

Avocado Price

Team Algoritma

4/5/2020

Contents

The main Problem	1
Example Dataset: Avocado Sales	1
Exploratory Data Analysis	3
Modeling with Regression Model	4
Evaluating Model Prediction	5

The main Problem

price elasticity of demand: Price elasticity is used to understand how supply or demand change given changes in price to understand the pattern of sales. For instance, some goods are very inelastic, that is, their price do not change very much given changes in supply or demand. For example: People need to buy gasoline to get to work, and so if oil price rise, people will likely still to buy just the same amount of gas. On the other hand some goods are very elastic, their price moves cause substantial changes in its demand or its supply. This report will try to formulated the relation of price and quantity in demand of avocado.

Price elasticity is addressing the relation between price and quantity of item sold: if we increase the price, say 1 currency unit, how much decrease in item sold that we should expect?

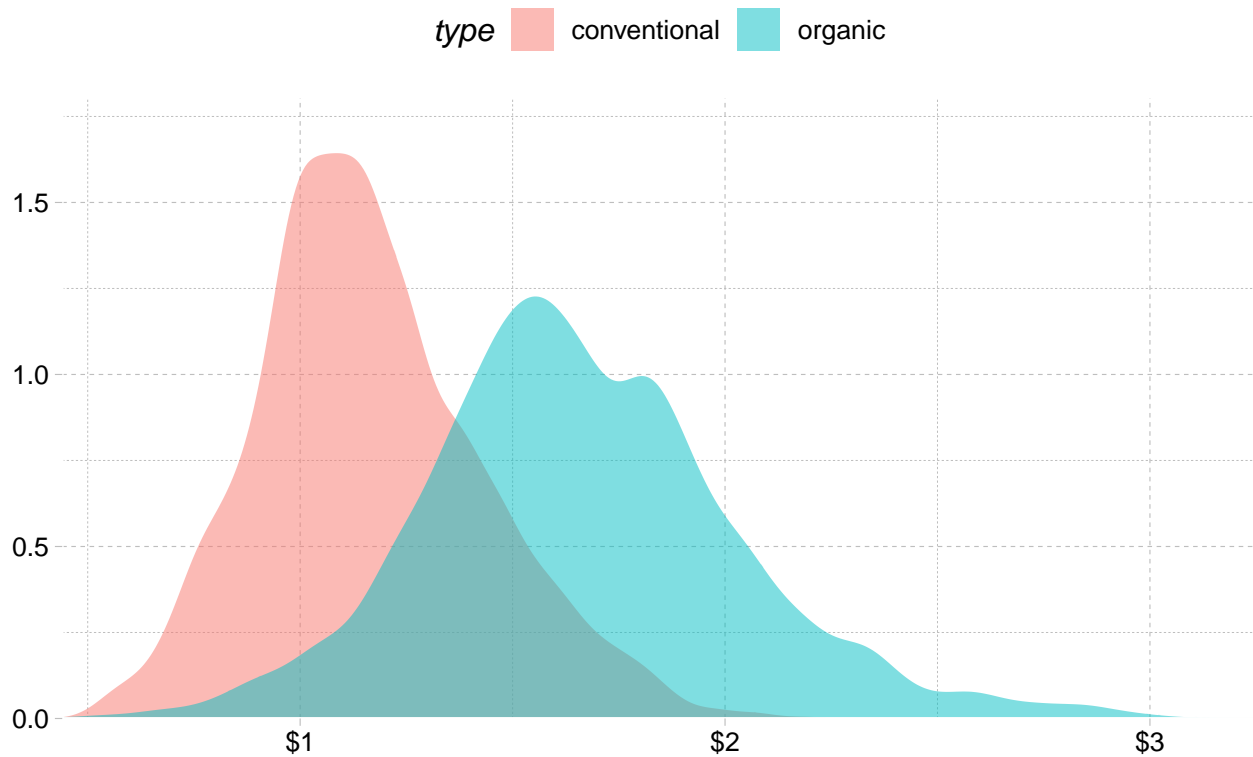
Example Dataset: Avocado Sales

For price elasticity example, we will use avocado sales Dataset. Here are example observation from the data:

region	date	type	price	quantity
Albany	2015-01-04	conventional	1.22	40873.28
Albany	2015-01-04	organic	1.79	1373.95
Albany	2015-01-11	conventional	1.24	41195.08
Albany	2015-01-11	organic	1.77	1182.56
Albany	2015-01-18	conventional	1.17	44511.28
Albany	2015-01-18	organic	1.93	1118.47
Albany	2015-01-25	conventional	1.06	45147.50
Albany	2015-01-25	organic	1.89	1115.89
Albany	2015-02-01	conventional	0.99	70873.60
Albany	2015-02-01	organic	1.83	1228.51

Distribution of Avocado Price

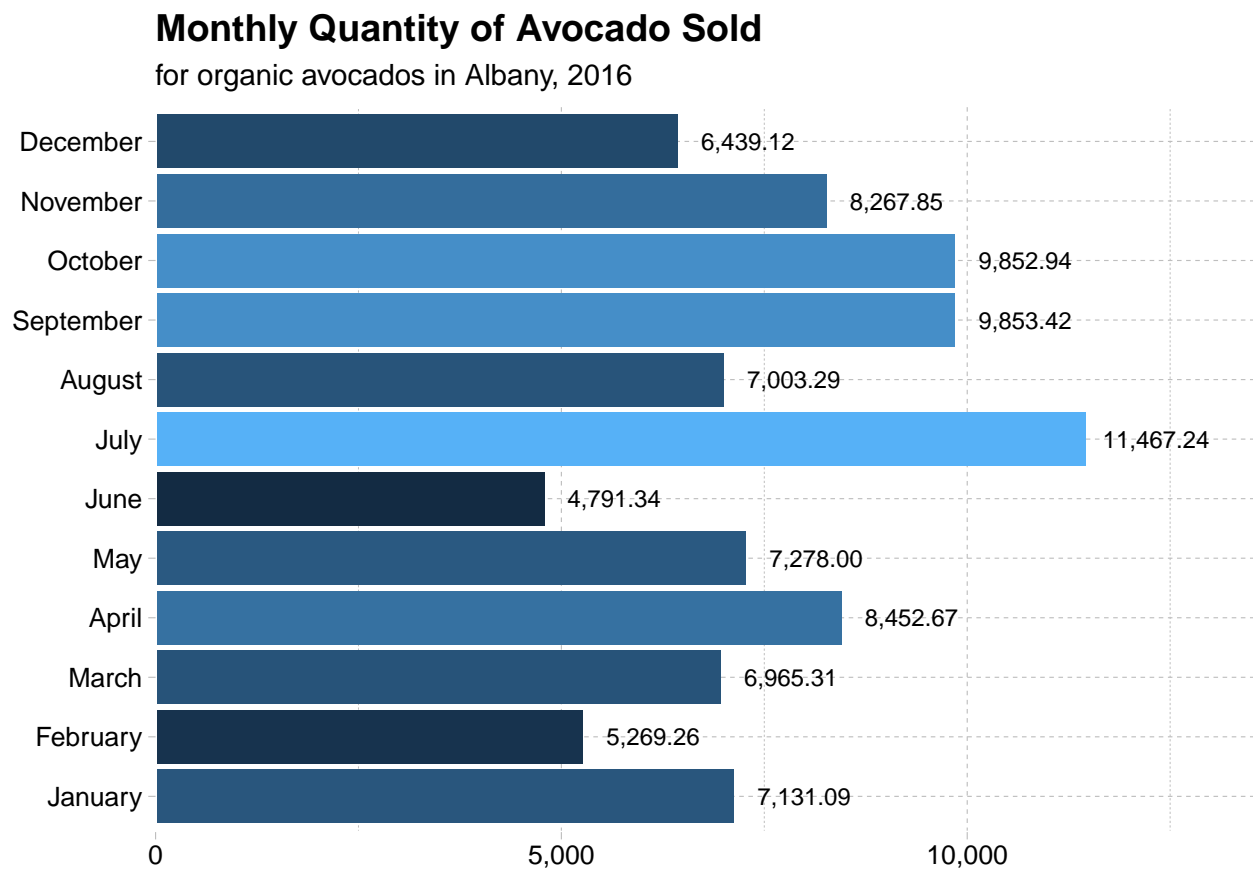
Organic vs Conventional Cultivation



We can look that organic avocado tend to more expensive. Seems make sense, because their cultivation is more expensive and we all like natural product and willing to pay a higher price for them. But the prices is does not only depend on the type. Let's look at avocado price from various regions in different years. As an example case, we will use the sales data in **Albany** for **organic** avocados in **2016**:

region	date	type	price	quantity
Albany	2016-01-03	organic	1.75	1145.50
Albany	2016-01-10	organic	1.83	1676.05
Albany	2016-01-17	organic	1.67	1496.73
Albany	2016-01-24	organic	1.83	1506.87
Albany	2016-01-31	organic	1.79	1305.94
Albany	2016-02-07	organic	1.81	1252.86
Albany	2016-02-14	organic	1.81	1247.80
Albany	2016-02-21	organic	1.79	1282.87
Albany	2016-02-28	organic	1.79	1485.73
Albany	2016-03-06	organic	1.92	1704.24

Exploratory Data Analysis

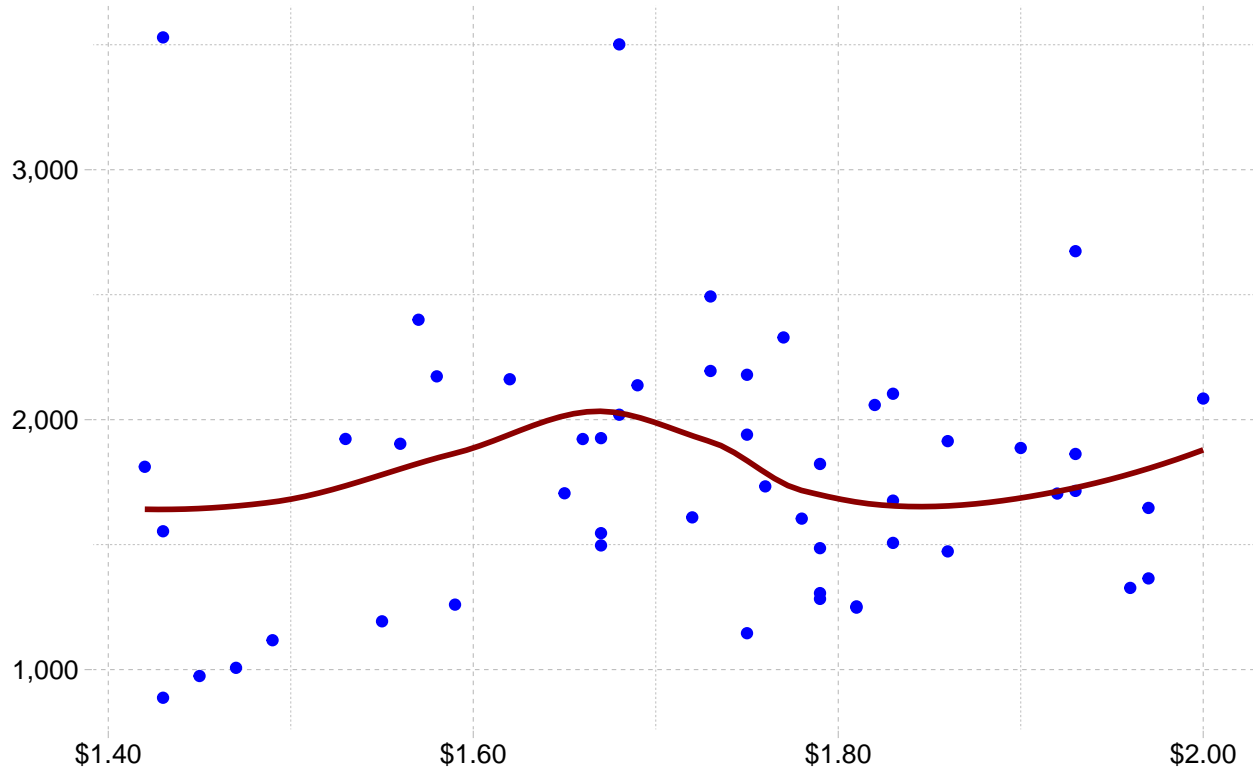


We can see the highest avocado sales is on **July** by **11,467.24**. Next, we want to build a simple model statistics called regression to predict the quantity given the price information. In exploring a regression

problem, one of the fundamental exploratory step is ensuring whether we really have an appropriate feature(s) to predict our target variables. One of the most common visualization is using scatter plot.

Price and Quantity of Avocados Sales

for organic avocados in Albany, 2016



Modeling with Regression Model

Model fitting process is—oftenly—very straightforward. But the challenge is more to how to communicate our model, instead of just make a model and doing prediction.

Since we use Ordinary Least Squares (OLS), one of its many perks is its interpretability. As shown in table below, from our model we could suggest that every 1 dollar increase in the price, will **increase** the total avocados consumption in a week (for **Albany** and **organicav** ocados) by **0.0002 Million**.

Terms	Coefficients	Standard Error	t-statistics	p-value
(Intercept)	1635.3810	1155.4322	1.415385	0.1672536
price	164.4529	676.4298	0.243119	0.8095679

Making prediction using our model is also very straight forward; and still, the challenge is how to communicate the prediction. We could either represent them as a table, or visualization if appropriate.

price	quantity	prediction
1.83	1506.87	1936.330
1.81	1252.86	1933.041
1.81	1247.80	1933.041
1.79	1282.87	1929.752
1.92	1704.24	1951.131

Evaluating Model Prediction

As a data scientist, we need to ensure that our model is already the best model that we could make; or even if our model is not perfect (which is very likely!), we could report how much is the error expectation. This is where model evaluation very useful. For example, our model give an Mean Average Error (MAE) value of 0 K in unseen dataset. This could be intepreted as upper and lower confidence that we could apply to our prediction, if we need any.