NAME: SHINDE Daji Rao Akash Shinde

COLLEGE: MVSR ENGINEERING COLLEGE

Dataset: Avocado Prices

From: Historical data on avocado prices and sales volume in multiple US markets

Context:

It is a well-known fact that Millennials LOVE Avocado Toast. It's also a well-known fact that all Millennials live in their parent's basements.

Clearly, they aren't buying home because they are buying too much Avocado Toast!

But maybe there's hope… if a Millennial could find a city with cheap avocados, they could live out the Millennial American Dream.

Content:

The table below 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. green skins) are not included in this table.

Some relevant columns in the dataset:

- Date - The date of the observation
- Average Price - 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

CONVENTIONAL AVOCADOS

My insights on solving the problem for <u>Conventional Avocados</u>

From Univariate Analysis of Average Price of Conventional Avocados

1) The maximum AveragePrice is 2.22

2) The minimum AveragePrice is 0.46

3)The mean AveragePrice is 1.15

4)Every region has same export for convention avocados

From Bivariate analysis of Conventional Avocados

Here, I calculated the correlation between AveragePrice and different volumes of Avocados (i.e. 4046U,4225U,4770U and Total Volume)

Correlation	Value	Conclusion
AveragePrice and PLU4045	-0.15	No Correlation
AveragePrice and PLU4225	-0.06	No Correlation
AveragePrice and PLU4770	-0.11	No Correlation
AveragePrice and Total Volume	-0.10	No Correlation

Here, I calculated the correlation between AveragePrice and different bags of Avocados (i.e. Small Bags, Large Bags, XLarge Bags and Total Bags)

Correlation	Value	Conclusion
AveragePrice and Small Bags	-0.08	No Correlation
AveragePrice and Large Bags	-0.10	No Correlation
AveragePrice and XLarge Bags	-0.02	No Correlation
AveragePrice and Total Bags	-0.08	No Correlation

Therefore, I can conclude that there is no correlation between Avocado's (AveragePrice Volume) and Avocado's (AveragePrice Bags)

From Bivariate analysis of Average Prices and Regions

1) The region with maximum AveragePrice is Greatlakes (1.40$)
2) The region with maximum AveragePrice is Philadelphia (0.72$)
3) The mean AveragePrice is 1.15$

From Bivariate analysis of Average Prices and years

Year	Average price in $
2015	1.07
2016	1.05
2017	1.29
2018	1.12

NOTE:

The average prices have been almost same every year.

From Bivariate analysis of Volume and years

Year	Total Volume	4046PLU	4225PLU	4770PLU
2015	1.5m	600k	600k	500k
2016	1.6m	500k	570k	550k
2017	1.6m	550k	500k	300k
2018	2m	700k	600k	370k

Note:

The total volumes sold have been increasing every year.

The sale of 4046PLU Avocados has been increasing every year.

The sale of 4225PLU Avocados decreases till 2017 and then starts increasing in 2018.

But, the sale of 4770PLU Avocados have been decreasing every year.

From Bivariate analysis of Bags and years

Year		Small Bags	Large Bags	XLarge Bags
2015		200k	470k	2000
2016		390k	1.25m	7500
2017		400k	1.25m	8000

2018		500k	1.75m	10200

Note:

The sale of small bags has been increasing every year.

The sale of Large bags has been increasing rapidly every year.

The sale of XLarge bags has been increasing but very slowly.

LINEAR REGRESSION:

Since, the dependent variable (i.e. Total Volume) is dependent on volume, bags, region and AveragePrice.

I performed Multiple Linear Regression and got the test score 0.68

ORGANIC AVOCADOS

My insights on solving the problem for Organic Avocados

From Univariate Analysis of Average Price of Organic Avocados

1) The maximum AveragePrice is 3.25

2) The minimum AveragePrice is 0.44

3)The mean AveragePrice is 1.65

4)Every region has same export for convention avocados (i.e.169) but, in Albany it is 168 and in WestNewMexico it is 166.

From Bivariate analysis of Conventional Avocados

Here, I calculated the correlation between AveragePrice and different volumes of Avocados (i.e. 4046U,4225U,4770U and Total Volume)

Correlation	Value	Conclusion
AveragePrice and PLU4045	-0.10	No Correlation
AveragePrice and PLU4225	-0.05	No Correlation
AveragePrice and PLU4770	-0.03	No Correlation
AveragePrice and Total Volume	-0.08	No Correlation

Here, I calculated the correlation between AveragePrice and different bags of Avocados (i.e. Small Bags, Large Bags, XLarge Bags and Total Bags)

Correlation	Value	Conclusion
AveragePrice and Small Bags	-0.04	No Correlation
AveragePrice and Large Bags	-0.15	No Correlation
AveragePrice and XLarge Bags	-0.01	No Correlation
AveragePrice and Total Bags	-0.06	No Correlation

Therefore, I can conclude that there is no correlation between Avocado's (AveragePrice Volume) and Avocado's (AveragePrice Bags)

From Bivariate analysis of Average Prices and Regions

1) The regions with maximum AveragePrice (i.e. 2.22$) are Hat trod Springfield and San Fransisco.
2) The regions with maximum AveragePrice (i.e. 1.27$) are Houston and South Central
3) The mean AveragePrice is 1.65$

From Bivariate analysis of Average Prices and years

Year	Average price in $
2015	1.67
2016	1.56
2017	1.73
2018	1.57

NOTE:

The average price has decreasing and increasing every year.

From Bivariate analysis of Volume and years

Year	Total Volume	4046PLU	4225PLU	4770PLU
2015	300k	8500	13000	300
2016	420k	6100	16000	400
2017	580k	6000	15000	100
2018	700k	6100	17500	50

Note:

The total volumes sold have been increasing every year.

The sale of 4046PLU Avocados has been decreasing every year.

The sale of 4225PLU Avocados increases till 2016, decreases till 2017 and then starts increasing in 2018.

But, the sale of 4770PLU Avocados have been decreasing every year.

From Bivariate analysis of Bags and years

Year		Small Bags	Large Bags	XLarge Bags
2015		5000	3500	0
2016		15000	8000	0
2017		25000	9500	2
2018		40000	9600	8

Note:

The sale of small bags has been increasing every year.

The sale of Large bags has been increasing every year.

The sale of XLarge bags has been increasing but very slowly.

LINEAR REGRESSION:

Since, the dependent variable (i.e. Total Volume) is dependent on volume, bags, region and AveragePrice.

I performed Multiple Linear Regression and got the test score 0.76