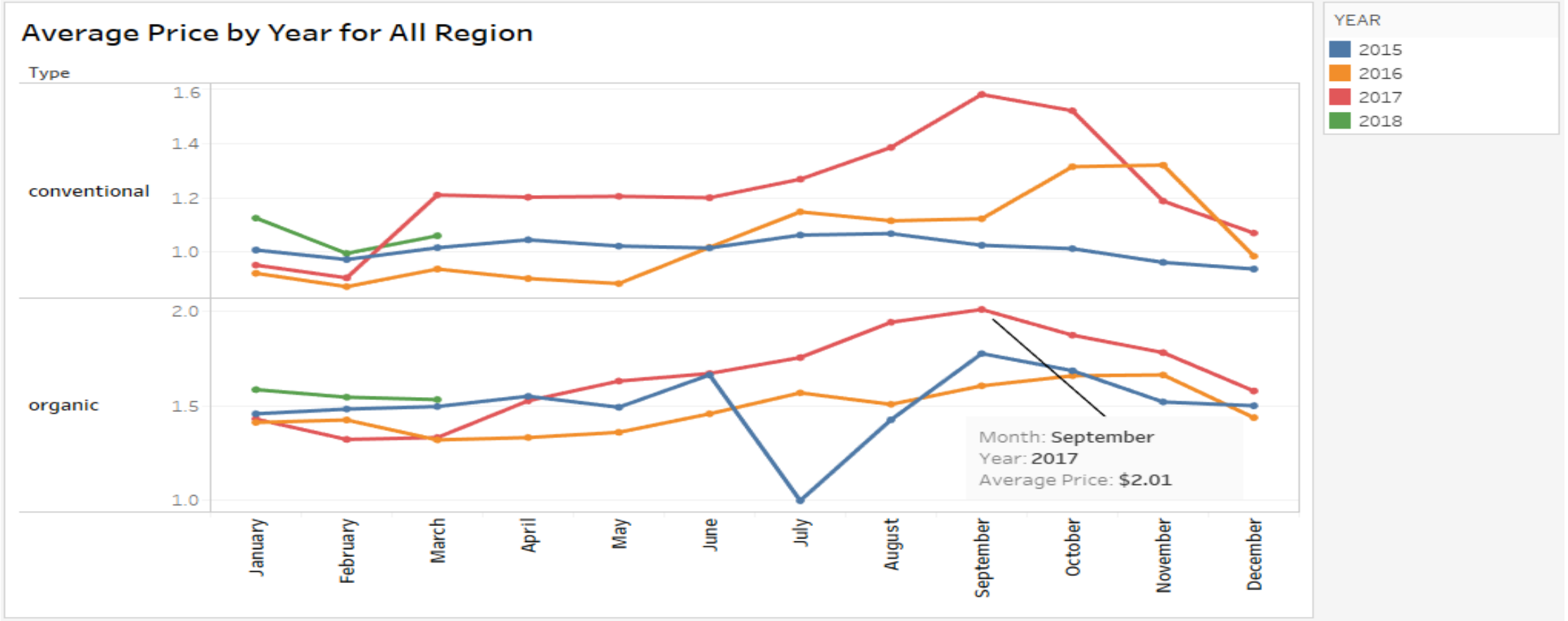
Data Visualization (Tableau)



Above graph type is tree map which shows type of avocado(organic or conventional), year, average price, total sales of Avocado, and volume of Avocado. Even if we observed yearly, the price for organic Avocado is noted higher than the conventional Avocado. The year of 2017 indicate the highest amount of sales and price as compared to the previous year for both organic and conventional Avocado.

What is the difference of price between organic and conventional avocado?

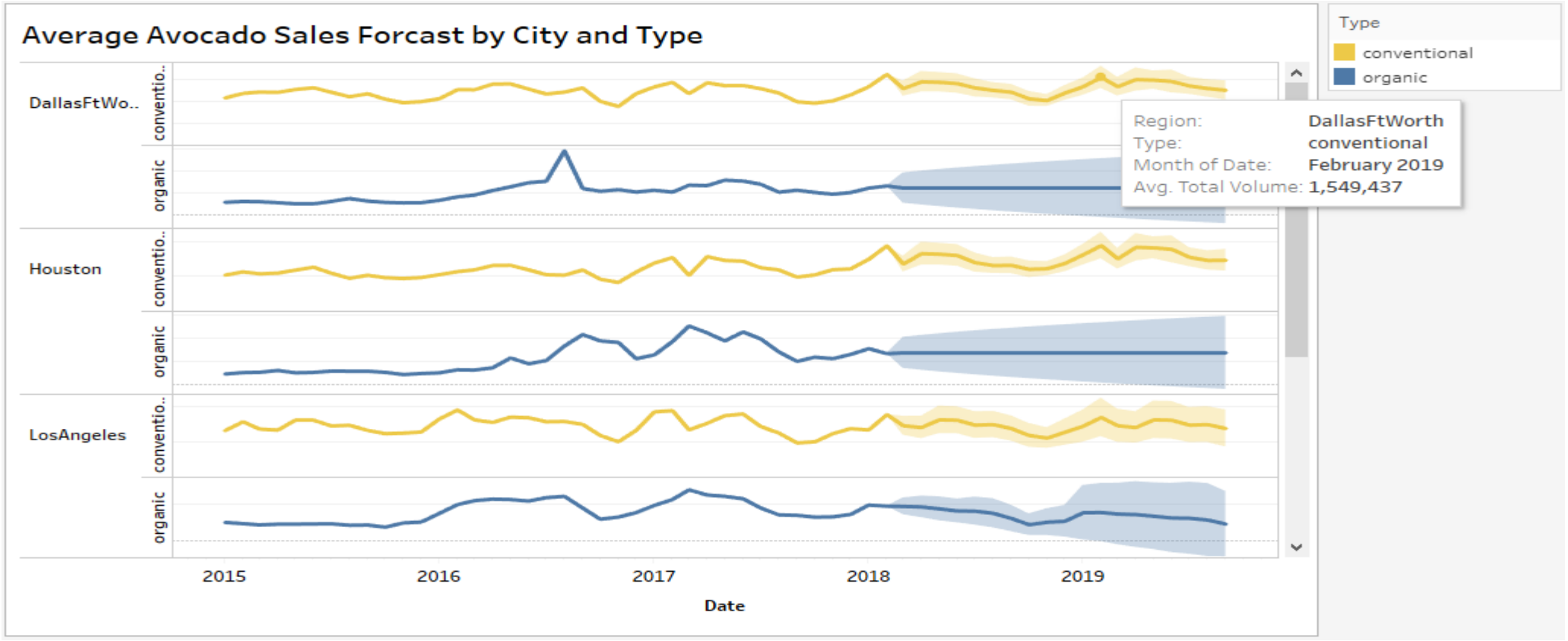
Data Visualization (Tableau)



Above line graph gives information about average price for full year and months as well. The price range for organic ranges from $1.0 to $2.0 and that of conventional is form $0.50 to $1.6. Though the charts are extremely volatile, the year 2015 looks very much stable as compared to 2017. For the year of 2016, it can be said that the prices for organic Avocados were stable and that of conventional Avocados were fluctuating.

What would be the average future Avocado sales?

Data Visualization (Tableau)



The highlighted area shows the future predictions of average volume of Avocado sales for top 5 cities for the period of April 2018 to Sep 2019 based on the previous data obtained(January 2015 to Mar 2018). There is a stable average sales forecast pattern in organic Avocado and fluctuating average sales forecast pattern in conventional Avocado.

Conclusion

Overall this data is providing information about Avocado sold during the period January 2015 to March 2018 for different regions. Along with sales it also shows data with types of Avocado, size of Avocado and bag pack size. It also gives information of average price and total sold volume so all information can be useful for prediction and analysis.

Thank You